STRONG HAAGERUP INEQUALITIES FOR FREE $\mathscr{R}$-DIAGONAL ELEMENTS

TODD KEMP* AND ROLAND SPEICHER**

ABSTRACT. In this paper, we generalize Haagerup's inequality [H] (on convolution norm in the free group) to a very general context of $\mathscr{R}$-diagonal elements in a tracial von Neumann algebra; moreover, we show that in this "holomorphic" setting, the inequality is greatly improved from its originial form. We give combinatorial proofs of two important special cases of our main result, and then generalize these techniques. En route, we prove a number of moment and cumulant estimates for $\mathscr{R}$-diagonal elements that are of independent interest. Finally, we use our strong Haagerup inequality to prove a strong ultracontractivity theorem, generalizing and improving the one in [Bi2].

1. INTRODUCTION

There is an interesting phenomenon which often occurs in holomorphic spaces. A theorem in the context of a function space (for example a family of norm-estimates, such as the L^p-bound of the Riesz projection, [R]) takes on a stronger form when restricted to a holomorphic subspace. L^p-bounds often shrink, and have meaningful extensions to the regime $p < 1$. For our purposes, the most relevant example is *Janson's strong hypercontractivity theorem* [Ja], discussed below. In algebraic terms, this theorem states that a certain semigroup has better properties when acting on the algebra generated by i.i.d. *complex* Gaussians than on the algebra generated by i.i.d. *real* Gaussians. The latter is a $*$-algebra while the former is far from one; we will exploit this difference in what follows.

In this paper, we will primarily be concerned with one prominent non-commutative norm inequality: the *Haagerup inequality*. It first arose in [H], where it was the main estimate used to foster an example of a non-nuclear C^*-algebra with the metric approximation property. In the context of [H], Haagerup's inequality takes the following form:

Theorem 1.1 ([H], Lemma 1.4)**.** *Let $\mathbb{F}_k$ be the free group on k generators, and let $f \in \ell^2(\mathbb{F}_k)$ be a function supported on words in $\mathbb{F}_k$ of length n. Then f acts as a convolutor on $\ell^2(\mathbb{F}_k)$, and its convolution norm $\|f\|_* = \sup_{\|g\|_2=1} \|f * g\|_2$ satisfies*

$$\|f\|_* \leq (n+1)\|f\|_2.$$

Note that the convolution product is just the usual product in the von Neumann algebra generated by the left-regular representation of $\mathbb{F}_k$ (known as the *free group factor* $L(\mathbb{F}_k)$), and so in the language of operator algebras the statement is that the (non-commutative) L^2-norm controls the operator norm on subspaces of uniform finite word-length, where the bound grows *linearly* with word-length.

The Haagerup inequality, and its decendents, have played important roles in several different fields. In the context of geometric group theory, the Haagerup inequality (and other constructions presented in [H]) have evolved into *a-T-menability* and are closely related to *Kazhdan's property (T)* [Va2]; in the context of Lie theory, Haagerup's inequality is related to *property RD* [Laf2]. It has proved useful for other operator algebraic applications: in [Laf1], Lafforgue uses the Haagerup inequality as a crucial tool in his proof of the Baum-Connes conjecture for cocompact lattices in $SL(3,\mathbb{R})$; in this context, the precise order of growth of the Haagerup constant is immaterial (so long as it is polynomial). On the other hand, the Haagerup inequality has proved useful in studying return probabilities and other statistics of random walks on groups (see [CPS, Va1]), where the exact form of the Haagerup constant is important.

Our main theorem, Theorem 1.3 below, is a strong Haagerup inequality in a general "holomorphic" setting – i.e. a non-self-adjoint algebra. In the special case of the free group factor, this amounts to considering convolution operators which involve only generators of the group, not their inverses; the resulting Haagerup inequality (Theorem 1.4 below) then has growth of order $\sqrt{n}$, where n is the word-length.

There are two main approaches to norm estimates in such a setting. A direct one (as used in the original approach of Haagerup) is to work directly in the concrete representation of the considered element as operator on a Hilbert space and try to estimate the operator norm by considering the action of the operator on vectors. A more indirect approach is to recover the operator norm as the limit of the L^p-norms as $p \to \infty$, and therefore try to get a combinatorial understanding of L^p-norms for $p = 2m$ even. It is the latter approach which we take. Thus, we need a good (at least asymptotic) understanding of the moments of the involved operators with respect to the underlying state. To our benefit, the moments of the generators of free groups possess a lot of structure: namely the generators are free in the sense of Voiculescu's free probability theory.

Our strong Haagerup inequality is actually derived in a much more general setting: algebras generated by free $\mathscr{R}$-diagonal elements. We therefore handle not only the original framework of Haagerup (in the form of free Haar unitaries), but also free circular elements, and a wealth of other non-normal operators.

There have been some predecessor of our strong Haagerup inequality for the general $\mathscr{R}$-diagonal case. Namely, the one-dimensional case was mainly addressed in [HL] and, in particular, in [Lar]. Furthermore, [Lar] contains a very specialized multi-dimensional case, where the considered operator is a product of identically-distributed free $\mathscr{R}$-diagonal elements. All these results relied on analytic techniques, using the theory of $\mathscr{R}$- and $\mathscr{S}$-transforms for probability measures on $\mathbb{R}$. However, in the genuine non-commutative case of polynomials in several non-commuting $\mathscr{R}$-diagonal elements, as we treat it here, such analytical tools are unavailable to us, and so our analysis will rely on the combinatorial machinery of free cumulants, as powered by free probability theory.

Our main tool is the moment-cumulant formula (Equation 2.5, below), which expresses the moments of the considered elements in a very precise combinatorial way in terms of free cumulants. This allows us to reduce the multi-dimensional case essentially to the one-dimensional case. (Note that this reduction is usually the hardest part in such inequalities.) Whereas in some cases (as for circular elements) this reduction directly yields the desired result, in other cases – namely when the cumulants of the $\mathscr{R}$-diagonal element may be negative (as it happens for Haar unitaries, i.e., in the free group situation) – we need an additional step. Our strategy is to replace the original $\mathscr{R}$-diagonal element a with a different $\mathscr{R}$-diagonal element b whose cumulants are positive and dominate the absolute values of the cumulants of a; this has to be done in such a way that we have control over both the L^2-norm and the operator norm of b in terms of the corresponding norms of a. The technique we develop will, we hope, have more general applicability.

Let us now give a precise definition of the arena for our Haagerup inequality. Section 2 contains brief introductions to all the terms used in what follows (and in the foregoing).

Definition 1.2. *Let I be any indexing set, and let $\{a_i : i \in I\}$ be $*$-free identically distributed $\mathscr{R}$-diagonal elements in a C^*-probability space with state φ; for convenience, let a be a fixed $\mathscr{R}$-diagonal element with the same $*$-distribution. Define $\mathcal{H}(a, I)$ to be the norm-closed (non-$*$) algebra generated by the a_i. For each*

$n \geq 0$, define $\mathcal{H}^{(n)}(a, I)$ as the Hilbert subspace of $L^2(\mathcal{H}(a, I), \varphi)$ generated by the set of elements of the form

$$T = \sum_{|\mathbf{i}|=n} \lambda_{\mathbf{i}} a_{\mathbf{i}},$$

*where $\mathbf{i} = (i_1, \ldots, i_n) \in I^n$, $\lambda_{\mathbf{i}} \in \mathbb{C}$ and only finitely many are non-zero, and $a_{\mathbf{i}} = a_{i_1} \cdots a_{i_n}$. We refer to $\mathcal{H}^{(n)}(a, I)$ as the n-**particle space** (relative to a, I).*

The motivation for considering the algebra $\mathcal{H}(a, I)$ comes from the first author's paper [Ke], and [Bi1]. If c is a circular element, then $L^2(\mathcal{H}(c, I), \varphi)$ is a free analogue of the *Segal-Bargmann space* of [Ba] – i.e. the space $\mathcal{H}L^2(\mathscr{H}, \gamma)$ of holomorphic functions on a Hilbert space $\mathscr{H}$ of dimension $|I|$, square-integrable with respect to a certain Gaussian measure γ. The Segal-Bargmann space is the framework for the complex wave representation of quantum mechanics. It played an important role in the constructive quantum field theory program in the mid- to late-twentieth century.

