

SOERGEL CALCULUS

BEN ELIAS AND GEORDIE WILLIAMSON

To Mikhail Khovanov and Raphaël Rouquier, who taught us generators and relations

ABSTRACT. The monoidal category of Soergel bimodules is an incarnation of the Hecke category, a fundamental object in representation 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 give a new proof and a generalization of Soergel’s classification of the indecomposable Soergel bimodules.

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 subsets $BwB \subset G$. 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

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

Received by the editors January 16, 2015 and, in revised form, March 24, 2016.
2010 *Mathematics Subject Classification*. Primary 20C33, 20F55, 20G05; Secondary 22E46.
The first-named author was supported by NSF Postdoctoral Fellowship DMS-1103862.

for $w \in W$ and $s \in S$. Writing $\mathbf{H}_{\mathbb{F}_q} \stackrel{\text{def}}{=} \mathbf{H} \otimes \mathbb{C}$ for the specialization of \mathbf{H} at $v^{-1} \mapsto \sqrt{|\mathbb{F}_q|} \in \mathbb{C}$, we have an isomorphism of algebras

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

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 relations

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

for each $s \in S$, and the braid relations

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

for each $s \neq t \in S$, 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 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 realized 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 suppose that G is a complex reductive group with Borel subgroup B and maximal torus $T \subset B$. The Hecke category (in its simplest geometric incarnation) is the additive subcategory

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

of semi-simple complexes 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 $\mathbf{IC}_w \stackrel{\text{def}}{=} \mathbf{IC}(\overline{BwB})$, the equivariant intersection cohomology complexes of $B \times B$ -orbits. There is a monoidal structure $*$ on $D_{B \times B}^b(G, \mathbb{C})$ given by convolution, and it preserves \mathcal{H} ; this follows because G/B is compact, and we can apply the decomposition theorem. 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 realized 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 with \mathfrak{h}^* in degree two (so R is zero in odd degrees). 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 seen as a functor to $R - \text{Bim}$, the category of graded R -bimodules.

Soergel’s first key observation is that hypercohomology

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

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 $\mathbf{IC}_s * \mathbf{IC}_t * \dots * \mathbf{IC}_u$ (for a reduced expression $st \dots u$ 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

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

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, $\mathbb{S}\text{Bim}$ 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 [Soe07] Soergel imitates the above definition of $\mathbb{S}\text{Bim}$ 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 $\mathbb{S}\text{Bim}$ defined as above, Soergel then constructs an isomorphism of $\mathbb{Z}[v^{\pm 1}]$ -algebras

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

In analogy to the Bott-Samelson description of intersection cohomology complexes in \mathcal{H} , Soergel proves that the indecomposable bimodules in $\mathbb{S}\text{Bim}$ (up to grading shift) are in bijection with W . More precisely, for each $w \in W$ an indecomposable bimodule B_w appears as the unique summand of $B_s \otimes B_t \otimes \dots \otimes B_u$ (for a reduced expression of w) which does not appear for a shorter expression, and these parametrize the indecomposable bimodules up to shift. These results are known as Soergel’s Categorification Theorem.

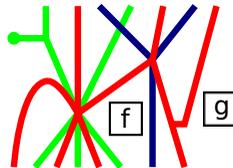
For a Weyl group, one can prove Soergel’s Categorification Theorem easily by transferring known facts about \mathcal{H} to $\mathbb{S}\text{Bim}$ using hypercohomology. Soergel’s proof for the general case is much trickier, but relies only on commutative algebra. Soergel proves his results for reflection faithful representations of a Coxeter system over an infinite field of characteristic $\neq 2$.

Soergel’s theory (or the Bott-Samelson description of \mathcal{H}) states that the objects of $\mathbb{S}\text{Bim}$ are generated by the objects B_s . Moreover, there are isomorphisms between objects in $\mathbb{S}\text{Bim}$ which lift the quadratic and braid relations of the Hecke algebra. (There is no Soergel bimodule B with $[B] = T_s$, so that one does not categorify Iwahori’s presentation given above but instead an analogous presentation using the Kazhdan-Lusztig generators; see Section 2.3.) Heuristically speaking, this is the categorical analogue of Iwahori’s algebraization of \mathbf{H} , on the level of objects. However, in \mathcal{H} or $\mathbb{S}\text{Bim}$ 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 $B_{\underline{w}} \stackrel{\text{def}}{=} B_s \otimes_R B_t \otimes_R \cdots \otimes_R B_u$ for an expression $\underline{w} = st \dots u$. They form a full monoidal subcategory of $R - \text{Bim}$, which we denote $\mathbb{B}\text{SBim}$. By definition, any Soergel bimodule is a direct sum of shifts of summands of bimodules in $\mathbb{B}\text{SBim}$. Said another way, the category of Soergel bimodules $\mathbb{S}\text{Bim}$ is the Karoubi envelope of (the additive, graded envelope of) $\mathbb{B}\text{SBim}$. 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 $\mathbb{B}\text{SBim}$.

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 [Elid]. 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 $\mathbb{B}\text{SBim}$, 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 planar diagram



represents a map from $rgrgrgbr$ to $grbgrbr$ (where r represents red, etc.).

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	$B_s \rightarrow R$	$f \otimes g \mapsto fg$
	deg 1	$R \rightarrow B_s$	$1 \mapsto \frac{1}{2}(\alpha_s \otimes 1 + 1 \otimes \alpha_s)$
	deg -1	$B_s B_s \rightarrow B_s$	$1 \otimes g \otimes 1 \mapsto \partial_s g \otimes 1$
	deg -1	$B_s \rightarrow B_s B_s$	$1 \otimes 1 \mapsto 1 \otimes 1 \otimes 1$
	deg 0	$\underbrace{B_s B_t \dots}_{m_{st}} \rightarrow \underbrace{B_t B_s \dots}_{m_{st}}$	
	deg f	$R \rightarrow R$	$1 \mapsto f$

It is worth noting that each of these generators (except the last) lives in a one-dimensional space (of bimodule morphisms of the given degree), so that our formulas merely pin down a choice of scalars. In the above, α_s denotes a fixed choice of equation for the hyperplane fixed by s , and ∂_s denotes the Demazure operator $\partial_s(f) \stackrel{\text{def}}{=} (f - sf)/\alpha_s$. 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 $B_s B_t \dots$ and $B_t B_s \dots$ 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.

For example, a subset of the relations, the *isotopy relations*, are sufficient to imply that any two diagrams constructed from these generators which are isotopic as embedded graphs are actually equal. The remaining relations can then be expressed as equalities between linear combinations of (isotopy classes of) graphs.

Remark 1.1. In a monoidal category relations are *local* in the plane. We may draw the relations as equalities between graphs on the planar disk, even though graphs on the disk do not represent morphisms. These relations can be “plugged in” as subgraphs of graphs on the planar strip to give equalities between actual morphisms.

Let us call a subset $J \subset S$ *finitary* if the corresponding parabolic subgroup is finite. In the Iwahori presentation of \mathbf{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 $\mathbb{B}SBim$, 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 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 one color isotopy relations (not drawn here; see [EK, (3.5) through (3.9)]), which guarantee that any diagram involving one color is isotopy invariant, follow from the axioms for Frobenius objects.

The remaining one color relations not involving polynomials are the following:

$$\begin{array}{c} \diagup \\ \diagdown \end{array} = \begin{array}{c} \diagdown \\ \diagup \end{array}, \quad \begin{array}{c} | \\ \bullet \end{array} = |, \quad \bigcirc = 0.$$

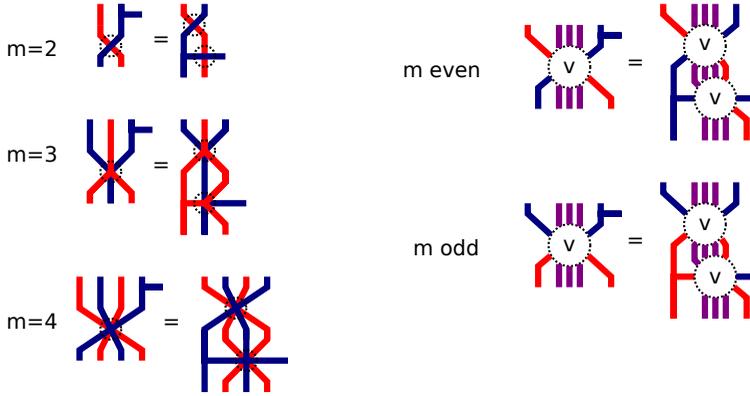
One has the following relations governing the interaction of edges and polynomials:

$$\begin{array}{c} \bullet \\ | \\ \bullet \end{array} = \alpha_s, \quad \begin{array}{c} f \\ | \end{array} = \begin{array}{c} s(f) \\ | \end{array} + \begin{array}{c} \bullet \\ | \\ \bullet \end{array} \partial_s f$$

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

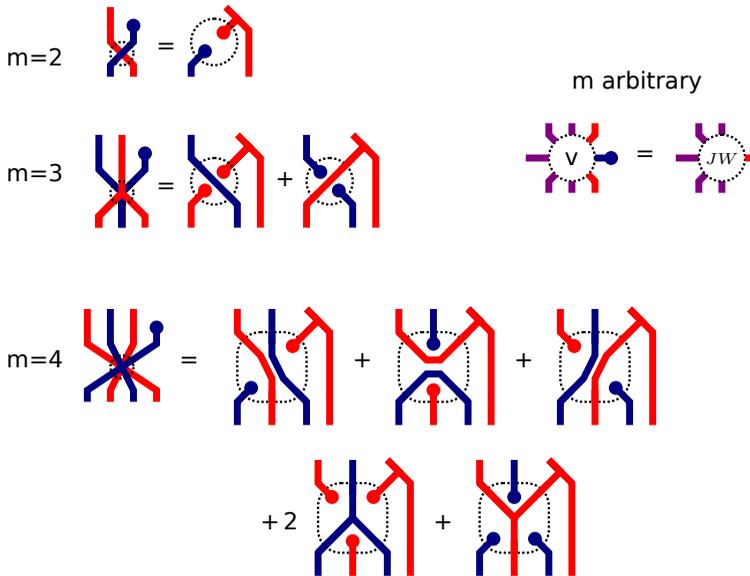
The first important two color relation (not drawn here; see [EK, (3.27)] for an example) is the cyclicity of the $2m_{st}$ -valent vertex, which is the last of the isotopy relations. (There is a subtlety if the Cartan matrix is not symmetric, which we ignore in the introduction.) The second relation is called *two color associativity*

(here $m = m_{st}$):



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

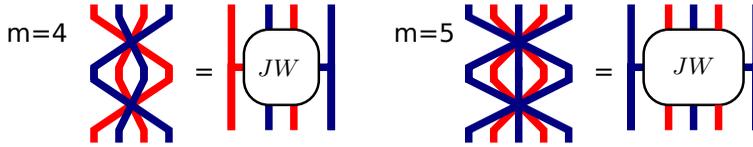
The third relation allows one to expand the composition of a dot and a $2m_{st}$ -valent vertex into a linear combination of diagrams in which the $2m_{st}$ -valent vertex does not occur. This relation is best understood using Jones-Wenzl projectors, as explained in [Elid], and is difficult to state 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.

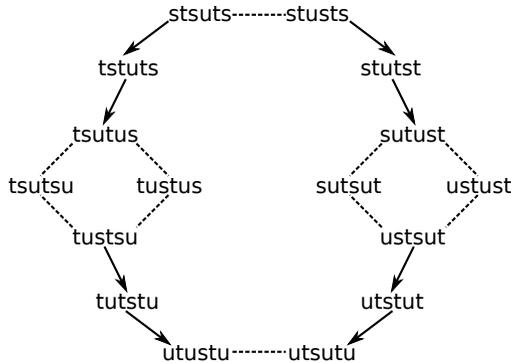
Together, the two color relations imply that the composition of two $2m_{st}$ -valent vertices is an idempotent endomorphism (corresponding to the projection to $B_{s,t}$

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, so we will digress on the topic of braidings. 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 braid groups “acting on categories”, or more generally of braided monoidal 2-categories, the role of the Yang-Baxter equation is played by a 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.

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 $su \leftrightarrow us$):



One should think that each expression above corresponds to a 1-morphism in a braided monoidal 2-category, obtained by applying the braid generators in a sequence determined by the expression. Each dashed edge is a 2-isomorphism coming from the monoidal structure, while each solid edge is a Yang-Baxter isomorphism, so that a path gives a 2-morphism by composition. The two dashed squares commute by the axioms of a monoidal 2-category. The Zamolodchikov tetrahedron equation is the requirement that the morphism obtained by following a non-trivial loop around this graph is the identity 2-morphism.

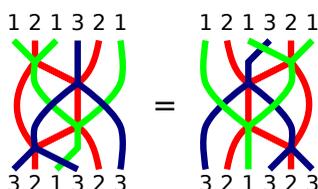
For any element $w \in W$ of type A_n , one can draw a “reduced expression graph” as above, and obtain a morphism for any path in this graph. The Zamolodchikov relation is already sufficient to imply that any non-trivial loop yields the identity

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

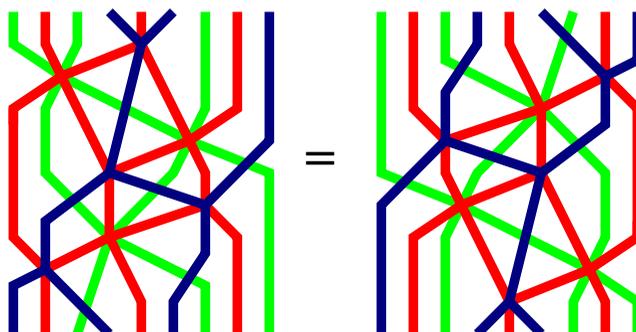
2-morphism, which explains its great importance. In fact, for any element of any Coxeter group, the loops in its reduced expression graph are (in a suitable sense) generated by the loops in the reduced expression graph of the longest element in any finite parabolic subgroup of rank 3. Thus, in addition 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 $m < \infty$. In the setting above, the $A_1 \times A_2$ and $A_1 \times A_1 \times A_1$ relations came for free from the monoidal structure, which is why we said the Zamolodchikov relation was sufficient in type A_n . In our setting, however, the $A_1 \times I_2(m)$ relations will be on equal footing with the Zamolodchikov relation.

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. However, for general reasons (explained later in the introduction), two paths with the same start and endpoint will be equal “modulo lower terms,” that is, modulo morphisms which factor through $B_{\underline{y}}$ for a sequence \underline{y} strictly shorter than \underline{w} .

The discussion above seems to have no relation to the category $\mathbb{B}SBim$, which is not a braided monoidal category. 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 (non-dashed edges in the) reduced expression graph of the longest element, such that the two paths from source to sink yield morphisms in $\mathbb{B}SBim$ which are equal on the nose! For example, we have placed the desired orientation on the non-dashed edges in the A_3 graph above (although some other orientations work too, such as the reversed orientation). Tracing out the two morphisms from the sink to source in the above graph yields the A_3 Zamolodchikov relation:



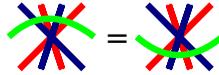
Entirely analogously one has the following Zamolodchikov relation for B_3 :



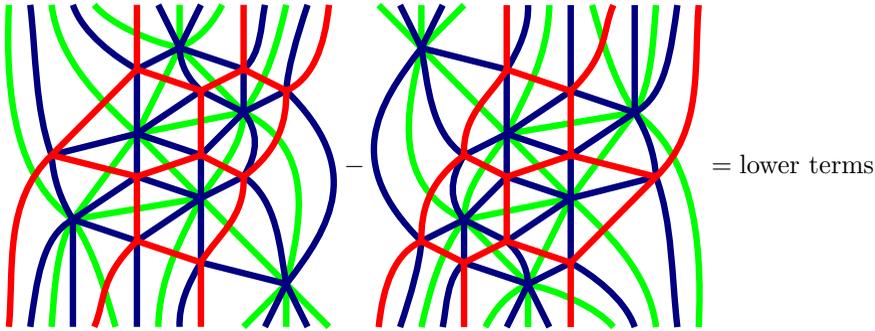
There is a “canonical” (and an “anti-canonical”) choice of orientation on a reduced expression graph for any element of any type A Coxeter group. This is an

old result of Manin-Schechtman [MS89], and some implications of this for morphisms between Bott-Samelson bimodules have been explored in [Elia]. However, the relationship between Manin-Schechtman 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 (up to reversal), and it yields the following relation in $\mathbb{B}\mathbb{S}\text{Bim}$:



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 non-trivial sum of lower terms. There is some relation of the form:



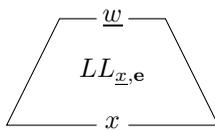
However, despite considerable effort, we have not been able to compute the lower terms which appear. The question of what these lower terms are could in principle be decided by computer, however the computation is impossible with our current algorithms and technology. This is the caveat mentioned earlier: we do not have a completely explicit presentation of the category $\mathbb{B}\mathbb{S}\text{Bim}$ 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 $\mathbb{B}\mathbb{S}\text{Bim}$, we construct a basis for morphisms in \mathcal{D} , which is sent to a basis for morphisms in $\mathbb{B}\mathbb{S}\text{Bim}$. 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 can be understood using the technique of localization, which we will also explore diagrammatically in this paper.

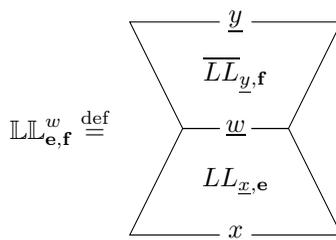
Let \underline{x} be an expression and \mathbf{e} a subexpression expressing w . That is, if $\underline{x} = s_1 \dots s_m$, then \mathbf{e} is a sequence $\mathbf{e}_1 \dots \mathbf{e}_m$ of 0's and 1's such that $w = s_1^{\mathbf{e}_1} \dots s_m^{\mathbf{e}_m}$. To this pair Libedinsky assigns a morphism $LL_{\underline{x}, \mathbf{e}} : B_{\underline{x}} \rightarrow B_{\underline{w}}$, 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 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 can be thought of as a kind of trapezoid (see Chapter 6)



where the upper expression is always reduced.

Now let \underline{x} and \underline{y} be fixed expressions. Following Libedinsky [Lib15], we introduce the *double leaves* morphism



where $\overline{LL}_{\underline{y}, \mathbf{f}}$ denotes the vertical flip of a light leaf morphism, and \mathbf{e} (resp. \mathbf{f}) is a subexpression of \underline{x} (resp. \underline{y}) which expresses w . Hence $\mathbb{L}\mathbb{L}_{\mathbf{e}, \mathbf{f}}^w$ is a morphism $B_{\underline{x}} \rightarrow B_{\underline{y}}$. As w is determined by \mathbf{e} and \mathbf{f} , we often omit it from the notation. Consider the set

$$\mathbb{L}\mathbb{L}_{\underline{x}, \underline{y}} \stackrel{\text{def}}{=} \bigcup_{w \in W} \{\mathbb{L}\mathbb{L}_{\mathbf{e}, \mathbf{f}}^w \mid \mathbf{e} \in M(\underline{x}, w), \mathbf{f} \in M(\underline{y}, w)\},$$

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

Theorem 1.2. $\mathbb{L}\mathbb{L}_{\underline{x}, \underline{y}}$ is an R -basis for $\text{Hom}_{\mathcal{D}}(B_{\underline{x}}, B_{\underline{y}})$.

