

HIGHEST WEIGHT CATEGORIES ARISING FROM KHOVANOV'S DIAGRAM ALGEBRA I: CELLULARITY

JONATHAN BRUNDAN AND CATHARINA STROPPEL

ABSTRACT. This is the first of four articles studying some slight generalisations H_m^n of Khovanov's diagram algebra, as well as quasi-hereditary covers K_m^n of these algebras in the sense of Rouquier, and certain infinite dimensional limiting versions K_m^∞ , $K_m^{\pm\infty}$ and K_∞ . In this article we prove that H_m^n is a cellular symmetric algebra and that K_m^n is a cellular quasi-hereditary algebra. In subsequent articles, we relate H_m^n , K_m^n and K_m^∞ to level two blocks of degenerate cyclotomic Hecke algebras, parabolic category \mathcal{O} and the general linear supergroup, respectively.

CONTENTS

1. Introduction	1
2. Diagrams	7
3. Cellularity of Khovanov's diagram algebra	12
4. The algebra K_Λ : a larger cellular algebra	19
5. The quasi-hereditary structure of K_Λ	25
6. Further results	31
References	36

1. INTRODUCTION

Quasi-hereditary algebras (and highest weight categories) play an important role in representation theory thanks to their nice homological properties. They are finite dimensional algebras with distinguished collections of projective indecomposable, standard and irreducible modules, all of which are labelled by the same indexing set. Each of these three collections of modules defines a distinguished basis for the Grothendieck group, and the transition matrices between these bases are triangular. Quasi-hereditariness implies finite global dimension, and so excludes many interesting algebras like group algebras or Hecke algebras. In [GL] Graham and Lehrer introduced the weaker notion of a cellular algebra which again comes along with three distinguished classes of modules. Cellular algebras often have a more combinatorial nature, and include many diagram algebras such as Temperley-Lieb algebras and Brauer algebras which appear in connection with knot invariants.

2000 *Mathematics Subject Classification*: 17B10, 16S37.

First author supported in part by NSF grant no. DMS-0654147.

Second author supported by the NSF and the Minerva Research Foundation DMS-0635607.

Khovanov’s categorification of the Jones polynomial is based on certain diagrammatically defined algebras denoted here by H_n^n , whose multiplication is defined via some 2-dimensional TQFT structure; see [K1], [K2]. This article is the first of a series of four articles studying some slight generalisations H_m^n of Khovanov’s diagram algebra, and their quasi-hereditary covers K_m^n which were introduced already in [CK], [S]. All of these are positively graded finite dimensional algebras; since $H_m^n \cong H_n^m$ and $K_m^n \cong K_n^m$ we can assume that $m \leq n$. By taking direct limits, we also define and study some interesting positively graded infinite dimensional algebras denoted K_m^∞ , $K_m^{\pm\infty}$ and K_∞^∞ .

Here is a very brief summary of the results to be proved about these algebras in this article and its sequels.

In Part I of the series (this article) we show that the algebras H_m^n are cellular symmetric algebras and the algebras K_m^n are cellular quasi-hereditary algebras. We also compute explicitly the q -decomposition matrices that describe the transition between the various distinguished bases for the graded Grothendieck groups. Analogous results are obtained for the limits K_m^∞ and $K_m^{\pm\infty}$ too, though because they are locally unital but not unital algebras we must say instead that their module categories are highest weight categories. Amusingly, it happens that K_m^∞ and $K_m^{+\infty}$ are also symmetric algebras, a combination which would be impossible in the finite dimensional setting; see also [CT] for another class of locally unital algebras with similar properties.

In Part II we study the appropriate analogues of Khovanov’s diagrammatically defined “projective functors” for our various families of algebras. In Khovanov’s work, these functors play the central role in the categorification of the Jones polynomial. They play an equally central but very different role in our work. For example, we use them to give an elementary proof that the algebras $K_m^n, K_m^\infty, K_m^{\pm\infty}$ and K_∞^∞ are Koszul and compute the associated Kazhdan-Lusztig polynomials explicitly; these polynomials are the entries of the inverse of the q -decomposition matrix mentioned in the previous paragraph. Projective functors also allow us to prove the double centraliser property that justifies our earlier statement that K_m^n is a quasi-hereditary cover of H_m^n .

In Part III, we give a new proof of a known result relating the algebra K_m^n to the blocks of a certain category $\mathcal{O}(m, n)$, namely, the parabolic analogue of the Bernstein-Gelfand-Gelfand category \mathcal{O} associated to the subalgebra $\mathfrak{gl}_m(\mathbb{C}) \oplus \mathfrak{gl}_n(\mathbb{C})$ in $\mathfrak{gl}_{m+n}(\mathbb{C})$ (we consider only integral central characters). In [S], the second author showed that the algebra K_m^n is isomorphic to the endomorphism algebra of a projective generator for the category of perverse sheaves on the Grassmannian of m -dimensional subspaces in an $(m+n)$ -dimensional complex vector space. This was proved by identifying the diagrammatically defined algebra K_m^n with a very different presentation of this endomorphism algebra due to Braden [B]. Combined with the Beilinson-Bernstein localisation theorem and the Riemann-Hilbert correspondence, this means that K_m^n is also the endomorphism algebra of a projective generator for a regular block of $\mathcal{O}(m, n)$. Our new approach to the proof of this statement bypasses geometry entirely: it is based instead on the generalised Schur-Weyl duality from [BK] which connects $\mathcal{O}(m, n)$ to degenerate cyclotomic Hecke algebras of level two.

We actually prove a more general result giving an equivalence between the entire category $\mathcal{O}(m, n)$ and a category introduced in [HK], [C]. As a by-product we recover an old result of Enright and Shelton formulated in a special case in [ES, Proposition 11.2]: singular blocks of $\mathcal{O}(m, n)$ are equivalent to regular blocks for smaller m and n . Moreover, we show that the generalised Khovanov algebras H_m^n are Morita equivalent to the blocks of degenerate cyclotomic Hecke algebras of level two, and prove a very recent conjecture of Khovanov and Lauda from [KL, §3.4] in the case of level two weights in finite type A.

In Part IV, we relate the limiting versions K_r^∞ to blocks of the general linear supergroup $GL(m|n)$, a setting where a geometric approach via some sort of localisation theorem is missing. More precisely, we prove that the (locally finite) endomorphism algebra of a projective generator for a block of $GL(m|n)$ of atypicality r is isomorphic to the algebra K_r^∞ . Combined with results from Part II, this implies that blocks of $GL(m|n)$ are Koszul. It also shows that all blocks of the same atypicality are equivalent, which is an old unpublished result of Serganova. The strategy for the proof in Part IV is similar in spirit to the argument in Part III, being based on a new “super” Schur-Weyl duality relating the general linear supergroup to the same degenerate cyclotomic Hecke algebras of level two mentioned before. By combining the results of Parts III and IV, one can deduce a proof of the so-called *super duality conjecture* of Cheng, Wang and Zhang [CWZ]. In fact this conjecture was the original motivation leading us to the limiting versions of Khovanov’s algebra.

Since the first version of this article became available, there have been several subsequent developments, some of which we mention briefly here. For an interpretation of Khovanov’s algebra in terms of the geometry of the Springer fibre we refer to [SW]. In that work, weights are interpreted as fixed points of a torus action on a Springer fibre, the oriented circle diagrams defined below correspond to fixed points lying inside the intersection $C_1 \cap C_2$ of two irreducible components hence index a basis of $H^*(C_1 \cap C_2)$, and Khovanov’s combinatorial multiplication rule arises from some convolution product. Our diagrammatics for parabolic category \mathcal{O} in Part III has recently been extended to the Hermitian symmetric pair (D_n, A_{n-1}) by Lejczyk [L]. Intriguingly the diagrams in [L] are very similar to ones appearing in another recent paper by Cox-De Visscher [CD] on combinatorial representation theory of Brauer algebras. This work of Cox-De Visscher also suggests that there should be a Morita equivalence between blocks of the walled Brauer algebras and certain idempotent truncations of the algebras K_∞ . We have recently constructed such an equivalence via an application of the results about $GL(m|n)$ from Part IV; see [BS] for details. Our diagrammatics for $GL(m|n)$ has been extended to cover blocks of the orthosymplectic supergroup $OSp(m|2n)$ by Gruson and Serganova [GS]. Finally we remark that a more direct proof of the super duality conjecture mentioned above has now been found by Cheng and Lam [CL].

We devote the remainder of this introduction to giving an informal guide to the construction of the algebras H_m^n , K_m^n and their limiting versions. We hope the baby examples sketched here make the precise definitions explained in the main body of the article more accessible. As a vector space, Khovanov’s algebra

H_n^n is usually identified with

$$\bigoplus_{ab} [\mathbb{C}[x]/(x^2)]^{\otimes (\text{the number of circles in } ab)}$$

where the direct sum is over all closed circle diagrams ab obtained by gluing a cup diagram a with n cups under a cap diagram b with n caps; see §2 for all this language. The summand of H_n^n corresponding to the circle diagram ab has a distinguished basis consisting of tensors in 1 and x . In the literature, this distinguished basis is often parametrised by decorating each circle in ab by an additional type, 1 or x . In this article, we instead encode circles of types 1 or x by orienting them so that they are anti-clockwise or clockwise, respectively. For example, we represent the distinguished basis $\{1, x\}$ of $H_1^1 = \mathbb{C}[x]/(x^2)$ by the closed oriented circle diagrams

$$1 \leftrightarrow \begin{array}{c} \curvearrowright \\ \hline \curvearrowleft \end{array} \quad x \leftrightarrow \begin{array}{c} \curvearrowleft \\ \hline \curvearrowright \end{array}$$




















In fact H_1^1 equals $\mathbb{C}[x]/(x^2)$ as an algebra, making it a deceptively simple example. Khovanov's definition of the multiplication on H_n^n in general involves a 2-dimensional TQFT arising from $\mathbb{C}[x]/(x^2)$. In this article, we have reformulated Khovanov's definition in a slightly different way. Here is an example giving the product of two basis vectors in H_2^2 :

Roughly, to calculate this product in our setup, the idea is to draw the first diagram under the second and then perform two “surgery procedures” following rules explained in detail in §3.

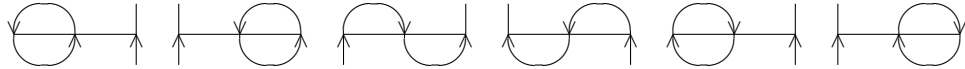
Like in the above pictures, a choice of orientation on a circle diagram gives rise to a “weight”, that is, a sequence of \vee 's and \wedge 's on the horizontal “number line” that separates cups from caps. So it makes sense to denote oriented circle diagrams by triples $a\lambda b$, where a is the cup diagram at the bottom, λ is the weight in the middle encoding the orientation, and b is the cap diagram at the top. The weights arising from H_n^n have n \vee 's and n \wedge 's; for example, the set of all weights for H_2^2 is $\Lambda := \{\vee\vee\wedge\wedge, \vee\wedge\vee\wedge, \wedge\vee\vee\wedge, \vee\wedge\wedge\vee, \wedge\vee\wedge\vee, \wedge\wedge\vee\vee\}$. There is a partial order \leq on Λ generated by swapping $\vee\wedge$ pairs. In the H_2^2 example displayed in the previous paragraph, the weight $\wedge\vee\wedge\vee$ on the right hand side is strictly greater than the weight $\vee\wedge\vee\wedge$ on the left hand side. This is a general phenomenon: it is *always* the case that the product $(a\lambda b)(c\mu d)$ of two basis vectors is a linear combination $(a\nu d)$ of basis vectors with $\lambda \leq \nu \leq \mu$. This leads to the cellular structure on H_n^n ; see §3.

For $m < n$, the generalised Khovanov algebra H_m^n has a very similar cellular basis parametrised by (no longer closed) oriented circle diagrams $a\lambda b$. The cup diagram a at the bottom now has m cups and $(n - m)$ rays going down to infinity, the weight λ in the middle encoding the orientation has m \vee 's and n \wedge 's, and the cap diagram b at the top has m caps and $(n - m)$ rays going up to

Table 1: Multiplication in K_1^1

K_1^1					
			0	0	0
	0	0	0		0
			0	0	0
	0	0			
	0	0	0		0

infinity. For example, H_1^2 has basis parametrised by the following six diagrams:



The multiplication on H_n^m can be defined by a slightly generalised surgery procedure; see §6 for details. We have not yet mentioned the natural \mathbb{Z} -grading on all our algebras. This can be defined simply by declaring that the degree of a basis vector is equal to the total number of clockwise cups and caps in its diagram. The above basis vectors for H_1^2 are of degrees 0, 0, 1, 1, 2, 2. As this example suggests, all H_m^n are symmetric algebras; see §6.

Now we can outline the construction of the quasi-hereditary covers K_m^n of H_m^n . These also have bases parametrised by oriented circle diagrams $a\lambda b$, and the weights λ decorating the number line again have m \vee 's and n \wedge 's. However we now allow cup and cap diagrams a and b with fewer cups and caps than required before, compensating by adding additional rays to infinity. There is also one new constraint on orientations: we do not allow two rays to infinity to be oriented $\vee\wedge$ in that order from left to right. The multiplication on K_m^n is somewhat harder to define than on H_m^n ; the basic idea going back to Khovanov and Braden is to construct K_m^n as a quotient of H_{m+n}^{m+n} . There is also a direct way to compute multiplication in K_m^n in terms of the generalised surgery procedure which is much quicker in practise, though from a theoretical point of view the quotient construction is better. We give an example illustrating the quotient construction in the next paragraph, referring the reader to §4 for details and a proof of cellularity and §6 for the quicker way to compute multiplication. In §5, we show K_m^n is a quasi-hereditary algebra and construct its projective indecomposable, standard and irreducible modules explicitly; all three families are parametrised by the same underlying set Λ of weights.

The algebra K_1^1 is a certain five dimensional algebra that is well known to category \mathcal{O} experts; see Table 1 for its full multiplication table in terms of diagrams. The entry

$$\begin{array}{c} \downarrow \\ \cup \\ \downarrow \end{array} \times \begin{array}{c} \uparrow \\ \cap \\ \uparrow \end{array} = \begin{array}{c} \circlearrowleft \\ \cap \\ \circlearrowright \end{array}$$

of this table is computed as follows. First “close” the diagrams to be multiplied by adding one extra vertex labelled \vee to the left end and one labelled \wedge to the right end of the number line, then adding some new anti-clockwise cups and caps connecting the rays into closed circles through these new vertices:

$$\begin{array}{c} \downarrow \\ \cup \\ \downarrow \end{array} \rightsquigarrow \begin{array}{c} \downarrow \\ \cup \\ \downarrow \\ \cup \\ \downarrow \end{array} \quad \begin{array}{c} \uparrow \\ \cap \\ \uparrow \end{array} \rightsquigarrow \begin{array}{c} \uparrow \\ \cap \\ \uparrow \\ \cap \\ \uparrow \end{array}$$

This produces the two closed oriented circle diagrams which we already multiplied together in H_2^2 above. Finally “open” the result to get back to a vector in K_1^1 by removing the left and right vertices (and the arcs passing through them) from all the diagrams in which the new vertices are still oriented \vee and \wedge , respectively, and mapping all other diagrams to zero:

$$\begin{array}{c} \downarrow \\ \cup \\ \downarrow \end{array} + \begin{array}{c} \uparrow \\ \cap \\ \uparrow \end{array} \rightsquigarrow \begin{array}{c} \downarrow \\ \cup \\ \downarrow \end{array} + 0$$

Now we sketch the definitions of limiting versions. First there are infinite dimensional versions H_m^∞ and $H_m^{\pm\infty}$ of H_m^n , defined by extending the number line infinitely far in both directions for H_m^∞ , infinitely far to the right for $H_m^{+\infty}$, or infinitely far to the left for $H_m^{-\infty}$. In the diagrams parametrising the bases of these algebras, weights have infinitely many \wedge 's and exactly m \vee 's, and cup/cap diagrams have exactly m cups/caps as before but infinitely many rays. Similarly, there are infinite dimensional versions K_m^∞ , $K_m^{\pm\infty}$ and K_∞ of K_m^n , the last of which has infinitely many \wedge 's and infinitely many \vee 's but finitely many cups and caps in its diagram basis. In all cases, the multiplications in these infinite dimensional versions can be computed effectively using the generalised surgery procedure from §6. However, it is usually best to view the algebras K_m^∞ , $K_m^{\pm\infty}$ and K_∞ as direct limits of finite dimensional K_n^m 's as is explained formally in §4, and then the algebras H_m^∞ and $H_m^{\pm\infty}$ emerge naturally as subalgebras. In fact it happens that $H_m^\infty = K_m^\infty$ and $H_m^{+\infty} = K_m^{+\infty}$, which is why we know that these algebras are both quasi-hereditary and symmetric.