There is a natural operator, the *Ornstein-Uhlenbeck operator* or *number operator* N on $L^2(\mathscr{H}, \gamma)$, which is related to the energy operator in quantum field theory. In the classical (Gaussian) context, the Ornstein-Uhlenbeck semigroup e^{-tN} satisfies a regularity property called *hypercontractivity*: for $1 < p \leq r < \infty$ the semigroup e^{-tN} is a contraction from $L^p(\mathscr{H}, \gamma)$ to $L^r(\mathscr{H}, \gamma)$ for large enough time t. When e^{-tN} is restricted to the Segal-Bargmann space and its holomorphic L^p generalizations, the time to contraction is shorter, as shown in [Ja] and generalized in [G]. This *strong hypercontractivity* demonstrates that contraction properties of the Ornstein-Uhlenbeck semigroup improve in the holomorphic category.

In [Bi2], Biane showed how to canonically generalize the Ornstein-Uhlenbeck operator to an operator N_0 in the setting of the free group factor, and proved that the resulting semigroup e^{-tN_0} is hypercontractive. He further showed that the semigroup e^{-tN_0} satisfies an even stronger condition called *ultracontractivity*: it continuously maps L^2 into L^∞ for *all* $t > 0$, and for small time $\|e^{-tN_0}\|_{2\to\infty}$ is of order $t^{-3/2}$. This result was proved using a version of the Haagerup inequality presented in [Bo1]. We should note that, although this result is for the free group factor, the n-particle spaces used in the proof are *not* the same as in Theorem 1.1, but are rather defined in terms of a generating family of *semicircular elements* defined in Section 2; nevertheless, the relevant Haagerup inequality *can* be proved from Theorem 1.1 using a central limit approach similar to the one in [VDN].

It is Biane's free ultracontractivity theorem, along with our intuition that norm-inequalities improve in holomorphic categories, that motivated us to consider the same type of Haagerup inequality for $\mathscr{R}$-diagonal elements. In the special case of a circular element c, the first author showed in [Ke] that in the holomorphic category (in this case the spaces $L^p(\mathcal{H}(c, I), \varphi)$) Biane's hypercontractivity result is trumped by Janson's strong hypercontractivity. The first author further spelled out precisely the holomorphic structure inherent in $\mathcal{H}(c, I)$. Our interpretation of an $\mathscr{R}$-diagonal element a as "holomorphic" is more vague. Nevertheless, the algebra $\mathcal{H}(a, I)$ is a triangular algebra much like the space of bounded Hardy functions H^∞ is (as a Banach algebra acting on $L^2(S^1)$). More importantly, the kinds of norm estimates used in [Ke] have natural analogues for $\mathscr{R}$-diagonal elements.

The following theorem, which is our strong version of Haagerup's inequality in the general $\mathscr{R}$-diagonal setting, is the main result of this paper.

Theorem 1.3. *Let a be an $\mathscr{R}$-diagonal element in a C^*-probability space. There is a constant $C_a < \infty$ such that for all $T \in \mathcal{H}^{(n)}(a, I)$,*

$$\|T\| \leq C_a \sqrt{n}\, \|T\|_2. \tag{1.1}$$

In general, C_a may be taken $\leq 515\sqrt{e}\, \|a\|^2/\|a\|_2^2$; if a has non-negative free cumulants, C_a may be taken $\leq \sqrt{e}\, \|a\|/\|a\|_2$.

Since both circular operators and Haar unitary operators are $\mathscr{R}$-diagonal, Theorem 1.3 yields surprisingly strong versions of the classical Haagerup inequality for both of them. For a circular element c, we will show that the (asymptotically) optimal inequality is $\|T\| \le \sqrt{e}\sqrt{n+1}\,\|T\|_2$, and since $\|c\| = 2$ while $\|c\|_2 = 1$, the constant in Theorem 1.3 is the best possible in this case. On the other hand, the free cumulants of a Haar unitary operator u are of alternating sign, and so the constant C_u from the above theorem may be as large as $515\sqrt{e}$. In fact, we will show the optimal inequality in this case is again the same as that for circular elements. (Regardless of the optimal constant, it is the fact that the estimate is $O(n^{1/2})$ rather than $O(n)$ which is striking.) Since the generators of a free group correspond to $*$-free Haar unitary elements in a free group factor, this result may be interpretted directly in terms of the free group as in Theorem 1.1, as follows.

Theorem 1.4. *Let $k \ge 2$, let $\mathbb{F}_k$ be the free group on k generators, and let $\mathbb{F}_k^+ \subset \mathbb{F}_k$ be the free* semigroup *(i.e. the set of all words in the generators, excluding their inverses). If $f \in \ell^2(\mathbb{F}_k^+) \subset \ell^2(\mathbb{F}_k)$ is supported on words of length n, then f acts (via the left-regular representation on the* full group $\mathbb{F}_k$*) as a convolutor, with convolution norm*

$$\|f\|_* \le \sqrt{e}\,\sqrt{n+1}\,\|f\|_2.$$

This paper is organized as follows. In section 2, we give a brief introduction to free probability theory and $\mathscr{R}$-diagonal elements, in addition to setting the standard notation we will use throughout the paper. In Section 3, we provide a concrete bijection in order to calculate the moments of a circular element c; in it we derive, using more elementary techniques, a formula for $\|c^n\|$, confirming results in [O] and [Lar]. We then use this calculation, together with more involved combinatorial techniques, to estimate the norm of an element in the n-particle space $\mathcal{H}^{(n)}(c, I)$ for arbitrary indexing set I, and thus prove a special case of Theorem 1.3 in the circular context. We use the same combinatorial bijection, in a different way, to prove Theorem 1.4 for Haar unitaries.

In Section 4, we show how to modify the techniques in Section 3 to prove Theorem 1.3 in general. In the process, we derive bounds on the growth of the free cumulants of $\mathscr{R}$-diagonal elements and, given an $\mathscr{R}$-diagonal a, show how to construct another $\mathscr{R}$-diagonal element b with all positive cumulants dominating the cumulants of a. We also show that the Haagerup inequality affiliated to the space $\mathcal{H}L^2(\nu_a)$ of holomorphic functions square integrable with respect to the Brown measure ν_a of a is consistent with Theorem 1.3, which shows that ν_a does carry some information about the mixed moments of a. Finally, in Section 5, we introduce a natural analogue of the Ornstein-Uhlenbeck semigroup affiliated with $\mathcal{H}(a, I)$, and prove a strong ultracontractivity theorem for it.

2. A Free Probability Primer

In this section we collect all the relevant results from free probability theory that will be used in what follows. Our descriptions will be brief, as this material is quite standard and is explained in depth in the book [NS3].

2.1. C^*-probability spaces.

Let $\mathscr{A}$ be a unital C^* algebra, and let φ be a faithful state on $\mathscr{A}$ (i.e. φ is a continuous linear functional on $\mathscr{A}$ and, for $a \in \mathscr{A}$, $\varphi(a^*a) \ge 0$ and vanishes only when $a = 0$). The pair $(\mathscr{A}, \varphi)$ is a *C^*-probability space*. Elements of $\mathscr{A}$ are *non-commutative random variables* (which we will often refer to simply as *random variables*). (Let us point out that random variables do not have to be self-adjoint, or even normal, in general.) The motivating example is afforded by the commutative von Neumann algebra $L^\infty(\Omega, \mathcal{F}, P)$ of a probability space. It comes equipped with the faithful state $\varphi = \int_\Omega \cdot\, dP$; the random variables in this context are bounded random variables in the usual sense.