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

- (1) Under general assumptions one has a functor $\mathcal{F}: \mathcal{D} \rightarrow \mathbb{S}\text{Bim}$ (we defined this functor earlier under the assumption that 2 was invertible). In fact, \mathcal{F} is always faithful when it is defined; see Proposition 5.27. If the category of Soergel bimodules is well behaved (i.e., if the Soergel Categorification Theorem holds), then \mathcal{F} is an equivalence of graded monoidal categories.
- (2) The category of Bott-Samelson 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.
- (3) For applications in representation theory it is desirable to have versions of the Hecke category over fields of positive characteristic. The diagrammatic category can be defined over a broad class of rings; for example it is defined

over \mathbb{Z} if W is crystallographic. Under mild assumptions the indecomposable objects in the diagrammatic category are parametrized (up to shift) by W , and the split Grothendieck group categorifies the Hecke algebra. This gives a new proof of Soergel’s Categorification Theorem whenever Soergel bimodules are well behaved, and indicates that the diagrammatic category is the correct replacement when they are not.

Variants of Soergel bimodules attached to (affine) Weyl groups play an important role in modular representation theory [Soe00, Fie11], when defined in positive characteristic. The diagrammatic category appears to be the correct general replacement for Soergel bimodules in these settings. The corresponding geometric replacement is the category of parity sheaves [JMW14], and under mild assumptions the (additive, monoidal, graded) category of parity sheaves and the diagrammatic category coincide [RW].

Similarly, one can define both Soergel bimodules and our diagrammatic category for other representations of W (the correct generality is a *realization*, defined in Section 3.1). When the realization is not faithful, Soergel bimodules behave poorly, while the diagrammatic category remains well behaved. So, to belabor the point once more, presenting the Hecke category by generators and relations and freeing it from the constraints of Soergel’s bimodule construction allows one to work in much greater generality. This has already found use in applications, for example [Elic].

1.6. Proving the main theorem. Before we describe the methods of proof in more detail, let us emphasize the major dichotomy which just arose in the previous section. There are two different categories at play here: the diagrammatic category \mathcal{D} , and the category of Bott-Samelson bimodules $\mathbb{B}\text{SBim}$. As mentioned in the previous section, these categories are not always equivalent, and when they are not, it is the category \mathcal{D} which behaves well (it categorifies the Hecke algebra). This paper proves facts about \mathcal{D} which are inspired by facts in $\mathbb{B}\text{SBim}$, but must be proven independently.

The crucial result from which all other results follow is Theorem 1.2, which states that double leaves form an R -basis for morphisms in the diagrammatic category \mathcal{D} . By studying double leaves as a cellular basis, one can prove the Soergel Categorification Theorem using elementary idempotent-lifting arguments. This is accomplished in Theorem 6.26. For this result, the assumption is made that \mathbb{k} be a complete local ring, in order that the Karoubi envelope of \mathcal{D} is Krull-Schmidt and has the desired idempotent-lifting properties. It also follows from Libedinsky’s results [Lib08] that the functor \mathcal{F} must be an equivalence when $\mathbb{B}\text{SBim}$ is well behaved (i.e., when \mathbb{k} is an infinite field of characteristic $\neq 2$, and the realization is reflection faithful), since then double leaves form an R -basis for morphisms in $\mathbb{B}\text{SBim}$ as well.

The technology of localization is developed in order to prove that double leaves are linearly independent over R . Meanwhile, the proof that double leaves span, which takes the entirety of Chapter 7, is a long and convoluted diagrammatic argument. Let us discuss localization first.

Definition 1.3. Let A be an algebra with an action of a group W . The *2-groupoid of W over A* , denoted $\Omega_A(W)$, is the A -linear monoidal category defined as follows. The objects are in bijection with W . For $w, x \in W$, there are no non-zero morphisms from w to x unless $w = x$, in which case $\text{End}(w) \cong A$. The monoidal structure is defined on objects by $w \otimes x \cong wx$; the corresponding map on morphism

spaces $\text{End}(w) \otimes_A \text{End}(x) \cong A \otimes_A A \rightarrow A \cong \text{End}(wx)$ is defined to be $a \otimes b \rightarrow aw(b)$ for $a, b \in A$.

In Soergel's algebraic study of Soergel bimodules [Soe07], a key role is played by *standard bimodules* R_w , one for each element $w \in W$. These are R -bimodules which generate a monoidal category equivalent to $\Omega_R(W)$ (so long as the realization of W is faithful). Every Soergel bimodule has a *standard filtration*, a filtration whose subquotients are standard bimodules with shifts. After changing base from R to its fraction field Q , these standard filtrations split, so that the Karoubi envelope of $\mathbb{B}\text{SBim} \otimes_R Q$ is actually isomorphic to the additive closure of $\Omega_Q(W)$. Thus, localization to Q drastically simplifies the structure of the category. More details on standard bimodules can be found in Section 3.4.

The category \mathcal{D} is also R -linear, and one wishes to prove that taking the Karoubi envelope of $\mathcal{D}_Q = \mathcal{D} \otimes_R Q$ will also result in the additive closure of $\Omega_Q(W)$. (Here, it does not matter if the realization is faithful, as \mathcal{D} is defined using the combinatorics of W , not the behavior of its chosen representation.) We developed a diagrammatic version of the 2-groupoid $\Omega_{\mathbb{k}}(W)$ of W over \mathbb{k} in the previous paper [EW] (the secret weapon in that paper is the topology of the Coxeter complex). Passing to a diagrammatic presentation of $\Omega_R(W)$ or $\Omega_Q(W)$ is easy; see Section 4. We call these diagrammatic categories \mathcal{D}^{std} and $\mathcal{D}_Q^{\text{std}}$ respectively. In Section 5.4, we define a third diagrammatic category (temporarily called $\text{Kar}(\mathcal{D}_Q)$) which combines \mathcal{D} and \mathcal{D}^{std} , augmenting these calculi by adding the extra morphisms which split the standard filtrations of the Bott-Samelson generators. We prove in Theorem 5.17 that $\text{Kar}(\mathcal{D}_Q)$ is, in fact, the Karoubi envelope of \mathcal{D}_Q , and is also isomorphic to the 2-groupoid $\mathcal{D}_Q^{\text{std}}$. Thus, diagrammatic localization is accomplished without any reference to bimodules.

In Section 6.2, we study the image of double leaves morphisms inside $\text{Kar}(\mathcal{D}_Q)$. We prove an upper-triangularity result which implies that the set of double leaves is sent to a linearly independent set of morphisms in $\text{Kar}(\mathcal{D}_Q)$, so that they must have been linearly independent (over R) to begin with. In fact, the inductive algorithm used to construct the light leaves morphisms is precisely tailored to this upper-triangularity property. This algorithm is explained in Section 6.1.

Remark 1.4. It is not terribly difficult for Soergel, in the algebraic context of $\mathbb{B}\text{SBim}$, to give a direct proof that morphism spaces are free as left R -modules. However, it is not obvious from the definitions of \mathcal{D} that morphisms are free as left R -modules. In particular, one cannot assume from the start that the base change functor $\mathcal{D} \rightarrow \mathcal{D}_Q$ is faithful. Thankfully, the diagrammatic localization argument above does imply that the R -span of double leaves is free as a left R -module. Thus, once we prove that double leaves span, we see that morphism spaces are free left R -modules and base change is faithful.

In Section 7 we give a diagrammatic argument that double leaves morphisms span. The real essence of the proof lies in Claims 7.9, 7.10, 7.13, 7.15, and 7.17 from Section 7.4. These five claims show, for each of the five generators of \mathcal{D} (in the list in Section 1.4, not counting multiplication by a polynomial), that placing this generator below a light leaf morphism yields something in the span of light leaves, modulo lower terms. The remainder of Section 7 builds up a scaffolding whereby these claims yield the inductive step of a complicated induction. This scaffolding relies on the general way in which light leaves are built inductively (e.g.

the formulas (6.1) and (6.2)), but not on the specifics of the construction. The key claims do rely on the specific construction (e.g. the form of the maps in Figure 2).

Remark 1.5. In recent work of the first author [Elib] (written several years after this paper first became available), a very similar proof is used to show that a certain inductively-defined collection of webs gives a cellular basis for the category of \mathfrak{sl}_n -webs. The “scaffolding” portion of the proof is cleaned up but is largely identical to the one in this paper, because the inductive formulation of the basis is essentially the same. However, the specifics are quite different. When we originally wrote the proof in Section 7 it was unsatisfying, as it seemed ad hoc and overly complicated. In hindsight, it seems to reflect some underlying generalities of monoidal cellular categories. One still hopes that the proof can be simplified, but it is not truly ad hoc.

1.7. Outline of the paper. This paper contains three parts.

Part 1: The first two sections give background on the Hecke algebra and Soergel bimodules.

Section 2: We recall the Hecke algebra, Kazhdan-Lusztig basis and Deodhar’s defect formula.

Section 3: We define standard and Soergel bimodules, state Soergel’s Categorification Theorem and discuss realizations and localization.

Part 2: In the next two sections, we define two diagrammatic categories.

Section 4: We define the diagrammatic presentation of standard modules.

Section 5: We recall results of [Elid] and define the diagrammatic category of Soergel bimodules.

Part 3: In the last two sections we study the diagrammatic categories and prove the equivalence to Soergel bimodules.

Section 6: We define Libedinsky’s light leaves and double leaves morphisms in the diagrammatic setting. We state the theorem that double leaves span, and deduce the main theorems of the paper.

Section 7: We prove that double leaves span.

Part 1. Background on the Hecke algebra and Soergel bimodules

2. THE HECKE ALGEBRA

Background on this chapter 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:

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

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

The numbers $m_{st} = m_{ts}$ associated to each pair of simple reflections determine the group W , and must satisfy $m_{st} \in \{2, 3, 4, \dots\}$ 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 $\ell : W \rightarrow \mathbb{Z}_{\geq 0}$.

For any subset $J \subset S$ the corresponding *parabolic subgroup* W_J is the subgroup generated by 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 element of S is isomorphic to S_2 , and the parabolic subgroup of a pair of elements 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:

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

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

It will be useful to work with a different normalization. If we set $H_s \stackrel{\text{def}}{=} vT_s$, then these relations become:

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

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

Notation 2.1. We will use an underlined letter $\underline{w} = (s_1, s_2, \dots, s_m)$ to denote a finite sequence of elements of S . Omission of the underline will denote the product $w = s_1 s_2 \dots s_m$ 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 $\underline{w} = s_1 s_2 \dots s_m$. The underline reminds us that the sequence of simple reflections, and not just their product in W , is important. Given $\underline{w} = s_1 s_2 \dots s_m$ we write $H_{\underline{w}} = H_{s_1} H_{s_2} \dots H_{s_m}$. 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 $H_e \stackrel{\text{def}}{=} 1$ for the identity of \mathbf{H} . These elements $\{H_w\}_{w \in W}$ form the *standard basis* of \mathbf{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 $h \mapsto \bar{h}$ 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 \mathbf{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 $\underline{w} = s_1 s_2 \dots s_m$ we set $\underline{H}_{\underline{w}} \stackrel{\text{def}}{=} \underline{H}_{s_1} \dots \underline{H}_{s_m}$. 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.

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 more complicated:

$$(2.7) \quad \underline{H}_s^2 = \underline{H}_s(v + v^{-1}) \quad \text{for all } s \in S,$$

$$(2.8) \quad \underbrace{\underline{H}_s \underline{H}_t \underline{H}_s \dots}_{k+1} = \sum_d c_{k,d} \underbrace{\underline{H}_{sts \dots}}_{d+1} \quad \text{for } 1 \leq k+1 \leq m_{st}, \quad \text{for all } s \neq t \in S.$$

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 $V_1^{\otimes k} \cong \bigoplus_d V_d^{\oplus c_{k,d}}$. Roughly speaking, this is because the Temperley-Lieb algebra at a root of unity “categorifies” the alternating product $\underline{H}_s \underline{H}_t \underline{H}_s \dots$ in the dihedral Hecke algebra. Relation (2.8) holds even when $m_{st} = \infty$. More details can be found in [Elid].

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

Example 2.5. The first few examples are:

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

2.4. The standard trace and the defect formula. A *trace* on \mathbf{H} is a $\mathbb{Z}[v^{\pm 1}]$ -linear map $\varepsilon: \mathbf{H} \rightarrow \mathbb{Z}[v^{\pm 1}]$ 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 combinatorial formula for the standard trace of a product $\underline{H}_{\underline{w}}$, known as the *defect formula*, which we now discuss.

A *subsequence* of $\underline{w} = s_1 s_2 \dots s_m$ is a sequence $\pi_1 \pi_2 \dots \pi_m$ 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 $\mathbf{e} = \mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_m$ 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 $x_0 = e, x_1, \dots, x_m$ defined by

$$x_i \stackrel{\text{def}}{=} s_1^{\mathbf{e}_1} s_2^{\mathbf{e}_2} \dots s_i^{\mathbf{e}_i}$$

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 $D(\text{own})$. We assign U to the index i if $x_{i-1}s_i > x_{i-1}$, and D if $x_{i-1}s_i < x_{i-1}$. Note that when $\mathbf{e}_i = 1$, then $x_i = x_{i-1}s_i$, so that e_i is labeled $U1$ if we moved up in the Bruhat order at time i , and $D1$ if we moved down. Meanwhile, if $\mathbf{e}_i = 0$, then $x_i = x_{i-1}$; we have glanced longingly in the direction of $x_{i-1}s_i$ but remained unmoved; we still label e_i as $U0$ or $D0$ based on where we might have gone. 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 difference between where we longed to go and where we actually went.

Example 2.6. 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 0, $(U1, D0, D0)$ 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 end-points 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 2.7. *For any expression \underline{w} we have*

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

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

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

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

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

We now come to the defect formula for the trace:

Corollary 2.8. *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 2.9. We continue Example 2.6 and check Lemma 2.7:

- Using relation (2.7) we see that

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

- Using direct calculation, or smoothness and (2.8), or Lemma 2.7 we obtain

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

The set of subsequences \mathbf{e} of a fixed expression \underline{w} is equipped with a partial order, the *path dominance order*. Let \mathbf{e} and \mathbf{f} be two 01-sequences and let their corresponding Bruhat strolls be x_0, x_1, \dots, x_m and y_0, y_1, \dots, 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 \mathbf{e} is greater than or equal to the end-point of \mathbf{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 anti-involution for which $\omega(\underline{H}_s) = \underline{H}_s$ and $\omega(v) = v^{-1}$. The standard trace gives rise to the *standard pairing* $\mathcal{H} \times \mathcal{H} \rightarrow \mathbb{Z}[v^{\pm 1}]$, defined by $(a, b) = \varepsilon(b\omega(a))$. This pairing is *semi-linear* 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, \underline{H}_s is self-biadjoint, i.e.,

$$(\underline{H}_s x, y) = (x, \underline{H}_s y), \quad (x \underline{H}_s, y) = (x, y \underline{H}_s).$$

Remark 2.10. 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 semi-linear pairings with self-biadjoint \underline{H}_s is isomorphic to the module of all $\mathbb{Z}[v^{\pm 1}]$ -linear traces. Any such pairing is determined by the values $\varepsilon(\underline{H}_{\mathbf{x}}) = (1, \underline{H}_{\mathbf{x}})$ over all sequences \mathbf{x} .

3. SOERTEL 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 \mathbb{k} be a commutative integral domain. A *realization* of (W, S) over \mathbb{k} is a free, finite rank \mathbb{k} -module \mathfrak{h} , together with subsets $\{\alpha_s^\vee \mid s \in S\} \subset \mathfrak{h}$ and $\{\alpha_s \mid s \in S\} \subset \mathfrak{h}^* = \text{Hom}_{\mathbb{k}}(\mathfrak{h}, \mathbb{k})$, satisfying:

- (1) $\langle \alpha_s^\vee, \alpha_s \rangle = 2$ for all $s \in S$;
- (2) 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 ;
- (3) the technical condition in (3.3) is satisfied (its description requires some background).

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

Remark 3.2. In a moment we will give examples of realizations of (W, S) . The reader can ignore condition (3) on a first reading.

Given a realization over \mathbb{k} and a homomorphism $\mathbb{k} \rightarrow \mathbb{k}'$ we obtain a realization $\mathbb{k}' \otimes_{\mathbb{k}} \mathfrak{h}$ over \mathbb{k}' 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 does not preserve faithfulness in general. For us, the ability to perform 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 $m_{st} = 2m, 3m, \dots$, 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 3.3. Some examples of realizations that we have in mind are the following:

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

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

(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 equation (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 \mathbb{k} 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 \mathbb{k}) 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 $\mathbb{k} = \mathbb{Z}[\phi]$, where $\phi = (1 + \sqrt{5})/2$ denotes the golden ratio. Let $\mathfrak{h} = \bigoplus_{s \in S} \mathbb{k}\alpha_s^\vee$ 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 \mathbb{k} . In particular, the finite reflection groups of types H_3 and H_4 have symmetric realizations over (any extension of) \mathbb{k} .

- (6) Let (W, S) be the affine Weyl group of type \widetilde{A}_n for $n \geq 2$. Let $\mathbb{k} = \mathbb{Z}[q, q^{-1}]$ 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:

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

(More precisely, this is the example when $n = 4$.) This gives a realization of W . Specializing q to an element of $\mathbb{C} \setminus \mathbb{R}$, one obtains a realization of W over \mathbb{C} which cannot be obtained by extension of scalars from a realization defined over \mathbb{R} . (This “deformed” Cartan matrix is the key to the first author’s work on quantum geometric Satake [Elic].)

Given a realization $(\mathfrak{h}, \{\alpha_s^\vee\}, \{\alpha_s\})$ of (W, S) over \mathbb{k} 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 $a_{ss} = 2$, one can construct the free \mathbb{k} -module $\mathfrak{h} = \bigoplus_{s \in S} \mathbb{k} \alpha_s^\vee$, 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 \mathbb{k} . 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 3.4. In Example 3.3 the realizations discussed in (1), (5) and (6) are minimal. The example in (3) is minimal if and only if the root system is semi-simple and of adjoint type (so that the coroot lattice coincides with \mathfrak{h}). Examples (2) and (4) are not minimal in general.

Remark 3.5. 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) . A thorough discussion of this can be found in the appendix to [Elid]. We provide a short discussion here.

Definition 3.6. Define the *2-colored quantum numbers* $[k]_x$ and $[k]_y$ inside the ring $\mathbb{Z}[x, y]$ inductively. One has $[0]_x = [0]_y = 0$, $[1]_x = [1]_y = 1$, and $[2]_x = x$, $[2]_y = y$. The other 2-colored quantum numbers are defined by the rules

$$(3.2a) \quad [2]_x [k]_y = [k + 1]_x + [k - 1]_x,$$

$$(3.2b) \quad [2]_y [k]_x = [k + 1]_y + [k - 1]_y.$$

When k is odd, $[k]_x = [k]_y$ and we shorten the notation to $[k]$.

Fix a pair $s, t \in S$ and let $x = a_{st}$ and $y = a_{ts}$. The condition that (st) has order exactly k is an algebraic condition on x and y . For instance, when α_s and α_t are linearly independent, the action of (st) on their span has order k when $[2k + 1] = 1$ and $[2k]_x = [2k]_y = 0$. In fact, this implies further that $2[k]_x = 2[k]_y = [2]_x [k]_y = [2]_y [k]_x = 0$. This suggests that one should set

$$(3.3) \quad [m_{st}]_x = [m_{st}]_y = 0.$$

This is the technical condition mentioned in Definition 3.1. While (3.3) is sufficient to imply that (st) has order dividing k on the span of the roots, it is independent of the condition that W acts on \mathfrak{h} . The reason that (3.3) is required is to ensure that 2-colored Jones-Wenzl projectors are rotation-invariant, as discussed in Section 5.2.