One final point. In the main body of the article, we will adopt a slightly different notation to this introduction, denoting the important algebras by H_Λ and K_Λ where Λ is the set (or “block”) of weights being considered. We do this because it provides a little extra flexibility, allowing us to incorporate additional combinatorial data into the notion of weight by labeling some of the vertices on the number line by \circ 's and \times 's. These extra vertices are “dummies” that play absolutely no role in this article but they play a key role in Parts III and IV:

the extra labels encode the combinatorics of singular blocks on the category \mathcal{O} and general linear supergroup sides. There are obvious isomorphisms (“delete all \circ 's and \times 's”) between the algebras H_Λ and K_Λ studied later on and the algebras H_m^n and K_m^n described in this introduction.

Notation. In the rest of the article we work over a fixed ground field \mathbb{F} , and gradings mean \mathbb{Z} -gradings.

2. DIAGRAMS

In this section we introduce the basic combinatorial notions of weights, blocks, cup diagrams and cap diagrams, and circle diagrams. We then explain how cup, cap and circle diagrams can be “oriented” by decorating them with an additional weight.

Weights and blocks. By a *number line* we mean a horizontal line containing some *vertices* indexed by a set of consecutive integers in increasing order from left to right. A *weight* is a diagram obtained by labeling each of the vertices on such a number line by \circ (nought), \times (cross), \vee (down) or \wedge (up) in such a way that outside of some finite subset of the vertices it is impossible to find two (not necessarily neighbouring) vertices that are labelled by $\vee \wedge$ in that order from left to right. As mentioned at the end of the introduction, the labels \circ and \times do not have any structural significance in this article (but are needed in the sequels), so the reader can safely ignore them on first reading.

A weight λ is *bounded* if there are only finitely many vertices. In that case, we let λ^* denote the weight obtained from λ by reversing the orientations on all the vertices, i.e. changing all \vee 's to \wedge 's and \wedge 's to \vee 's. In any case, we let λ^\frown denote the weight obtained from λ by rotating the diagram through 180° . For example:

$$\begin{aligned} \lambda &= \times \text{---} \circ \text{---} \wedge \text{---} \vee \text{---} \vee \text{---} \times \text{---} \wedge \text{---} \circ \text{---} \vee \\ \lambda^* &= \times \text{---} \circ \text{---} \vee \text{---} \wedge \text{---} \wedge \text{---} \times \text{---} \vee \text{---} \circ \text{---} \wedge \\ \lambda^\frown &= \wedge \text{---} \circ \text{---} \vee \text{---} \times \text{---} \wedge \text{---} \wedge \text{---} \vee \text{---} \circ \text{---} \times \end{aligned}$$

The *Bruhat order* on the set of weights is the partial order \leq generated by the basic operation of swapping a \vee and an \wedge ; getting bigger in the Bruhat order means \vee 's move to the right. We define an equivalence relation \sim on the set of all weights by declaring that $\lambda \sim \mu$ if μ can be obtained from λ by permuting \vee 's and \wedge 's (and doing nothing to \circ 's and \times 's). In other words, $\lambda \sim \mu$ if and only if there exists a weight ν such that $\lambda \leq \nu \leq \mu$ in the Bruhat order. We refer to the \sim -equivalence classes of weights as *blocks*.

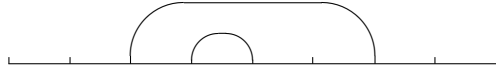
Cup and cap diagrams. A *cup diagram* is a diagram obtained by attaching rays and finitely many cups to some subset of the vertices on a number line, so that rays join vertices down to infinity, cups are lower semi-circles joining pairs

of vertices, and no rays and/or cups intersect. Here is an example:



Two cup diagrams are *equal* if their underlying number lines are the same and all their rays and cups are incident with exactly the same vertices (regardless of the precise shapes of the cups). A vertex that is not at the end of some cup or ray will be called a *free vertex*, indicated on the diagram by a vertical mark.

The notion of a *cap diagram* is entirely similar, replacing cups with caps, i.e. upper semi-circles, and drawing rays up to infinity rather than down. For example, the following is a cap diagram:

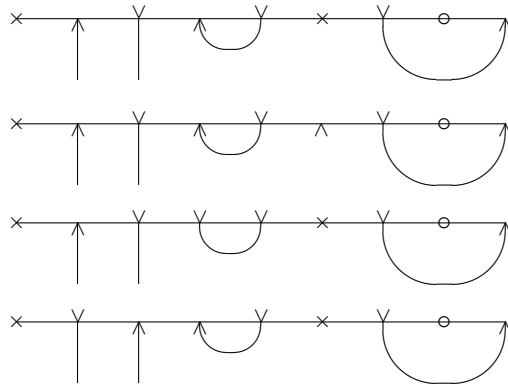


A *closed* cup or cap diagram means one containing no rays. We denote the mirror image of a cup or cap diagram c in the number line by c^* , which is a cap or cup diagram, respectively. Also c^\frown denotes the diagram obtained by rotating c through 180° .

If c is a cup diagram and λ is a weight with the same underlying number line, we can glue c under λ to obtain a new diagram denoted $c\lambda$. We call $c\lambda$ an *oriented cup diagram* if

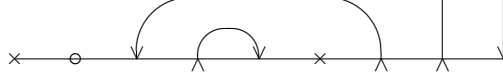
- ▶ each free vertex in c is labelled either \circ or \times in the weight λ ;
- ▶ the vertices at the ends of each cup in c are labelled by exactly one \vee and one \wedge in the weight λ ;
- ▶ each vertex at the top of a ray in c is labelled either \vee or \wedge in the weight λ ;
- ▶ it is impossible to find two rays in c whose top vertices are labelled $\vee \wedge$ in that order from left to right in the weight λ .

For example, the first of following diagrams is oriented, the other three are not (and note we now omit the vertical marks for free vertices):



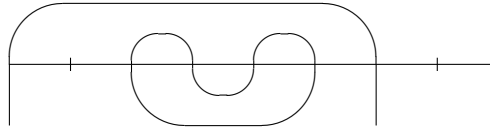
Instead, if c is a cap diagram and λ is a weight with the same underlying number line, we can glue λ under c to obtain a diagram denoted λc . We call

λc an *oriented cap diagram* if $c^*\lambda$ is an oriented cup diagram according to the previous definition. The following is an example of an oriented cap diagram:



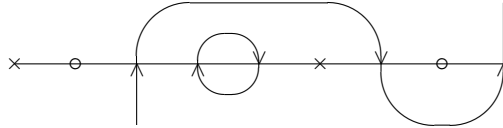
We describe a cup or a cap in any such oriented diagram as *anti-clockwise* or *clockwise* according to whether its leftmost vertex is labelled \vee or \wedge (equivalently, its rightmost vertex is labelled \wedge or \vee). For example, in the above oriented cap diagram, the small cap is clockwise and the big cap is anti-clockwise.

Circle diagrams. By a *circle diagram* we mean a diagram of the form ab obtained by gluing a cup diagram a underneath a cap diagram b , assuming that a and b have the same underlying number lines and their free vertices are in all the same positions. For example:



Any such diagram is a union of *circles* and *lines* (both viewed up to homeomorphism), e.g. the above circle diagram contains one circle and two lines. We call it a *closed circle diagram* if it involves only circles and no lines. Also a *small circle* means a circle consisting of just one cap and one cup.

An *oriented circle diagram* means a diagram $a\lambda b$ obtained by gluing an oriented cup diagram $a\lambda$ underneath an oriented cap diagram λb . Here is an example:



We refer to a circle in an oriented circle diagram as

- ▶ an *anti-clockwise circle* or a circle of *type 1* if the leftmost vertex on the circle is labelled by \vee (equivalently, the rightmost vertex is labelled by \wedge);
- ▶ a *clockwise circle* or a circle of *type x* if the leftmost vertex on the circle is labelled by \wedge (equivalently, the rightmost vertex is labelled by \vee).

Degrees. The *degree* of an oriented cup diagram $a\lambda$ or an oriented cap diagram λb means the total number of clockwise cups and caps that it contains. The *degree* of an oriented circle diagram $a\lambda b$ is defined by

$$\deg(a\lambda b) := \deg(a\lambda) + \deg(\lambda b). \tag{2.1}$$

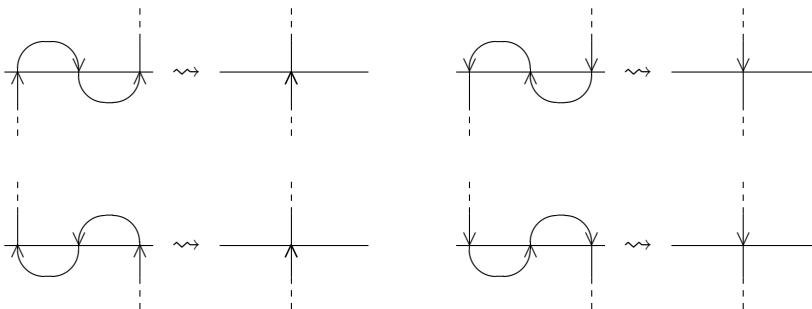
Equivalently, $\deg(a\lambda b)$ is the sum of the degrees of the circles and lines in the diagram, where the degree of a single circle or line in an oriented circle diagram means its total number of clockwise cups and caps. For example, the oriented circle diagram displayed at the end of the previous subsection is of degree 3.

Lemma 2.1. *The degree of an anti-clockwise circle in an oriented circle diagram is one less than the total number of caps that it contains. The degree of a clockwise circle is one more than the total number of caps that it contains.*

Proof. Proceed by induction on the number of caps contained in the circle. For the base case, circles with just one cap are small circles and the lemma is clear:

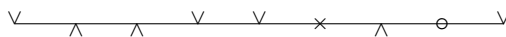
$$\deg \left(\begin{array}{c} \circlearrowleft \\ \text{---} \end{array} \right) = 0, \quad \deg \left(\begin{array}{c} \circlearrowright \\ \text{---} \end{array} \right) = 2.$$

For the induction step, remove a kink using one of the following straightening rules:

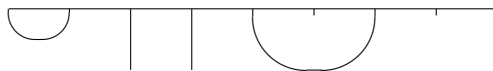


The result is a new oriented circle drawn on a number line with two fewer vertices than before. Moreover, the new circle is oriented in the same way as the original circle, and it either has one less clockwise cup and one less anti-clockwise cap, or it has one less anti-clockwise cup and one less clockwise cap. This information suffices to complete the proof of the induction step. \square

The cup diagram associated to a weight. We will often use weights to parametrise cup diagrams via the following important construction: for each weight λ there is a unique cup diagram denoted $\underline{\lambda}$ with the property that $\underline{\lambda}\lambda$ is an oriented cup diagram of degree 0, i.e. all of its cups are anti-clockwise. To compute $\underline{\lambda}$ in practice, find a pair of vertices labelled $\vee \wedge$ in order from left to right that are neighbours in the sense that there are only \circ 's, \times 's or vertices already joined by cups in between. Join these new vertices together with a cup then repeat the process until there are no more such $\vee \wedge$ pairs (this will happen after finally many repetitions). Finally draw rays down to infinity at all the remaining \wedge 's and \vee 's. For example, if λ is



then $\underline{\lambda}$ equals



Similarly, $\bar{\lambda} := (\underline{\lambda})^*$ denotes the unique cap diagram such that $\lambda\bar{\lambda}$ is an oriented cap diagram of degree 0, i.e. all its caps are anti-clockwise.

Lemma 2.2. *If $\lambda \sim \mu$ and $\underline{\lambda} = \underline{\mu}$ then $\lambda = \mu$.*

Proof. As $\underline{\lambda} = \underline{\mu}$, all the cups in the oriented cup diagrams $\underline{\lambda}\lambda$ and $\underline{\lambda}\mu$ are anti-clockwise. Hence the labels on the vertices at the ends of each cup of $\underline{\lambda}$ are the same in both λ and μ . On the remaining vertices, all \wedge 's must be to the left of all \vee 's. So since $\lambda \sim \mu$ all these remaining labels are also the same in λ and μ . \square

For weights λ and μ , we use the notation $\mu \subset \lambda$ to indicate that $\mu \sim \lambda$ and $\underline{\mu}\lambda$ is an oriented cup diagram. Sometimes we write $\lambda \supset \mu$ instead of $\mu \subset \lambda$, so we have that $\lambda \supset \mu$ if $\lambda \sim \mu$ and $\lambda\bar{\mu}$ is an oriented cap diagram. Using the first statement from the following lemma, it is easy to check that the relation \subset just defined is reflexive and anti-symmetric; its transitive closure is the Bruhat order \leq . (Perhaps now is a good moment to recall that our cup diagrams have only finitely many cups; this is crucial for everything in the rest of the section.)

Lemma 2.3. *If $\mu \subset \lambda$ then $\mu \leq \lambda$ in the Bruhat order. Moreover,*

- (i) *if $a\lambda$ is an oriented cup diagram then $a = \underline{\alpha}$ for a unique weight α with $\alpha \subset \lambda$;*
- (ii) *if λb is an oriented cap diagram then $b = \bar{\beta}$ for a unique weight β with $\lambda \supset \beta$;*
- (iii) *if $a\lambda b$ is an oriented circle diagram then $a = \underline{\alpha}$ and $b = \bar{\beta}$ for unique weights α, β with $\alpha \subset \lambda \supset \beta$.*

Proof. Let $a\lambda$ be an oriented cup diagram. We claim that $a = \underline{\alpha}$ for some weight α such that $\alpha \leq \lambda$. To see this, reverse the orientations on the vertices at the ends of each clockwise cup in the diagram $a\lambda$ to obtain a new oriented cup diagram involving only anti-clockwise cups. This new diagram is of the form $a\alpha$ for some weight α , and since it is of degree 0 we necessarily have that $a = \underline{\alpha}$. Whenever a clockwise cup is converted to an anti-clockwise cup during this process, a pair of vertices labelled $\wedge\vee$ change to $\vee\wedge$, so the weight gets smaller in the Bruhat ordering. Hence we have that $\alpha \leq \lambda$ and the claim follows.