In classical probability theory, any random variable X has a probability distribution ν_X – a measure on $\mathbb{C}$ which, among other things, determines the moments of X:

$$\int_\Omega X(\omega)^n \overline{X(\omega)}^m \, dP(\omega) = \int_{\mathbb{C}} z^n \overline{z}^m \, d\nu_X(z, \bar{z}).$$

In the case of a real random variable X, ν_X is supported in $\mathbb{R}$ and we have $\int X^n \, dP = \int_{\mathbb{R}} t^n d\nu_X(t)$. At least in the case of bounded random variables, these moment conditions uniquely determine the distribution, which is a compactly-supported probability measure. The same holds true for normal elements in a C^*-probability space – if a is normal then there is a unique probability measure ν_a on $\mathbb{C}$ which satisfies

$$\varphi\left(a^n (a^*)^m\right) = \int_{\mathbb{C}} z^n \overline{z}^m \, d\nu_a(z, \bar{z}), \tag{2.1}$$

and the measure ν_a is compactly supported. Indeed, $\operatorname{supp} \nu_a$ is the spectrum of a, and the measure can be constructed using the spectral theorem: $\nu_a = \varphi \circ E^a$ where E^a is the spectral resolution of a in $\mathscr{A}$.

If a is not a normal element, then there is no measure satisfying Equation 2.1; more generally, given two elements in $\mathscr{A}$ that do not commute, there is no measure which represents their joint probability distribution (this is one way to state the Heisenberg uncertainty principle). In the case where $(\mathscr{A}, \varphi)$ is a tracial W^*-probability space ($\mathscr{A}$ is a von Neumann algebra, φ is a faithful normal tracial state) however, there is a best-approximation of a probability distribution called the *Brown measure*, introduced in [Br]. If a is normal, then its Brown measure coincides with its spectral measure, and so the Brown measure is also denoted ν_a. The Brown measure of a always satisfies the moment condition $\varphi(a^n) = \int_{\mathbb{C}} z^n \, d\nu_a(z, \bar{z})$, however it does not respect mixed-moments.

2.2. **The free group factors.** Free probability was invented by Voiculescu in [Vo] in order to import tools from classical probability theory into the study of *the free group factors* (specifically to address the still-open question of whether different free group factors are isomorphic).

Let $k \geq 2$, and let $\mathbb{F}_k$ denote the free group on k generators $u_1, u_2, \ldots, u_k$. (We will also allow $k = \infty$ to denote the free group with countably-many generators.) The *kth free group factor* $L(\mathbb{F}_k)$ is the von Neumann algebra generated by the left-regular representation of $\mathbb{F}_k$ on $\ell^2(\mathbb{F}_k)$. (Note: if $g \in \mathbb{F}_k$, then the image of g in $L(\mathbb{F}_k)$ is an operator with $g^* = g^{-1}$.) There is a natural state φ_k defined on $L(\mathbb{F}_k)$ induced by the indicator function $\mathbb{1}_e$ of the identity $e \in \mathbb{F}_k$. This state is faithful, normal, and tracial, making $(L(\mathbb{F}_k), \varphi_k)$ into a W^*-probability space.

There is a canonical representation of the free group factor on the full Fock space. Let $\mathscr{H}$ be a *real* Hilbert space, and let $\mathscr{H}_{\mathbb{C}} = \mathbb{C} \otimes \mathscr{H}$ be its complexification. The *full Fock space* of $\mathscr{H}$ is $\mathcal{F}(\mathscr{H}) = \bigoplus_{j=0}^\infty (\mathscr{H}_{\mathbb{C}})^{\otimes j}$, where $\oplus$ and $\otimes$ are the Hilbert space direct sum and tensor product, and $(\mathscr{H}_{\mathbb{C}})^{\otimes 0}$ is defined to be the $\mathbb{C}$-span of an abstract unit vector Ω (not in $\mathscr{H}$) called the *vacuum vector*.

For each $h \in \mathscr{H}$, the *creation operator* $l(h)$ in $\mathscr{B}(\mathcal{F}(\mathscr{H}))$ is uniquely defined by its action $l(h)(h_1 \otimes \cdots \otimes h_j) = h \otimes h_1 \otimes \cdots \otimes h_j$ on $(\mathscr{H}_{\mathbb{C}})^{\otimes j}$ (and $l(h)\Omega = h$). The adjoint $l(h)^*$ is called the *annihilation operator*, and is given by $l(h)^*(h_1 \otimes h_2 \otimes \cdots \otimes h_j) = \langle h_1, h \rangle h_2 \otimes \cdots \otimes h_j$ (and $l(h)^*\Omega = 0$). The operator $l(h)$ is not normal (if $h \neq 0$), but it is natural to consider (2 times) the real part $X(h) = l(h) + l(h)^*$. For any k-dimensional real Hilbert space $\mathscr{H}$, the von Neumann algebra generated by $\{X(h) : h \in \mathscr{H}\}$ is isomorphic to $L(\mathbb{F}_k)$. What's more, under this isomorphism, the state φ_k conjugates to the *vacuum expectation state* $\tau(X) = \langle X\Omega, \Omega \rangle$.

Let $e_1, \ldots, e_k$ be an orthonormal basis for $\mathscr{H}$. The algebra $W^*\{X(h) : h \in \mathscr{H}\} \cong L(\mathbb{F}_k)$ is, of course, generated by the set $\{X(e_1), \ldots, X(e_k)\}$. It is important to note that the isomorphism does *not* carry the generators $u_1, \ldots, u_k$ in $\mathbb{F}_k \subset L(\mathbb{F}_k)$ to the generators $X(e_1), \ldots, X(e_k)$. Indeed, the two generating sets give two different, and important, families of non-commutative random variables: *Haar unitary* and *semicircular* elements, which we will discuss below. In both cases, the relationship between different generators is a model of a non-commutative version of independence called *freeness*.

2.3. **Free cumulants and free independence.** A normal random variable in a C^*-probability space is indistinguishable from a classical bounded complex random variable (indeed, one can construct a random variable with any given distribution ν as the identity function in the space $L^\infty(\nu)$.) The important classical notion of independence of random variables, however, has no direct analog for pairs of non-commuting random variables. The notion of *free independence* or *freeness*, introduced in [Vo], is a substitute which is, in many ways, better.

Let $\pi = \{V_1, \ldots, V_r\}$ be a partition of the set $\{1, \ldots, n\}$. The partition is called *crossing* if for some $i \neq j$ there are numbers $p < q < p' < q'$ with $p, p' \in V_i$ and $q, q' \in V_j$. (*Notation*: we say $p \sim_\pi q$ if p, q are in the same block of the partition π. Thus, π is crossing iff there are $p < q < p' < q'$ with $p \sim_\pi p'$, $q \sim_\pi q'$, and $p' \nsim_\pi q$.) A *non-crossing partition* is one which is not crossing. We represent a partition by connecting numbers in the same block V_i of the partition. Figure 1 gives four examples of non-crossing partitions of the set $\{1, \ldots, 6\}$.