If either x or y is a non-zero-divisor, then (3.3) is equivalent to the statement that xy satisfies the minimal polynomial of the algebraic integer $4 \cos^2(\frac{\pi}{m})$. If this algebraic integer does not exist in \mathbb{k} , then (W, S) does not admit a realization over \mathbb{k} . Any Coxeter system (of finite rank) admits a realization over some ring of integers. Finally, we introduce one other technical condition.

Definition 3.7. We call a realization *balanced* if for every $s, t \in S$ one has $[m_{st} - 1]_x = [m_{st} - 1]_y = 1$. We refine this notion by calling the realization *even-balanced* (resp. *odd-balanced*) if this property holds when m_{st} is even (resp. odd). The opposite of even-balanced is *even-unbalanced*.

The familiar Cartan matrices of Weyl groups are balanced. However, the Cartan matrix of type A_2 is not balanced when viewed as a realization of G_2 . The exotic Cartan matrices for type \tilde{A}_n , $n \geq 3$, given in example (6) above are not balanced, except when $q = 1$. Being balanced is equivalent to the existence of an unambiguous notion of positive roots in \mathfrak{h}^* ; when the realization is symmetric, being balanced is similar to the condition that simple roots form an obtuse angle. Faithful realizations are almost always even-balanced; any minimal even-unbalanced realization is not faithful. Once again, a thorough discussion of these technicalities can be found in [Elid].

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 \mathbb{k} -algebra with $\text{deg } \mathfrak{h}^* = 2$. Then W acts on \mathfrak{h}^* via the contragredient representation $(s(\gamma) = \gamma - \langle \alpha_s^\vee, \gamma \rangle \alpha_s)$ 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 $R - \text{Mod}$ and $R - \text{Bim}$ denote the category of graded R -modules and bimodules respectively. We view $R - \text{Mod}$ and $R - \text{Bim}$ as graded categories; that is, as categories enriched in graded \mathbb{k} -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 $\text{Hom}_0(M, N)$.

3.2. Technicalities. It is important to remember the key dichotomy in this paper: we will be discussing two separate categories. Fix a realization of (W, S) , and consider the ring R defined above. In this section we will introduce Soergel’s monoidal category $\mathbb{S}\text{Bim}$, which is a full subcategory of $R - \text{Bim}$. In Section 5 we will define a diagrammatic category \mathcal{D} by generators and relations, whose morphism spaces will be enriched in $R - \text{Bim}$. 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 $\mathbb{S}\text{Bim}$ behaves well, one can construct an equivalence from \mathcal{D} to $\mathbb{S}\text{Bim}$. However there are certain situations (for example when the characteristic of \mathbb{k} 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.² This is one of the advantages of the diagrammatic approach.

Remark 3.8. The very existence of a realization is already an assumption on the base ring \mathbb{k} : namely, that it contains certain algebraic integers.

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 3.3. 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.

Assumption 3.9 (Demazure Surjectivity). *The map $\alpha_s: \mathfrak{h} \rightarrow \mathbb{k}$ is surjective, for all $s \in S$. Evaluation at α_s^\vee gives a surjective map $\mathfrak{h}^* \rightarrow \mathbb{k}$, 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, the first condition in Assumption 3.9 guarantees that $\alpha_s \neq 0$, so that $s(\delta) = \delta - \alpha_s \neq \delta$.

If 2 is invertible in \mathbb{k} , then Demazure Surjectivity holds, because $a_{ss} = 2$. If m_{st} is odd, then both α_s and α_t (and α_s^\vee and α_t^\vee) are surjective, because the algebraic integer $4 \cos^2(\frac{\pi}{m_{st}})$ is invertible in any ring which contains it (see [Elid]). Even when the ideal in \mathbb{k} 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, and the assumption that the realization is even-balanced, 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 $\text{Kar}(\mathcal{D})$ require that \mathbb{k} 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 \mathbb{k} is not a complete local ring (for example \mathbb{Z}) we do not know 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 one can easily discuss indecomposable objects.

Definition 3.10. Following Soergel [Soe07], we say that a realization \mathfrak{h} over a field \mathbb{k} 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 3.3.

²Another natural replacement is the Braden-MacPherson-Fiebig theory of sheaves on moment graphs [BM01, Fie08, Fie10].

Soergel’s theory gives techniques to study SBim defined for a reflection faithful representation over an infinite field \mathbb{k} 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 full generality of his results is unknown.

Finally, the assumption that the realization is balanced will drastically simplify both the study of bimodules and the study of diagrammatics. The realizations studied by Soergel have all been balanced. If the realization is even-unbalanced, then any diagrammatic encoding of Soergel bimodules will not be cyclic. We will not attempt to address this case. If the realization is odd-unbalanced, the diagrammatics have extra complications, but are still tractable. We discuss this case in remarks and in Section 5.6, though mostly we will assume the realization is odd-balanced for the sake of simplicity.

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

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

The numerator and denominator are both s -anti-invariant, 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. Note that $\delta - s(\delta) = \alpha_s$.

Claim 3.11. *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 \mathbb{k} is a domain R is also, and therefore $g = g'$, so that $h = h'$.

For any $\lambda \in \mathfrak{h}^*$, $\langle \alpha_s^\vee, \lambda \rangle \delta - \lambda \in \mathfrak{h}^*$ 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 of R^s -bimodules, whose kernel is precisely R^s . It is easy to show that ∂_s satisfies the *twisted Leibniz rule*

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

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

Claim 3.11 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 $(g, h) = \partial_s(gh)$. Thus R has the structure of a graded Frobenius extension of R^s . A *graded Frobenius extension* is a Frobenius extension of graded algebras where the Frobenius trace map (in this case ∂_s) is homogeneous of some degree.

The coproduct 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 3.12. 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 \mathbb{k} . This was the choice used in the introduction.

Remark 3.13. 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 $\partial_s: \mathfrak{h}^* \rightarrow \mathbb{k}$ 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 3.14. If the realization is odd-balanced, then Demazure operators associated to $s \in S$ satisfy the braid relations. Otherwise, they do not. See [Elid] 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 graded 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 $\underline{w} = s_1 s_2 \dots s_d$ the corresponding *Bott-Samelson bimodule* is the tensor product

$$B_{\underline{w}} \stackrel{\text{def}}{=} B_{s_1} \otimes B_{s_2} \otimes \dots \otimes B_{s_d}$$

viewed as an R -bimodule under left and right multiplication. The Bott-Samelson bimodule $B_{\underline{w}}$ is isomorphic to $R \otimes_{s_1} R \otimes_{s_2} R \otimes \dots \otimes_{s_d} R(d)$. We let $\mathbb{B}\text{SBim}$ denote the full monoidal subcategory of $R - \text{Bim}$ whose objects are Bott-Samelson bimodules (where, as before, morphism spaces are graded \mathbb{k} -modules). Finally, we let $\mathbb{S}\text{Bim}$ denote the idempotent closure or Karoubi envelope of the additive, graded closure of $\mathbb{B}\text{SBim}$, 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 $\mathbb{S}\text{Bim}$ 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 in general, 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 \mathbb{k} -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 $R_{\underline{w}} \stackrel{\text{def}}{=} R_{s_1} \otimes R_{s_2} \otimes \dots \otimes R_{s_d}$. 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 three 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 three 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 \mathbb{k} -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 section that the realization is faithful. We have

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

In other words, $StdBim$ is isomorphic to the graded 2-groupoid for W over R (see Definition 1.3). 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_x and R_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 \mathbb{k} , and can be defined without any restrictions on \mathbb{k} . Presenting this 2-groupoid as a monoidal category by generators and relations is surprisingly interesting (see [EW] and Section 4).

Because R is free as a left R^s -module, all Bott-Samelson bimodules will be free as left R -modules. Thus 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. In fact (under certain assumptions) Hom spaces are free as left (or right) R -modules, which is part of Soergel’s Categorification Theorem below.

Now we discuss the maps between Soergel bimodules and standard bimodules. There is an injection of bimodules $R(-1) \rightarrow B_s$ 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 $B_s \rightarrow R_s(1)$ sending $f \otimes g \mapsto fs(g)$. Conversely, there is a surjection $B_s \rightarrow R(1)$ 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 3.15. Just as Δ_s has a canonical description, so too does this element $\delta \otimes 1 - 1 \otimes \delta$. The trace map $\partial_s: R \rightarrow R^s$ 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:

$$(3.4) \quad 0 \rightarrow R(-1) \rightarrow B_s \rightarrow R_s(1) \rightarrow 0,$$

$$(3.5) \quad 0 \rightarrow R_s(-1) \rightarrow B_s \rightarrow R(1) \rightarrow 0.$$

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.

As we approach the statement of the Soergel Categorification Theorem, we assume that \mathfrak{h} is a realization over an infinite field \mathbb{k} of characteristic $\neq 2$, and that \mathfrak{h} is reflection faithful, as these are the conditions under which Soergel proved his theorems. As mentioned above, the results may hold in more generality.

Let \mathfrak{h} be such a 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 $\text{ch}(B)$ 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 3.16. *Let \mathfrak{h} be a reflection faithful realization of W over an infinite field of characteristic $\neq 2$. 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 $\text{End}_0(B_{\underline{w}})$, 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 \mathbb{k} 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 non-trivial summand B_w , while for a certain finite characteristic $B_{\underline{w}}$ is indecomposable, meaning that $B_w = B_{\underline{w}}$. This paper provides the tools needed to compute with the algebra $\text{End}_0(B_{\underline{w}})$, but so far we are only able to make very basic statements about its representation theory.

3.5. Categorification. We denote the split Grothendieck group of an additive category \mathcal{C} 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 $v[M] \stackrel{\text{def}}{=} [M(1)]$.

The following is *Soergel's Categorification Theorem* [Soe07, Satz 1.10 and Satz 5.15].

Theorem 3.17. *Let \mathfrak{h} be a reflection faithful realization of W over an infinite field of characteristic $\neq 2$. There is a unique isomorphism of $\mathbb{Z}[v^{\pm 1}]$ -algebras:*

$$\begin{aligned} \varepsilon : \mathbf{H} &\xrightarrow{\sim} [\mathbb{S}\text{Bim}] \\ \underline{H}_s &\mapsto [B_s]. \end{aligned}$$

Given Soergel bimodules B and B' the graded rank of $\text{Hom}(B, B')$ 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} as defined in Section 2.4.

We will refer to a reflection faithful realization of W over an infinite field of characteristic $\neq 2$ as a *realization for which Soergel's Categorification Theorem holds*. If Soergel's algebraic results are proven in more generality, then the corresponding statements in this paper will hold in that generality as well.

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

$$(3.6) \quad B_s \otimes B_s \cong B_s(1) \oplus B_s(-1).$$

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 [Elid]. 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 $\{\underline{H}_{\underline{w}} \mid w \in W\}$ 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 3.16 shows that $\{[B_{\underline{w}}] \mid w \in W\}$ is upper triangular in the basis $\{[B_w] \mid w \in W\}$ for $[\mathbb{S}\text{Bim}]$. 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

$$\text{ch} : [\mathbb{S}\text{Bim}] \rightarrow \mathbf{H}$$

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

Conjecture 3.18. *If \mathbb{k} is a field of characteristic 0, then $\text{ch}(B_w) = \underline{H}_w$.*

It is immediate that Soergel’s conjecture implies the Kazhdan-Lusztig positivity conjectures (see [KL79]). 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 [EW14] for realizations (not necessarily reflection faithful) defined over \mathbb{R} which satisfy a certain positivity property.

Remark 3.19. This statement of Soergel’s conjecture is known to fail when \mathbb{k} 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]).

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 2.10, any Hom space between Bott-Samelson bimodules can be understood after adjunction in terms of $\text{Hom}(R, B_w)$. Soergel’s Categorification Theorem combined with Corollary 2.8 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.

3.6. Localization. As \mathbb{k} is a domain, so is R . Let Q denote the fraction field of R , which is an (ungraded) field living over \mathbb{k} . One can extend any R -module to a Q -module in the usual way. We have seen above 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 3.20. *For any $\underline{w} = s_1 s_2 \dots s_d$ one has an isomorphism of right Q -modules*

$$B_{\underline{w}, Q} \stackrel{\text{def}}{=} B_{\underline{w}} \otimes_R Q = Q \otimes_{Q^{s_1}} Q \otimes_{Q^{s_2}} \dots \otimes_{Q^{s_d}} Q.$$

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 $i : R \otimes_{R^s} Q \hookrightarrow Q \otimes_{Q^s} Q$ and it is enough to show that this is surjective. However, any $0 \neq f \in R$ divides the s -invariant polynomial $fs(f)$ and in $Q \otimes_{Q^s} Q$ we have

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

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 $\mathbb{B}\text{SBim}_Q$ denote the essential image of $\mathbb{B}\text{SBim}$ inside the category of Q -bimodules. By the above lemma, it is the full subcategory of Q -bimodules

generated by the bimodules $B_{s,Q} \stackrel{\text{def}}{=} Q \otimes_{Q_s} Q$. Let $\mathbb{S}\text{Bim}_Q$ denote its Karoubi envelope. We define StdBim_Q similarly.

As Q -bimodules we have a splitting

$$(3.7) \quad B_{s,Q} \cong Q \oplus Q_s.$$

One way to observe this is that the sequences (3.4) and (3.5) are split. In fact, they “split each other”: the composition $R \rightarrow B_s \rightarrow R$ 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 (which we assume for the rest of this section) one has

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

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

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

$$\mathbb{S}\text{Bim} \rightarrow \mathbb{S}\text{Bim}_Q$$

as *localization*. It plays an important role in what follows, essentially because $\mathbb{S}\text{Bim}_Q$ 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 $\mathbb{S}\text{Bim}$ based on what their restrictions are to simple summands in $\mathbb{S}\text{Bim}_Q$ after localization. Given an explicit morphism f in $\mathbb{B}\mathbb{S}\text{Bim}$, we will be able to write down explicitly its restriction to each standard summand of the source and target in $\mathbb{S}\text{Bim}_Q$ and, in doing so, will be able to conclude whether or not $f = 0$.

Remark 3.21. 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.

When the realization is not faithful, localization (and other “geometric” techniques used by Soergel) will not be as useful. Thankfully, our diagrammatic approach to localization, begun in Section 4, will still function properly for the diagrammatic analog of Soergel bimodules.

Part 2. Diagrammatics

4. DIAGRAMMATICS FOR STANDARD BIMODULES

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

Fix (W, S) . In the paper [EW], we give a diagrammatic monoidal presentation for the 2-groupoid $\Omega_{\mathbb{k}}(W)$ of W over \mathbb{k} . We recall that definition, and add polynomials to obtain $\Omega_R(W)$, the 2-groupoid of W over R . When the realization is faithful, $\Omega_R(W)$ is equivalent to $StdBim$, and $\Omega_{\mathbb{k}}(W)$ is equivalent to $StdBim_0$. Note that the 2-groupoid, as an abstract monoidal category, depends on W but not on S ; it is determined by a group, not a group with presentation. However, our combinatorial description of $\Omega_R(W)$ requires a choice S of simple reflections.

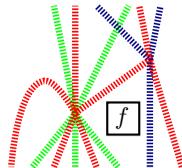
We introduce a general diagrammatic notion of a graph, which will be used for standard graphs in this section, and S -graphs in Section 5.

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 $\mathbb{R} \times [0, 1]$. All vertices of the graph lie on the interior of the strip. Edges may terminate at a vertex as usual in a graph; they may terminate on the boundary of the strip (such a place is called a *boundary point*, and is not counted as a vertex); or they may not terminate, instead forming a closed loop without any vertices. 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 closed loop, is an edge 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 $s \neq t \in S$ 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 S -graphs to be defined later. Here is an example where $S = \{r, b, g\}$ (for red, blue and green) and $m_{rg} = 5, m_{rb} = 3, m_{bg} = 2$:



Definition 4.3. Let \mathcal{D}^{std} denote the \mathbb{k} -linear additive graded monoidal category defined as follows. Objects are (direct sums of grading shifts of) sequences \underline{x} of elements of S , with monoidal structure given by concatenation. If we wish to emphasize that a sequence \underline{x} is being viewed as an object of \mathcal{D}^{std} , we may write $\underline{x}^{\text{std}}$ instead. 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 $\text{Hom}_{\mathcal{D}^{\text{std}}}(\underline{x}, \underline{y})$ will be the free \mathbb{k} -module generated by standard graphs

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:

$$(4.5) \quad (A_1 \times I_2(m)) \quad \text{[Diagram: A central vertex with } m \text{ edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]} = \text{[Diagram: A central vertex with } m \text{ edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]}$$

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

$$(4.6) \quad (A_1 \times A_1 \times A_1) \quad \text{[Diagram: A central vertex with three edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored red, green, and blue.]} = \text{[Diagram: A central vertex with three edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored red, green, and blue.]}$$

$$(4.7) \quad (A_3) \quad \text{[Diagram: A central vertex with six edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]} = \text{[Diagram: A central vertex with six edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]}$$

$$(4.8) \quad (B_3) \quad \text{[Diagram: A central vertex with six edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]} = \text{[Diagram: A central vertex with six edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]}$$

$$(4.9) \quad (H_3) \quad \text{[Diagram: A central vertex with six edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]} = \text{[Diagram: A central vertex with six edges radiating outwards, each edge having a small loop at its midpoint. The edges are colored in a repeating sequence of red, green, and blue.]}$$

This concludes the definition of \mathcal{D}^{std} .

Some of these relations, like (4.5) and (4.7), have been drawn with edges which do not terminate on the top or bottom boundary, but instead in “mid-air”. These are disk graphs.

Definition 4.5. Analogous to graphs with boundary on the planar strip, one can define a *disk graph with boundary*, which is an isotopy class of a graph with boundary, properly embedded in the planar disk. It obeys all the same rules as a graph with boundary on the planar strip, only now one cannot distinguish between a bottom or a top boundary, as the boundary lies on the circle.

Standard disk graphs are not actually morphisms in \mathcal{D}^{std} , as they have no source or target, but they are “local” portions of morphisms, and can be glued into larger graphs embedded in the planar strip. Relations in a monoidal category are local, so they can be stated using disk graphs rather than graphs on the strip. Henceforth,

all uses of disk graphs will be stated aloud; the reader should assume that any graph is, by default, a graph in the planar strip.

Definition 4.6. If we do not allow any grading shifts, and only allow \mathbb{k} -linear combinations of standard graphs without polynomials, we get a monoidal category $\mathcal{D}_0^{\text{std}}$ consisting entirely of degree 0 maps.

Remark 4.7. When working with $\mathcal{D}_0^{\text{std}}$, the base ring \mathbb{k} is entirely arbitrary. When defining \mathcal{D}^{std} one needs a realization of W to obtain the ring R but no additional restrictions on R are necessary. The action of W need not even be faithful.

Theorem 4.8. *The category $\mathcal{D}_0^{\text{std}}$ is equivalent to the 2-groupoid of W over \mathbb{k} , and the category \mathcal{D}^{std} is equivalent to the 2-groupoid of W over R . In other words, $\text{Hom}_{\mathcal{D}^{\text{std}}}(\underline{x}, \underline{y})$ is zero unless $x = y$, in which case it is R .*

This is the main result of [EW]; see Theorem 6.6. Using biadjunction and various isomorphisms, it reduces to the fact that standard diagrams with empty boundary are polynomial multiples of the empty diagram. The proof is 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 contractible space related to the Coxeter complex.