If we assume now $\mu \subset \lambda$ and apply the claim from the previous paragraph to $a := \underline{\mu}$, we get that $\underline{\mu} = \underline{\alpha}$ for some weight α with $\alpha \leq \lambda$. Since $\mu \sim \lambda \sim \alpha$ we actually have that $\underline{\mu} = \alpha$ by Lemma 2.2. Hence $\mu \subset \lambda$ implies $\mu \leq \lambda$ as required. Next we prove (i). If $a\lambda$ is any oriented cup diagram, the claim in the previous paragraph shows that $a = \underline{\alpha}$ for some $\alpha \subset \lambda$. For the uniqueness, suppose also that $a = \underline{\beta}$ for $\beta \subset \lambda$. Then $\alpha \sim \lambda \sim \beta$ and $\underline{\alpha} = \underline{\beta}$, hence $\alpha = \beta$ by Lemma 2.2. The proof of (ii) is similar. Then (iii) follows immediately from (i) and (ii). \square

Defect. For a weight λ , the *defect* $\text{def}(\lambda) \in \{0, 1, 2, \dots\}$ means the (necessarily finite) number of cups in the cup diagram $\underline{\lambda}$. The defect of a block Λ is then

$$\text{def}(\Lambda) := \sup\{\text{def}(\lambda) \mid \lambda \in \Lambda\} \in \{0, 1, 2, \dots\} \cup \{\infty\}. \quad (2.2)$$

It is easy to see that $\text{def}(\Lambda)$ is simply the number of \wedge 's or the number of \vee 's in the weights belonging to Λ , whichever is smaller.

Lemma 2.4. *Let Λ be a block and fix $\lambda \in \Lambda$.*

- (i) *The number of weights μ such that $\mu \supset \lambda$ is equal to $2^{\text{def}(\lambda)}$.*

- (ii) The number of weights μ such that $\mu \subset \lambda$ is finite if and only if $\text{def}(\Lambda) < \infty$.
- (iii) For each fixed degree $j \in \mathbb{Z}$, the number of weights μ such that $\mu \subset \lambda$ and $\text{deg}(\underline{\mu}\lambda) = j$ is finite (even if $\text{def}(\Lambda) = \infty$).

Proof. (i) Start from the oriented cap diagram $\lambda\bar{\lambda}$. Choose some subset S of the caps in this diagram and reverse the orientations of the vertices at the ends of each cap in S . The result is an oriented cap diagram of the form $\mu\bar{\lambda}$ for some weight μ in the same block as λ . Moreover, every weight μ with $\mu \supset \lambda$ is obtained in this way from some choice of S . It remains to observe that there are $2^{\text{def}(\lambda)}$ different choices for S because $\bar{\lambda}$ has $\text{def}(\lambda)$ caps.

(ii) In view of Lemma 2.3(i), the number of weights μ such that $\mu \subset \lambda$ is the same as the number of cup diagrams a such that $a\lambda$ is an oriented cup diagram. If $\text{def}(\Lambda) = \infty$ then λ has infinitely many \wedge 's and \vee 's. So there exist oriented cup diagrams of the form $a\lambda$ having an arbitrary large number of cups. Conversely, if $\text{def}(\Lambda) < \infty$ then either λ has finitely many \wedge 's or finitely many \vee 's (or both). In the former case, suppose $a\lambda$ is an oriented cup diagram. At each of the finitely many vertices labelled \wedge in λ , there must be either a ray or the leftmost end of a cup or the rightmost end of a cup in a . As rays are not allowed to cross cups, the cup diagram a is uniquely determined by λ and this additional information, so there are only finitely many possibilities for a (at most three to the power the number of \wedge 's in λ). The argument when λ has finitely many \vee 's is analogous.

(iii) In view of (ii), we only need to consider the case that $\text{def}(\Lambda) = \infty$. We need to show that there are only finitely many cup diagrams a such that $a\lambda$ is oriented and $\text{deg}(a\lambda) = j$. To see this, note that far to the left of λ there are only \wedge 's and far to the right there are only \vee 's. So if $a\lambda$ is oriented the \wedge 's far to the left can only be at the end of a ray or at the leftmost end of a clockwise cup, and similarly the \vee 's far to the right can only be at the end of a ray or at the rightmost end of a clockwise cup. Given a fixed bound j on the total number of clockwise cups in a , there are still only finitely many possible a 's. \square

3. CELLULARITY OF KHOVANOV'S DIAGRAM ALGEBRA

Throughout the section, assume that Λ is a *Khovanov block* of rank $n \geq 0$, meaning that it is a block consisting of bounded weights that have exactly n \wedge 's and n \vee 's. We are going to associate a certain algebra H_Λ to Λ . The construction of this algebra is due to Khovanov [K2, §2.4], but we will explain it in detail since our viewpoint is rather more combinatorial in nature than Khovanov's. Using the new viewpoint, we then prove that H_Λ is cellular.

The underlying graded vector space. Define Λ° to be the subset of Λ consisting of all weights with the property that at each vertex the number of \vee 's labeling vertices to the left or equal to the given vertex is greater than or equal to the number of \wedge 's labeling these vertices. In other words,

$$\begin{aligned} \Lambda^\circ &= \{\alpha \in \Lambda \mid \underline{\alpha} \text{ is a closed cup diagram}\} \\ &= \{\beta \in \Lambda \mid \bar{\beta} \text{ is a closed cap diagram}\}. \end{aligned}$$

We note that $|\Lambda^\circ| = \frac{1}{n+1} \binom{2n}{n}$, the n th Catalan number; see e.g. [St, ex.6.19]. Let H_Λ denote the vector space with basis

$$\{(\underline{\alpha\lambda\bar{\beta}}) \mid \text{for all } \alpha \in \Lambda^\circ, \lambda \in \Lambda, \beta \in \Lambda^\circ \text{ such that } \alpha \subset \lambda \supset \beta\}. \quad (3.1)$$

Equivalently, by Lemma 2.3(iii), the basis of H_Λ is

$$\{(a\lambda b) \mid \text{for all closed oriented circle diagrams } a\lambda b \text{ with } \lambda \in \Lambda\}. \quad (3.2)$$

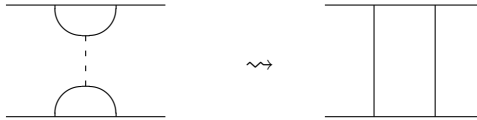
We put a grading on H_Λ by defining the degree of the basis vector $(a\lambda b)$ to be $\deg(a\lambda b)$; see (2.1). Writing $e_\lambda := (\underline{\lambda\lambda\bar{\lambda}})$ for short, the vectors

$$\{e_\lambda \mid \lambda \in \Lambda^\circ\} \quad (3.3)$$

form a basis for the degree 0 component of the graded vector space H_Λ .

Multiplication. Now we define a multiplication making the graded vector space H_Λ into a positively graded associative algebra. Take two closed oriented circle diagrams $a\lambda b$ and $c\mu d$ with $\lambda, \mu \in \Lambda$. If $b^* \neq c$, we simply define the product $(a\lambda b)(c\mu d)$ to be zero. Now assume that $b^* = c$. Draw the diagram $a\lambda b$ underneath the diagram $c\mu d$ to obtain a new sort of diagram with two number lines and a symmetric middle section containing b under c . Then iterate the *surgery procedure* explained in the next paragraph to convert this diagram into a disjoint union of diagrams none of which have any cups or caps left in their middle sections. Finally, collapse each of the resulting diagrams by identifying their top and bottom number lines to obtain a disjoint union of some new closed oriented circle diagrams. The product $(a\lambda b)(c\mu d)$ is then defined to be the sum of the corresponding basis vectors of H_Λ .

The surgery procedure. Choose a symmetric pair of a cup and a cap in the middle section of a diagram that can be connected without crossings. Forgetting orientations for a while, cut open the cup and the cap then stitch the loose ends together to form a pair of vertical line segments:



This surgery either combines two circles into one or splits one circle into two. Finally re-orient the new circle(s) according to the orientation of the old circle(s) using the following rules (remembering $1 =$ anti-clockwise, $x =$ clockwise)

$$1 \otimes 1 \mapsto 1, \quad 1 \otimes x \mapsto x, \quad x \otimes 1 \mapsto x, \quad x \otimes x \mapsto 0, \quad (3.4)$$

$$1 \mapsto 1 \otimes x + x \otimes 1, \quad x \mapsto x \otimes x. \quad (3.5)$$

to obtain a disjoint union of zero, one or two new oriented diagrams replacing the old diagram. For instance, the rule $1 \otimes 1 \mapsto 1$ means that two anti-clockwise circles transform to one anti-clockwise circle. The rule $1 \mapsto 1 \otimes x + x \otimes 1$ means that one anti-clockwise circle transforms to give a disjoint union of two different diagrams each having one anti-clockwise circle and one clockwise circle. Finally the rule $x \otimes x \mapsto 0$ means that two clockwise circles produce no diagram at all at the end of the surgery procedure.

Khovanov’s definition via TQFTs. In order to explain why the above construction makes H_Λ into a well-defined associative algebra, we need to explain in a little more detail the relationship to Khovanov’s original setup. Let 2-Cob be the cobordism category whose objects are closed oriented 1-manifolds and whose morphisms are homeomorphism classes of oriented 2-cobordisms. Let Vec be the category of vector spaces over \mathbb{F} . The commutative Frobenius algebra $\mathbb{F}[x]/(x^2)$ whose multiplication and comultiplication are defined by the formulae (3.4)–(3.5) defines a 2-dimensional TQFT, that is, a symmetric monoidal functor

$$\mathcal{F} : 2\text{-Cob} \rightarrow \text{Vec}. \quad (3.6)$$

We refer the reader to [K] for the detailed construction of \mathcal{F} , just noting here that it takes a disjoint union $X_1 \sqcup \cdots \sqcup X_r$ of r copies of S^1 to the tensor product of r copies of the vector space $\mathbb{F}[x]/(x^2)$. The monomials $t_1 \otimes \cdots \otimes t_r$ with each $t_i \in \{1, x\}$ give a canonical basis for the latter vector space.

Now suppose we are given a closed circle diagram $\underline{\alpha}\bar{\beta}$ with $\alpha, \beta \in \Lambda^\circ$. Letting r denote the number of circles in this diagram, we pick a homeomorphism from $\underline{\alpha}\bar{\beta}$ to a disjoint union $X_1 \sqcup \cdots \sqcup X_r$ of r copies of S^1 . This choice induces an isomorphism between the vector space $\mathcal{F}(\underline{\alpha}\bar{\beta})$ and the tensor product of r copies of $\mathbb{F}[x]/(x^2)$. Now the basis $\{t_1 \otimes \cdots \otimes t_r \mid t_1, \dots, t_r \in \{1, x\}\}$ for the latter tensor product pulls back to give us a basis for $\mathcal{F}(\underline{\alpha}\bar{\beta})$ too. It is natural to parametrise this basis simply by decorating the i th circle in the diagram $\underline{\alpha}\bar{\beta}$ by the additional *type* $t_i \in \{1, x\}$; equivalently, we orient the i th circle so that it is an anti-clockwise or clockwise circle according to whether $t_i = 1$ or $t_i = x$. In this way, the basis for the vector space $\mathcal{F}(\underline{\alpha}\bar{\beta})$ is parametrised by the set of closed oriented circle diagrams

$$\{\underline{\alpha}\lambda\bar{\beta} \mid \text{for all } \lambda \in \Lambda \text{ such that } \alpha \subset \lambda \supset \beta\}.$$

Now it makes sense to identify the vector space H_Λ from (3.1) with the vector space

$$\bigoplus_{\alpha, \beta \in \Lambda^\circ} \mathcal{F}(\underline{\alpha}\bar{\beta}), \quad (3.7)$$

since both spaces have distinguished bases parameterised by the same set.

Under this identification, the multiplication on H_Λ described above corresponds to the direct sum of linear maps

$$m_{\alpha, \beta, \gamma, \delta} : \mathcal{F}(\underline{\alpha}\bar{\beta}) \otimes \mathcal{F}(\underline{\gamma}\bar{\delta}) \rightarrow \mathcal{F}(\underline{\alpha}\bar{\delta})$$

for all $\alpha, \beta, \gamma, \delta \in \Lambda^\circ$ defined as follows. If $\beta \neq \gamma$, then $m_{\alpha, \beta, \gamma, \delta} = 0$. If $\beta = \gamma$, then $m_{\alpha, \beta, \gamma, \delta}$ is the linear map arising by applying the functor \mathcal{F} to the cobordism $\underline{\alpha}\bar{\beta} \sqcup \underline{\gamma}\bar{\delta} \rightarrow \underline{\alpha}\bar{\delta}$ constructed by composing “pair of pants” shaped cobordisms going from two to one or from one to two circles lifting the sequence of surgery procedures explained above. The advantage of this more topological point of view is that it is now clear that the multiplication is well defined independent of the order that the surgery procedures are performed, since the resulting cobordism is independent of the choice of order up to homeomorphism; see [K2, p.679] for a careful justification of the latter statement. In a similar fashion, the proof that the multiplication now defined is associative reduces

to checking that two corresponding cobordisms are homeomorphic, something which is apparent from their construction.

We still need to prove that the multiplication preserves the grading, a fact which is also proved on [K2, p.680] from the topological point of view. We have to verify that each surgery procedure preserves the total number of clockwise cups and caps in a diagram. In view of Lemma 2.1, this amounts to showing that

$$\#(\text{caps}) - \#(\text{anti-clockwise circles}) + \#(\text{clockwise circles})$$

is the same at the end of each surgery procedure as it was at the start. When two circles go to one circle, there is one less cap, one less anti-clockwise circle and the same number of clockwise circles at the end compared to the beginning. When one circle goes to two circles, there is one less cap, the same number of anti-clockwise circles and one more clockwise circle at the end compared to the beginning. The desired invariance follows directly.

Cellular algebra structure. We have now constructed the graded associative algebra H_Λ . It is clear from the definition that

$$e_\alpha(a\lambda b) = \begin{cases} (a\lambda b) & \text{if } \underline{\alpha} = a, \\ 0 & \text{otherwise,} \end{cases} \quad (a\lambda b)e_\beta = \begin{cases} (a\lambda b) & \text{if } b = \bar{\beta}, \\ 0 & \text{otherwise,} \end{cases}$$

for $\alpha, \beta \in \Lambda^\circ$ and any basis vector $(a\lambda b) \in H_\Lambda$. This implies that the vectors $\{e_\alpha \mid \alpha \in \Lambda^\circ\}$ are mutually orthogonal idempotents whose sum is the identity in H_Λ . The resulting direct sum decomposition

$$H_\Lambda = \bigoplus_{\alpha, \beta \in \Lambda^\circ} e_\alpha H_\Lambda e_\beta \quad (3.8)$$

is just the decomposition from (3.7) in different notation, and the summand $e_\alpha H_\Lambda e_\beta$ has basis $\{(\underline{\alpha}\lambda\bar{\beta}) \mid \text{for all } \lambda \in \Lambda \text{ such that } \alpha \subset \lambda \supset \beta\}$. Observe also that the linear maps $*$: $H_\Lambda \rightarrow H_\Lambda$ and \frown : $H_\Lambda \rightarrow H_\Lambda$ defined by

$$(a\lambda b)^* := (b^* \lambda a^*), \quad (3.9)$$

$$(a\lambda b)^\frown := (b^\frown \lambda^\frown a^\frown) \quad (3.10)$$

are graded algebra anti-isomorphisms, where $\Lambda^\frown := \{\lambda^\frown \mid \lambda \in \Lambda\}$.

Theorem 3.1. *Let $(a\lambda b)$ and $(c\mu d)$ be basis vectors of H_Λ . Then,*

$$(a\lambda b)(c\mu d) = \begin{cases} 0 & \text{if } b \neq c^*, \\ s_{a\lambda b}(\mu)(a\mu d) + (\dagger) & \text{if } b = c^* \text{ and } a\mu \text{ is oriented,} \\ (\dagger) & \text{otherwise,} \end{cases}$$

where

- (i) (\dagger) denotes a linear combination of basis vectors of H_Λ of the form $(a\nu d)$ for $\nu > \mu$;
- (ii) the scalar $s_{a\lambda b}(\mu) \in \{0, 1\}$ depends only on $a\lambda b$ and μ , but not on d ;
- (iii) if $b = \bar{\lambda} = c^*$ and $a\mu$ is oriented then $s_{a\lambda b}(\mu) = 1$.