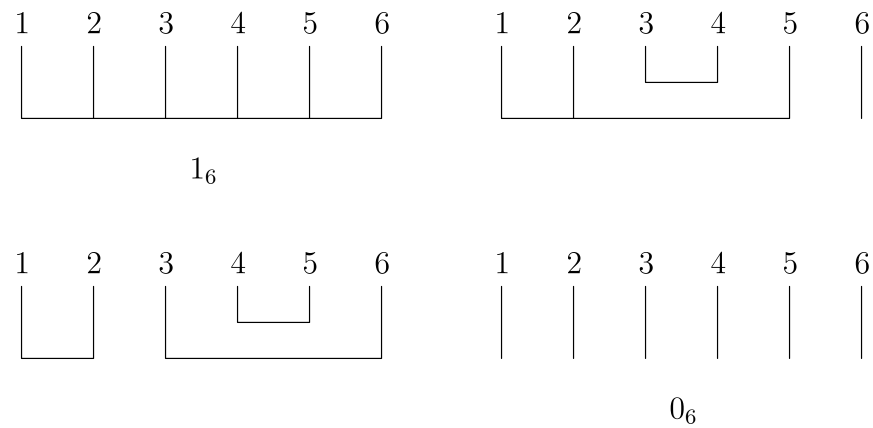


FIGURE 1. Four elements of $NC(6)$, including the minimal and maximal elements 0_6 and 1_6.

An often useful characterization of non-crossing partitions is given by the following recursive definition: a partition π of $\{1, \ldots, n\}$ is non-crossing iff at least on block of π is an interval $V = \{k, k+1, \ldots, k+r\}$, and the partition $\pi - V$ of the ordered set $\{1, \ldots, k-1, k+r+1, \ldots, n\}$ is non-crossing. For example, in the upper-right partition in Figure 1, the block $\{3, 4\}$ is an interval, and removing this block we have the partition $\{\{1, 2, 5\}, \{6\}\}$ of the set $\{1, 2, 5, 6\}$. Both these remaining blocks are intervals.

The set of non-crossing partitions of $\{1, \ldots, n\}$, denoted $NC(n)$, is partially-ordered under reverse refinement. It is a lattice, in fact, with minimal element 0_n and maximal element 1_n as in Figure 1. The Möbius function μ_n of this lattice is well-known (see [Kr]). In particular, $\mu_n(0_n, 1_n) = (-1)^{n-1} C_{n-1}$, where C_n are the *Catalan numbers*

$$C_n = \frac{1}{n} \binom{2n}{n-1}. \tag{2.2}$$

More generally, for any $\sigma \in NC(n)$,

$$|\mu_n(\sigma, 1_n)| \leq 4^{n-1}. \tag{2.3}$$

(The proof can be found contained in the proof of Proposition 13.15 in [NS3].) It is worth noting that $C_n \leq 4^n$ (and indeed $C_n \asymp 4^n$).

Let $(\mathscr{A}, \varphi)$ be a C^*-probability space. Let $n > 0$ and let π be a partition in $NC(n)$. For each block $V = \{i_1, \ldots, i_k\}$ in π, define the function $\varphi_V : \mathscr{A}^n \to \mathbb{C}$ by $\varphi_V[a_1, \ldots, a_n] = \varphi(a_{i_1} \cdots a_{i_k})$. Then define $\varphi_\pi : \mathscr{A}^n \to \mathbb{C}$ by $\varphi_\pi[a_1, \ldots, a_n] = \prod_{V \in \pi} \varphi_V[a_1, \ldots, a_n]$. Finally, define the *free cumulants* of $(\mathscr{A}, \varphi)$ to be the functionals $\{\kappa_\pi : \pi \in NC(n) \text{ for some } n > 0\}$ by

$$\kappa_\pi[a_1, \ldots, a_n] = \sum_{\substack{\sigma \in NC(n) \\ \sigma \leq \pi}} \varphi_\sigma[a_1, \ldots, a_n]\, \mu_n(\sigma, \pi), \tag{2.4}$$

for each $\pi \in NC(n)$. An immediate consequence of this definition is that the moments can be recovered from the free cumulants,

$$\varphi_\pi[a_1, \ldots, a_n] = \sum_{\substack{\sigma \in NC(n) \\ \sigma \leq \pi}} \kappa_\sigma[a_1, \ldots, a_n].$$

(Indeed, this is the motivation for the inclusion of the coefficients $\mu_n(\sigma, \pi)$ in the definition of κ_π, for the Möbius function is the convolution-inverse of the Zeta-function for the lattice $NC(n)$.) As a special case, we have the formula

$$\varphi(a_1 a_2 \cdots a_n) = \sum_{\pi \in NC(n)} \kappa_\pi[a_1, \ldots, a_n]. \tag{2.5}$$

Free cumulants allow a very easy statement of the definition of free independence, or freeness, of random variables. Let κ_n denote the free cumulant κ_{1_n}. (These cumulants in fact contain all information about the cumulants, since all others can be built up block-wise by multiplication.) Elements $a_1, \ldots, a_n$ in $\mathscr{A}$ are called *free* if, for $j \geq 2$ and $1 \leq i_1, \ldots, i_j \leq n$, $\kappa_j[a_{i_1}, \ldots, a_{i_j}] = 0$ whenever there is at least one pair $1 \leq \ell, m \leq j$ with $i_\ell \neq i_m$. In other words, *random variables are free if all their mixed free cumulants vanish*.

One can calculate that the generators $u_1, \ldots, u_n$ of $\mathbb{F}_n \subset L(\mathbb{F}_n)$ are free, as are the generators $X(e_1), \ldots, X(e_n)$ in the Fock-space representation of $L(\mathbb{F}_n)$; hence, this notion generalizes freeness from the free group context. This approach mirrors the classical theory of cumulants in the method of moments (where the lattice considered is the lattice of *all* partitions). All of the usual probabilistic constructions work: given any countable list of probability measures ν_j, there is a C^* probability space in which there are free random variables with distributions ν_j (one can construct the reduced free-product C^* algebra of the $L^\infty(\nu_j)$, for example).

2.4. **$\mathscr{R}$-diagonal elements.** As commented above, the operators $X(e_j)$ in the Fock-space representation of $L(\mathbb{F}_n)$ are semicircular elements: $s = X(e_j)$ has as distribution ν_s with

$$d\nu_s(t) = \frac{1}{2\pi}\sqrt{4 - t^2}\,\mathbb{1}_{[-2,2]}\, dt.$$

The cumulants of a semicircular element s can be calculated quite easily (see [NS3]): $\kappa_n[s, \ldots, s] = \delta_{n2}$. (In this sense, they are analogues of standard normal random variables, whose classical cumulants are the same.)

Let s_1, s_2 be two free semicircular random variables. The operator $c = (s_1 + is_2)/\sqrt{2}$ (where $i = \sqrt{-1}$) is called a *circular element*. It is non-normal, and so does not have a probability distribution. (It's Brown measure is known, however, to be the uniform measure on the closed unit disc in $\mathbb{C}$.) The $*$-cumulants of a circular element (i.e. the free cumulants of tuples of operators all of the form c or c^*) have a particularly nice form. If $\varepsilon_j \in \{1, *\}$ then $\kappa_n[c^{\varepsilon_1}, \ldots, c^{\varepsilon_n}] = 0$ for $n \neq 2$, and in fact only $\kappa_2[c, c^*] = \kappa_2[c^*, c] = 1$ are nonzero.

Consider also a generator $u = u_j$ of $\mathbb{F}_k$. Note that $\varphi_k(u^n) = \delta_{n0}$, and the same holds true for $u^* = u^{-1}$. The spectral measure of u is thus the Haar measure on the unit circle, and such random variables are called *Haar unitary*. The $*$-cumulants of a Haar unitary are not as restricted as those of a circular, but they follow a similar pattern. The only nonvanishing cumulants κ_n have n even, and must have alternating u and u^* arguments:

$$\kappa_{2n}[u, u^*, \ldots, u, u^*] = \kappa_{2n}[u^*, u, \ldots, u^*, u] = (-1)^{n-1} C_{n-1},$$

the same as the Möbius coefficents $\mu_n(0_n, 1_n)$ of $NC(n)$ (and this is no coincidence).