Definition 4.9. We define a functor $\mathcal{F}^{\text{std}}: \mathcal{D}^{\text{std}} \rightarrow \text{StdBim}$. Given an expression $\underline{x} = st\dots u$, \mathcal{F}^{std} sends the corresponding object to $R_{\underline{x}} = R_s \otimes R_t \otimes \dots \otimes R_u$. 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 $\mathcal{F}_0^{\text{std}}: \mathcal{D}_0^{\text{std}} \rightarrow \text{StdBim}_0$.

Claim 4.10. *The functors \mathcal{F}^{std} and $\mathcal{F}_0^{\text{std}}$ 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 StdBim_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}^{std} is well defined. □

Theorem 4.11. *The functors $\mathcal{F}_0^{\text{std}}$ and \mathcal{F}^{std} are faithful and essentially surjective. When the realization is faithful, they are equivalences.*

Proof. \mathcal{F}^{std} and $\mathcal{F}_0^{\text{std}}$ are obviously essentially surjective. Theorem 4.8 implies that $\mathcal{D}_0^{\text{std}}$ consists only of isomorphisms, which are sent to isomorphisms in StdBim_0 , so $\mathcal{F}_0^{\text{std}}$ is faithful. \mathcal{F}^{std} sends R isomorphically to R inside the endomorphisms of the empty sequence, so it acts faithfully on this endomorphism ring. Since any morphism space is related to the endomorphisms of the empty sequence by isomorphism and adjunction, \mathcal{F}^{std} is faithful.

Suppose that the realization is faithful. Whenever a Hom space in StdBim is non-zero, the source and the target sequence express the same element in W , and the Hom space is generated by the isomorphism which sends the 1-tensor to the 1-tensor. This isomorphism is clearly in the image of \mathcal{F}^{std} and $\mathcal{F}_0^{\text{std}}$, so they are full. □

When the realization is not faithful, \mathcal{F}^{std} is not full: there is an isomorphism $R_{\underline{w}} \cong R_{\underline{v}}$ whenever wv^{-1} is in the kernel of the action. In this case, \mathcal{D}^{std} will

be more useful to us than *StdBim*, because it really encodes all of W , and not just the quotient which acts faithfully. In general, our diagrammatics will encode only those bimodule morphisms which are generic, which depend on W but not on specific features of the realization.

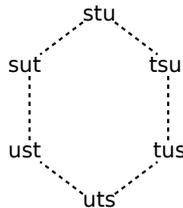
4.2. Rex moves and rex graphs. For each $w \in W$, let Γ_w denote the set of all reduced expressions for w . This can be given the structure of a connected graph, the *rex graph*, where two reduced expressions are connected by an edge if they differ by a single application of a braid relation (2.2). 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 $m_{st} = 2$, and those for which $m_{st} > 2$. Let $\bar{\Gamma}_w$ 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 , or its analog $\underline{x}^{\text{std}}$ in \mathcal{D}^{std} . 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 $2m_{st}$ -valent vertex, which is a morphism from $\underline{x}^{\text{std}}$ to $\underline{y}^{\text{std}}$. 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}^{std} essentially encodes a study of the graph of all expressions, reduced and non-reduced, for w ; this point is made clear in [EW], and may become more obvious after the remainder of this section.

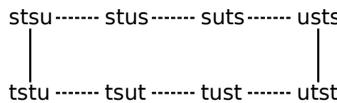
The relations of \mathcal{D}^{std} which do not involve cups and caps, namely (4.4) through (4.9), all come from loops (i.e., paths from a vertex to itself) in Γ_w . That is, each side of the relation comes from a path from \underline{x} to \underline{x}' in Γ_w , and the relation states that the corresponding morphisms are equal. Equivalently, once (4.4) is known, the relation states that the loop $\underline{x} \rightarrow \underline{x}' \rightarrow \underline{x}$ 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.

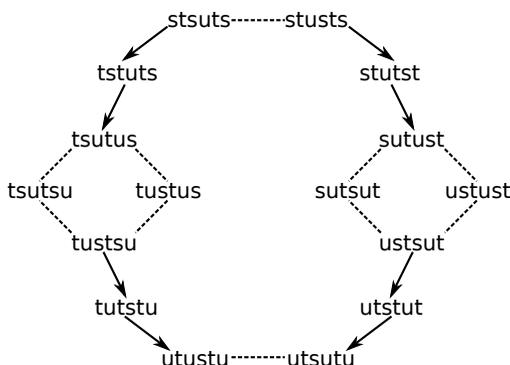
Example 4.12. We give two examples of type $A_1 \times I_2(m)$. The labeling of indices in S should be obvious. Type $A_1 \times A_1 \times A_1$:



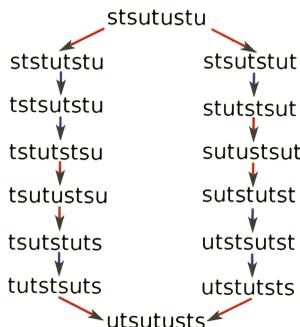
Type $A_1 \times A_2$:



Example 4.13. Here is Γ_{w_0} for type A_3 . The orientation on the arrows is the Manin-Schechtman orientation discussed in Section 1.4.3 of the introduction.

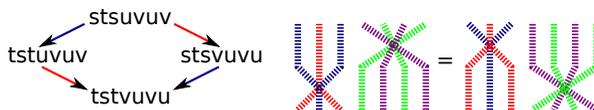


Example 4.14. Here is $\bar{\Gamma}_{w_0}$ for type B_3 . This can be deduced from (4.8), and we let the avid reader do the same for H_3 . The red and blue edges correspond to different parabolic subsets.



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}^{std} , this corresponds to the fact that distant pictures commute in a diagrammatic category. Disjoint squares need not involve disjoint colors.

Example 4.15. A disjoint square.



It is known [Ron09, Chapter 2, §5] that these loops generate the set of all loops in Γ_w , for any w in any Coxeter group W . That is, whenever W has a finite rank 3 parabolic subgroup J and a rex for w contains a rex for the longest element $w_J \in W_J$ as a subexpression, one will obtain an embedding of Γ_{w_J} into Γ_w , yielding a loop in Γ_w . Together with disjoint squares, these loops generate π_1 of the graph.



FIGURE 1. The vertices in an S -graph

This corresponds topologically to the fact that π_2 of the completed 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}^{std} . The theory of reduced expressions is enough to say that diagrams in \mathcal{D}^{std} without cups and caps form a one-dimensional space.

Theorem 4.8 is the correct generalization of this fact for graphs of non-reduced expressions.

5. DIAGRAMMATICS FOR SOERGEL BIMODULES

In this section 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 an adaptation of Soergel’s Categorification Theorem for \mathcal{D} , and a proof that \mathcal{F} is an equivalence when Soergel’s Categorification Theorem holds for Soergel bimodules. The proof will be given in later sections.

Fix a realization \mathfrak{h} of (W, S) over \mathbb{k} , as in Section 3.1, with S finite (though see Remark 5.8). As before, we let $R = \mathbb{k}[\mathfrak{h}]$ denote the coordinate ring of the reflection representation, α_s the simple root associated to $s \in S$, and ∂_s the Demazure operator. We let Q denote the field of fractions of R .

For reasons of simplicity, we will assume the realization is balanced in the first pass. Later, in Section 5.6, we treat the unbalanced case. For reasons to become clear in Section 5.2, we must always assume our realization is even-balanced.

5.1. Generators and relations.

Definition 5.1. An S -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 three types (see Figure 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.

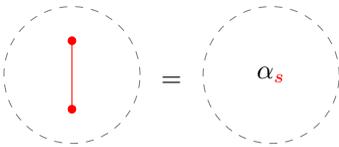
The *degree* of an S -graph is the sum of the degree of each vertex and the degree of each box.

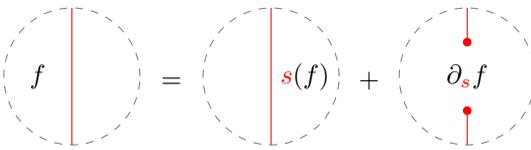
When there is no ambiguity we refer to an S -graph merely as a “graph”, even though it is an isotopy class of embedded graph.

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

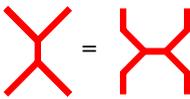
Definition 5.2. Let \mathcal{D} (or if there is ambiguity, \mathcal{D}_S) denote the \mathbb{k} -linear monoidal category defined as follows. Objects are sequences \underline{w} , with monoidal structure given by concatenation. If we wish to emphasize that a sequence \underline{x} is to be viewed as an element of \mathcal{D} , we may write $\underline{x}^{\mathcal{D}}$ instead. The space $\text{Hom}_{\mathcal{D}}(\underline{w}, \underline{y})$ is the free \mathbb{k} -module generated by S -graphs with bottom boundary \underline{w} and top boundary \underline{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:

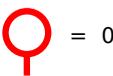
(5.1) 

(5.2) 

The one color relations:

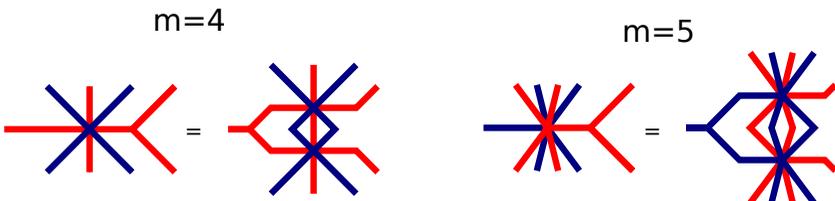
(5.3) 

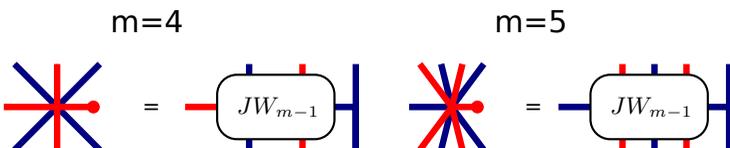
(5.4) 

(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. These relations are drawn using disk graphs.

(5.6) 

(5.7) 

In equation (5.7) above, the *Jones-Wenzl morphism* JW_{m-1} is a \mathbb{k} -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}^{std} earlier, with the exception of H_3 . Some of these relations are drawn using disk graphs.

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

$$(5.8) \quad (A_1 \times I_2(m)) \quad \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} = \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array}$$

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

$$(5.9) \quad (A_1 \times A_1 \times A_1) \quad \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} = \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array}$$

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

$$(5.10) \quad (A_3) \quad \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} = \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array}$$

$$(5.11) \quad (B_3) \quad \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} = \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array}$$

$$(5.12) \quad (H_3) \quad \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} - \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} = \text{lower terms}$$

This concludes the definition of \mathcal{D} .

Definition 5.3. Fix a reduced expression \underline{w} for w . Inside $\bigoplus_{\underline{x}} \text{Hom}_{\mathcal{D}}(\underline{x}, \underline{w})$, the right ideal $I_{\underline{w}}$ (bottom ideal, if you think diagrammatically) spanned by diagrams with a boundary dot on top is called the ideal of *lower terms*. Similarly, the left ideal inside $\bigoplus_{\underline{x}} \text{Hom}_{\mathcal{D}}(\underline{w}, \underline{x})$ spanned by diagrams with a bottom boundary dot is also called the ideal of lower terms.

There are various other descriptions of lower terms, which will be equivalent once the main results of this paper are proven: lower terms factor through shorter expressions; they have strictly negative-positive decompositions (see Section 7.1); they act by zero on the top summand after localization. A general discussion of lower terms will be given once the ideas have been developed. Whenever lower terms are discussed, an element $w \in W$ is understood.

Remark 5.4. The “lower terms” in (5.12) are lower terms in the above sense, and we also assume they have a strictly negative-positive decomposition. Exactly what the lower terms are in (5.12) is currently unknown. In Section 5.3 we will be defining a functor \mathcal{F} from \mathcal{D} to $\mathbb{S}\text{Bim}$, and in order for this functor to be well defined, an analogous relation must hold in $\mathbb{S}\text{Bim}$. Thus, if such a relation holds in $\mathbb{S}\text{Bim}$, then there is a unique way to define the lower terms in (5.12) such that \mathcal{F} is well defined. We discuss how the lower terms could be computed in Section 5.5.

If our realization is one for which Soergel’s Categorification Theorem holds for $\mathbb{S}\text{Bim}$, then one can deduce the existence of a relation of the form (5.12) in $\mathbb{S}\text{Bim}$. If not, we do not know whether a relation of the form (5.12) holds in $\mathbb{S}\text{Bim}$, though we conjecture that it does. As this paper is incomplete with respect to H_3 anyway, we will ignore the question of whether such a relation exists for degenerate realizations of Coxeter groups with parabolic subgroups of type H_3 .

Remark 5.5. As a monoidal category, \mathcal{D} can be defined with generators and relations. The generating morphisms are as listed in the introduction. As for \mathcal{D}^{std} in Section 4, one should include cups and caps as generators, and impose biduality and cyclicity relations, as is standard in diagrammatic categories (see, for instance, [Lau10]). However, relation (5.4) implies that cups and caps can be constructed out of dots and trivalent vertices, so that they are not needed as generators. When constructing a functor to $\mathbb{B}\mathbb{S}\text{Bim}$, we will also need to check these unwritten isotopy relations.

We now explore some basic properties of \mathcal{D} .

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. Eventually, Theorem 6.12 will imply 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*:

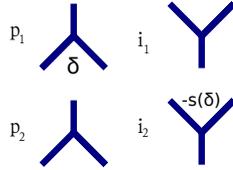
$$(5.13) \quad \boxed{f} \quad \Big| \quad = \quad \Big| \quad \boxed{f}$$

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. See, for instance, the discussion in [EK, Section 4.1].

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

$$(5.14) \quad ss \cong s(1) \oplus s(-1).$$

One simply needs to find two inclusion maps $i_1, i_2: s \rightarrow ss$ 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: $p_1 i_1 = p_2 i_2 = \mathbb{1}_s$, $p_1 i_2 = p_2 i_1 = 0$, and $\mathbb{1}_{ss} = i_1 p_1 + i_2 p_2$. One can choose these maps as follows:



This splitting is not canonical, as there may be multiple choices for δ . However, the space of maps in degree -1 is one-dimensional, so that i_1 and p_2 are canonical.

Remark 5.6. In fact, the choice of splitting is even more general than a choice of δ . For instance one may have

$$(5.15) \quad p_1 \quad f \quad + \quad \text{dot} \quad i_2 \quad g \quad + \quad \text{dot}$$

so long as $f + g = -\alpha_s$. Using (5.2), the example above has $f = \delta - \alpha_s$ and $g = -\delta$. 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 5.7. 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 Section 6 will no longer span all morphisms. Thus Demazure Surjectivity is still very important for proper behavior of the diagrammatic category.

Remark 5.8. Given an inclusion of Coxeter systems $S_1 \subset S_2$, there is a natural functor $\iota: \mathcal{D}_1 \rightarrow \mathcal{D}_2$ which sends an S -graph for S_1 to itself, viewed as an S -graph for S_2 (reinterpreting boxes via the map $R_1 \rightarrow R_2$). 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.

5.2. Jones-Wenzl morphisms. Presenting the Bott-Samelson category in rank 2 is the topic of [Elid]. 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 (no pun intended!), satisfying the properties stated in this section and the next, then the reader need not consult [Elid]. We assume the reader is familiar with quantum numbers and Temperley-Lieb algebras; background can be found in [Elid, Chapter 4].

The Temperley-Lieb algebra TL is typically defined over $\mathbb{Z}[q, q^{-1}]$, where the circle evaluates to $-[2] = -(q + q^{-1})$. It can be defined over the subring $\mathbb{Z}[q + q^{-1}] \subset \mathbb{Z}[q, q^{-1}]$ generated by $[2]$, a subring which contains all the quantum numbers. The effect of specializing q to a root of unity can be understood algebraically for the subring $\mathbb{Z}[q + q^{-1}]$ as well, even though this subring does not contain q . Precisely, q being a primitive $2m$ -th root of unity is equivalent to the statements that $[m] = 0$, $[k] \neq 0$ for $0 < k < m$, and $[m - 1] = 1$.

The Temperley-Lieb algebra with m strands may, after base change, contain an idempotent known as the *Jones-Wenzl projector* JW_m . Using diagrammatics for TL , the Jones-Wenzl projector can be described as a linear combination of crossingless matchings.

Example 5.9.

$$\begin{aligned}
 JW_1 &= \left| \right. \\
 JW_2 &= \left| \right| + \frac{1}{[2]} \begin{array}{c} \cup \\ \cap \end{array} \\
 JW_3 &= \left| \right| \left| \right| + \frac{[2]}{[3]} \begin{array}{c} \cup \\ \cap \end{array} + \frac{[2]}{[3]} \begin{array}{c} \cup \\ \cap \end{array} \\
 &\quad + \frac{1}{[3]} \begin{array}{c} \diagdown \\ \diagup \end{array} + \frac{1}{[3]} \begin{array}{c} \diagup \\ \diagdown \end{array}
 \end{aligned}$$

The Jones-Wenzl projector JW_m is only defined when certain quantum binomial coefficients are invertible (like the denominators above). However, special things happen for JW_{m-1} when q is specialized to a primitive $2m$ -th root of unity (in the sense above). In this case, JW_{m-1} is defined without any further specialization (no denominators are necessary), it is negligible (i.e., it is in the kernel of some pairing on the algebra), and it is rotation-invariant in a sense which will be described later in this section. For example, when $m = 4$, $[3] = 1$ and JW_3 has no denominators; “rotation” will swap the two diagrams with coefficient $[2]$, and will cycle the three diagrams with coefficient 1.

There is a variant on the Temperley-Lieb algebra described in [Elid, Section 4.3], namely the *two-colored Temperley-Lieb algebra* $2TL$, which is defined over the base ring $\mathbb{Z}[x, y]$. One should think of the elements x and y as two different “flavors” of the scalar $[2]$ from $\mathbb{Z}[q, q^{-1}]$. Any crossingless matching with m strands will divide the planar strip into $m + 1$ regions, which can be colored (say, red and blue) so that colors always alternate across a strand. Let us fix the color of the rightmost region to be blue; this will determine all the other colors in any crossingless matching. The two-colored Temperley-Lieb algebra is defined exactly as the Temperley-Lieb algebra except that, when resolving circles, a red circle surrounded by a blue region

evaluates to $-x$, while a blue circle surrounded by a red region evaluates to $-y$. Again, Jones-Wenzl projectors may exist after inverting some scalars in $\mathbb{Z}[x, y]$, and the first few examples are depicted below.

Example 5.10.

$$\begin{aligned}
 JW_1 &= \text{[Diagram: vertical strip with red and blue regions]} \\
 JW_2 &= \text{[Diagram: vertical strip with red, blue, red regions]} + \frac{1}{x} \text{[Diagram: blue disk with red boundary]} \\
 JW_3 &= \text{[Diagram: vertical strip with red, blue, red, blue regions]} + \frac{y}{xy-1} \text{[Diagram: blue disk with red boundary and red strip]} \\
 &\quad + \frac{1}{xy-1} \text{[Diagram: blue disk with red boundary and diagonal red region]} + \frac{x}{xy-1} \text{[Diagram: red disk with blue boundary]} \\
 &\quad + \frac{1}{xy-1} \text{[Diagram: red disk with blue boundary and diagonal blue region]}
 \end{aligned}$$

The coefficients appearing in Jones-Wenzl projectors are rational functions in the two-colored quantum numbers $[k]_x$ and $[k]_y$, which were defined in Definition 3.6. Note that $2TL$ can be defined in two “flavors”, the other one having red as the color of the rightmost region. To obtain the Jones-Wenzl projectors for the other flavor from those pictured above, swap the colors and swap x and y .