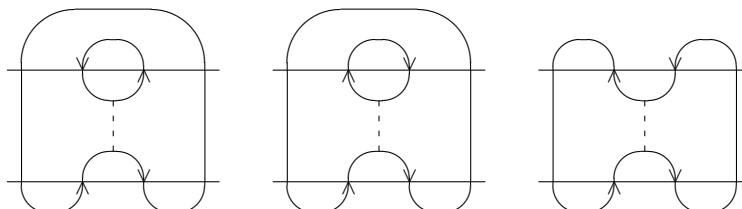
Proof. We already know that $(a\lambda b)(c\mu d) = 0$ if $b \neq c^*$, so assume $b = c^*$ from now on. Consider a single iteration of the surgery procedure in the algorithm

computing $(a\lambda b)(c\mu d)$. Let τ be the top weight (the weight on the top number line) of this diagram at the start of the surgery procedure. We claim that

- (1) the top weight of each diagram obtained at the end of the surgery procedure is greater than or equal to τ in the Bruhat order;
- (2) the total number of diagrams produced with top weight equal to τ is either zero or one, independent of the cap diagram d ;
- (3) if $a\tau$ is oriented and the cap to be cut is anti-clockwise then exactly one diagram is produced with top weight equal to τ .

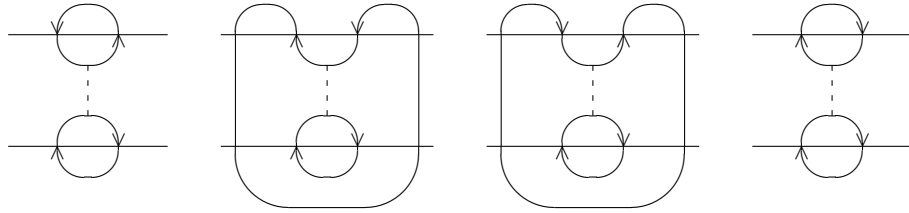
Before proving these things, let us explain how the theorem follows. By applying (1) repeatedly, starting with $\tau = \mu$ at the first step, it follows that $(a\lambda b)(c\mu d)$ is a linear combination of $(a\nu d)$'s for $\nu \geq \mu$. Assuming $a\mu$ is oriented, (2) applied repeatedly implies that the coefficient $s_{a\lambda b}(\mu)$ of the basis vector $(a\mu d)$ in the product is zero or one independent of the cap diagram d . This proves (i) and (ii). If in addition $b = \bar{\lambda}$ (so that *all* caps to be cut are anti-clockwise) then (3) applied repeatedly implies that $s_{a\lambda b}(\mu) = 1$, as in (iii). Therefore, it just remains to prove (1)–(3). To do this, we analyse three different situations depending on the orientations of the cap and the cup to be cut (both of which clearly do not depend on d).

Case one: the cap to be cut is clockwise. In this case, we show that the top weight of each diagram obtained at the end of the surgery procedure is strictly greater than τ , from which (1)–(3) follow immediately. If the circle containing the cap to be cut is anti-clockwise, the cap is necessarily a concave part of the circumference of that circle, and we are in one of three basic situations represented by the following pictures:



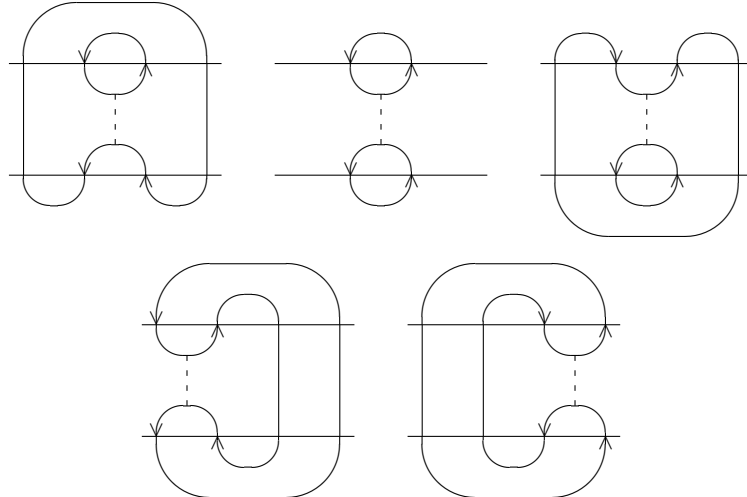
(We stress that these and subsequent pictures should be interpreted only up to homeomorphism; in particular the circles represented in the pictures may well cross both number lines many more times than indicated.) In the first of these cases, the orientation of each vertex lying on the anti-clockwise circle containing the cap to be cut gets reversed when re-orienting, hence the top weight gets strictly larger in the Bruhat order as required. For the second case, the orientation of each vertex from the anti-clockwise circle containing the cap to be cut gets reversed. Since this circle necessarily crosses the top number line, we see again that the top weight increases. For the final case, two anti-clockwise circles are produced when the initial surgery is performed. Then these circles are re-oriented so that orientation of all vertices lying on one or other of these anti-clockwise circles gets switched, so the top weight again goes up in the Bruhat order. Instead, assume the circle containing the cap to be cut is clockwise, so the cap is a convex part of the circumference. There are then

four basic situations:



In the first two of these, the orientation of every vertex lying on the anti-clockwise circle containing the cup to be cut gets switched, so the top weight gets bigger. The last two involve the rule $x \otimes x \mapsto 0$, so there is nothing to check.

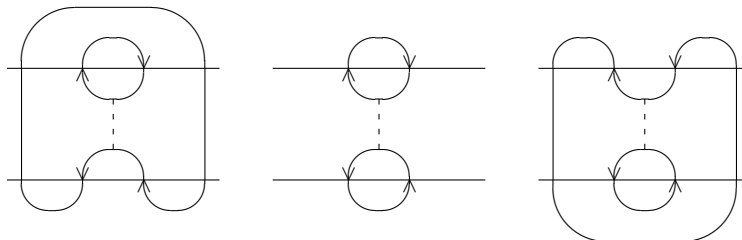
Case two: both the cap and the cup to be cut are anti-clockwise. Here we show there is exactly one new diagram produced having top weight equal to τ , together possibly with one other diagram with a strictly larger top weight. This is again enough to establish (1)–(3). There are five basic configurations to consider:



In the first three of these configurations, no vertices change orientation during the procedure, so we get precisely one diagram at the end, and it still has top weight equal to τ . For both the remaining two, we obtain a pair of nested circles, one anti-clockwise and the other clockwise, when the surgery is performed. This produces one new diagram with the top weight equal to τ . Re-orienting the diagram produces another new diagram obtained from this by reversing the orientations of both of these circles. Since the outside circle is anti-clockwise and the inside one is clockwise, the top weight of the resulting diagram is indeed strictly greater than τ in the Bruhat order.

Case three: the cap to be cut is anti-clockwise but the cup to be cut is clockwise. In this case, the outcome depends on whether the circle containing the cap to be cut crosses the top number line (something which is again independent of d) or not. We show if it does not cross the top number line then just one diagram, with top weight τ , is produced by the surgery procedure, and if it does cross the top number line then we only get diagrams with a strictly larger top

weight. This is enough to prove (1)–(2); then we will argue further in the next paragraph to get (3). To see these things, there are three remaining situations to be considered:



The first of these (in which the bottom circle definitely crosses the top number line) produces no diagrams at all as it is a situation when $x \otimes x \mapsto 0$. For the second two, the orientation of every vertex from the anti-clockwise circle containing the cap to be cut gets switched, so we get a diagram with strictly bigger top weight if the bottom circle crosses the top number line, and we get a diagram with top weight equal to τ if it does not.

We still need to complete the proof of (3) in case three. To do that, we must show that if $a\tau$ is oriented then the circle containing the cap to be cut definitely does not cross the top number line, hence by the previous paragraph exactly one diagram is definitely produced with top weight τ . Let σ be the bottom weight of the diagram at the start of the surgery procedure. Also let e be the cap diagram in the middle section of the diagram, so that the middle section of the diagram looks like e glued under e^* with the ends of matching rays joined together to form vertical line segments. Because the entire diagram is oriented at the start of the surgery procedure, both the diagrams $a\sigma e$ and $e^*\tau d$ are oriented. So under our assumption that $a\tau$ is oriented, we get that $a\sigma e$ and $a\tau e$ are oriented circle diagrams. Let i be the integer indexing the leftmost vertex of the cap to be cut. This vertex has different orientations in $a\sigma e$ and $a\tau e$. Hence all the other vertices that are joined to the vertex i have different orientations in $a\sigma e$ and $a\tau e$ too. Finally observe that the vertices in σ and τ at the ends of each vertical line segment in the middle section necessarily have the same orientation. Hence there cannot be any such line segment in the circle containing the cap to be cut, i.e. this circle does not cross the top number line. \square

Corollary 3.2. *The product $(a\lambda b)(c\mu d)$ of two basis vectors of H_Λ is a linear combination of vectors of the form $(a\nu d)$ for $\nu \in \Lambda$ with $\lambda \leq \nu \leq \mu$.*

Proof. By Theorem 3.1(i), $(a\lambda b)(c\mu d)$ is a linear combination of $(a\nu d)$'s for various $\nu \geq \mu$ and $(d^*\mu c^*)(b^*\lambda a^*)$ is a linear combination of $(d^*\nu a^*)$'s for various $\nu \geq \lambda$. Applying the anti-automorphism $*$ from (3.9) to the latter statement gives that $(a\lambda b)(c\mu d)$ is a linear combination of $(a\nu d)$'s for various $\nu \geq \lambda$ too. \square

Corollary 3.3. *Khovanov's algebra H_Λ is a cellular algebra in the sense of Graham and Lehrer [GL] with cell datum $(\Lambda, M, C, *)$ where*

- (i) $M(\lambda)$ denotes $\{\alpha \in \Lambda^\circ \mid \alpha \subset \lambda\}$ for each $\lambda \in \Lambda$;
- (ii) C is defined by setting $C_{\alpha, \beta}^\lambda := (\underline{\alpha\lambda\bar{\beta}})$ for $\lambda \in \Lambda$ and $\alpha, \beta \in M(\lambda)$;

(iii) $*$ is the anti-automorphism from (3.9).

Before we prove the corollary let us first recall the relevant definitions from [GL]. A *cellular algebra* means an associative unital algebra H together with a *cell datum* $(\Lambda, M, C, *)$ such that

- (1) Λ is a partially ordered set and $M(\lambda)$ is a finite set for each $\lambda \in \Lambda$;
- (2) $C : \bigcup_{\lambda \in \Lambda} M(\lambda) \times M(\lambda) \rightarrow H, (\alpha, \beta) \mapsto C_{\alpha, \beta}^\lambda$ is an injective map whose image is a basis for H ;
- (3) the map $*$: $H \rightarrow H$ is an algebra anti-automorphism such that $(C_{\alpha, \beta}^\lambda)^* = C_{\beta, \alpha}^\lambda$ for all $\lambda \in \Lambda$ and $\alpha, \beta \in M(\lambda)$;
- (4) if $\mu \in \Lambda$ and $\gamma, \delta \in M(\lambda)$ then for any $x \in H$ we have that

$$xC_{\gamma, \delta}^\mu \equiv \sum_{\gamma' \in M(\mu)} r_x(\gamma', \gamma) C_{\gamma', \delta}^\mu \pmod{H(> \mu)}$$

where the scalar $r_x(\gamma', \gamma)$ is independent of δ and $H(> \mu)$ denotes the subspace of H generated by $\{C_{\gamma'', \delta''}^\nu \mid \nu > \mu, \gamma'', \delta'' \in M(\nu)\}$.

Proof of Corollary 3.3. Condition (1) is clear as $|\Lambda^\circ|$ itself is a finite set. Condition (2) is a consequence of the definition (3.1). We have already asserted that $*$ is an anti-automorphism giving (3). Finally to verify (4) it suffices to consider the case that $x = C_{\alpha, \beta}^\lambda$ for some $\lambda \in \Lambda$ and $\alpha, \beta \in M(\lambda)$. If $\beta = \gamma$ and $\alpha \subset \mu$, then Theorem 3.1(i)–(ii) shows that

$$C_{\alpha, \beta}^\lambda C_{\gamma, \delta}^\mu \equiv s_{\underline{\alpha}\lambda\bar{\beta}}(\mu) C_{\alpha, \delta}^\mu \pmod{H_\Lambda(> \mu)}$$

where $s_{\underline{\alpha}\lambda\bar{\beta}}(\mu)$ is independent of δ ; otherwise, we have that

$$C_{\alpha, \beta}^\lambda C_{\gamma, \delta}^\mu \equiv 0 \pmod{H_\Lambda(> \mu)}.$$

Taking

$$r_x(\gamma', \gamma) := \begin{cases} s_{\underline{\alpha}\lambda\bar{\beta}}(\mu) & \text{if } \gamma' = \alpha, \beta = \gamma \text{ and } \alpha \subset \mu, \\ 0 & \text{otherwise,} \end{cases}$$

we deduce that (4) holds. \square

Remark 3.4. In fact H_Λ is an example of a *graded* cellular algebra in the sense of [HM, §2]; the degree of $\alpha \in M(\lambda)$ is the number of clockwise cups in the diagram $\underline{\alpha}\lambda$. The same remark applies to Corollary 4.6 and Theorem 6.2 below.

Remark 3.5. Assume for this remark that $n > 0$. Khovanov showed in [K2, Proposition 32] that the algebra H_Λ is a symmetric algebra; see also Theorem 6.3 below. Since H_Λ is obviously not a semisimple algebra, it follows easily that H_Λ has infinite global dimension. In particular, this implies that H_Λ is *not* a quasi-hereditary algebra in the sense of Cline, Parshall and Scott [CPS1].

4. THE ALGEBRA K_Λ : A LARGER CELLULAR ALGEBRA

Fix a block Λ throughout the section. We are going to construct a graded algebra K_Λ which, if Λ is a Khovanov block, we will show in Part II is a quasi-hereditary cover of Khovanov's algebra H_Λ . We split the definition of the multiplication on K_Λ into two cases: the bounded case (when K_Λ is finite dimensional) and the general case (when it may be infinite dimensional).

The underlying graded vector space. We begin by defining K_Λ to be the vector space with basis

$$\{(\underline{\alpha}\lambda\bar{\beta}) \mid \text{for all } \alpha, \lambda, \beta \in \Lambda \text{ such that } \alpha \subset \lambda \supset \beta\}. \quad (4.1)$$

Equivalently, by Lemma 2.3(iii), this basis is the set

$$\{(a\lambda b) \mid \text{for all oriented circle diagrams } a\lambda b \text{ with } \lambda \in \Lambda\}. \quad (4.2)$$

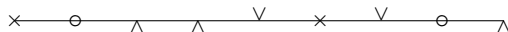
We stress the differences between this and Khovanov's algebra H_Λ : first, we now allow more general (perhaps even infinite) blocks Λ ; second, the basis is parametrised by arbitrary oriented circle diagrams whereas Khovanov's algebra only involves closed ones. We put a grading on K_Λ by defining the degree of the basis vector $(a\lambda b)$ to be $\deg(a\lambda b)$ as in (2.1). For $\lambda \in \Lambda$, we write simply e_λ for $(\underline{\lambda}\lambda\bar{\lambda})$. The vectors

$$\{e_\lambda \mid \lambda \in \Lambda\} \quad (4.3)$$

give a basis for the degree 0 component of K_Λ .