This connection between two widely known classes of non-selfadjoint random variables (circulars and Haar unitaries) motivated the second author, in [NS1], to introduce *$\mathscr{R}$-diagonal elements*. A random variable a in a C^*-probability space is $\mathscr{R}$-diagonal if its only novanishing cumulants are the alternating ones $\kappa_{2n}[a, a^*, \ldots, a, a^*]$ and $\kappa_{2n}[a^*, a, \ldots, a^*, a]$. (The notation $\mathscr{R}$-diagonal derives from a characterization of such elements in terms of the multivariate *$\mathscr{R}$-transform*, a combinatorial free version of the logarithmic Fourier transform in classical probability theory.)

Note that an $\mathscr{R}$-diagonal element's odd cumulants vanish. (The term *even element* is used in this context, but is usually formulated in terms of *mixed* moments, so we do not use it for $\mathscr{R}$-diagonal elements.) From Equations 2.4 and 2.5 we see vanishing of odd cumulants is equivalent to vanishing of odd moments. If a is $\mathscr{R}$-diagonal, its *determining sequences* are $(\alpha_n[a])_{n=1}^{\infty}$ and $(\beta_n[a])_{n=1}^{\infty}$ defined by

$$\begin{aligned} \alpha_n[a] &= \kappa_{2n}[a, a^*, \ldots, a, a^*], \\ \beta_n[a] &= \kappa_{2n}[a^*, a, \ldots, a^*, a]. \end{aligned} \tag{2.6}$$

If a is in a tracial probability space (better yet if φ restricted to the algebra generated by a and a^* is tracial), then $\alpha_n[a] = \beta_n[a]$; in any case, these sequences contain all the information about the cumulants (and therefore mixed moments) of a and a^*.

$\mathscr{R}$-diagonal elements form a large class of (mostly) non-normal elements about which a great deal is known. In a sense, they are non-normal analogues of rotationally invariant distributions in $\mathbb{C}$; namely, the distribution of an $\mathscr{R}$-diagonal element is not changed if is multiplied by a free Haar unitary. This results in a special polar decomposition and relations with maximization problems for free entropy [NS3, NSS, HP]. Our main theorem (Theorem 1.3) supports the point of view that $\mathscr{R}$-diagonal elements can be considered as non-normal versions of holomorphic variables.

Finally, we comment that there is a precise description of the Brown measure of an $\mathscr{R}$-diagonal element in terms of its $\mathscr{S}$-transform (another formal power-series associated to the moments of a). The following theorem shows that $\mathscr{R}$-diagonal elements have rotationally-invariant Brown measures with nice densities. Let $\times_p$ denote the polar Cartesian product (i.e. $[x, y] \times_p [0, 2\pi)$ is the closed annulus with inner-radius x and outer-radius y).

Theorem 2.1 (Corollary 4.5 in [HL])**.** *If a is $\mathscr{R}$-diagonal and is not a scalar multiple of a Haar unitary, then its Brown measure ν_a is supported on $\left(\|a^{-1}\|_2^{-1}, \|a\|_2\right] \times_p [0, 2\pi)$ if a is invertible, and on the disc $[0, \|a\|_2] \times_p [0, 2\pi)$ if it is not. Moreover, ν_a is rotationally-invariant with density*

$$d\nu_a(r, \theta) = f(r)\, dr\, d\theta,$$

where f is strictly positive on $\left(\|a^{-1}\|_2^{-1}, \|a\|_2\right]$ or $[0, \|a\|_2]$ and has an analytic continuation to a neighbourhood of this interval in $\mathbb{C}$.

3. Circular and Haar Unitary Elements

In this section, we prove Theorem 1.3 in the special case that a is circular c or Haar unitary u. Our proof in Section 4 subsumes this one, but the techniques are new and motivate the later proof. In Section 3.1, we introduce an important combinatorial structure which underlies the moments and cumulants of circular and Haar unitary elements (and, in some sense, all $\mathscr{R}$-diagonal elements), and use it to give a new proof that the $*$-moments of the powers of a circular element are the *Fuss-Catalan numbers*, defined in Equation 3.5 below. The main ideas of the construction in this section are due to Drew and Heather Armstrong, and we thank them for their contribution. Since the result was already known, we leave much of the detail here to the pictures. In Section 3.2, we use the asymptotics of the Fuss-Catalan numbers to demonstrate the strong Haagerup inequality for algebras generated by free circular elements. We conclude with Section 3.3 by showing that the same combinatorial structure appears in the direct calculation of moments in the Haar unitary case, and thence prove Theorem 1.4.

3.1. **The powers of a circular element.** Let c be a (variance 1) circular element in a C^*-probability space $(\mathscr{A}, \varphi)$. The moments of c^n were calculated first by Oravecz [O] and Larsen [Lar], each using a different approach to iterated free convolution of the $\mathscr{R}$-transform of c. We will reproduce their results here, using more elementary combinatorial techniques.

From Equation 2.5, we have

$$\varphi[(c^n(c^n)^*)^m] \quad = \sum_{\pi\in NC(2nm)} \kappa_\pi[c_{n,m}], \tag{3.1}$$

where $c_{n,m}$ is the list

$$c_{n,m} = \overbrace{\underbrace{c,\ \ldots, c,}_{n}\ \underbrace{c^*, \ldots, c^*,}_{n} \ldots, \underbrace{c,\ \ldots, c,}_{n}\ \underbrace{c^*, \ldots, c^*}_{n}}^{2m \text{ groups}}\ . \tag{3.2}$$

Since c is circular, its only nonzero free cumulants are $\kappa_2[c, c^*] = 1$ and $\kappa_2[c^*, c] = 1$, hence the only nonzero terms in the above sum are those for which the partition π is a pair partition $\pi \in NC_2(2mn)$ (each block is of size 2), and for which each c is paired to a c^* in $c_{n,m}$. We call such pairings $*$-*pairings*, and denote the set of $*$-pairings in $NC_2(2mn)$ by $NC_2^*(n, m)$. Pictured below are two examples of elements in $NC_2^*(3, 4)$.

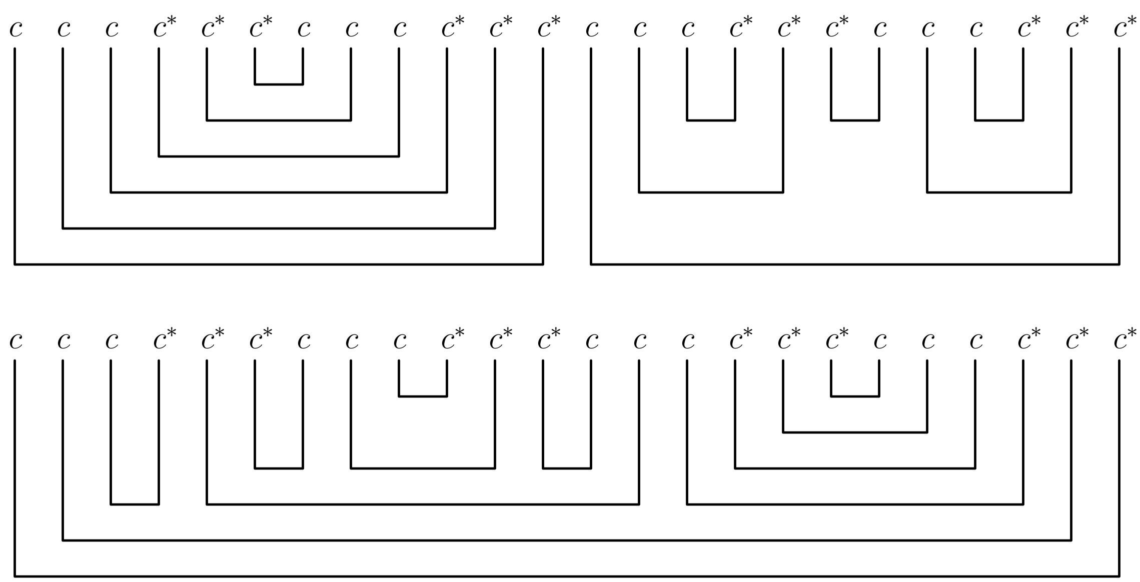


FIGURE 2. Two $*$-pairings in $NC_2^*(3, 4)$.