One can ask for the specialization of $\mathbb{Z}[x, y]$ which corresponds philosophically to specializing q to a primitive $2m$ -th root of unity, and there are several choices one might make. The weakest choice is to set $[m]_x = [m]_y = 0$, which will be sufficient to imply that both flavors of JW_{m-1} are defined without further specialization. Additional assumptions are required for JW_{m-1} to be negligible and to have certain rotational properties, and these will be discussed soon.

Choose two indices $s, t \in S$ to correspond to red and blue; let $m = m_{st}$. Let us change base to $2TL \otimes_{\mathbb{Z}[x, y]} \mathbb{k}$ via the map $x \mapsto a_{st}$ and $y \mapsto a_{ts}$. By the “technical condition” (3.3), $[m]_x = [m]_y = 0$ and the Jones-Wenzl projector JW_{m-1} is defined over \mathbb{k} . Given a two-colored crossingless matching, one can obtain an S -graph of degree 0 by deformation retracting each colored region into a tree composed out of trivalent and univalent vertices, and coloring the tree appropriately. We call this *taking the retract on the strip*. For example, JW_1 would become the identity map of st . This operation extends to an algebra homomorphism from $2TL$ to the degree zero part of an endomorphism ring inside \mathcal{D} . For example, the fact that a red circle in a blue region evaluates to $-x$ can be shown using (5.2), (5.1), and (5.5) until the scalar $\partial_t(\alpha_s) = a_{st}$ appears. What happens to Jones-Wenzl projectors can be seen in [Elid, Example 5.16].

In this paper, we will use a deformation retract in a slightly different way, to take a two-colored crossingless matching and obtain an S -graph of degree $+2$ on the planar disk, as in Definition 4.5. Where before the leftmost region of a crossingless matching gave rise to two boundary points in the deformation retract (one on bottom and one on top), now they consolidate into one boundary point on the side of the disk. We call this *taking the retract on the disk*. For example, here is what happens to the Jones-Wenzl projectors.

Example 5.11.

$$\begin{aligned}
 JW_1 &= \text{red dot} \text{---} \text{blue dot} & JW_2 &= \text{blue dot} \text{---} \text{blue dot} + \frac{1}{x} \text{red dot} \\
 JW_3 &= \text{red dot} \text{---} \text{blue dot} + \frac{y}{xy-1} \text{trivalent vertex} + \frac{x}{xy-1} \text{trivalent vertex} \\
 &+ \frac{1}{xy-1} \text{trivalent vertex} + \frac{1}{xy-1} \text{trivalent vertex}
 \end{aligned}$$

A black box labeled JW within an S -graph will always represent JW_{m-1} , viewed as a degree +2 linear combination of S -graphs on the disk, via the implicit choice of coloring. We call it the JW -box. An example appears in the relation (5.7).

An implication of (5.7) and (5.6) is the following relation (see [Elid, Claim 6.17]):

$$(5.16) \quad \begin{matrix} m=4 \\ \text{2m-valent vertex} \end{matrix} = \begin{matrix} \text{JW box} \\ \text{with 4 lines} \end{matrix} \quad \begin{matrix} m=5 \\ \text{2m-valent vertex} \end{matrix} = \begin{matrix} \text{JW box} \\ \text{with 5 lines} \end{matrix}$$

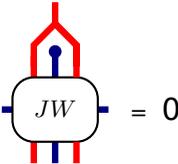
The right-hand side of (5.16) is the degree 0 S -graph which would be obtained by taking the retract on the strip. In general, the retract on the strip can be obtained from the retract on the disk by composing with trivalent vertices on either side, as in the right-hand side of (5.16). Because the retract on the strip is a homomorphism, the right-hand side of (5.16) is an idempotent in \mathcal{D} , so that the $2m$ -valent vertex is “half an idempotent”.

Consider the boundary of the retract on the disk of a two-colored crossingless matching. It alternates between red and blue, and thus has two symmetries: color-preserving rotation (rotation by “two notches”, which sends blue boundary points to blue boundary points), and color-reversing rotation (rotation by “one notch”, which sends blue boundary points to red ones) followed by a color swap. One might ask when the JW -box is preserved by these symmetries. This depends on whether $m = m_{st}$ is even or odd. If m is even, then the additional condition $[m-1] = 1$ is required in order that JW_{m-1} be fixed by color-preserving rotation, and will also suffice for it to be fixed by color-reversing rotation followed by a color swap. If m is odd, then JW_{m-1} is always fixed by color-preserving rotation, and will be fixed by color-reversing rotation followed by a color swap precisely when $[m-1]_x = [m-1]_y = 1$. Using the examples above, the reader can convince themselves of these rotational facts for $m \leq 4$. A proof and further discussion can be found in [Elid, Sections 4.2 and 4.3].

The $2m$ -valent vertex in \mathcal{D} is cyclic (because we have defined morphisms to be isotopy invariant). However, it turns out that this is inconsistent unless the JW -box is fixed by color-preserving rotation (more precisely, the $2m$ -valent vertex would be equal to a non-trivial multiple of itself, so it must vanish). Thus, we cannot define our category \mathcal{D} as above when the realization is even-unbalanced (see Definition 3.7), and hope for any reasonable behavior! When the realization is odd-unbalanced, \mathcal{D} is well behaved, but the failure to be fixed by color-reversing

rotation followed by color swap will lead to lack of symmetry between the two flavors of JW -boxes. The best way to deal with this is using some additional bookkeeping, which we describe in Section 5.6. Until then, let us assume for simplicity that the realization is balanced.

The defining properties of the Jones-Wenzl projectors in TL correspond to the following two properties of the JW -box. The first is that the coefficient of a certain graph is 1, as can be seen from the examples above. This is the S -graph on the disk which corresponds to the identity map on the planar strip. The second property is called *death by pitchfork*:

(5.17)  = 0

Remark 5.12. The usual Temperley-Lieb category describes morphisms between iterated tensor products of the standard representation V_1 of quantum \mathfrak{sl}_2 . If one changes base to a ground ring where the quantum numbers $[k]$ are invertible for $0 < k < m$, one can define (as linear combinations of crossingless matchings) the idempotents which project to each summand in the direct sum decomposition

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

where V_d represents the irreducible representation of highest weight d . After applying the retract on the strip, one obtains idempotents in \mathcal{D} , which can be used to categorify (2.8); see [Elid, Section 5.4.2].

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, the death by pitchfork relation will eventually be used to show that morphism spaces from the image of the idempotent in (5.16) have the “correct” size, and this will be used to categorify the dihedral relations of the Hecke algebra in Section 6.

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 5.13. Let \mathcal{F} be the \mathbb{k} -linear monoidal functor from \mathcal{D} to $\mathbb{B}SBim$ defined as follows. The object $\underline{x}^{\mathcal{D}}$ 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 [Elid, Section 6.2.2], though in a convoluted and not particularly useful form.

Claim 5.14. *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 two colors, and the dihedral case was checked in [Elid]. 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 computer check used a localization technique, which we will discuss in Section 5.5. The case of H_3 can be checked by computer, once the appropriate relation is found (with the caveat of Remark 5.4). \square

5.4. Localization.

Definition 5.15. 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 $\mathcal{D}_Q \rightarrow \mathbb{B}\text{SBim}_Q$ which extends \mathcal{F} under base change.

Because of arguments akin to Lemma 3.20, 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 .

Recall that the bimodule B_s is equipped with short exact sequences (3.4) and (3.5) which split after localization to Q . Each short exact sequence “splits” the other, in the sense that following one map from (3.4) and then one map from (3.5) gives an endomorphism of Q or of Q_s which is multiplication by α_s , and is therefore invertible. We have already seen two of the maps: 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:

$$(5.18) \quad \begin{array}{c} \text{||||} \\ | \\ \text{||||} \end{array} f \otimes g \mapsto f \otimes g = fs(g) \otimes 1 \quad \begin{array}{c} \text{||||} \\ | \\ \text{||||} \end{array} 1 \otimes 1 \mapsto \delta \otimes 1 - 1 \otimes \delta$$

Definition 5.16. Let $\text{Kar}(\mathcal{D}_Q)$ (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 $\text{Kar}(\mathcal{D}_Q)$ will have objects which are sequences of normal indices and reflection indices. We typically disambiguate by abusing notation and writing $\underline{x}^{\mathcal{D}}$ for a sequence of normal indices and $\underline{x}^{\text{std}}$ 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:

$$(5.19) \quad \begin{array}{c} \text{||||} \\ | \\ \text{||||} \end{array} = \begin{array}{c} \bullet \\ | \\ \bullet \\ | \\ \text{||||} \end{array}$$

$$(5.20) \quad \begin{array}{c} \text{||||} \\ | \\ \text{||||} \end{array} = \begin{array}{c} \bullet \\ | \\ \bullet \\ | \\ \text{||||} \end{array} - \begin{array}{c} \bullet \\ | \\ \bullet \\ | \\ \text{||||} \end{array}$$

$$(5.21) \quad \begin{array}{c} \text{---} \\ \text{---} \end{array} \text{---} \text{---} \text{---} \text{---} \text{---} = \begin{array}{c} | \\ | \\ | \end{array} = \begin{array}{c} \text{---} \\ \text{---} \end{array} \text{---} \text{---} \text{---} \text{---} \text{---}$$

$$(5.22) \quad \begin{array}{c} \text{---} \\ \text{---} \end{array} \text{---} \text{---} = \text{---} \text{---} \text{---} \quad \begin{array}{c} \text{---} \\ \text{---} \end{array} \text{---} \text{---} = \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} = \begin{array}{c} \text{---} \\ \text{---} \end{array} \text{---} \text{---} \quad \text{---} \text{---} \text{---} = \text{---} \text{---} \text{---}$$

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 $\text{Kar}(\mathcal{D}_Q)$ further, and prove the following theorem.

Theorem 5.17. *Assume Demazure Surjectivity. The diagrammatic category temporarily denoted $\text{Kar}(\mathcal{D}_Q)$ above is equivalent to the Karoubi envelope of \mathcal{D}_Q . It is also equivalent to the base change $\mathcal{D}_Q^{\text{std}} = \mathcal{D}^{\text{std}} \otimes_R Q$.*

Unfortunately, $\text{Kar}(\mathcal{D}_Q)$ is not cyclic! The bivalent vertex 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 5.18. 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 semi-simple perverse sheaves on the flag variety, and taking global sections is a well-behaved functor on this semi-simple category, so that cyclicity happens to be preserved. However, taking global sections of non-semi-simple sheaves tends to forget structure, and break the compatibility with the biadjunction functor. If one more appropriately models non-semi-simple perverse sheaves as complexes of semi-simple 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 semi-simple perverse sheaves, and these descend to (3.4) and (3.5). 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 $\text{Kar}(\mathcal{D}_Q)$ up to sign. A mixed graph has some solid edges and some dashed edges. A mixed graph without any dashed edges is an S -graph or *solid graph*, and a mixed graph without any solid edges is a standard graph or *dashed graph*. At the moment, there is no notion of a dashed $2m$ -valent vertex, but

we will define such a morphism 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 :

$$(5.23) \quad \begin{array}{c} \bullet \\ | \\ \text{---} \\ | \\ \bullet \end{array} = 0 \quad \begin{array}{c} | \\ | \\ | \\ | \\ \bullet \end{array} = 0$$

Remark 5.19. 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 $\frac{1}{\alpha_s} \begin{array}{c} | \\ | \\ | \\ | \\ \text{---} \\ | \\ | \\ | \\ | \\ \bullet \end{array}$, which is complementary to the idempotent $\frac{1}{\alpha_s} \begin{array}{c} \bullet \\ | \\ \text{---} \\ | \\ \bullet \end{array}$ which already existed in \mathcal{D}_Q . Therefore, $s^{\mathcal{D}} \cong s^{\text{std}}(1) \oplus \mathbb{1}(1)$. (We include the gradings for those who wish to use the homogeneous fraction field rather than the full fraction field. Remember that $\mathbb{1} \cong \mathbb{1}(2)$ so that the grading effectively 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.

$$(5.24) \quad \begin{array}{c} | \\ | \\ | \\ | \\ | \\ | \\ | \\ | \\ | \\ | \end{array} = \frac{1}{\alpha_s} \begin{array}{c} | \\ | \\ | \\ | \\ \text{---} \\ | \\ | \\ | \\ | \\ \bullet \end{array} + \frac{1}{\alpha_s} \begin{array}{c} \bullet \\ | \\ \text{---} \\ | \\ \bullet \end{array}$$

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

Claim 5.20. *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. \square

Claim 5.21. *The dashed cups and caps are redundant, being equal to the following maps. The only relations necessary are (5.19) and (5.20).*

$$(5.25) \quad \begin{array}{c} \text{---} \\ \cup \end{array} = \frac{1}{\alpha_s} \begin{array}{c} \cup \end{array} \quad \begin{array}{c} \cup \\ \text{---} \end{array} = \frac{1}{\alpha_s} \begin{array}{c} \cup \end{array}$$

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:

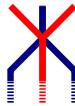
$$\begin{array}{c} \text{---} \\ \cup \end{array} = \frac{1}{\alpha_s} \begin{array}{c} \cup \end{array} = \frac{1}{\alpha_s} \begin{array}{c} \cup \\ | \\ \bullet \end{array} - \frac{1}{\alpha_s} \begin{array}{c} \cup \\ | \\ \bullet \end{array} \xrightarrow{0} = \frac{1}{\alpha_s} \begin{array}{c} \cup \\ | \\ \bullet \end{array} = \begin{array}{c} \cup \end{array}$$

\square

Fix a category \mathcal{C} specified with generators and relations, an object $M \in \mathcal{C}$, and an idempotent $e \in \text{End}(M)$. 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 $\iota: X \rightarrow M$ and a map $\pi: M \rightarrow X$. One adds two new relations, which state that $\iota\pi = e \in \text{End}(M)$ and $\pi\iota = \mathbb{1} \in \text{End}(X)$. This presentation clearly gives a category $\mathcal{C}(M, e)$ equipped with a fully faithful map $\mathcal{C} \rightarrow \mathcal{C}(M, e)$ 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 $\text{Kar}(\mathcal{D}_Q)$ is the category obtained from \mathcal{D}_Q by adjoining the complements of $\mathbb{1}$ in $s^{\mathcal{D}}$ 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 $\text{Kar}(\mathcal{D}_Q)$ is the Karoubi envelope of \mathcal{D}_Q , it is enough to show that $\text{Kar}(\mathcal{D}_Q)$ is idempotent complete. Because each $s^{\mathcal{D}}$ decomposes into s^{std} 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 $\mathcal{D}_Q^{\text{std}}$.

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 [Elid]). 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:

$$(5.26) \quad \text{Diagram} = (-1)^m \text{Diagram}$$

Therefore:

$$(5.27) \quad \text{Diagram} = \frac{1}{\rho} \text{Diagram}$$

Thus the right-hand side 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 5.22. Let Std 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 $\underline{x}^{\text{std}}$ 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 discussion we have:

$$(5.28) \quad \text{Diagram} = \text{Diagram}$$

Proposition 5.23. *The functor Std is well defined, and is an equivalence of categories.*

This proposition implies Theorem 5.17. We have already showed that Std is essentially surjective, because $\text{Kar}(\mathcal{D}_Q)$ is additively generated by reflection indices. We need to check that Std is well defined, full and faithful.

Proof that Std is well defined. We have already checked the isotopy relations in $Std\text{Bim}$, 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):

$$\begin{aligned}
 \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} &= \frac{1}{\alpha_s} \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} = \frac{1}{2} \frac{1}{\alpha_s^2} \left(\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \right) = -\frac{1}{2} \frac{1}{\alpha_s^2} \left(\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \right) \\
 &= -\frac{1}{\alpha_s^2} \begin{array}{c} \text{---} \\ \text{---} \end{array} = -\frac{1}{\alpha_s^2} \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array}
 \end{aligned}$$

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}^{std} , with bivalent vertices on top. Bivalent vertices are invertible, so this checks the relation. \square

Now we can apply any relations in \mathcal{D}^{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

$$\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \end{array}$$

Therefore

$$(5.29) \quad \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} = 0$$

Lemma 5.24. *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) We can assume that any two regions (i.e., connected components of the complement of the graph) can be connected by a path which only crosses dashed lines. This can be done by taking a solid line which separates two regions 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 cannot form a circle, for this would violate the first step). 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 immediate from (1) and the sliding relation (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}^{std} reduce to the empty diagram.

□

Corollary 5.25. *Std is full.*

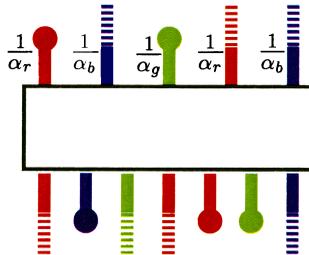
Proof. Apply the reduction of the lemma to a map between reflection indices. There can be no boundary dots. A solid edge cannot connect to anything except a bivalent vertex any more. Thus every solid edge can be removed with (5.19) (at the cost of adding a polynomial). Any map with only dashed edges is clearly in the image of *Std*. □

We can define a functor $\text{Kar}(\mathcal{D}_Q) \rightarrow \text{StdBim}$ extending \mathcal{F} . This functor acts on bivalent vertices as in (5.18). Clearly this functor intertwines *Std* and \mathcal{F}_{std} . Since \mathcal{F}_{std} is faithful, this would imply that *Std* is faithful, and is therefore an equivalence. This concludes the proof of Proposition 5.23 and Theorem 5.17.

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

5.5. Computation using localization. Each object $\underline{x}^{\mathcal{D}}$ in \mathcal{D} splits in $\text{Kar}(\mathcal{D}_Q)$ as the sum of $Q_{\mathbf{e}}$ over all subsequences \mathbf{e} of \underline{x} . Here, $Q_{\mathbf{e}}$ denotes the sequence of reflection indices which is the tensor product of s_k^{std} when $\mathbf{e}_k = 1$ and $\mathbb{1}$ when $\mathbf{e}_k = 0$. Our notation $Q_{\mathbf{e}}$ is slightly abusive, confusing an object in $\text{Kar}(\mathcal{D}_Q) \cong \mathcal{D}_Q^{\text{std}}$

with a localized standard bimodule, but the alternative notations were not as clear. Similarly, we may write $Q_{\mathbf{e}} \cong Q_w$ when $(\underline{x}, \mathbf{e})$ expresses w , to help indicate precisely when these objects are isomorphic. Given a graph depicting a morphism from a sequence \underline{x}^D of length d to a sequence $\underline{x}'^{D'}$ of length d' , we can localize to obtain a $2^{d'} \times 2^d$ matrix of maps between reflection sequences. Since Hom spaces between objects in $\mathcal{D}_Q^{\text{std}}$ 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 $\text{Kar}(\mathcal{D}_Q)$ 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 $Q_{\mathbf{e}} \rightarrow Q_{\mathbf{e}'}$ 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 S -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 $\text{Kar}(\mathcal{D}_Q)$.

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.

It will be proven later in this paper that the passage $\mathcal{D} \rightarrow \mathcal{D}_Q$ is faithful, and hence so is the passage $\mathcal{D} \rightarrow \text{Kar}(\mathcal{D}_Q)$. Not knowing this, the equality of two matrices only implies that the original maps are equal modulo R -torsion. However, localization 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 S -graphs have the same image under \mathcal{F} by localizing $\mathcal{D} \rightarrow \text{Kar}(\mathcal{D}_Q)$ 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 non-zero 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 5.26. *Std is faithful.*

Proof. In fact, the techniques we have developed allow us to construct a quasi-inverse for *Std*. We construct a functor \mathcal{G} from $\text{Kar}(\mathcal{D}_Q)$ to $\mathcal{D}_Q^{\text{std}}$ 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 *Std*. \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 $\mathcal{D}_Q^{\text{std}}$ corresponding to a solid graph.