Multiplication in the bounded case. We assume in this subsection that Λ consists of bounded weights. In that case, our strategy to define the multiplication on K_Λ is to first construct an isomorphism of graded vector spaces $\text{cl} : K_\Lambda \rightarrow H_\Delta/I_\Delta$ for some Khovanov block Δ and some homogeneous ideal I_Δ of H_Δ . Using this isomorphism, we will then simply lift the multiplication on the quotient algebra H_Δ/I_Δ to the vector space K_Λ . We already want to point out that actual computations needed in practice can be done in a quicker way; see §6 for details.

Let p be the number of \wedge 's and q be the number of \vee 's labeling the vertices of each weight in Λ . For $\lambda \in \Lambda$, we define its *closure* $\text{cl}(\lambda)$ to be the weight obtained from λ by adding p new vertices labelled by \vee to the left end of the number line and q new vertices labelled by \wedge to the right end. For example, the closure of



is



Let Δ be the unique block containing the weights $\text{cl}(\lambda)$ for all $\lambda \in \Lambda$. Note that Δ is a Khovanov block, so it makes sense to consider Khovanov's algebra H_Δ from §3. The map

$$\text{cl} : \Lambda \hookrightarrow \Delta \quad (4.4)$$

just defined is injective with image $\text{cl}(\Lambda)$ contained in Δ° .

Lemma 4.1. *If $\lambda \in \Delta \setminus \text{cl}(\Lambda)$ and $\mu \geq \lambda$ then $\mu \in \Delta \setminus \text{cl}(\Lambda)$.*

Proof. The set $\Delta \setminus \text{cl}(\Lambda)$ consists of all weights in Δ either with some \wedge labeling one of its leftmost p vertices or some \vee labeling one of its rightmost q vertices. The lemma is obvious from this description and the definition of the Bruhat order. \square

Now suppose that a is a cup diagram with $a = \underline{\alpha}$ for $\alpha \in \Lambda$. We define its closure $\text{cl}(a)$ by setting

$$\text{cl}(a) := \underline{\text{cl}(\alpha)}. \quad (4.5)$$

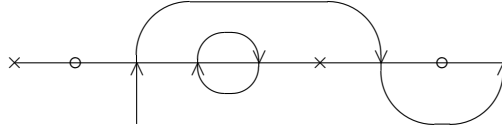
To compute $\text{cl}(a)$ in practice, let r denote the number of cups in a . First add p new vertices to the left end and q new vertices to the right end of the number line in a . Then connect the leftmost r of these new vertices to the rightmost r of the new vertices by adding r nested cups. Finally convert the leftmost $(p-r)$ rays of a into cups incident with the remaining $(p-r)$ new vertices on the left hand side and convert the rightmost $(q-r)$ rays of a into cups incident with the remaining $(q-r)$ new vertices on the right hand side. This explicit description implies for any $\lambda \in \Lambda$ that $a\lambda$ is an oriented cup diagram if and only if $\text{cl}(a\lambda) := \text{cl}(a)\text{cl}(\lambda)$ is a closed oriented cup diagram. Moreover, since only anti-clockwise cups are added when passing from $a\lambda$ to $\text{cl}(a\lambda)$, we have that $\deg(\text{cl}(a\lambda)) = \deg(a\lambda)$. Similarly, if b is a cap diagram with $b = \overline{\beta}$ for $\beta \in \Lambda$, we set

$$\text{cl}(b) := \overline{\text{cl}(\beta)}, \tag{4.6}$$

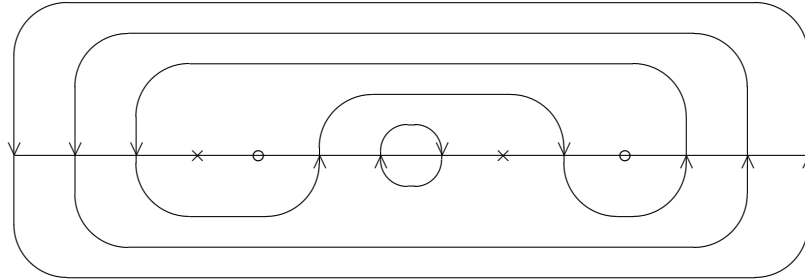
so that $\text{cl}(b)^* = \text{cl}(b^*)$. We then get for any oriented circle diagram $a\lambda b$ with $\lambda \in \Lambda$ that

$$\text{cl}(a\lambda b) := \text{cl}(a)\text{cl}(\lambda)\text{cl}(b) \tag{4.7}$$

is a closed oriented circle diagram of the same degree as $a\lambda b$. For example, if $a\lambda b$ equals



then $\text{cl}(a\lambda b)$ is the closed oriented circle diagram



Conversely, given a closed oriented circle diagram with underlying weight belonging to $\text{cl}(\Lambda)$, there is an obvious way to “open” the diagram, deleting the leftmost p and rightmost q vertices in the process, to obtain an oriented circle diagram $a\lambda b$ with $\lambda \in \Lambda$ such that $\text{cl}(a\lambda b)$ is the initial closed oriented circle diagram. This proves the following lemma.

Lemma 4.2. *The map $a\lambda b \mapsto \text{cl}(a\lambda b)$ is a degree preserving bijection between the set of oriented circle diagrams with underlying weight belonging to Λ and the set of closed oriented circle diagrams with underlying weight belonging to $\text{cl}(\Lambda)$.*

Now we can complete the definition of the graded algebra K_Λ when Λ consists of bounded weights. Lemma 4.1 and Corollary 3.2 imply that the subspace I_Λ of H_Δ spanned by the vectors

$$\{(a\lambda b) \mid \text{for all closed oriented circle diagrams } a\lambda b \text{ with } \lambda \in \Delta \setminus \text{cl}(\Lambda)\}$$

is a two-sided ideal of H_Δ . In view of Lemma 4.2, the vectors

$$\{(\text{cl}(a\lambda b)) + I_\Lambda \mid \text{for all oriented circle diagrams } a\lambda b \text{ with } \lambda \in \Lambda\}$$

give a basis for the quotient algebra H_Δ/I_Λ . We deduce that the map

$$\text{cl} : K_\Lambda \rightarrow H_\Delta/I_\Lambda, \quad (a\lambda b) \mapsto (\text{cl}(a\lambda b)) + I_\Lambda \quad (4.8)$$

is an isomorphism of graded vector spaces. Using this, we transport the multiplication on the algebra H_Δ/I_Λ to the vector space K_Λ , to make K_Λ into a well-defined positively graded algebra in its own right. In other words, to compute the product $(a\lambda b)(c\mu d)$ of two basis vectors of K_Λ , we first close both diagrams, then compute their product in Khovanov's algebra H_Δ modulo the ideal I_Λ , then apply the inverse map cl^{-1} to get back to an element of K_Λ .

Multiplication in the general case. It just remains to define the multiplication on K_Λ in the unbounded cases too, by taking a direct limit of the algebras already defined. To set this up, we need one more piece of notation. For weights λ and μ we write $\lambda \prec \mu$ if

- ▶ λ is a bounded weight;
- ▶ the set I indexing the vertices of λ is a subset of the set J indexing the vertices of μ ;
- ▶ the labels on each vertex of μ indexed by I are the same as the labels on the corresponding vertices of λ ;
- ▶ amongst all vertices of μ indexed by the set $J \setminus I$ it is impossible to find two that are labelled $\vee \wedge$ in that order from left to right.

For example the following weights λ and μ (whose vertices are indexed by $\{1, 2, 3\}$ and by \mathbb{Z} , respectively) satisfy $\lambda \prec \mu$:



Given two blocks Γ and Λ we write $\Gamma \prec \Lambda$ if there exist weights $\lambda \in \Gamma$ and $\mu \in \Lambda$ such that $\lambda \prec \mu$. In that case, for every $\lambda \in \Gamma$ there is a unique element denoted $\text{ex}_\Gamma^\Lambda(\lambda) \in \Lambda$, the *extension* of λ from Γ to Λ , with the property that $\lambda \prec \text{ex}_\Gamma^\Lambda(\lambda)$. This defines an injection

$$\text{ex}_\Gamma^\Lambda : \Gamma \hookrightarrow \Lambda. \quad (4.9)$$

If $\Gamma \prec \Omega \prec \Lambda$ then $\Gamma \prec \Lambda$ and $\text{ex}_\Gamma^\Lambda = \text{ex}_\Omega^\Lambda \circ \text{ex}_\Gamma^\Omega$. Moreover, Λ is the union of the finite subsets $\text{ex}_\Gamma^\Lambda(\Gamma)$ for all $\Gamma \prec \Lambda$. In other words, the maps ex_Γ^Λ induce a canonical isomorphism $\varinjlim \Gamma \xrightarrow{\sim} \Lambda$ between the direct limit of all $\Gamma \prec \Lambda$ and the set Λ .

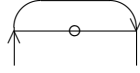
Now suppose that a is a cup diagram with $a = \underline{\alpha}$ and b is a cap diagram with $b = \overline{\beta}$ for $\alpha, \beta \in \Gamma$. We define their extensions $\text{ex}_\Gamma^\Lambda(a)$ and $\text{ex}_\Gamma^\Lambda(b)$ by setting

$$\text{ex}_\Gamma^\Lambda(a) := \underline{\text{ex}_\Gamma^\Lambda(\alpha)}, \quad \text{ex}_\Gamma^\Lambda(b) := \overline{\text{ex}_\Gamma^\Lambda(\beta)}. \quad (4.10)$$

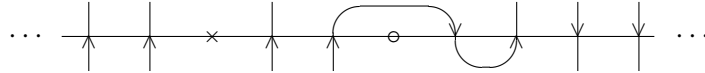
Then for an oriented circle diagram $a\lambda b$ with $\lambda \in \Gamma$, we define its extension to be the oriented circle diagram

$$\text{ex}_\Gamma^\Lambda(a\lambda b) := \text{ex}_\Gamma^\Lambda(a)\text{ex}_\Gamma^\Lambda(\lambda)\text{ex}_\Gamma^\Lambda(b). \quad (4.11)$$

Informally, $\text{ex}_\Gamma^\Lambda(a\lambda b)$ is obtained from the diagram $a\lambda b$ by first extending the weight λ to $\text{ex}_\Gamma^\Lambda(\lambda)$, then extending the cup and cap diagrams a and b by converting as many as possible of their rays into anti-clockwise cups and caps through the new vertices, then adding new rays through any remaining vertices labelled \wedge or \vee . For instance if Γ and Λ are the blocks generated by the weights λ and μ from the previous example and $a\lambda b$ equals



then $\text{ex}_\Gamma^\Lambda(a\lambda b)$ is



In this way, we obtain a degree preserving injective linear map

$$\text{ex}_\Gamma^\Lambda : K_\Gamma \hookrightarrow K_\Lambda, \quad (a\lambda b) \mapsto (\text{ex}_\Gamma^\Lambda(a\lambda b)). \quad (4.12)$$

If $\Gamma \prec \Omega \prec \Lambda$, then $\text{ex}_\Gamma^\Lambda = \text{ex}_\Omega^\Lambda \circ \text{ex}_\Gamma^\Omega$, and also K_Λ is the sum of the subspaces $\text{ex}_\Gamma^\Lambda(K_\Gamma)$ for all $\Gamma \prec \Lambda$. So the maps ex_Γ^Λ induce a canonical isomorphism of graded vector spaces $\varinjlim K_\Gamma \xrightarrow{\sim} K_\Lambda$, where the direct limit is again taken over all $\Gamma \prec \Lambda$.

Lemma 4.3. *Given blocks $\Gamma \prec \Omega \prec \Lambda$, the map $\text{ex}_\Gamma^\Omega : K_\Gamma \rightarrow K_\Omega$ is an algebra homomorphism.*

Proof. It suffices to consider the four cases in which $\text{ex}_\Gamma^\Omega : \Gamma \rightarrow \Omega$ adds one new vertex labelled either \vee or \wedge to the weights in Γ , either at the left or the right end. In all four cases, for any oriented circle diagram $a\lambda b$ with $\lambda \in \Gamma$, the closure of $\text{ex}_\Gamma^\Omega(a\lambda b)$ is obtained from the closure of $a\lambda b$ by adding one more anti-clockwise circle. More precisely, if ex_Γ^Ω adds one vertex labelled \wedge to the left or one vertex labelled \vee to the right then $\text{cl}(\text{ex}_\Gamma^\Omega(a\lambda b))$ is $\text{cl}(a\lambda b)$ with one additional small anti-clockwise circle inside, and if ex_Γ^Ω adds one vertex labelled \vee to the left or one vertex labelled \wedge to the right then $\text{cl}(\text{ex}_\Gamma^\Omega(a\lambda b))$ is $\text{cl}(a\lambda b)$ with one additional large anti-clockwise circle outside. Since the additional circle always sits in the same position independent of the diagram $a\lambda b$, it follows that it makes no difference to the multiplication rule, because $1 \otimes 1 \mapsto 1$ in the definition of Khovanov's algebra. \square

In view of Lemma 4.3, the direct limit $\varinjlim K_\Gamma$ is a graded associative (but in general not unital) algebra. Now transport the multiplication on $\varinjlim K_\Gamma$ to the vector space K_Λ using the canonical isomorphism constructed just before Lemma 4.3, to make K_Λ into a well-defined graded associative algebra. In other words, to compute the product $(a\lambda b)(c\mu d)$ of two basis vectors of K_Λ , first choose some $\Gamma \prec \Lambda$ so that $(a\lambda b)$ and $(c\mu d)$ are in the image of the map ex_Γ^Λ , then “truncate” by applying $(\text{ex}_\Gamma^\Lambda)^{-1}$ and compute the product in the finite dimensional algebra K_Γ , then finally apply ex_Γ^Λ to extend back to an element of K_Λ .

Cellular algebra structure. We have now defined the multiplication making the graded vector space K_Λ into a graded associative algebra in all cases. The

definition gives quite easily that

$$e_\alpha(a\lambda b) = \begin{cases} (a\lambda b) & \text{if } \underline{\alpha} = a, \\ 0 & \text{otherwise,} \end{cases} \quad (a\lambda b)e_\beta = \begin{cases} (a\lambda b) & \text{if } b = \bar{\beta}, \\ 0 & \text{otherwise,} \end{cases}$$

for $\alpha, \beta \in \Lambda$ and any basis vector $(a\lambda b) \in K_\Lambda$. So the vectors $\{e_\alpha \mid \alpha \in \Lambda\}$ give a set of mutually orthogonal idempotents in K_Λ . If $|\Lambda| < \infty$, the sum of these idempotents is the identity element of the finite dimensional algebra K_Λ . In general, K_Λ might not be a unital algebra but we still have at least that

$$K_\Lambda = \bigoplus_{\alpha, \beta \in \Lambda} e_\alpha K_\Lambda e_\beta. \quad (4.13)$$

The summand $e_\alpha K_\Lambda e_\beta$ of K_Λ has basis

$$\{(\underline{\alpha}\lambda\bar{\beta}) \mid \text{for all } \lambda \in \Lambda \text{ such that } \alpha \subset \lambda \supset \beta\}.$$

This set is finite thanks to Lemma 2.4(i), so each $e_\alpha K_\Lambda e_\beta$ is finite dimensional even if K_Λ itself is infinite dimensional. Note also for $\Gamma \prec \Lambda$ that the map $\text{ex}_\Gamma^\Lambda : K_\Gamma \rightarrow K_\Lambda$ sends the idempotent $e_\gamma \in K_\Gamma$ for $\gamma \in \Gamma$ to the idempotent $e_\lambda \in K_\Lambda$ where $\lambda := \text{ex}_\Gamma^\Lambda(\gamma)$. Observe finally that there are graded algebra anti-isomorphisms

$$* : K_\Lambda \rightarrow K_\Lambda, \quad (a\lambda b) \mapsto (b^* \lambda a^*), \quad (4.14)$$

$$\curvearrowright : K_\Lambda \rightarrow K_{\Lambda^\curvearrowright}, \quad (a\lambda b) \mapsto (b^\curvearrowright \lambda^\curvearrowright a^\curvearrowright), \quad (4.15)$$

just like (3.9). Composing $*$ and \curvearrowright , we deduce that $K_\Lambda \cong K_{\Lambda^\curvearrowright}$ as graded algebras.