Since $\kappa_\pi[c_{n,m}] = 1$ whenever $\pi \in NC_2^*(n, m)$ and equals 0 otherwise, Equation 3.1 reduces to

$$\|c^n\|_{2m}^{2m} \quad = \sum_{\pi\in NC_2^*(n,m)} 1 \quad = \quad |NC_2^*(n, m)|. \tag{3.3}$$

A non-crossing partition can be represented linearly as in Figures 1 and 2, or equivalently on a circle, as seen below in Figure 3. As such, we can describe the problem of counting the elements in $NC_2^*(n,m)$ in the following medieval terms:

Knights and Ladies of the Round Table. King Arthur's Knights wish to bring their Ladies to a meeting of the Round Table. There are $k = nm$ Knights (including Arthur himself) and each has one Lady. Arthur wishes to seat everyone so that men and women alternate in groups of n, and in such a way that each Lady can converse with her Knight across the table without any conversations crossing. How many possible seating plans are there?

Letting c stand for "Knight" and c^* stand for "Lady," the pictures in Figure 3 (which are the circular representations of the pairings from Figure 2) represent allowable seating plans. (To be clear: King Arthur's seat is fixed at the top; without this constraint we overcount by a factor of m.)

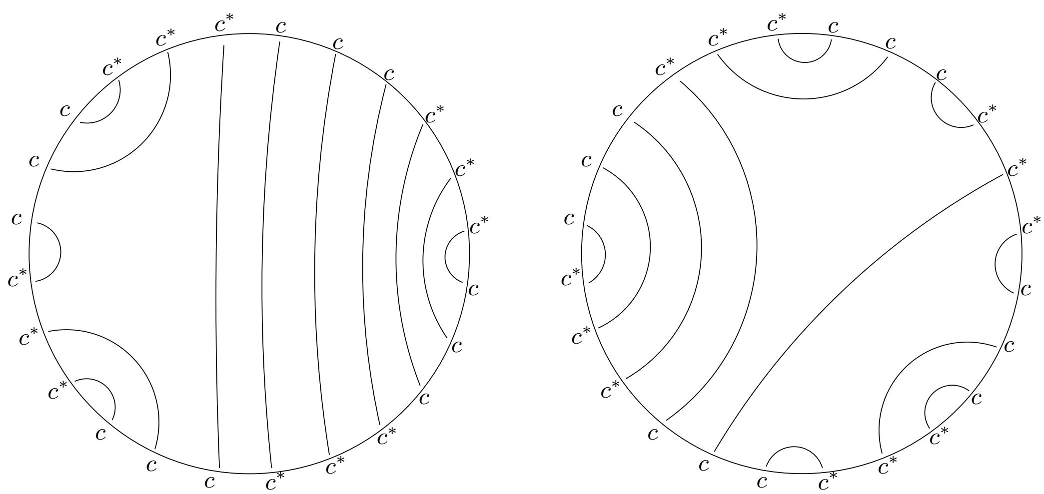


FIGURE 3. The $*$-pairings from Figure 2, in circular form.

A related counting problem asks for pairings of the pattern $c_{n,m}$ where we relax the condition that each c must be paired to a c^*, but still required that *no two elements in a single n-block are paired together*. Denote the set of all such non-crossing pairings as $\mathscr{T}(n,m)$ (so $NC_2^*(n,m) \subset \mathscr{T}(n,m)$). As discussed in [BiS], this problem is the combinatorial counterpart to another moment problem, this time dealing with a semicircular element s. Of course, since s is selfadjoint, $(s^n(s^*)^n)^m = s^{2nm}$, and calculating these moments is routine. Instead, the number of pairings in $\mathscr{T}(n,m)$ equals the moment $\varphi(T_n(s)^{2m})$, where T_n are the *Tchebyshev polynomials*. While we do not have a nice schema for calculating $|\mathscr{T}(n,m)|$ explicitly (which we do for $|NC_2^*(n,m)|$ below), functional calculus for selfadjoint operators immediately yields that $\varphi(T_n(s)^{2m})^{1/2m} \to n+1$ as $m \to \infty$ — the norm $\|T_n(s)\|$ is *linear* in n, rather than in $\sqrt{n}$ as in Theorem 1.3 above. This difference in size precisely reflects the improvement of Haagerup's inequality from $O(n)$ to $O(n^{1/2})$ behaviour for circular elements, and indeed for all $\mathscr{R}$-diagonal elements as discussed in Section 4.

As to the *Knights and Ladies of the Round Table* problem, let us introduce some notation which will be useful throughout what follows.

Notation 3.1. *Label the entries in $c_{n,m}$ with decreasing indices n through 1 in each block of c's and increasing indices 1 through n in each block of c^*'s.*

$$c_{n,m} = \underbrace{\underset{n}{c}, \underset{n-1}{c}, \ldots, \underset{2}{c}, \underset{1}{c}, \underset{1}{c^*}, \underset{2}{c^*}, \ldots, \underset{n-1}{c^*}, \underset{n}{c^*}}_{k=1}, \ldots, \underbrace{\underset{n}{c}, \underset{n-1}{c}, \ldots, \underset{2}{c}, \underset{1}{c}, \underset{1}{c^*}, \underset{2}{c^*}, \ldots, \underset{n-1}{c^*}, \underset{n}{c^*}}_{k=m} \tag{3.4}$$

The point of this labeling is that any pairing in $NC_2^*(n,m)$ must pair each c to a c^* with the same index; indeed, if a $\underset{j}{c}$ were paired to a $\underset{j'}{c}^*$ for $j \neq j'$, there would be unequal numbers of c's and c^*'s between the two; hence, since all must be paired off, one must be paired outside the interval between the two, producing a crossing. We may note further that any non-crossing pairing which respects the labels in Equation 3.4 is, in fact, a $*$-pairing, and so enumerating $NC_2^*(n,m)$ amounts to counting the non-crossing pairings which respect those labels. Using this observation, we proceed to define a bijection from $NC_2^*(n,m)$ to a set we can enumerate.

The above labeling allows us to dissect any $\pi \in NC_2^*(n,m)$ into n different partitions $\Phi_1, \ldots, \Phi_n$. This is accomplished as follows. Label the m intervals $c^n(c^*)^n$ in $c_{n,m}$ by $k = 1, 2, \ldots, m$, as in Notation 3.1 and Figure 4. Within each interval k, we identify (for each j) $\underset{j}{c}$ with $\underset{j}{c}^*$; this identification is represented graphically by connecting them with dotted-arc above the partition diagram. Now, restricting our attention just to those c's and c^*'s with index j, we can follow the pairings made by $\pi \in NC_2(n,m)$ through the identifications; we then say that k and k' are (π, j)-*connected* if there is a path connecting them, as in Figure 4.