This proof that $\text{Kar}(\mathcal{D}_Q) \cong \mathcal{D}_Q^{\text{std}}$ was entirely diagrammatic, and no mention of bimodules was required. In particular, it still works when the realization is not faithful; the “localization” of \mathcal{D} lives in the 2-groupoid of W regardless. As a result, we have the following proposition.³

Proposition 5.27. *Assume Demazure Surjectivity, so that the functor $\mathcal{F}: \mathcal{D} \rightarrow \mathbb{S}\text{Bim}$ is well defined. Then \mathcal{F} is faithful.*

Proof. The composition of functors $\mathcal{D} \rightarrow \mathbb{S}\text{Bim} \rightarrow \mathbb{S}\text{Bim}_Q \cong \text{StdBim}_Q$, which applies \mathcal{F} and then localizes, is isomorphic to the composition $\mathcal{D} \rightarrow \mathcal{D}_Q \rightarrow \text{StdBim}_Q$, which applies formal diagrammatic localization *Std* and then the functor \mathcal{F}^{std} from Section 4. Both *Std* and \mathcal{F}^{std} are faithful, and so is their composition. Therefore \mathcal{F} is faithful. \square

Thus faithfulness of \mathcal{F} holds even for non-faithful realizations, where Soergel bimodules are very poorly understood. Note that this proof does not assume that localization for Soergel bimodules $\mathbb{S}\text{Bim} \rightarrow \mathbb{S}\text{Bim}_Q$ is faithful; we expect this to be true in broad generality, but it is not in the literature.

5.6. Unbalanced realizations. Now we discuss the diagrammatic alterations which must be made to accommodate the case of non-balanced realizations. This was discussed for dihedral groups in the appendix to [Elid], where two separate diagrammatic conventions are proposed to deal with the new bookkeeping required. We follow the second suggested convention from that appendix.

³Thanks to Macky Shotaro for this question.

Fix a dihedral parabolic subgroup with $m = m_{st} < \infty$. Relation (5.7) implies that there is a close connection between $2m$ -valent vertices and two-colored Jones-Wenzl projectors. However, this begs the question: which Jones-Wenzl projector? After all, the left-blue-aligned Jones-Wenzl projector is not equal to the rotation of the left-red-aligned one. A careful examination of (5.7) and (5.16) shows that different choices of Jones-Wenzl projector must be made, depending on the location of the dot and the orientation of the $2m$ -valent vertex.

If some rescaling of (5.7) is to hold for any positioning of the dot and the $2m$ -valent vertex, then the Jones-Wenzl projector must satisfy the death by pitchfork property (5.17). In particular, JW_{m-1} must be negligible, and must have some rotational eigenvalue. This is not possible when the realization is even-unbalanced (see [Elid] for more details). While one may be able to design a diagrammatic calculus for even-unbalanced realizations, we will not attempt to do so. When m is even and $[m - 1] = 1$, all the two color relations above hold as stated.

The $2m$ -valent vertex is supposed to correspond, under the functor \mathcal{F} , to some non-zero morphism between bimodules, living in a one-dimensional space of morphisms. Let $\underline{w}_{(s)}$ denote the reduced expression $\dots tsts$ of length m ending in s , and let $\underline{w}_{(t)}$ denote the reduced expression $\dots stst$ of length m ending in t . There is a unique bimodule map $f_{(s)}: BS(\underline{w}_{(t)}) \rightarrow BS(\underline{w}_{(s)})$ which sends the 1-tensor to the 1-tensor, and a unique map $f_{(t)}: BS(\underline{w}_{(s)}) \rightarrow BS(\underline{w}_{(t)})$ which does the same. However, when the realization is not balanced, these maps are not rotations of each other (by one strand), and this is the underlying issue. However, these maps are individually invariant under color-preserving rotations (just like JW_{m-1} , even in the odd-unbalanced case), so we may draw each unambiguously as some kind of $2m$ -valent vertex. We label the vertices (s) or (t) to distinguish the two. (We only draw the case when m is odd, but nothing prevents drawing the even case too.)



The purple strand is meant to encode the appropriate sequence of alternating red and blue strands. To reiterate, when the blue-centered $2m$ -valent vertex is oriented such that its upper-right strand is blue (as in the picture above), then it corresponds under \mathcal{F} to a morphism which preserves the 1-tensor. When the blue-centered $2m$ -valent vertex is oriented such that its upper-right strand is red, one differs from this map by an invertible scalar λ . In [Elid], this scalar $\lambda = [m - 1]_y$ is discussed at additional length. Therefore, one has the rotational relation:

$$(5.30) \quad \begin{array}{c} \text{Diagram with blue center and blue upper-right strand} \\ \text{Diagram with red center and red upper-right strand} \end{array} = \lambda \quad \begin{array}{c} \text{Diagram with red center and red upper-right strand} \\ \text{Diagram with blue center and blue upper-right strand} \end{array}$$

When we place a dot on a $2m$ -valent vertex, we obtain a relation similar to (5.7).

(5.31)

In each case, the version of the Jones-Wenzl projector used is the one whose coefficient of the identity is 1, when color-aligned as above. By rotating these pictures, one knows how to place a dot on any strand in either $2m$ -valent vertex. To check that these relations make sense, observe that both sides send the 1-tensor to the 1-tensor.

To give the correct versions of (5.6) and (5.16), one should label the $2m$ -valent vertices in such a way that both sides preserve the 1-tensor:

(5.32)

(5.33)

There are additional, color-switched versions of each of these relations.

Finally, to give the correct version of the three color relations, one should again ensure that both sides preserve the 1-tensor. Here is the A_3 relation:

(5.34)

We will use the term *rex move* to refer to an S -graph built from $2m$ -valent vertices, yielding a morphism of degree zero between reduced expressions. When working with unbalanced diagrams, there is a scalar ambiguity that appears when defining a rex move, determined by the choice of central color on each $2m$ -valent vertex. Our convention is that one will always choose a coloring so that the 1-tensor is preserved by the rex move.

For the remainder of this paper, we will work solely with the balanced case, and thus will not need the extra confusion of labeled $2m$ -valent vertices. In the rest of the paper, it is roughly the case that diagrams are only important “up to

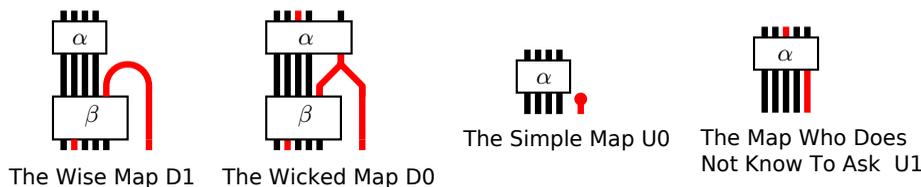
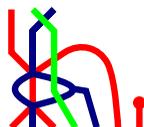


FIGURE 2. The four maps

- 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 $LL_{\underline{w}, \mathbf{e}}$ is $+1$ for each $U0$ and -1 for each $D0$ and hence agrees with the 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 $LL_{\underline{x}, \mathbf{e}}$ for $\underline{x} = rbrgbr$ 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 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 Section 7) we will have to show how one set of choices “spans” all the other possible choices.

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

Now suppose that $(\underline{x}, \mathbf{e})$ has length k , and the first $k - 1$ terms are all $U1$, thus forming a rex for w_{k-1} . In the construction of $LL_{\underline{x}, \mathbf{e}}$ we may choose LL_{k-1} to be the identity map, meaning that $LL_{\underline{x}, \mathbf{e}} = \phi_k$.

This observation leads to another inductive perspective on light leaves:

$$(6.2) \quad \begin{array}{c} \text{---} \quad \text{---} \\ | \quad | \\ \boxed{LL_{k+n}} \\ | \quad | \quad | \quad | \quad | \\ \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array} = \begin{array}{c} \boxed{LL_{n+l(w_k)}} \\ | \quad | \\ \boxed{1} \quad \boxed{1} \\ | \quad | \\ \boxed{LL_k} \\ | \quad | \quad | \quad | \\ \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array}$$

Let $(\underline{x}, \mathbf{e})$ have length $k + n$ for $n \geq 0$. Having performed $LL_{\underline{x}_{\leq k}, \mathbf{e}_{\leq k}}$, the output is a reduced expression \underline{w}_k . What happens in steps $k + 1$ through $k + n$ can be viewed as a light leaf for a different sequence $(\underline{y}, \mathbf{f})$. Here $\underline{y} = \underline{w}_k \underline{x}_{>k}$ is the concatenation of \underline{w}_k and the remainder of \underline{x} , and has length $\ell(w_k) + n$; $\mathbf{f} = (1, 1, \dots, 1, \mathbf{e}_{\geq k})$ has all $U1$ on the \underline{w}_k portion of \underline{y} .

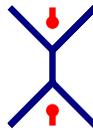
We will write $LL_{\underline{x}}^w$ to indicate a set consisting of one fixed choice of $LL_{\underline{x}, \mathbf{e}}$ for each subexpression \mathbf{e} of \underline{x} expressing w . When we speak of the “span” of $LL_{\underline{x}}^w$, we mean all morphisms obtained as linear combinations of LL maps, with polynomials appearing in the leftmost 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:

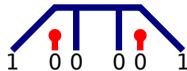
$$\begin{array}{c} \text{---} \quad \text{---} \\ | \quad | \\ \text{---} \quad \text{---} \end{array} = \begin{array}{c} \text{---} \quad \text{---} \\ | \quad | \\ \text{---} \quad \text{---} \end{array} + \begin{array}{c} \text{---} \quad \text{---} \\ | \quad | \\ \text{---} \quad \text{---} \end{array} - \begin{array}{c} \text{---} \quad \text{---} \\ | \quad | \\ \text{---} \quad \text{---} \end{array}$$

0 1 1 1 1 0 1 0 1

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 $LL_{\underline{x}, \mathbf{e}}$ should genuinely form a basis for $\text{Hom}_{\mathcal{D}}(\underline{x}, 1)$.

6.2. Localizing light leaves. Let us fix a light leaves map $LL_{\underline{x}, \mathbf{e}}$, with \mathbf{e} expressing w . Let \underline{w} be the target of the map. Now consider what happens after the passage $\mathcal{D} \rightarrow \text{Kar}(\mathcal{D}_Q)$. The target \underline{w} splits up into summands $Q_{\mathbf{h}}$ for $\mathbf{h} \subset \underline{w}$, with a unique summand isomorphic to Q_w , namely $\underline{w}^{\text{std}}$. To distinguish this from other copies of Q_w , we write it as $Q_{\underline{w}}$, as one would in StdBim . For each subsequence \mathbf{f} of \underline{x} which also expresses w we have a summand $Q_{\mathbf{f}} \subset \underline{x}$ isomorphic to Q_w , and this is sent by $LL_{\underline{x}, \mathbf{e}}$ into the unique summand $Q_{\underline{w}} \subset \underline{w}$ 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 $LL_{\underline{x}, \mathbf{e}}$.

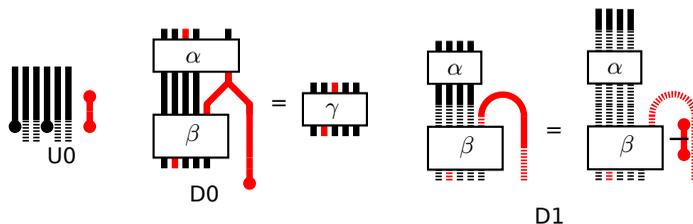


FIGURE 3. Calculating $p_{\mathbf{e}}^{\mathbf{e}}$

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 (this was defined at the end of Section 2.4). 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 $(\underline{x}_{\leq k}, \mathbf{e}_{\leq k})$. We write v_k for the corresponding element with \mathbf{e} replaced by \mathbf{f} .

Remember that $LL_{\underline{x}, \mathbf{e}}$ is defined inductively, beginning with $LL_k = LL_{\underline{x}_{\leq k}, \mathbf{e}_{\leq k}} \otimes \mathbb{1}$. 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 $v_k \leq w_k$. Thus $LL_{\underline{x}, \mathbf{e}}$ will act as zero on $Q_{\mathbf{f}}$ unless $v_k \leq w_k$ for every k , which is exactly the condition for $\mathbf{f} \leq \mathbf{e}$ 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 $k \leq \ell(\underline{x})$ define a polynomial α_k as follows:

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

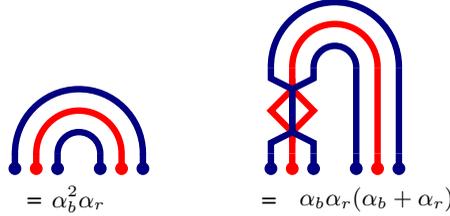
Note that $w_{k-1}(\alpha_{s_k})$ is the coefficient one obtains if one places α_{s_k} in the region just to the right of \underline{w}_k , as an element of $Q_{\mathbf{e}}$, and pulls it to the far left region. We claim that $p_{\mathbf{e}}^{\mathbf{e}} = \prod_{k=1}^{\ell(\underline{x})} \alpha_k$.

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 3). 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 $\mathbb{1}$ 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 3.

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 $\prod_{k=1}^{\ell(\underline{w})} \alpha_k$. \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 $\mathbf{f} = (0, 0, 0, 0, 0, 0)$:



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 $LL_{\underline{x}}^w$ be a set consisting of one light leaves map $LL_{\underline{x},\mathbf{e}}: \underline{x} \rightarrow \underline{w}$ for each subexpression \mathbf{e} expressing w . Now consider the maps $\underline{x}^{\mathcal{D}} \rightarrow Q_w$ in $\text{Kar}(\mathcal{D}_Q)$ obtained by postcomposing $LL_{\underline{x},\mathbf{e}}$ with the projection $\underline{w}^{\mathcal{D}} \rightarrow Q_w$. These maps form a basis for $\text{Hom}_{\text{Kar}(\mathcal{D}_Q)}(\underline{w}^{\mathcal{D}}, Q_w)$. Moreover, the original set $LL_{\underline{x}}^w$ is linearly independent over R as a subset of $\text{Hom}_{\mathcal{D}}(\underline{x}, \underline{w})$.*

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

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

Consider a light leaves map $LL_{\underline{x},\mathbf{e}}: \underline{x} \rightarrow \underline{w}$, where \underline{w} is a rex for w . Flipping this diagram upside-down, we get a map $\overline{LL}_{\underline{w},\mathbf{e}}: \underline{w} \rightarrow \underline{x}$. By the results of the previous section, $\overline{LL}_{\underline{w},\mathbf{e}}$ behaves nicely after localization, giving a non-zero map $Q_{\underline{w}} \rightarrow Q_{\mathbf{f}}$ only when $\mathbf{f} \leq \mathbf{e}$. The coefficients appearing are not actually $p_{\mathbf{f}}^{\mathbf{e}}$, because the polynomials $\frac{1}{\alpha_s}$ should appear only in the projection maps from $s^{\mathcal{D}}$ to $\mathbb{1}$ 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 $LL_{\underline{x},\mathbf{e}}: \underline{x} \rightarrow \underline{w}$ and $\overline{LL}_{\underline{y},\mathbf{f}}: \underline{w} \rightarrow \underline{y}$. We define the corresponding *double leaves map* to be the composition

$$\mathbb{L}\mathbb{L}_{\mathbf{f},\mathbf{e}} \stackrel{\text{def}}{=} \overline{LL}_{\underline{y},\mathbf{f}} \circ LL_{\underline{x},\mathbf{e}}.$$

We write $\mathbb{L}\mathbb{L}_{\mathbf{f},\mathbf{e}}^w$ when we want to emphasize the element w , which we say that the double leaf *factors through*. 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 \underline{x} and projection to each standard summand \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 non-zero and is a product of roots, obeying a simple formula independent of the choice of LL maps.

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

Proof. After localization, the space of maps $\underline{x} \rightarrow \underline{y}$ is a direct sum of $\text{Hom}(Q_{\mathbf{e}}, Q_{\mathbf{f}})$ for each pair of subsequences. Place a partial order on pairs (\mathbf{e}, \mathbf{f}) , where $(\mathbf{e}_1, \mathbf{f}_1) \leq (\mathbf{e}_2, \mathbf{f}_2)$ if $\mathbf{e}_1 \leq \mathbf{e}_2$ and $\mathbf{f}_1 \leq \mathbf{f}_2$. Then the $\mathbb{L}\mathbb{L}$ maps satisfy upper-triangularity with respect to this partial order, with an invertible diagonal. \square

Remark 6.10. Proposition 6.9 was also proven by Libedinsky in the context of Soergel bimodules in his paper [Lib15].

Remark 6.11. 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 Section 7, but would introduce its own annoyances. Most notably, rotation takes a map $\underline{w} \rightarrow \underline{y}$ and returns a map $\omega(\underline{y}) \rightarrow \omega(\underline{w})$, 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 $\omega(\underline{w})$, which is actually a rex for w^{-1} , and this requires additional bookkeeping. Rotation will be more obviously useful in Section 7. 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 6.12. *The set $\mathbb{L}\mathbb{L}_{\underline{x}, \underline{y}}$ forms a free R -basis for $\text{Hom}(\underline{x}, \underline{y})$ in \mathcal{D} .*

Proposition 6.13. *The set $LL_{\underline{x}}^e$ forms a free R -basis for $\text{Hom}(\underline{x}, \mathbb{1})$ in \mathcal{D} .*

Corollary 6.14. *Hom spaces in \mathcal{D} are free graded R -modules.*

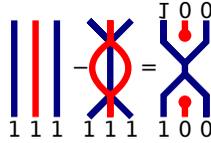
Remark 6.15. Proposition 6.13 is a special case of Theorem 6.12, because when $\underline{y} = \emptyset$ we must have $w = e$, and $\overline{LL}_{\underline{y}}$ is the empty diagram.

The final section of this paper, Section 7, contains an elementary diagrammatic (though unpleasant) proof of these results. Of course it is enough to show that $\mathbb{L}\mathbb{L}$ spans, as we have already shown that this set is linearly independent over R . The reader is now equipped to read that section, which does not use anything beyond this section.

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

In defining $\mathbb{L}\mathbb{L}_{\underline{x}, \underline{y}}$, many choices were made in the construction of the particular LL maps being used. However, Theorem 6.12 applies regardless of which choices are made. This implies that any choice of $\mathbb{L}\mathbb{L}$ spans any other. However, it does *not* imply that one choice of $LL_{\underline{x}}^w$ will span another, because the set $LL_{\underline{x}}^w$ is only supposed to give a basis for morphisms modulo lower terms.

Example 6.16. Let s, t be the indices in type A_2 . There are two choices for the map $LL_{st,111}$, as below. The difference between them is not an LL map, but it is an $\mathbb{L}\mathbb{L}$ map.



6.4. Cellularity. We assume Theorem 6.12, so that Hom spaces are free left R -modules and localization is injective on Hom spaces.

Claim 6.17. 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 (or equivalently target). Then \mathcal{X}_w is a 2-sided ideal in \mathcal{D} .

Proof. Left to the reader. □

Claim 6.18. The light leaves map $\mathbb{L}\mathbb{L}_{\mathbf{e},\mathbf{f}}^w$ 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 $\mathbb{L}\mathbb{L}_I$ denote the span of all $\mathbb{L}\mathbb{L}$ 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 6.19. When I is an ideal, $\mathbb{L}\mathbb{L}_I$ 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 $\mathbb{L}\mathbb{L}_I$ is equal to $\mathcal{X}_{W \setminus I}$, for the latter is clearly a 2-sided ideal. The inclusion $\mathbb{L}\mathbb{L}_I \subset \mathcal{X}_{W \setminus I}$ follows from the previous claim.