Theorem 4.4. *Let $(a\lambda b)$ and $(c\mu d)$ be basis vectors of K_Λ . Then,*

$$(a\lambda b)(c\mu d) = \begin{cases} 0 & \text{if } b \neq c^*, \\ s_{a\lambda b}(\mu)(a\mu d) + (\dagger) & \text{if } b = c^* \text{ and } a\mu \text{ is oriented,} \\ (\dagger) & \text{otherwise,} \end{cases}$$

where

- (i) (\dagger) denotes a linear combination of basis vectors of K_Λ of the form $(a\nu d)$ for $\nu > \mu$;
- (ii) the scalar $s_{a\lambda b}(\mu) \in \{0, 1\}$ depends only on $a\lambda b$ and μ (but not on d);
- (iii) if $b = \bar{\lambda} = c^*$ and $a\mu$ is oriented then $s_{a\lambda b}(\mu) = 1$.

Proof. Assume to start with that the weights in Λ are bounded. We have already seen that $(a\lambda b)(c\mu d) = 0$ if $b^* \neq c$, so assume that $b^* = c$. In that case, $\text{cl}(b)^* = \text{cl}(b^*) = \text{cl}(c)$, so we get by Theorem 3.1(i) that $(\text{cl}(a\lambda b))(\text{cl}(c\mu d)) = s_{\text{cl}(a\lambda b)}(\text{cl}(\mu))(\text{cl}(a\mu d)) + (\dagger)$ if $\text{cl}(a\mu)$ is oriented, and $(\text{cl}(a\lambda b))(\text{cl}(c\mu d)) = (\dagger)$ otherwise, where (\dagger) denotes a linear combination of $(\text{cl}(a)\nu'\text{cl}(d))$'s for $\nu' > \text{cl}(\mu)$. If $\nu' > \text{cl}(\mu)$ then we either have that $\nu' = \text{cl}(\nu)$ for some $\nu > \mu$ or $(\text{cl}(a)\nu'\text{cl}(d)) \in I_\Lambda$. Moreover, $\text{cl}(a\mu)$ is oriented if and only if $a\mu$ is oriented. So applying the map $(\text{cl})^{-1}$, we have shown that

$$(a\lambda b)(c\mu d) = \begin{cases} s_{a\lambda b}(\mu)(a\mu d) + (\dagger) & \text{if } a\mu \text{ is oriented,} \\ (\dagger) & \text{otherwise,} \end{cases}$$

where $s_{a\lambda b}(\mu) := s_{\text{cl}(a\lambda b)}(\text{cl}(\mu)) \in \{0, 1\}$ and (\dagger) denotes a linear combination of basis vectors of K_Λ of the form $(a\nu d)$ for $\nu > \mu$.

Since $s_{a\lambda b}(\mu)$ depends only on $a\lambda b$ and μ by Theorem 3.1(ii), this is what we wanted for (i)–(ii). Moreover, if $b = \bar{\lambda} = c^*$ and $a\mu$ is oriented as in (iii), then we get that $s_{a\lambda b}(\mu) = 1$ by Theorem 3.1(iii).

It just remains to pass from the bounded case to the unbounded case by taking direct limits. This is done by a similar argument to the preceding one and causes no additional problems. \square

Corollary 4.5. *The product $(a\lambda b)(c\mu d)$ of two basis vectors of K_Λ is a linear combination of vectors of the form $(a\nu d)$ for $\nu \in \Lambda$ with $\lambda \leq \nu \leq \mu$.*

Proof. Argue as in the proof of Corollary 3.2. \square

In the next corollary, we need to exclude the situations when $|\Lambda| = \infty$, but only because cellular algebras are unital algebras by definition.

Corollary 4.6. *Assuming $|\Lambda| < \infty$, the algebra K_Λ is a cellular algebra in the sense of Graham and Lehrer with cell datum $(\Lambda, M, C, *)$ where*

- (i) $M(\lambda)$ denotes $\{\alpha \in \Lambda \mid \alpha \subset \lambda\}$ for each $\lambda \in \Lambda$;
- (ii) C is defined by setting $C_{\alpha, \beta}^\lambda := (\underline{\alpha\lambda\bar{\beta}})$ for $\lambda \in \Lambda$ and $\alpha, \beta \in M(\lambda)$;
- (iii) $*$ is the anti-automorphism from (4.14).

Proof. Argue as in the proof of Corollary 3.3. \square

5. THE QUASI-HEREDITARY STRUCTURE OF K_Λ

Now we begin the study of the representation theory of the algebras K_Λ for any block Λ in a systematic way. We define cell modules for K_Λ , giving explicit closed formulae for the q -decomposition numbers which simultaneously describe composition multiplicities of cell modules and cell filtration multiplicities of projective indecomposable modules. We deduce in almost all cases that the category $\text{Rep}(K_\Lambda)$ of finite dimensional graded K_Λ -modules is a graded highest weight category.

Graded modules. We begin by fixing some conventions for working with graded modules over a locally unital graded algebra $K = \bigoplus_{i \in \mathbb{Z}} K_i$. Here *locally unital* means that K is an associative (but not necessarily unital) algebra with a given system $\{e_\lambda \mid \lambda \in \Lambda\}$ of mutually orthogonal idempotents such that

$$K = \bigoplus_{\alpha, \beta \in \Lambda} e_\alpha K e_\beta.$$

Of course if K is finite dimensional, all this is just saying that K is a unital algebra with identity element $\sum_{\lambda \in \Lambda} e_\lambda$. (Alternatively, a locally unital algebra such as K could be viewed as a small category with Λ as its set of objects, and $\text{Hom}(\beta, \alpha) := e_\alpha K e_\beta$ for each $\alpha, \beta \in \Lambda$.) By a K -module we always mean a left K -module M such that

$$M = \bigoplus_{\lambda \in \Lambda} e_\lambda M.$$

If K is finite dimensional, this is just saying that M is a unital module in the usual sense.

If $M = \bigoplus_{j \in \mathbb{Z}} M_j$ is a *graded* K -module, i.e. $K_i M_j \subseteq M_{i+j}$, then we write $M\langle j \rangle$ for the same module but with new grading defined by $M\langle j \rangle_i := M_{i-j}$. For graded modules M and N , we define

$$\mathrm{Hom}_K(M, N) := \bigoplus_{j \in \mathbb{Z}} \mathrm{Hom}_K(M, N)_j \quad (5.1)$$

where $\mathrm{Hom}_K(M, N)_j$ denotes all homogeneous K -module homomorphisms of degree j , meaning that they map M_i into N_{i+j} for each $i \in \mathbb{Z}$. Let $\mathrm{Mod}_f(K)$ be the category of all graded K -modules $M = \bigoplus_{j \in \mathbb{Z}} M_j$ that are locally finite dimensional and bounded below, i.e. each M_j is finite dimensional and $M_j = 0$ for $j \ll 0$. Morphisms in $\mathrm{Mod}_f(K)$ mean homogeneous K -module homomorphisms of degree 0, so that $\mathrm{Mod}_f(K)$ is an abelian category. The space $\mathrm{Hom}_K(M, N)_j$ can be also described as all morphisms from $M\langle j \rangle$ to N in the category $\mathrm{Mod}_f(K)$.

Irreducible and projective K_Λ -modules. Now fix an arbitrary block Λ and note by (4.13) that K_Λ is locally unital with respect to the system of idempotents $\{e_\lambda \mid \lambda \in \Lambda\}$. Let $(K_\Lambda)_{>0}$ be the sum of all components of the graded algebra K_Λ of strictly positive degree, so

$$K_\Lambda / (K_\Lambda)_{>0} \cong \bigoplus_{\lambda \in \Lambda} \mathbb{F} \quad (5.2)$$

as an algebra, with a basis given by the images of all the idempotents e_λ . The image of e_λ spans a one dimensional graded K_Λ -module which we denote by $L(\lambda)$. Thus, $L(\lambda)$ is a copy of the field concentrated in degree 0, and $(a\mu b) \in K_\Lambda$ acts on $L(\lambda)$ as multiplication by 1 if $a\mu b = \underline{\lambda\lambda\bar{\lambda}}$, or as zero otherwise. The modules

$$\{L(\lambda)\langle j \rangle \mid \lambda \in \Lambda, j \in \mathbb{Z}\} \quad (5.3)$$

give a complete set of isomorphism classes of irreducible graded K_Λ -modules. For any graded K_Λ -module M we let M^\circledast denote its graded dual, which means that $(M^\circledast)_j := \mathrm{Hom}_{\mathbb{F}}(M_{-j}, \mathbb{F})$ and $x \in K_\Lambda$ acts on $f \in M^\circledast$ by $(xf)(m) := f(x^*m)$, where x^* denotes the image of x under (4.14). Clearly we have that

$$L(\lambda)^\circledast \cong L(\lambda) \quad (5.4)$$

for each $\lambda \in \Lambda$.

For $\lambda \in \Lambda$, let $P(\lambda) := K_\Lambda e_\lambda$. This is a graded K_Λ -module with basis

$$\{\underline{\nu\mu\bar{\lambda}} \mid \text{for all } \nu, \mu \in \Lambda \text{ such that } \nu \subset \mu \supset \lambda\}.$$

By Lemma 2.4, $P(\lambda)$ is locally finite dimensional and bounded below, and it is globally finite dimensional if and only if $\mathrm{def}(\Lambda) < \infty$. So $P(\lambda)$ belongs to the category $\mathrm{Mod}_f(K_\Lambda)$. In fact, it is the projective cover of $L(\lambda)$ in this category, and the modules

$$\{P(\lambda)\langle j \rangle \mid \lambda \in \Lambda, j \in \mathbb{Z}\} \quad (5.5)$$

give a full set of projective indecomposable modules in $\mathrm{Mod}_f(K_\Lambda)$.

Grothendieck groups and Cartan matrix. Let $\mathrm{Rep}(K_\Lambda)$ and $\mathrm{Proj}(K_\Lambda)$ be the full subcategories of $\mathrm{Mod}_f(K_\Lambda)$ consisting of finite dimensional modules and finitely generated projective modules, respectively. The Grothendieck

groups $[\text{Rep}(K_\Lambda)]$ and $[\text{Proj}(K_\Lambda)]$ are the free \mathbb{Z} -modules with bases given by the isomorphism classes of the irreducible modules from (5.3) and the projective indecomposable modules from (5.5), respectively. If we view all our Grothendieck groups instead as $\mathbb{Z}[q, q^{-1}]$ -modules so that

$$q^j[M] := [M\langle j \rangle],$$

then $[\text{Rep}(K_\Lambda)]$ and $[\text{Proj}(K_\Lambda)]$ become the free $\mathbb{Z}[q, q^{-1}]$ -modules with bases $\{[L(\lambda)] \mid \lambda \in \Lambda\}$ and $\{[P(\lambda)] \mid \lambda \in \Lambda\}$, respectively.

Let $\mathbb{Z}((q))$ denote the ring of formal Laurent series in q (the localisation of $\mathbb{Z}[[q]]$ at q). Introduce the q -Cartan matrix

$$C_\Lambda(q) = (c_{\lambda, \mu}(q))_{\lambda, \mu \in \Lambda} \quad (5.6)$$

where

$$c_{\lambda, \mu}(q) := \sum_{j \in \mathbb{Z}} q^j \dim \text{Hom}_{K_\Lambda}(P(\lambda), P(\mu))_j \in \mathbb{Z}((q)). \quad (5.7)$$

Writing $[\text{Rep}(K_\Lambda)]^\wedge$ and $[\text{Proj}(K_\Lambda)]^\wedge$ for the completions of $[\text{Rep}(K_\Lambda)]$ and $[\text{Proj}(K_\Lambda)]$ to topological $\mathbb{Z}((q))$ -modules, there is a unique continuous $\mathbb{Z}((q))$ -module homomorphism

$$j : [\text{Proj}(K_\Lambda)]^\wedge \rightarrow [\text{Rep}(K_\Lambda)]^\wedge, \quad [P(\mu)] \mapsto \sum_{\lambda \in \Lambda} c_{\lambda, \mu}(q)[L(\lambda)], \quad (5.8)$$

which makes sense because $P(\mu)$ is locally finite dimensional. Since

$$\text{Hom}_{K_\Lambda}(P(\lambda), P(\mu)) = \text{Hom}_{K_\Lambda}(K_\Lambda e_\lambda, K_\Lambda e_\mu) = e_\lambda K_\Lambda e_\mu$$

and $e_\lambda K_\Lambda e_\mu$ has basis $\{(\underline{\lambda\nu\bar{\mu}}) \mid \nu \in \Lambda \text{ such that } \lambda \subset \nu \supset \mu\}$, we have that

$$c_{\lambda, \mu}(q) = \sum_{\lambda \subset \nu \supset \mu} q^{\deg(\underline{\lambda\nu\bar{\mu}})}. \quad (5.9)$$

Recalling Lemma 2.4, it follows that $c_{\lambda, \mu}(q)$ is in fact a polynomial in q , and its constant coefficient is 1 if $\lambda = \mu$ or 0 otherwise. So the Cartan matrix $C_\Lambda(q)$ is invertible in the ring of $\Lambda \times \Lambda$ matrices with entries in $\mathbb{Z}((q))$. This implies that the map j from (5.8) is an isomorphism. From now on we will use this isomorphism to simply *identify* $[\text{Proj}(K_\Lambda)]^\wedge$ with $[\text{Rep}(K_\Lambda)]^\wedge$. (If $|\Lambda| < \infty$ then it is not necessary to complete in this way since in that case the restriction of j is a $\mathbb{Z}[q, q^{-1}]$ -module isomorphism $j : \text{Proj}(K_\Lambda) \xrightarrow{\sim} \text{Rep}(K_\Lambda)$ by (5.17) below.)

Graded cell modules and q -decomposition numbers. Now we introduce *cell modules* following [GL]. For $\mu \in \Lambda$, define $V(\mu)$ to be the vector space on basis

$$\{(\underline{\lambda\mu}) \mid \text{for all } \lambda \in \Lambda \text{ such that } \lambda \subset \mu\}. \quad (5.10)$$

Equivalently, by Lemma 2.3(i), this basis is the set

$$\{(c\mu) \mid \text{for all oriented cup diagrams } c\mu\}.$$

We put a grading on $V(\mu)$ by defining the degree of the vector $(c\mu)$ to be $\deg(c\mu)$ as in (2.1). Note by Lemma 2.4 that $V(\mu)$ is finite dimensional if and only if $\text{def}(\Lambda) < \infty$, though it is always locally finite dimensional and bounded below.

We make $V(\mu)$ into a graded K_Λ -module by declaring for any basis vector $(a\lambda b)$ of K_Λ that

$$(a\lambda b)(c\mu| := \begin{cases} s_{a\lambda b}(\mu)(a\mu| & \text{if } b^* = c \text{ and } a\mu \text{ is oriented,} \\ 0 & \text{otherwise.} \end{cases} \quad (5.11)$$

where $s_{a\lambda b}(\mu) \in \{0, 1\}$ is the scalar defined by Theorem 4.4. We remark that the proofs of Theorems 3.1 and 4.4 give an explicit algorithm to compute these scalars in practise. We have not yet checked that (5.11) makes $V(\mu)$ into a well-defined module, but that will be explained in the proof of the next theorem.