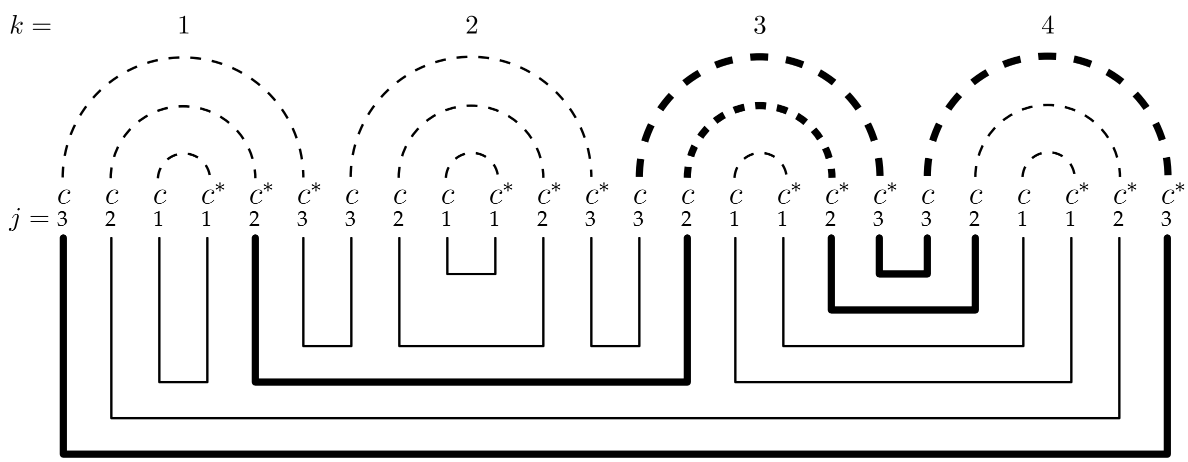


FIGURE 4. In the above $*$-pairing π, $1, 3$ are $(\pi, 3)$-connected, while $1, 4$ are $(\pi, 2)$-connected.

Thus, we can define a partition $\Phi_j(\pi)$ of the set $\{1, 2, \ldots, m\}$ to have blocks given by (π, j)-connectedness. (That is, k, k' are in the same block of $\Phi_j(\pi)$ iff k, k' are (π, j)-connected.) It is easy to see that $\Phi_j(\pi)$ is a non-crossing partition, since π is non-crossing. (In fact, letting $\pi|_j$ be the restriction of π to those elements with index j, $\Phi_j(\pi)$ is just the push-forward of $\pi|_j$ under the map which sends elements of the interval k to the number $k \in \{1, \ldots, m\}$; this is a monotonic function, and so the resulting partition is non-crossing.) Figure 5 shows the $\Phi_j(\pi)$'s resulting from this construction for the $*$-pairings π in Figure 2.

So, we have a mapping $\mathscr{P}\colon NC_2(n,m) \to NC(m)^n$, given by $\mathscr{P}(\pi) = (\Phi_1(\pi), \ldots, \Phi_n(\pi))$. What's more, as is evidenced by the examples in Figure 5, $\Phi_{j+1}(\pi)$ is a refinement of $\Phi_j(\pi)$ for each j. This follows from the nested structure of the π-connecting paths which results from the non-crossing requirement. A precise proof of this refinement property requires a careful but elementary argument tracing the paths, and is left to the reader. Recall that $NC(m)$ is a lattice under reverse-refinement; hence, we have $\Phi_1(\pi) \leq \Phi_2(\pi) \leq \cdots \leq \Phi_n(\pi)$.

Denote by $NC(m)^{(n)}$ the set of all multichains (non-decreasing sequences) of length n in $NC(m)$; thus $\mathscr{P}$ maps $NC_2^*(n,m)$ into $NC(m)^{(n)}$. In fact, this map is a bijection; it is relatively straightforward to exhibit its inverse. The idea (heuristically) is to "fatten up" each connecting line on the right-hand side of Figure 5, and assign pairings by ignoring the top connections (separating the

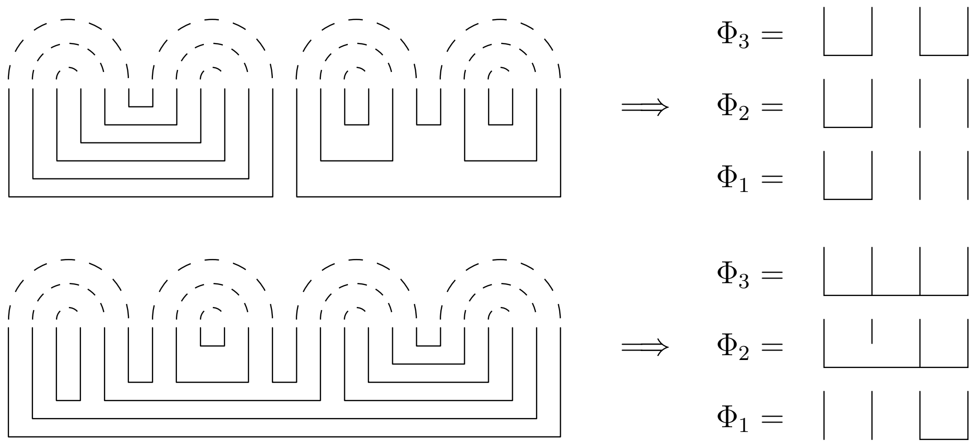


FIGURE 5. The partitions Φ_3, Φ_2, Φ_1 corresponding to the two $*$-pairings in Figure 2.

j-labelled c and c^* in each k-interval). This is done for each j separately, and is demonstrated in Figure 6.

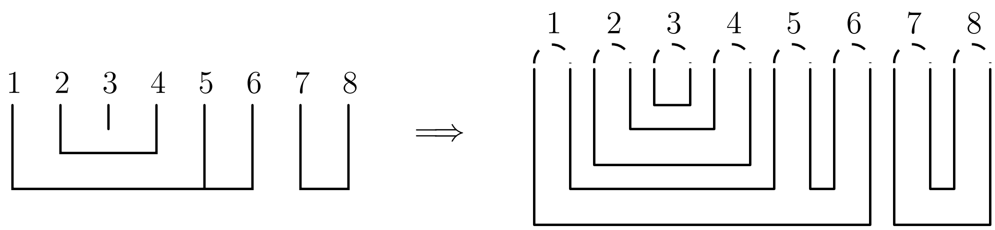


FIGURE 6. A "fattened" partition in $NC(8)$.

To be a little more precise: given a partition Φ in $NC(m)$, for each block $\{k_1 < k_2 < \cdots < k_{r-1} < k_r\}$ in Φ we produce a partial pairing of the pattern $c_{n,m}$ by pairing the $\underset{j}{c}$ in the k_1th interval with the $\underset{j}{c^*}$ in the k_rth interval, then moving back and pairing the $\underset{j}{c}$ in the k_rth interval with the $\underset{j}{c^*}$ in the k_{r-1}th interval, pairing the $\underset{j}{c}$ in the k_{r-1}th interval with the $\underset{j}{c^*}$ in the k_{r-2}th interval, and so forth. If we then have n partitions $\Phi_1, \ldots, \Phi_n$ in $NC(m)$, we follow the above procedure n times (using Φ_1 to pair the 1-labelled c's and c^*'s, Φ_2 to pair the 2-labelled ones, and so on).

This yields a map $\mathscr{Q}\colon NC(m)^{(n)} \to \{\text{pair partitions of } c_{n,m}\}$. A fairly straighforward (but lengthy) argument shows that, since each Φ_j is non-crossing, and (more importantly) since the earlier Φ_j's are refinements of the later ones, the partition $\mathscr{Q}(\Phi_1, \ldots, \Phi_n)$ is actually non-crossing. It respects the labels of the c's and c^*'s by definition, and hence $\mathscr{Q}\colon NC(m)^{(n)} \to NC_2^*(n,m)$.

It is then a simple matter (and is in fact clear from the above diagrams) that $\mathscr{P}$ and $\mathscr{Q}$ are inverses of one another. Hence, $|NC_2^*(n,m)| = |NC(m)^{(n)}|$. At this point, we have reproduced the results of Oravecz and Larsen: the set $NC(m)^{(n)}$ is a well-studied combinatorial structure, and its enumeration was calculated by Edelman in [E]. The next result follows.