Now suppose that $\phi \in \text{Hom}(\underline{w}, \underline{y})$ is in $\mathcal{X}_{W \setminus I}$, and write ϕ as a linear combination of $\mathbb{L}\mathbb{L}$ maps. Unless ϕ is zero, some $\mathbb{L}\mathbb{L}_{\mathbf{f},\mathbf{e}}^w$ has a non-zero coefficient. Choose w, \mathbf{e} , and \mathbf{f} successively such that each is maximal in the Bruhat order (for w) or path dominance order (for \mathbf{e} and \mathbf{f}) relative to the constraint that there is a non-zero coefficient for $\mathbb{L}\mathbb{L}_{\mathbf{f},\mathbf{e}}$ in ϕ . Because of upper triangularity (like in the proof of Proposition 6.9), 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 non-zero map on $Q_{\mathbf{e}}$, implying that $w \in I$. Since this is true for each maximal choice of w , we see that $\phi \in \mathbb{L}\mathbb{L}_I$. □

If \underline{w} and \underline{w}' are rexes for $w \in W$, then any rex move $\underline{w} \rightarrow \underline{w}'$ will induce the identity map on the unique Q_w summand after localization. Thus the difference between two rex moves is in \mathcal{X}_w . For obvious reasons, this difference is also in \mathcal{X}_v for any $v \not\leq w$. Thus the claim above implies that the two rex moves are equal modulo $\mathbb{L}\mathbb{L}_{<w}$.

Now consider $\mathbb{L}\mathbb{L}_{\underline{x},\underline{x}}$. Recall that $\mathbb{L}\mathbb{L}_{\mathbf{f},\mathbf{e}}^w = \overline{LL}_{\underline{x},\mathbf{f}} \circ LL_{\underline{x},\mathbf{e}}$, factoring through \underline{w} in the middle. Suppose we compose them in the opposite order, to get an endomorphism of \underline{w} . We have an $\mathbb{L}\mathbb{L}$ basis for endomorphisms of \underline{w} as well, and there is a *unique* light leaves morphism which induces a non-zero 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 $LL_{\underline{x},\mathbf{e}} \circ \overline{LL}_{\underline{x},\mathbf{f}}$.

Claim 6.20. *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 $\mathbb{L}\mathbb{L}_{\mathbf{h},\mathbf{g}}\mathbb{L}\mathbb{L}_{\mathbf{f},\mathbf{e}}: \underline{x} \rightarrow \underline{z}$ is equal to $\psi(\mathbf{g}, \mathbf{f})\mathbb{L}\mathbb{L}_{\mathbf{h},\mathbf{e}}$ modulo $\mathbb{L}\mathbb{L}_{<w}$.*

Proof. $\mathbb{L}\mathbb{L}_{\mathbf{h},\mathbf{g}}\mathbb{L}\mathbb{L}_{\mathbf{f},\mathbf{e}}$ 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 \underline{w} , plus maps in $\mathbb{L}\mathbb{L}_{<w}$. Therefore the overall composition is equal to $\psi(\mathbf{g}, \mathbf{f})\overline{LL}_{\underline{z},\mathbf{h}}LL_{\underline{x},\mathbf{e}}$ modulo $\mathbb{L}\mathbb{L}_{<w}$, as desired. \square

Claim 6.21. *Let $a: \underline{y} \rightarrow \underline{z}$ be an arbitrary morphism, and let $\mathbb{L}\mathbb{L}_{\mathbf{e},\mathbf{f}}$ be a light leaves map $\underline{x} \rightarrow \underline{y}$ factoring through the rex \underline{w} . Then $a\mathbb{L}\mathbb{L}_{\mathbf{e},\mathbf{f}} = \sum_{\mathbf{f}'} r_a(\mathbf{f}, \mathbf{f}')\mathbb{L}\mathbb{L}_{\mathbf{e},\mathbf{f}'}$ modulo $\mathbb{L}\mathbb{L}_{<w}$. 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 $a\mathbb{L}\mathbb{L}_{\mathbf{e},\mathbf{f}} = a\overline{LL}_{\underline{y},\mathbf{f}}LL_{\underline{w},\mathbf{e}}$, and consider $a\overline{LL}_{\underline{y},\mathbf{f}}$ as a map from $\underline{w} \rightarrow \underline{z}$. The space of maps $\underline{w} \rightarrow \underline{z}$ modulo $\mathbb{L}\mathbb{L}_{<w}$ 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. \square

Definition 6.22. Let $\iota: \mathcal{D} \rightarrow \mathcal{D}^{\text{op}}$ denote the anti-involution which preserves objects and flips diagrams upside-down.

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

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

Proposition 6.23. *The category \mathcal{D} is cellular, with cellular basis $\mathbb{L}\mathbb{L}$ (for any appropriate choice of $\mathbb{L}\mathbb{L}$ maps) and anti-involution 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 . The interactions between this cellular structure and the monoidal structure have not been explored.

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

$$\mathcal{D}^J \stackrel{\text{def}}{=} \mathcal{D}/\mathbb{L}\mathbb{L}_{W \setminus J}.$$

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 $\underline{w}^{\mathcal{D}}$ in \mathcal{D}^J for any rex \underline{w} are canonically isomorphic. (As discussed in the last section, the difference between any two rex moves $\underline{w} \rightrightarrows \underline{w}'$ lies in $\mathbb{L}\mathbb{L}_{<w} \subset \mathbb{L}\mathbb{L}_{W \setminus J}$.) Similarly for any rex \underline{w} we have

$$(6.3) \quad \text{End}_{\mathcal{D}^J}(\underline{w}) \cong R.$$

Given any $w \in W$ we set

$$\mathcal{D}^{\geq w} \stackrel{\text{def}}{=} \mathcal{D}^{\{y \mid y \geq w\}}$$

and the above remarks show that, for any rex \underline{w} , the image of \underline{w} does not depend on the choice of reduced expression up to canonical isomorphism. Given any expression \underline{x} , it follows from Theorem 6.12 that $\text{Hom}_{\mathcal{D} \geq w}(\underline{x}, \underline{w})$ is a free left R -module with basis 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 2.7 that we have the identity

$$(6.4) \quad H_{\underline{x}} = \sum \text{grk Hom}_{\mathcal{D} \geq w}(\underline{x}, \underline{w}) H_w,$$

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

We would like to extend this “character” map to the Karoubi envelope $\text{Kar}(\mathcal{D})$. The problem is that for an arbitrary \mathbb{k} and $B \in \text{Kar}(\mathcal{D})$, the R -module $\text{Hom}_{\mathcal{D} \geq w}(\underline{x}, \underline{w})$ is projective (as a 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 \mathbb{k} is a local ring. By Nakayama’s lemma and its graded version, direct summands of free graded R -modules are graded free.

Definition 6.24. We define the *diagrammatic character* by

$$\begin{aligned} \text{ch} : [\text{Kar}(\mathcal{D})] &\rightarrow \mathbf{H} \\ [B] &\mapsto \sum_{w \in W} \text{grk Hom}_{\text{Kar}(\mathcal{D} \geq w)}(B, \underline{w}) H_w. \end{aligned}$$

The diagrammatic character is obviously a homomorphism of abelian groups. If one equips $[\text{Kar}(\mathcal{D})]$ with the usual $\mathbb{Z}[v^{\pm 1}]$ -module structure, where $v[B] = [B(1)]$ for any object B , then it is easy to check that $\text{ch}(v[B]) = \text{ch}([B(1)]) = v \text{ch}([B])$. Hence ch is a homomorphism of $\mathbb{Z}[v^{\pm 1}]$ -modules. It is immediate from (6.4) that

$$(6.5) \quad \text{ch}(\underline{xy}) = H_{\underline{xy}} = H_{\underline{x}} H_{\underline{y}} = \text{ch}(\underline{x}) \text{ch}(\underline{y}),$$

and so ch is a homomorphism of algebras on the $\mathbb{Z}[v^{\pm 1}]$ -submodule of $[\text{Kar}(\mathcal{D})]$ generated by the isomorphism classes of objects in \mathcal{D} . In the next section we will see that ch is an isomorphism of algebras if \mathbb{k} is a complete local ring.

6.6. Soergel’s theorem. We present a proof of a generalization of Soergel’s Categorification Theorem (Theorems 3.16 and 3.17). This proof applies directly to \mathcal{D} , but implies the corresponding theorem for Soergel bimodules via Theorem 6.30. The proof is quite formal, relying only on general facts about Krull-Schmidt categories and Theorem 6.12. We find our proof conceptually simpler than Soergel’s original proof, although the complexity of the diagrammatic arguments in Section 7 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 \mathbb{k} is a complete local ring. It is known that any \mathbb{k} -linear idempotent complete additive category such that all Hom spaces are finitely generated is Krull-Schmidt. (This follows from the fact that any finite \mathbb{k} -algebra is either local or admits a non-trivial idempotent.) Theorem 6.12 shows that this condition is met for degree zero morphisms in \mathcal{D} . We conclude:

Lemma 6.25. *If \mathbb{k} is a complete local ring, then the category $\text{Kar}(\mathcal{D})$ is Krull-Schmidt.*

The following is a diagrammatic variant of Soergel’s theorem, classifying the indecomposable Soergel bimodules:

Theorem 6.26. *Assume that \mathbb{k} is a complete local ring. Then for all $w \in W$ there exists a unique summand B_w of \underline{w} which is not isomorphic to a shift of a summand of \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 $\text{Kar}(\mathcal{D})$ is isomorphic to a shift of B_w for some $w \in W$. Hence one has a bijection:*

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

$$w \mapsto B_w.$$

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

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

After localizing, each e_i acts as an idempotent on $\underline{w}^{\mathcal{D}} \cong Q_w \oplus \bigoplus_{\mathbf{e} \neq (1, \dots, 1)} Q_{\mathbf{e}}$. 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 characterized as the unique idempotent with a non-zero coefficient of $\mathbb{L}\mathbb{L}_{\mathbf{e}, \mathbf{f}}$, where $\mathbf{e} = \mathbf{f} = (1, \dots, 1)$.) We define B_w to be the image of e_1 in $\text{Kar}(\mathcal{D})$. Hence for all \underline{w} we have constructed an indecomposable object B_w in $\text{Kar}(\mathcal{D})$.

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

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

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

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

for some (homogeneous) coefficients $\gamma_{\mathbf{e}, \mathbf{f}} \in R$, where the sum is over subexpressions \mathbf{e}, \mathbf{f} of \underline{x} expressing w . Now assume that for all such subexpressions \mathbf{e} and \mathbf{f} with $d(\mathbf{e}) + d(\mathbf{f}) = 0$ (d denotes the defect) we have

$$\mathbb{L}\mathbb{L}_{\underline{x}, \mathbf{e}} \circ e \circ \overline{\mathbb{L}\mathbb{L}}_{\underline{x}, \mathbf{f}} \in \mathfrak{m} \subset \mathbb{k} = \text{End}_{\mathcal{D} \geq w}^0(\underline{w}),$$

where \mathfrak{m} denotes the maximal ideal of \mathbb{k} . Then by expanding $e^3 = e$ 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

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

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

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

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

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

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

and recall the summand $B_w \subset \underline{w}$ 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 $m \in \mathbb{Z}$. \square

Corollary 6.27. *Assume that \mathbb{k} is a complete local ring. The diagrammatic character*

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

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

Proof. It is immediate from Theorem 6.26 that:

- (1) $[\text{Kar}(\mathcal{D})]$ is spanned by the classes $[\underline{x}]$ 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 (6.5) we conclude that ch is a homomorphism. Using the definition of the diagrammatic character and the construction of the objects B_x we have

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

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

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

Corollary 6.28. *Assume that \mathbb{k} is a complete local ring. The map $\underline{H}_s \mapsto [B_s]$ defines a homomorphism $\mathbf{H} \rightarrow [\text{Kar}(\mathcal{D})]$.*

Remark 6.29. Note that we have proven the Soergel Categorification Theorem using the character map, which we managed to prove was an algebra homomorphism without ever mentioning the categorification of the relations in the Hecke algebra. This is quite different from the proof for the dihedral case in [Elid], and it is more general. Because it would otherwise be almost invisible, let us explain how the dihedral relation of the Hecke algebra from Example 2.5 has been categorified.

Consider an alternating reduced expression $\underline{w} = stst\dots$ for the longest element in a finite dihedral subgroup, having length $m_{st} < \infty$. The idempotent projecting to B_w is (5.16). Relation (5.17) allows one to compute that many light leaves vanish when precomposed with the idempotent projecting to B_w . This was used in [Elid] to confirm that the diagrammatic character of B_w is actually the Kazhdan-Lusztig basis element \underline{H}_w . Repeating the same process for the other reduced expression $tsts\dots$, one obtains the same diagrammatic character, and can check that the $2m_{st}$ -valent vertex induces isomorphisms on the surviving Hom spaces. Thus these two indecomposable objects are isomorphic, via the $2m_{st}$ -valent vertex. This isomorphism is what categorifies the dihedral relation in the Hecke algebra; the two sides

of each equality in Example 2.5 can be thought of as expressions for the summand B_w inside the two different reduced expressions.

6.7. The equivalence to bimodules. In this section we assume that \mathfrak{h} is a realization for which Soergel’s Categorification Theorem holds. We prove that our diagrammatic category is equivalent to Soergel bimodules.

By Corollary 6.28 the map $\mathbf{H} \rightarrow [\mathcal{D}] : \underline{H}_s \mapsto [B_s]$ 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 $\varepsilon_{\mathcal{D}} : \mathbf{H} \rightarrow \mathbb{Z}[v^{\pm 1}] : h \mapsto (h, 1)_{\mathcal{D}}$. Because the degree of any light leaves map is given by the defect of the corresponding subexpression, it follows from Proposition 6.13 and Corollary 2.8 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 $\mathcal{F} : \mathcal{D} \rightarrow \mathbb{B}\mathbb{S}\text{Bim}$. It induces a monoidal functor on the idempotent completions $\mathcal{F} : \text{Kar}(\mathcal{D}) \rightarrow \mathbb{S}\text{Bim}$.

Theorem 6.30. *Under the above assumptions $\mathcal{F} : \text{Kar}(\mathcal{D}) \rightarrow \mathbb{S}\text{Bim}$ is an equivalence of monoidal categories.*

Proof. Because idempotent completion preserves equivalences, it is enough to show that $\mathcal{F} : \mathcal{D} \rightarrow \mathbb{B}\mathbb{S}\text{Bim}$ 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, $2m_{st}$ -valent vertices, and polynomials generate all morphisms between Bott-Samelson bimodules. (This theorem could also be proven easily using the localization techniques of this paper.) It follows that \mathcal{F} is full. Now by the above discussion and the fact that our realization is such that Soergel’s Categorification Theorem holds, the graded dimensions of the homomorphism spaces in \mathcal{D} and $\mathbb{B}\mathbb{S}\text{Bim}$ coincide. 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. \square

7. DOUBLE LEAVES SPAN

This section 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 section we will be interested in embedded graphs, not isotopy classes thereof. We will abusively use the term *S-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 displayed in Section 1.4, tensored with identity maps. Recall that 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 \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 $LL_{\underline{x},e}$ 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*.

$$\begin{array}{ccccc}
 \bullet & \cup & \times & \bullet & \cap \\
 + & + & = & - & -
 \end{array}$$

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 \mathbb{k} -linear combination of negative-positive graphs is called a *negative-positive decomposition*.

Any map of the form $LL_{\underline{x},e}$ is constructed out of non-positive generators. Thus every double leaves map LL is a negative-positive map. Among other things, we must prove 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 actually a statement about any morphism obtained from a 2-colored non-identity crossingless matching by taking the retract on the strip. Temperley-Lieb diagrams have well-known cup-cap decompositions, which translate to the negative-positive decompositions here. □

Claim 7.2. *Any strictly non-positive map from \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 $ss \rightarrow s$. 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. □

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. □

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, $\beta - \beta'$ is in the span of diagrams having 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 (see Remark 5.4 for the H_3 relation). 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 $\beta - \beta' = (\beta - \beta_1) + (\beta_1 - \beta_2) + \dots + (\beta_k - \beta')$, where each successive difference corresponds to a single transformation. The result follows. □

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

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

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) to obtain 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 as in Claim 7.3, 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. □

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 combinations 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* (with respect to \underline{w}). When $\underline{w} = \emptyset$, the ideal $I_{\underline{w}}$ is zero.

We may not use localization in our proof that $\mathbb{L}\mathbb{L}$ spans, because localization would kill any R -torsion which may exist. 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 summand in $\text{Kar}(\mathcal{D}_Q)$ which is the joint kernel of the top boundary dots.

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 $\underline{w} \rightarrow \underline{w}'$, clearly sends $I_{\underline{w}} \rightarrow I_{\underline{w}'}$, because we can “pull” top-dots through $2m$ -valent vertices. The color-reversed $2m$ -valent vertex gives a map $B_{\underline{w}'} \rightarrow B_{\underline{w}}$ sending $I_{\underline{w}'} \rightarrow I_{\underline{w}}$. 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 $\text{Hom}(\underline{x}, \underline{w})/I_{\underline{w}}$ and $\text{Hom}(\underline{x}, \underline{w}')/I_{\underline{w}'}$ 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 $\text{Hom}(\underline{x}, \underline{w})/I_{\underline{w}} \rightarrow \text{Hom}(\underline{x}, \underline{w}')/I_{\underline{w}'}$, 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 care which rex \underline{w} we choose, 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 $\mathbb{L}\mathbb{L}$ 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 let \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 $\text{Hom}(\underline{x}, \underline{w})/I_{\underline{w}}$, under the action of R on the left.*

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

Before proving this proposition, we will prove the main theorem using it. 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 of double leaves 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}^{std} was shown in [EW] only using non-trivial arguments involving homotopy groups and the topology of the Coxeter complex.

7.3. Reduction to working modulo lower terms.

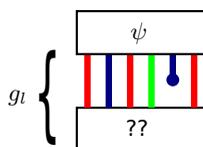
Proof of Theorem 6.12 (assuming Proposition 7.6). Fix sequences \underline{x} and \underline{y} and a morphism $\phi: \underline{x} \rightarrow \underline{y}$. We want to write ϕ as an R -linear combination of diagrams of the form $\overline{LL}_{\underline{y}, \mathbf{f}} LL_{\underline{x}, \mathbf{e}}$, 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 $g: \underline{x} \rightarrow \underline{w}$ and $f: \underline{w} \rightarrow \underline{y}$ for some reduced expression \underline{w} . We will prove the

statement by induction on w . That is, we assume that any morphism $\phi = fg$ which factors through a rex \underline{v} for $v < w$ is in the span of $\mathbb{L}\mathbb{L}$. 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 $g_l \in I_w$; similarly, f_w is in the span of $\overline{LL}_{\underline{y},\mathbf{f}}$ and $f_l \in \overline{I}_w$. This decomposition is guaranteed by Proposition 7.6. The composition $f_w g_w$ is clearly in the span of $\mathbb{L}\mathbb{L}$. 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 $\mathbb{L}\mathbb{L}$ for any ψ (the argument with f_l is the same, upside-down).