Theorem 5.1. *For $\lambda \in \Lambda$, enumerate the $2^{\text{def}(\lambda)}$ distinct elements of the set $\{\mu \in \Lambda \mid \mu \supset \lambda\}$ as $\mu_1, \mu_2, \dots, \mu_n = \lambda$ so that $\mu_i > \mu_j$ implies $i < j$. Let $M(0) := \{0\}$ and for $i = 1, \dots, n$ define $M(i)$ to be the subspace of $P(\lambda)$ generated by $M(i-1)$ and the vectors*

$$\{(c\mu_i\bar{\lambda}) \mid \text{for all oriented cup diagrams } c\mu_i\}.$$

Then

$$\{0\} = M(0) \subset M(1) \subset \dots \subset M(n) = P(\lambda)$$

is a filtration of $P(\lambda)$ as a K_Λ -module such that

$$M(i)/M(i-1) \cong V(\mu_i)\langle \text{deg}(\mu_i\bar{\lambda}) \rangle$$

for each $i = 1, \dots, n$.

Proof. Corollary 4.5 implies that each $M(i)$ is a K_Λ -submodule of $P(\lambda)$. Since $P(\lambda)$ has basis $\{(c\mu\bar{\lambda}) \mid \text{for all oriented circle diagrams } c\mu\bar{\lambda}\}$, the map

$$f_i : V(\mu_i)\langle \text{deg}(\mu_i\bar{\lambda}) \rangle \rightarrow M(i)/M(i-1), \quad (c\mu_i| \mapsto (c\mu_i\bar{\lambda}) + M(i-1))$$

is a vector space isomorphism, and it is homogeneous of degree 0 thanks to (2.1). Recalling Theorem 4.4, we use this isomorphism to transport the K_Λ -module structure on $M(i)/M(i-1)$ to the vector space $V(\mu_i)$, to make $V(\mu_i)$ into a well-defined K_Λ -module such that

$$(a\lambda b)(c\mu_i| := \begin{cases} s_{a\lambda b}(\mu_i)(a\mu_i| & \text{if } b^* = c \text{ and } a\mu_i \text{ is oriented,} \\ 0 & \text{otherwise.} \end{cases}$$

This is the same as the formula (5.11) with $\mu = \mu_i$, so the map f_i is a K_Λ -module isomorphism. Since any $\mu \in \Lambda$ arises as $\mu = \mu_i$ for some λ and some i (e.g. one can take $\lambda = \mu$ and $i = n$), we have proved at the same time that $V(\mu)$ itself is a well-defined K_Λ -module. \square

Theorem 5.2. *For $\mu \in \Lambda$, let $N(j)$ be the submodule of $V(\mu)$ spanned by all graded pieces of degree $\geq j$. Then*

$$V(\mu) = N(0) \supseteq N(1) \supseteq N(2) \supseteq \dots$$

is a filtration of $V(\mu)$ as a K_Λ -module such that

$$N(j)/N(j+1) \cong \bigoplus_{\substack{\lambda \subset \mu \text{ with} \\ \text{deg}(\lambda\mu) = j}} L(\lambda)\langle j \rangle$$

for each $j \geq 0$. Moreover, we have that $N(j) = 0$ for $j \gg 0$, i.e. $V(\mu)$ is finite dimensional, if and only if $\text{def}(\Lambda) < \infty$.

Proof. Since K_Λ is positively graded, it is immediate that each $N(j)$ is a submodule. The quotient $N(j)/N(j+1)$ has basis

$$\{(\underline{\lambda}\mu| + N(j+1) \mid \lambda \in \Lambda \text{ such that } \lambda \subset \mu \text{ and } \deg(\underline{\lambda}\mu) = j\}.$$

We claim that the subspace of $N(j)/N(j+1)$ spanned by $(\underline{\lambda}\mu| + N(j+1)$ is isomorphic to $L(\lambda)\langle j \rangle$. To see that, we just need to observe for any basis vector $(a\nu b)$ of K_Λ and $\lambda \subset \mu$ that

$$(a\nu b)(\underline{\lambda}\mu| \equiv \begin{cases} (\underline{\lambda}\mu| & \text{if } a\nu b = \underline{\lambda}\lambda\bar{\lambda}, \\ 0 & \text{otherwise,} \end{cases}$$

working modulo $N(j+1)$. This is clear if $\deg(a\nu b) > 0$ or if $b \neq \bar{\lambda}$ by (5.11). So we may assume that $a\nu b = \underline{\lambda}\lambda\bar{\lambda}$ and need to show that $s_{\underline{\lambda}\lambda\bar{\lambda}}(\mu) = 1$. This follows from Theorem 4.4(iii). The final statement of the theorem is immediate from Lemma 2.4(ii). \square

For $\lambda, \mu \in \Lambda$, we define

$$d_{\lambda, \mu}(q) := \begin{cases} q^{\deg(\underline{\lambda}\mu)} & \text{if } \lambda \subset \mu, \\ 0 & \text{otherwise.} \end{cases} \quad (5.12)$$

We call the resulting matrix

$$D_\Lambda(q) = (d_{\lambda, \mu}(q))_{\lambda, \mu \in \Lambda} \quad (5.13)$$

the *q-decomposition matrix* because by Theorems 5.1 and 5.2 we have that

$$[V(\mu)] = \sum_{\lambda \in \Lambda} d_{\lambda, \mu}(q)[L(\lambda)], \quad (5.14)$$

$$[P(\lambda)] = \sum_{\mu \in \Lambda} d_{\lambda, \mu}(q)[V(\mu)], \quad (5.15)$$

both equalities written in $[\text{Rep}(K_\Lambda)]^\wedge$. Recall from Lemma 2.3 that $\lambda \subset \mu$ implies $\lambda \leq \mu$. So the *q-decomposition matrix* $D_\Lambda(q)$ is upper unitriangular when rows and columns are ordered in some way refining the Bruhat order. Here is an example:

	$\vee\vee\wedge\wedge$	$\vee\wedge\vee\wedge$	$\wedge\vee\vee\wedge$	$\vee\wedge\wedge\vee$	$\wedge\vee\wedge\vee$	$\wedge\wedge\vee\vee$
$\vee\vee\wedge\wedge$	1	q	0	0	q	q^2
$\vee\wedge\vee\wedge$	0	1	q	q	q^2	0
$\wedge\vee\vee\wedge$	0	0	1	0	q	0
$\vee\wedge\wedge\vee$	0	0	0	1	q	0
$\wedge\vee\wedge\vee$	0	0	0	0	1	q
$\wedge\wedge\vee\vee$	0	0	0	0	0	1

(5.16)

On comparing (5.12) with (5.9) and using (2.1), we get the following familiar factorisation of the Cartan matrix $C_\Lambda(q)$:

$$C_\Lambda(q) = D_\Lambda(q)D_\Lambda(q)^T. \quad (5.17)$$

Highest weight categories. We are ready to prove the following important result. Note here it is essential to exclude $\text{def}(\Lambda) = \infty$, since in that case the cell modules are not Artinian.

Theorem 5.3. *Assuming $\text{def}(\Lambda) < \infty$, the category $\text{Rep}(K_\Lambda)$ is a positively graded highest weight category with duality in the sense of Cline, Parshall and Scott. For $\lambda \in \Lambda$ (the weight poset), the irreducible, standard, costandard, projective indecomposable and injective indecomposable objects are the modules $L(\lambda), V(\lambda), V(\lambda)^\otimes, P(\lambda)$ and $P(\lambda)^\otimes$, respectively.*

Before proving the theorem, we give the definition of a positively graded highest weight category with duality following [CPS1, Definition 3.1], [CPS2, (1.1)] and [CPS3, (1.2)] as follows. This means a \mathbb{Z} -category \mathcal{C} in the sense of [AJS, Appendix E] with shift operators denoted $M \mapsto M\langle j \rangle$ for each $j \in \mathbb{Z}$ (like the category of finite dimensional graded modules over a graded algebra) together with an interval-finite poset Λ such that

- (1) \mathcal{C} is an \mathbb{F} -linear Artinian category equipped with a duality, that is, an equivalence of categories $\otimes : \mathcal{C} \rightarrow \mathcal{C}^{\text{op}}$, such that $(M\langle j \rangle)^\otimes \cong M^\otimes\langle -j \rangle$ for $M \in \mathcal{C}$ and $j \in \mathbb{Z}$;
- (2) for each $\lambda \in \Lambda$ there is a given object $L(\lambda) \cong L(\lambda)^\otimes \in \mathcal{C}$ such that $\{L(\lambda)\langle j \rangle \mid \lambda \in \Lambda, j \in \mathbb{Z}\}$ is a complete set of representatives for the isomorphism classes of irreducible objects in \mathcal{C} ;
- (3) each $L(\lambda)$ has a projective cover $P(\lambda) \in \mathcal{C}$ such that all composition factors of $P(\lambda)$ are isomorphic to $L(\mu)\langle j \rangle$'s for $\mu \in \Lambda$ and $j \geq 0$;
- (4) writing $V(\lambda)$ for the largest quotient of $P(\lambda)$ with the property that all composition factors of its radical are isomorphic to $L(\mu)\langle j \rangle$'s for $\mu < \lambda$, the object $P(\lambda)$ has a filtration with $V(\lambda)$ at the top and all other factors isomorphic to $V(\nu)\langle j \rangle$'s for $\nu > \lambda$.

Proof of Theorem 5.3. We have to verify the above properties for $\text{Rep}(K_\Lambda)$. Property (1) is clear since all the objects are finite dimensional modules. We have already noted that the modules $L(\lambda)\langle j \rangle$ for $\lambda \in \Lambda$ and $j \in \mathbb{Z}$ give a complete set of irreducible graded K_Λ -modules giving (2). For (3), we have already noted that $P(\lambda)$ is the projective cover of $L(\lambda)$, so we are done by (5.8) and (5.9). Finally, for (4), Theorem 5.1 shows that the projective cover $P(\lambda)$ of $L(\lambda)$ has a filtration with top quotient isomorphic to the cell module $V(\lambda)$ and all other quotients isomorphic to $V(\mu)\langle j \rangle$'s for $\mu > \lambda$ and $j \in \mathbb{Z}$. Note in particular this means that $V(\lambda)$ has a unique irreducible quotient, isomorphic to $L(\lambda)$. It just remains to check that the cell module $V(\lambda)$ defined earlier is isomorphic to the module $V(\lambda)$ defined in (4). In the notation of Theorem 5.1, $V(\lambda)$ is isomorphic to the quotient $Q := P(\lambda)/M(n-1)$, and Theorem 5.2 shows that all composition factors of the radical of Q are isomorphic to $L(\mu)\langle j \rangle$'s for $\mu < \lambda$. Moreover, $M(n-1)$ has a filtration with sections isomorphic to $V(\mu)\langle j \rangle$'s for $\mu > \lambda$, so all irreducible quotients of $M(n-1)$ are isomorphic to $L(\mu)\langle j \rangle$'s for $\mu > \lambda$. This completes the proof. \square

If $|\Lambda| < \infty$, so that K_Λ is a finite dimensional algebra, Theorem 5.3 is well known to be equivalent to the following statement; see [CPS1, Theorem 3.6] and [AJS, Proposition E.4].

Corollary 5.4. *Assuming $|\Lambda| < \infty$, the algebra K_Λ is a positively graded quasi-hereditary algebra in the sense of Cline, Parshall and Scott.*

6. FURTHER RESULTS

In this section, we begin by explaining an effective direct way to compute the multiplication in K_Λ , bypassing the closure and extension constructions from §4. Then we generalise the definition of Khovanov's algebra slightly, introducing an analogous cellular algebra H_Λ for any block Λ (not just Khovanov blocks). Finally we show that H_Λ is a symmetric algebra, extending a known result for Khovanov blocks, and discuss briefly some other aspects of the representation theory of generalised Khovanov algebras.

The generalised surgery procedure. We now introduce a generalisation of the surgery procedure from §3 which makes sense for arbitrary (not necessarily closed) circle diagrams. Suppose we are given a consistently oriented diagram consisting of an oriented cup diagram at the bottom, a symmetric middle section, and then an oriented cap diagram at the top. Pick a symmetric pair of a cup and a cap in the middle section that can be connected without crossings. Cut open the cup and cap and stitch the loose ends together exactly as in the original surgery procedure. If the cup and cap both belonged to circles, we re-orient the resulting diagram(s) following exactly the same rules (3.4)–(3.5) as before. However it is now also possible that one or both of the cup and cap belonged to a line, not a circle, before the cut was made. In that case, we supplement the rules (3.4)–(3.5) with the following extra rules, in which y denotes a line and 1 and x denote anti-clockwise and clockwise circles as before:

$$1 \otimes y \mapsto y, \quad x \otimes y \mapsto 0, \quad y \mapsto x \otimes y, \quad (6.1)$$

$$y \otimes y \mapsto \begin{cases} y \otimes y & \text{if both rays from one of the lines are oriented } \wedge \\ & \text{and both rays from the other line are oriented } \vee; \\ 0 & \text{otherwise.} \end{cases} \quad (6.2)$$

So, the rule $1 \otimes y \mapsto y$ indicates that an anti-clockwise circle and a line convert to a single line, the rule $y \mapsto x \otimes y$ indicates that a single line converts to a clockwise circle and a line, the rule $y \otimes y \mapsto y \otimes y$ indicates that two lines convert to two (different) lines, and so on. It is important to note that there is only ever one admissible choice for the orientation of vertices lying on lines; in particular, the orientations of vertices at the ends of rays are never changed by the generalised surgery procedure.

Theorem 6.1. *For any block Λ , the product $(a\lambda b)(b^*\mu d)$ of basis vectors from K_Λ can be computed as follows.*

- (i) *First draw $a\lambda b$ underneath $b^*\mu d$ and stitch corresponding rays together to create a new diagram with a symmetric middle section; corresponding rays in λb and $b^*\mu$ are necessarily oriented in the same direction so this makes sense.*
- (ii) *Then iterate the generalised surgery procedure just described in any order that makes sense to create a disjoint union of diagrams, none of which has any cup-cap pairs left in its symmetric middle section.*

- (iii) Finally identify the top and bottom number lines of each of the resulting diagrams to obtain a disjoint union of some new oriented circle diagrams. The desired product $(a\lambda b)(b^*\mu d)$ is the sum of the corresponding basis vectors of K_Λ .

Before proving the theorem, we illustrate the statement by using it to compute the following product:

$$\times \quad (6.3)$$

Putting the first diagram under the second and stitching corresponding rays together according to (i) gives:

$$(6.4)$$

Then we make two iterations of the generalised surgery procedure according to (ii):

Finally we identify top and bottom number lines as in (iii) to get the answer:

We leave it to the reader to check that this is indeed the desired product according to the original definition of multiplication.

Proof of Theorem 6.1. It is enough to consider the case that Λ consists of bounded weights. We require an obvious extension of the notion of the closure of a diagram from (4.7) to diagrams with a symmetric middle section. For example, the closure of the diagram (6.4) is:

$$(6.5)$$

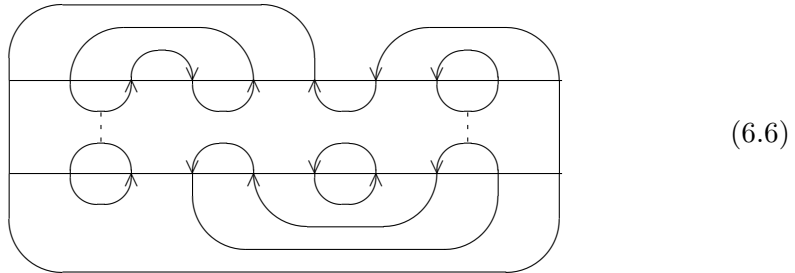
We will refer to the vertices in the diagram before closing as *inner vertices* and the additional vertices that get added as *outer vertices*. In (6.5), we have omitted the \vee 's labelling outer vertices on the left hand side and the \wedge 's labelling outer vertices on the right hand side; this convention helps to distinguish inner and outer vertices in practise.