Consider ψg_l . Now $g_l \in I_w$ 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 \underline{z} denote the sequence consisting of \underline{w} with the dotted index removed. This morphism factors through \underline{z} , and the identity of \underline{z} has a decomposition which factors through reduced expressions for elements $v \leq \underline{z}$. Subsequences of \underline{z} are strict subsequences of \underline{w} , so $v < w$. Now induction implies that ψg_l is in the span of $\mathbb{L}\mathbb{L}$, as desired. □

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 matter, because 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 subexpression \mathbf{e} of \underline{x} expression w , and with target \underline{w} . Every map $\underline{x} \rightarrow \underline{w}$ of maxwidth $\leq M$ is in the left R -span of the $LL_{\underline{x},\mathbf{e}}$ modulo I_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 $\underline{x} \rightarrow \underline{w}$ of maxwidth $\leq M$ is in the \mathbb{k} -span of $X_{\underline{x}}$ modulo I_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 $\underline{x} \rightarrow \underline{w}$ 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 $\underline{x} \rightarrow \underline{w}$. 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 $\mathbf{e} = (1, 1, \dots, 1)$. To show (L_M) we need to show that any single rex move, with polynomials only in the leftmost region, will span morphisms consisting of compositions of 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 $(X_M) \implies (L_M)$. We need to show that each element of $X_{\underline{x}}$ is in the span of our particular fixed 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.

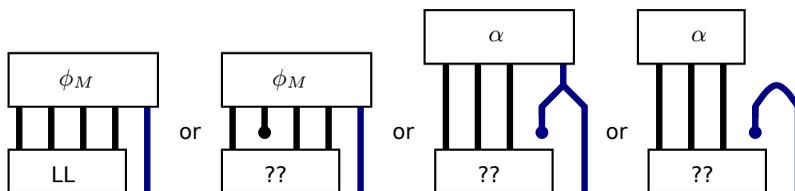
Suppose that we are building a new light leaves map for some $(\underline{x}, \mathbf{e})$ expressing w , using arbitrary choices, but we make a single error. While performing a rex move, we accidentally insert some map with a strictly negative-positive decomposition. We call a graph of this form an “ 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 into 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 intermediate reduced expressions \underline{w}_k and \underline{w}_k^s , but this can also be accounted for by changing rex moves, as follows. One can replace the identity of \underline{w}_k with a rex move fg , where $g: \underline{w}_k \rightarrow \underline{w}'_k$ and $f: \underline{w}'_k \rightarrow \underline{w}_k$; this does not change the morphism modulo error terms. Then, viewing g as part of α in ϕ_k and f as part of β in ϕ_{k+1} (if D) or α in ϕ_{k+1} (if U), we can effectively replace \underline{w}_k with \underline{w}'_k . (This also works for \underline{w}_k^s , because two such expressions differ by rex moves which do not involve the final s strand.) 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 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:



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). \square

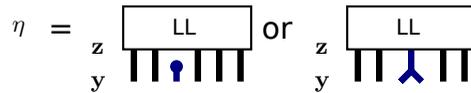
$(L_{M-1}) \implies (X_M)$. Now let us fix \underline{w} with $\ell(\underline{w}) < M$, and prove that (L_{M-1}) implies (X_M) for \underline{w} . Consider a diagram of maxwidth $\leq M$. If the maxwidth M is never attained, then we may use induction, so let us assume that width M is attained at least 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 $\underline{x} \rightarrow \underline{w}$ 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 $\underline{x} \rightarrow \underline{w}$ 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 $I_{\underline{w}}$.

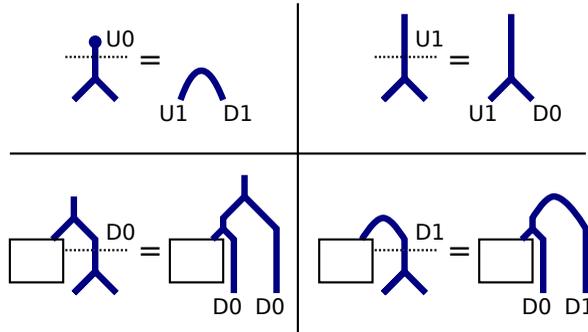
Suppose that we can show these two claims. Then we may simplify the diagram as follows. Consider the first interval from the top where width M appears. If that extends all the way to the bottom of the map, we can use Claim 7.8. If the interval ends and we return to width $M - 1$, we can apply Claim 7.7 to replace that region with an LL map (using (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 $\underline{z} \rightarrow \underline{w}$ 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:



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:

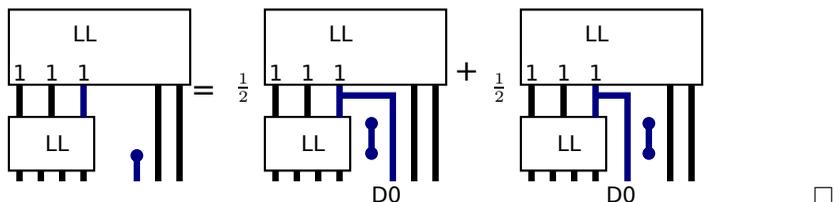


□

Claim 7.10. *Suppose that we add a negative (bottom) dot after the k -th strand of an $LL_{\underline{z},\mathbf{e}}$ of our choice, yielding a map from \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 $LL_{\underline{z},\mathbf{e}}$ appropriately, we can assume that

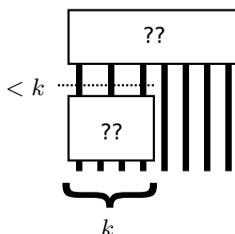
s occurs on the right of w_k . Now the statement follows from the equality below, because both terms on the RHS are in $X_{\underline{y}}$. We use the decomposition of Remark 6.4 and relation (5.2). Remember that the use of $\frac{1}{2}$ is unnecessary, as any dual basis will do.



Remark 7.11. While this is the only use in this section of the assumption of De-mazure Surjectivity, it is significant. Without this assumption, double leaves will not form a basis in the simplest counterexample: maps $ss \rightarrow s$.

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 $\underline{x} \rightarrow \underline{w}$ 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 \underline{v} with the dashed line through it (having length $< k$). By an application of Lemma 7.5 we can assume without loss of generality that \underline{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 $I_{\underline{v}}$, 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 \underline{v} , and $g_{w,0}$ is a linear combination of light leaves which do not begin with $U1$'s, meaning that somewhere on \underline{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 $X_{\underline{w}}$ 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 $0 < k \leq l$ or $l + m \leq k < M$, we can split the diagram into $LL_{\leq k}$ and the remainder as in Remark 6.4, placing the new $2m$ -valent vertex below whichever half is appropriate. Therefore, if any such $LL_{\leq k}$ 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: $\underline{x}_{\leq l}$ is a reduced expression and $\mathbf{e}_{\leq l}$ is all $U1$ (this includes the possibility $l = 0$); unless $l + m = M$, $\underline{x}_{\leq M-1}$ is a reduced expression and $\mathbf{e}_{\leq M-1}$ 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 $(\underline{x}, \mathbf{e})_{\leq l}$ 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 $LL_{\leq k}$, which has width $\leq M - 1$, 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 $I_{\underline{w}}$.

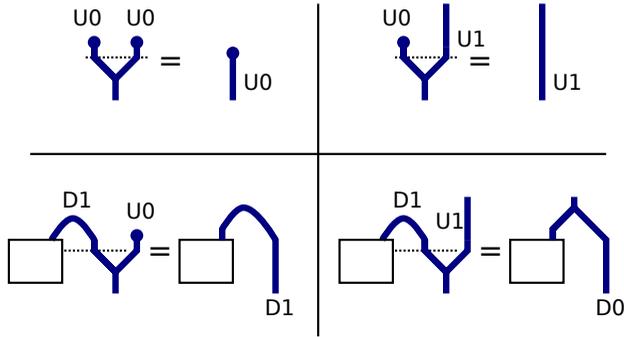
Hence, we can choose any desired reduced expression \underline{w}_l for w_l . Now $w_l = uv$ for $v \in W_{s,t}$ the dihedral group, and u a minimal right coset representative of $W_{s,t}$. We can choose a rex for w_l of the form \underline{uv} . Now the remaining terms of \underline{x} are all s or 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 $\underline{vx}_{>l}$, preserving \underline{u} . Suppose that they are so in $LL_{\underline{x}, \mathbf{e}}$. Then the reduction to light leaves form does not depend on \underline{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 [Elid] to be the right size, which implies the result; alternatively, one could do a case-by-case analysis. □

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 .

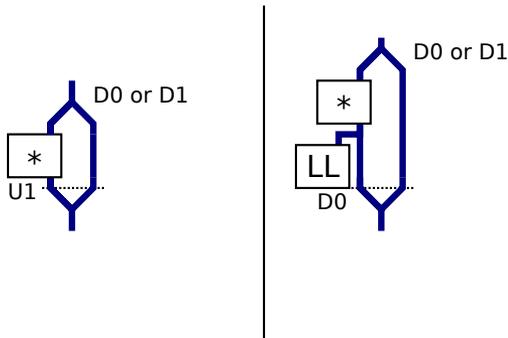
We have now proven Claim 7.8. In order to prove Claim 7.7, we need only put a positive generator below a map in $X_{\underline{y}}$ for $\ell(\underline{y}) = M$.

Claim 7.15. *Assume (L_{M-1}) . Suppose that $\ell(\underline{y}) = M$ and we add a positive (splitting) trivalent below a map in $X_{\underline{y}}$ to obtain a map $\underline{x} \rightarrow \underline{w}$. Then the result is in the span of $X_{\underline{x}}$, modulo $I_{\underline{w}}$.*

Proof. Suppose that the trivalent vertex is added to \underline{y}_k and \underline{y}_{k+1} . There are eight 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 e_k is $U1$ or $D0$, the result locally looks like the following:



By “ $D0$ or $D1$ ” we mean that the topmost trivalent in these graphs is as pictured when e_{k+1} is $D0$, and has a further dot on top when e_{k+1} is $D1$; this extra dot will not affect our discussion. The dashed line indicates where width M is reached. The asterisk box in these graphs is some rex move between two rexes which both have s on the far right.

In fact, we believe both these diagrams to be zero. Let us pause to state a computational conjecture, relying on the form of the H3 relation, among other things.

Conjecture 7.16. *For any rex move $\beta: \underline{w} \rightarrow \underline{w}'$, where both \underline{w} and \underline{w}' have s on the far right, one has:*

$$(7.1) \quad \begin{array}{c} \underline{w}' \\ \beta \\ \underline{w} \end{array} \begin{array}{c} s \\ \text{---} \\ \end{array} = \begin{array}{c} \underline{w}' \\ \beta \\ \underline{w} \end{array} \begin{array}{c} \text{---} \\ s \\ \end{array}$$

This would imply, using (5.3) and (5.5), that the following is zero:

$$(7.2) \quad \begin{array}{c} \beta \\ \text{---} \end{array} \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} = 0$$

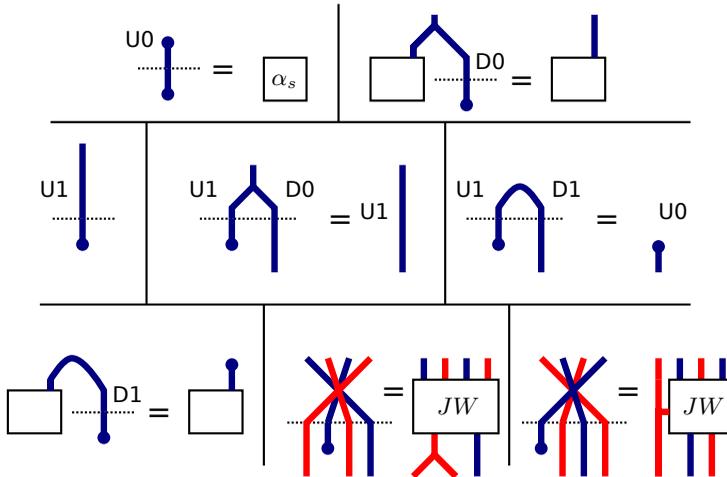
Let us continue the proof without assuming this conjecture.

Which rex move appears in the asterisk box? If the rex move does not involve the rightmost s -colored strand, then clearly the result is zero by (5.3) and (5.5). Since two rex moves are equal modulo lower terms, we may assume that the asterisk box actually contains a strictly negative map followed by a strictly positive map. Moreover, Lemma 7.4 guarantees that both the top and bottom of the asterisk box has a boundary dot.

We will now use these boundary dots to reduce the maxwidth of the diagram below M . The bottom dot (resp. top dot) appears either in one of the initial $k - 1$ strands, or on the final s -colored strand. If the dot appears on the final s -colored strand, then it will pull into the nearby trivalent vertex by (5.4). Otherwise, using rectilinear isotopy, the dot can be pulled past this trivalent vertex. After performing these operations to the top and bottom dots, the resulting diagram never factors through a sequence of width M , and therefore (L_{M-1}) implies that it can be expressed in the span of light leaves. \square

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 $\underline{x} \rightarrow \underline{w}$. 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 $LL_{\underline{y}, \mathbf{e}}$. Let us draw a number of possibilities for what can happen to the dot:



In the first row, \mathbf{e}_l is U_0 or D_0 and is quickly taken care of. When \mathbf{e}_l is U_1 , 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 D_0 or D_1 and is taken care of. Finally, if \mathbf{e}_l is D_1 , 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 JW -box 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. \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 $(\underline{y}, \mathbf{e})_{<M-1}$ is a reduced expression with all $U1$'s, and \mathbf{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 $(L_{M-1}) \implies (X_M)$. \square

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 emphasizing the importance of generators and relations, and Jean Michel for help speeding up his programs. Both authors would like to thank the referee for many detailed comments. The first author was supported by NSF grants DMS-524460 and DMS-524124 and DMS-1103862.

REFERENCES

- [BM01] T. Braden and R. MacPherson, *From moment graphs to intersection cohomology*, Math. Ann. **321** (2001), no. 3, 533–551, DOI 10.1007/s002080100232. MR1871967
- [Deo90] V. V. Deodhar, *A combinatorial setting for questions in Kazhdan-Lusztig theory*, Geom. Dedicata **36** (1990), no. 1, 95–119, DOI 10.1007/BF00181467. MR1065215
- [EK] B. Elias and M. Khovanov, *Diagrammatics for Soergel categories*, Int. J. Math. Math. Sci. (2010), Art. ID 978635, 58. MR3095655
- [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, *Light ladders and clasp conjectures*, Preprint, arXiv:1510.06840.
- [Elic] B. Elias, *Quantum satake in type A: part I*, Preprint, arXiv:1403.5570.
- [Elid] B. Elias, *The two-color Soergel calculus*, Preprint, arXiv:1308.6611.
- [EW] B. Elias and G. Williamson, *Diagrammatics for Coxeter groups and their braid groups*, Preprint, arXiv:1405.4928.
- [EW14] B. Elias and G. Williamson, *The Hodge theory of Soergel bimodules*, Ann. of Math. (2) **180** (2014), no. 3, 1089–1136, DOI 10.4007/annals.2014.180.3.6. MR3245013
- [Fen83] R. A. Fenn, *Techniques of geometric topology*, London Mathematical Society Lecture Note Series, vol. 57, Cambridge University Press, Cambridge, 1983. MR787801
- [Fie08] P. Fiebig, *Sheaves on moment graphs and a localization of Verma flags*, Adv. Math. **217** (2008), no. 2, 683–712, DOI 10.1016/j.aim.2007.08.008. MR2370278
- [Fie10] P. Fiebig, *Lusztig's conjecture as a moment graph problem*, Bull. Lond. Math. Soc. **42** (2010), no. 6, 957–972, DOI 10.1112/blms/bdq058. MR2740015
- [Fie11] P. Fiebig, *Sheaves on affine Schubert varieties, modular representations, and Lusztig's conjecture*, J. Amer. Math. Soc. **24** (2011), no. 1, 133–181, DOI 10.1090/S0894-0347-2010-00679-0. MR2726602
- [Hum90] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge Studies in Advanced Mathematics, vol. 29, Cambridge University Press, Cambridge, 1990. MR1066460
- [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** (1964), 215–236 (1964). MR0165016
- [JMW14] D. Juteau, C. Mautner, and G. Williamson, *Parity sheaves*, J. Amer. Math. Soc. **27** (2014), no. 4, 1169–1212, DOI 10.1090/S0894-0347-2014-00804-3. MR3230821
- [Kac90] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, Cambridge, 1990. MR1104219

- [KL79] D. Kazhdan and G. Lusztig, *Representations of Coxeter groups and Hecke algebras*, Invent. Math. **53** (1979), no. 2, 165–184, DOI 10.1007/BF01390031. MR560412
- [Lau10] A. D. Lauda, *A categorification of quantum $\mathfrak{sl}(2)$* , Adv. Math. **225** (2010), no. 6, 3327–3424, DOI 10.1016/j.aim.2010.06.003. MR2729010
- [Lib08] N. Libedinsky, *Sur la catégorie des bimodules de Soergel* (French, with French summary), J. Algebra **320** (2008), no. 7, 2675–2694, DOI 10.1016/j.jalgebra.2008.05.027. MR2441994
- [Lib10] N. Libedinsky, *Presentation of right-angled Soergel categories by generators and relations*, J. Pure Appl. Algebra **214** (2010), no. 12, 2265–2278, DOI 10.1016/j.jpaa.2010.02.026. MR2660912
- [Lib15] N. Libedinsky, *Light leaves and Lusztig’s conjecture*, Adv. Math. **280** (2015), 772–807, DOI 10.1016/j.aim.2015.04.022. MR3350234
- [MS89] Yu. I. Manin and V. V. Schechtman, *Arrangements of hyperplanes, higher braid groups and higher Bruhat orders*, Algebraic number theory, Adv. Stud. Pure Math., vol. 17, Academic Press, Boston, MA, 1989, pp. 289–308. MR1097620
- [Ron09] M. Ronan, *Lectures on buildings*, University of Chicago Press, Chicago, IL, 2009. Updated and revised. MR2560094
- [RW] S. Riche and G. Williamson, *Tilting modules and the p -canonical basis*, Preprint, arXiv:1512.08296.
- [Soe90] W. Soergel, *Kategorie \mathcal{O} , perverse Garben und Moduln über den Koinvarianten zur Weylgruppe* (German, with English summary), J. Amer. Math. Soc. **3** (1990), no. 2, 421–445, DOI 10.2307/1990960. MR1029692
- [Soe97] W. Soergel, *Kazhdan-Lusztig polynomials and a combinatoric[s] for tilting modules*, Represent. Theory **1** (1997), 83–114 (electronic), DOI 10.1090/S1088-4165-97-00021-6. MR1444322
- [Soe00] W. Soergel, *On the relation between intersection cohomology and representation theory in positive characteristic*, J. Pure Appl. Algebra **152** (2000), no. 1-3, 311–335, DOI 10.1016/S0022-4049(99)00138-3. Commutative algebra, homological algebra and representation theory (Catania/Genoa/Rome, 1998). MR1784005
- [Soe07] W. Soergel, *Kazhdan-Lusztig-Polynome und unzerlegbare Bimoduln über Polynomringen* (German, with English and German summaries), J. Inst. Math. Jussieu **6** (2007), no. 3, 501–525, DOI 10.1017/S1474748007000023. MR2329762
- [Spr98] T. A. Springer, *Linear algebraic groups*, 2nd ed., Progress in Mathematics, vol. 9, Birkhäuser Boston, Inc., Boston, MA, 1998. MR1642713
- [Wes09] B. W. Westbury, *Invariant tensors and cellular categories*, J. Algebra **321** (2009), no. 11, 3563–3567, DOI 10.1016/j.jalgebra.2008.07.004. MR2510062
- [Wil] G. Williamson, *Some examples of parity sheaves*, Oberwolfach reports, 5 pages.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OREGON, EUGENE, OREGON 97403
E-mail address: belias@uoregon.edu

MAX-PLANCK-INSTITUT FÜR MATHEMATIK, 53111 BONN, GERMANY
E-mail address: geordie@mpim-bonn.mpg.de