According to the definition of multiplication in K_Λ , to compute the product $(a\lambda b)(b^*\mu d)$ we should first close both diagrams, then draw one under the other and iterate the surgery procedure. Finally at the end we identify top and bottom number lines as usual, discarding any diagrams corresponding to basis vectors in the ideal I_Λ , then remove the outer vertices in the remaining diagrams to get the final answer. If at any point during the sequence of surgery procedures some diagram is produced with an \wedge labeling an outer vertex on the left or a \vee labeling an outer vertex on the right, we can simply omit that diagram from the outset, since it can only ever create diagrams belonging to the ideal I_Λ at the end. There are three types of surgery procedure needed in this process:

- (a) surgeries involving caps which pass through two outer vertices;
- (b) surgeries involving caps which pass through one outer vertex;
- (c) surgeries involving caps which pass through no outer vertices.

We order the sequence of surgery procedures so that all the ones of type (a) are performed first, followed by type (b), followed by type (c).

The surgeries of type (a) are rather trivial: they only ever involve the rule $1 \otimes 1 \mapsto 1$ and no vertices change orientation during these operations. Now assume all type (a) surgeries have been performed and consider the surgeries of type (b). For example if we are in the situation of (6.3) and have just performed the one required type (a) surgery, we obtain the following diagram in which two type (b) surgeries are needed:



The surgeries of type (b) are obviously in bijective correspondence with the pairs of rays that need to be stitched together according to (i) from the statement of the theorem. In fact, after ignoring diagrams with an \wedge labelling an outer vertex on the left or a \vee labelling an outer vertex on the right, there is a unique diagram obtained after performing all type (b) surgeries and this diagram coincides with the closure of the diagram obtained by stitching rays together according to (i); e.g. in the running example we get exactly the diagram (6.5) on making the two type (b) surgeries to (6.6). This follows in general by considering the types of lines that the rays in b and b^* can possibly belong to, and considering their closures in each case.

So now we have performed all surgeries of types (a) and (b), and have obtained a diagram which is simply the closure of the diagram obtained after

performing (i). Finally, we need to perform the surgeries of type (c). These are in bijective correspondence with the generalised surgeries required in (ii). To complete the proof of the theorem, we run the two sequences of surgeries (from (ii) and from (c)) in parallel. The point is that the closures of the diagrams obtained after each generalised surgery from (ii) are exactly the same as the diagrams obtained after performing the corresponding surgery of type (c) (ignoring diagrams with an \wedge on the left or a \vee on the right). This follows by considering the various cases that occur in a single generalised surgery procedure and comparing each with its closure. We omit the details. \square

The generalised Khovanov algebra. Let Λ be any block and recall the definition (2.2) of $\text{def}(\Lambda)$. Let

$$\Lambda^\circ := \{\lambda \in \Lambda \mid \text{def}(\lambda) = \text{def}(\Lambda)\} \quad (6.7)$$

denote the subset of Λ consisting of all weights of *maximal defect*. If Λ is a Khovanov block this is the same as the set Λ° defined at the beginning of §3. Define the *generalised Khovanov algebra* H_Λ by setting

$$H_\Lambda := \bigoplus_{\alpha, \beta \in \Lambda^\circ} e_\alpha K_\Lambda e_\beta. \quad (6.8)$$

If $|\Lambda| < \infty$, we have simply that $H_\Lambda = eK_\Lambda e$ where $e := \sum_{\lambda \in \Lambda^\circ} e_\lambda \in K_\Lambda$. In general, H_Λ has basis

$$\{(\underline{\alpha}\lambda\bar{\beta}) \mid \text{for all } \alpha \in \Lambda^\circ, \lambda \in \Lambda, \beta \in \Lambda^\circ \text{ such that } \alpha \subset \lambda \supset \beta\}. \quad (6.9)$$

Moreover, the degree 0 component of H_Λ has basis

$$\{e_\lambda \mid \lambda \in \Lambda^\circ\}, \quad (6.10)$$

recalling (4.3). The multiplication in H_Λ can be computed easily using Theorem 6.1 and the generalised surgery procedure; moreover the rule $y \otimes y \mapsto y \otimes y$ from (6.2) is never needed in this maximal defect situation.

If Λ is a Khovanov block it is straightforward to see that H_Λ as defined here is the same as Khovanov's algebra H_Λ from §3. If $\text{def}(\Lambda) = \infty$ then $H_\Lambda = \{0\}$; *we will exclude this boring case for the remainder of the section*. Finally if $|\Lambda| = \infty$ and $\text{def}(\Lambda) < \infty$ it often happens that $H_\Lambda = K_\Lambda$. More precisely, using notation from the introduction, it is the case that

$$H_\Lambda = K_\Lambda \quad (6.11)$$

if $K_\Lambda \cong K_m^\infty \cong K_\infty^m$ or $K_\Lambda \cong K_m^{+\infty} \cong K_{-\infty}^m$, but not if $K_\Lambda \cong K_m^{-\infty} \cong K_{+\infty}^m$. (The isomorphisms $K_m^\infty \cong K_\infty^m$, $K_m^{+\infty} \cong K_{-\infty}^m$ and $K_m^{-\infty} \cong K_{+\infty}^m$ noted here come from (4.14) and (4.15).)

The results from §3 extend almost at once to the generalised Khovanov algebra H_Λ for any Λ . For example the anti-automorphism from (4.14) restricts to an anti-automorphism

$$* : H_\Lambda \rightarrow H_\Lambda, \quad (a\lambda b) \mapsto (b^*\lambda a^*). \quad (6.12)$$

The following theorem generalises Corollary 3.3.

Theorem 6.2. *Assuming $|\Lambda| < \infty$, the algebra H_Λ is a cellular algebra in the sense of Graham and Lehrer with cell datum $(\Lambda, M, C, *)$ where*

- (i) $M(\lambda)$ denotes $\{\alpha \in \Lambda^\circ \mid \alpha \subset \lambda\}$ for each $\lambda \in \Lambda$;
- (ii) C is defined by setting $C_{\alpha,\beta}^\lambda := (\underline{\alpha}\lambda\bar{\beta})$ for $\lambda \in \Lambda$ and $\alpha, \beta \in M(\lambda)$;
- (iii) $*$ is the anti-automorphism from (6.12).

Proof. Argue as in the proof of Corollary 3.3 using Theorem 4.4. \square

Symmetric algebra structure. Now we prove that the generalised Khovanov algebra H_Λ is a symmetric algebra. In fact, we prove that it possesses a *symmetrising form* $\tau : H_\Lambda \rightarrow \mathbb{F}$ such that the induced bilinear form $H_\Lambda \otimes H_\Lambda \rightarrow \mathbb{F}, a \otimes b \mapsto \tau(ab)$ is non-degenerate and symmetric. For Khovanov blocks, this was proved already by Khovanov in [K2, Proposition 32].

For a basis vector $(a\lambda b) \in H_\Lambda$, we let $(a\lambda b)^\#$ denote the basis vector of H_Λ whose diagram is obtained by reversing the orientation of every circle in the diagram $b^*\lambda a^*$. The map $(a\lambda b) \mapsto (a\lambda b)^\#$ is an involution on the basis of H_Λ such that $\deg(a\lambda b)^\# = 2 \operatorname{def}(\Lambda) - \deg(a\lambda b)$. In particular, this shows that the top degree component of H_Λ is in degree $2 \operatorname{def}(\Lambda)$ and by (6.10) it has basis $\{e_\lambda^\# \mid \lambda \in \Lambda^\circ\}$.

Theorem 6.3. *Let $\tau : H_\Lambda \rightarrow \mathbb{F}$ be the linear map such that $(a\lambda b) \mapsto 1$ if $\deg(a\lambda b) = 2 \operatorname{def}(\Lambda)$ and $(a\lambda b) \mapsto 0$ otherwise. Then*

$$\tau((a\lambda b)(c\mu d)) = \begin{cases} 1 & \text{if } (c\mu d) = (a\lambda b)^\#, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, τ is a symmetrising form and H_Λ is a symmetric algebra.

Proof. We have that $(c\mu d) = (a\lambda b)^\#$ if and only if

- $d = a^*$ and $c = b^*$, i.e. the circle diagrams ab and cd are mirror images of each other;
- the mirror image pairs of circles in ab and cd are oriented in opposite ways in $a\lambda b$ and in $c\mu d$.

It is clear that $\tau((a\lambda b)(c\mu d)) = 0$ unless $d = a^*$ and $c = b^*$. Assume therefore that $d = a^*$ and $c = b^*$. Let m be the number of circles in ab and $n := \operatorname{def}(\Lambda)$.

The first surgery procedure performed to a mirror image pair of circles in ab and cd turns two circles into one circles. So after making exactly m surgery procedures, one to each mirror image pair of circles in ab and cd , there will be a total of m circles, and still $(n - m)$ more surgery procedures to do. Now $\tau((a\lambda b)(c\mu d)) \neq 0$ only if there are exactly n clockwise circles at the end. To get n circles at all, each of the remaining $(n - m)$ surgeries have to turn one circle into two. The final circles then get oriented via the rules $1 \mapsto 1 \otimes x + x \otimes 1$ or $x \mapsto x \otimes x$. It follows that to get n clockwise circles at the end, the m circles obtained after the first m surgeries have to be all of type x , i.e. each mirror image pair of circles must indeed have been oriented in opposite ways in $a\lambda b$ and $c\mu d$ at the start. \square

Representations of the generalised Khovanov algebra. To conclude the article, we indicate briefly how to deduce the basic results about the representation theory of H_Λ from the analogous results already established for K_Λ . To

do this, we exploit the exact functor

$$e : \text{Mod}_f(K_\Lambda) \rightarrow \text{Mod}_f(H_\Lambda) \quad (6.13)$$

mapping a module M to $eM := \bigoplus_{\lambda \in \Lambda^\circ} e_\lambda M$ and defined by restriction on morphisms. When $|\Lambda| < \infty$ this is just the obvious truncation functor arising from the idempotent $e = \sum_{\lambda \in \Lambda^\circ} e_\lambda$. For $\lambda \in \Lambda$, we have that $eL(\lambda) \neq 0$ if and only if $\lambda \in \Lambda^\circ$, and $\{eL(\lambda) \mid \lambda \in \Lambda^\circ\}$ is a complete set of irreducible H_Λ -modules (up to degree shifts). The modules $\{eV(\mu) \mid \mu \in \Lambda\}$ are the *cell modules* for H_Λ in the sense of Graham and Lehrer arising from the cellular structure from Theorem 6.2. Finally we refer to the modules $\{eP(\lambda) \mid \lambda \in \Lambda\}$ as *Young modules*, by analogy with the modular representation theory of symmetric groups. If $\lambda \in \Lambda^\circ$ then $eP(\lambda)$ is the projective cover of $eL(\lambda)$. Applying the exact functor e , the results on transition matrices in the graded Grothendieck group from §5 easily imply similar statements for H_Λ .

REFERENCES

- [AJS] H. Andersen, J. Jantzen and W. Soergel, Representations of quantum groups at a p th root of unity and of semisimple groups in characteristic p : independence of p , *Astérisque* **220** (1994), 1–321.
- [B] T. Braden, Perverse sheaves on Grassmannians, *Canad. J. Math.* **54** (2002), 493–532.
- [BK] J. Brundan and A. Kleshchev, Schur-Weyl duality for higher levels, *Selecta Math.* **14** (2008), 1–57.
- [BS] J. Brundan and C. Stroppel, Morita equivalence of walled Brauer and Khovanov arc algebras, in preparation.
- [C] Y. Chen, Categorification of level two representations of quantum \mathfrak{sl}_n via generalized arc rings; [arXiv:math/0611012](https://arxiv.org/abs/math/0611012).
- [CK] Y. Chen and M. Khovanov, An invariant of tangle cobordisms via subquotients of arc rings; [arXiv:math/0610054](https://arxiv.org/abs/math/0610054).
- [CL] S.-J. Cheng and N. Lam, Irreducible characters of general linear superalgebra and super duality, *Comm. Math. Phys.* **298** (2010), 645–672.
- [CWZ] S.-J. Cheng, W. Wang and R.B. Zhang, Super duality and Kazhdan-Lusztig polynomials, *Trans. Amer. Math. Soc.* **360** (2008), 5883–5924.
- [CPS1] E. Cline, B. Parshall and L. Scott, Finite dimensional algebras and highest weight categories, *J. Reine Angew. Math.* **391** (1988), 85–99.
- [CPS2] E. Cline, B. Parshall and L. Scott, Duality in highest weight categories, *Contemp. Math.* **82** (1989), 7–22.
- [CPS3] E. Cline, B. Parshall and L. Scott, The homological dual of a highest weight category, *Proc. London Math. Soc.* **68** (1994), 296–316.
- [CT] J. Chuang and W. Turner, Cubist algebras, *Advances Math.* **217** (2008), 1614–1670.
- [CD] A. Cox and M. De Visscher, Diagrammatic Kazhdan-Lusztig theory for the (walled) Brauer algebra; [arXiv:1009.4064](https://arxiv.org/abs/1009.4064).
- [ES] T. Enright and B. Shelton, Categories of highest weight modules: applications to classical Hermitian symmetric pairs, *Mem. Amer. Math. Soc.* **367** (1987), 1–94.
- [GL] J. J. Graham and G. I. Lehrer, Cellular algebras, *Invent. Math.* **123** (1996), 1–34.
- [GS] C. Gruson and V. Serganova, Cohomology of generalized supergrassmannians and character formulae for basic classical Lie superalgebras; [arXiv:0906.0918](https://arxiv.org/abs/0906.0918).
- [HM] J. Hu and A. Mathas, Graded cellular bases for the cyclotomic Khovanov-Lauda-Rouquier algebras of type A , *Advances Math.* **225** (2010), 598–642.
- [HK] R. Huerfano and M. Khovanov, Categorification of some level two representations of \mathfrak{sl}_n , *J. Knot Theory Ramifications* **15** (2006), 695–713.
- [K1] M. Khovanov, A categorification of the Jones polynomial, *Duke Math. J.* **101** (2000), 359–426.

- [K2] M. Khovanov, A functor-valued invariant of tangles, *Alg. Geom. Topology* **2** (2002), 665–741.
- [KL] M. Khovanov and A. Lauda, A diagrammatic approach to categorification of quantum groups I, *Represent. Theory* **13** (2009), 309–347.
- [K] J. Kock, *Frobenius Algebras and 2-D Topological Quantum Field Theories*, LMS Student Texts **59**, 2003.
- [L] T. Lejczyk, *A Graphical Description of (A_{n-1}, D_n) Kazhdan-Lusztig Polynomials*, Diplomarbeit, Universität Bonn, 2010.
- [St] R. Stanley, *Enumerative Combinatorics*, vol. 2, Cambridge Studies in Advanced Mathematics 62, CUP, 1999.
- [S] C. Stroppel, Perverse sheaves on Grassmannians, Springer fibres and Khovanov homology, *Compositio Math.* **145** (2009), 954–992.
- [SW] C. Stroppel and B. Webster, 2-block Springer fibers: convolution algebras and coherent sheaves, to appear in *Comm. Math. Helv.*; [arXiv:0802.1943](https://arxiv.org/abs/0802.1943).

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

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BONN, 53115 BONN, GERMANY
E-mail address: stroppel@math.uni-bonn.de