Corollary 3.2. *For all positive integers n and m, the number of $*$-pairings $|NC_2^*(n,m)|$ is equal to $|NC^{(n)}(m)| = C_m^{(n)}$, where $C_m^{(n)}$ are the* Fuss-Catalan numbers

$$C_m^{(n)} = \frac{1}{m}\binom{m(n+1)}{m-1}. \tag{3.5}$$

Note, in particular, that setting $n = 1$ yields the Catalan numbers $C_m^{(1)} = C_m = \frac{1}{m}\binom{2m}{m-1}$ from Equation 2.2, which count the set $NC(m)$. The Fuss-Catalan numbers were also computed in a similar context in [BJ], where the central objects of study, the *Fuss-Catalan algebras* (a generalization of the Temperly-Lieb algebras) are generated by diagrams like Figure 3, and hence the dimensions of the algebras (the number of essentially different such diagrams) are the numbers $C_m^{(n)}$.

3.2. **The Haagerup inequality in $\mathcal{H}(c, I)$.** From Equation 3.3 and Corollary 3.2, we have calculated the $2m$-norms of the powers of a circular element,

$$\|c^n\|_{2m} = \left[C_m^{(n)}\right]^{1/2m} = \left[\frac{1}{m}\binom{m(n+1)}{m-1}\right]^{1/2m}. \tag{3.6}$$

In particular, the 2-norm is $\|c^n\|_2 = 1$. We can calculate the norm $\|c^n\|$ by taking the limit as $m \to \infty$, which may be computed using Stirling's formula. The result is

$$\|c^n\|^2 = \lim_{m\to\infty}\left[\frac{1}{m}\binom{m(n+1)}{m-1}\right]^{1/m} = \frac{(n+1)^{n+1}}{n^n} = \left(1+\frac{1}{n}\right)^n (n+1) \le e\,(n+1). \tag{3.7}$$

Now, in line with Theorem 1.3, consider the algebra $\mathcal{H}(c) = \mathcal{H}(c, \{1\})$, the norm-closed algebra generated by c. In this case, the n-particle space $\mathcal{H}^{(n)}(c)$ is spanned by c^n, and hence Equation 3.7 immediately yields the following strong Haagerup inequality.

Proposition 3.3. *For $n \ge 0$ and $T \in \mathcal{H}^{(n)}(c)$,*

$$\|T\| \le \sqrt{e}\,\sqrt{n+1}\,\|T\|_2.$$

In fact, we can use similar techniques to achieve the same inequality for the algebra $\mathcal{H}(c, I)$ for any countable indexing set I. This jump, from 1 to many (even infinite) dimensions is usually the hardest part of such analyses; we will see below that the freeness does all the work for us. Note, the algebra $\mathcal{H}(c, I)$ is canonically isomorphic to the 0-holomorphic space $\mathcal{H}_0(\mathscr{H}_\mathbb{C})$ in [Ke] and the free Segal-Bargmann space $\mathscr{C}_{hol}(\mathscr{H})$ in [Bi1], where $\mathscr{H}_\mathbb{C}$ is a complex Hilbert space of dimension $|I|$.

Let $T \in \mathcal{H}^n(c, I)$, so that $T = \sum_{|\mathbf{i}|=n} \lambda_\mathbf{i} c_\mathbf{i}$ for some scalars $\lambda_\mathbf{i} \in \mathbb{C}$ satisfying a summability condition guaranteeing that $\|T\|_2 < \infty$ (see Equation 3.10 below), where $c_\mathbf{i} = c_{i_1}\cdots c_{i_n}$. By the definition of $\mathcal{H}(c, I)$, the generating elements c_{i_k} are variance 1 and $c_{i_k}, c_{i_{k'}}$ are $*$-free whenever $i_k \ne i_{k'}$. Then we have the following multinomial expansion for the mth moment of TT^*:

$$\begin{aligned}\|T\|_{2m}^{2m} &= \varphi[(TT^*)^m] \\ &= \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)}\cdots\lambda_{\mathbf{i}(m)}\overline{\lambda_{\mathbf{j}(1)}}\cdots\overline{\lambda_{\mathbf{j}(m)}}\,\varphi\left(c_{\mathbf{i}(1)}c^*_{\mathbf{j}(1)}\cdots c_{\mathbf{i}(m)}c^*_{\mathbf{j}(m)}\right).\end{aligned} \tag{3.8}$$

In particular, setting $m = 1$,

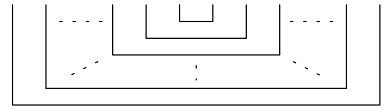
$$\|T\|_2^2 = \sum_{|\mathbf{i}|=|\mathbf{j}|=n} \lambda_\mathbf{i}\overline{\lambda_\mathbf{j}}\,\varphi\left(c_\mathbf{i}c_\mathbf{j}^*\right).$$

The expression $\varphi(c_\mathbf{i}c_\mathbf{j}^*)$ is a mixed moment of length $2n$, and can (by Equation 2.5) be expressed in terms of the cumulants of the $c_\mathbf{i}$:

$$\varphi\left(c_\mathbf{i}c_\mathbf{j}^*\right) = \sum_{\pi\in NC(2n)} \kappa_\pi[c_{i_1}, \ldots, c_{i_n}, c^*_{j_n}, \ldots, c^*_{j_1}].$$

As the c_i are circular (and so only the cumulants $\kappa_2[c, c^*] = \kappa_2[c^*, c] = 1$ are nonzero), only pair partitions π which match c's to c^*'s contribute to the sum. Any such partition is in $NC_2^*(n, 1)$, which contains only the partition ϖ

ϖ =



(the fact that there is only one follows from the calculation in Section 3.1 that $|NC_2^*(n,1)| = C_1^{(n)} = 1$). So, we have

$$\|T\|_2^2 = \sum_{|\mathbf{i}|=|\mathbf{j}|=n} \lambda_{\mathbf{i}} \overline{\lambda_{\mathbf{j}}}\, \kappa_\varpi[c_{\mathbf{i}}, c_{\mathbf{j}}^*]. \tag{3.9}$$

A note on notation: in Equation 3.9, the $c_{\mathbf{i}}$ and $c_{\mathbf{j}}^*$ stand for *lists* of length n, not products of n elements; i.e. there are implied commas. We will use this convention whenever such expressions appear as arguments of cumulants in what follows. To be clear, for the pairing ϖ above, we have

$$\kappa_\varpi[c_{\mathbf{i}}, c_{\mathbf{j}}^*] = \kappa_\varpi[c_{i_1}, \ldots, c_{i_n}, c_{j_n}^*, \ldots, c_{j_1}^*] = \kappa_2[c_{i_1}, c_{j_1}^*] \cdot \kappa_2[c_{i_2}, c_{j_2}^*] \cdots \kappa_2[c_{i_n}, c_{j_n}^*].$$

Now following Equation 3.9, since the c_{i_ℓ} are $*$-free, $\kappa_\varpi[c_{\mathbf{i}}, c_{\mathbf{j}}^*] = 0$ unless each block of ϖ contains like-indexed elements – i.e. unless $\mathbf{i} = \mathbf{j}$, in which case $\kappa_\varpi = 1$. Thus, we have the Pythagoreon formula

$$\|T\|_2^2 = \sum_{|\mathbf{i}|=n} |\lambda_{\mathbf{i}}|^2. \tag{3.10}$$

Following suit, for general $m > 1$ we have

$$\varphi\left(c_{\mathbf{i}(1)} c_{\mathbf{j}(1)}^* \cdots c_{\mathbf{i}(m)} c_{\mathbf{j}(m)}^*\right) = \sum_{\pi \in NC(2nm)} \kappa_\pi[c_{\mathbf{i}(1)}, c_{\mathbf{j}(1)}^*, \ldots, c_{\mathbf{i}(m)}, c_{\mathbf{j}(m)}^*].$$

Once again, since the $c_{i(k)_\ell}$ are circular elements, the only partitions π which contribute to the sum are those which pair c's with c^*'s – i.e. $\pi \in NC_2^*(n,m)$. This, with Equation 3.8, yields

$$\|T\|_{2m}^{2m} = \sum_{\pi \in NC_2^*(n,m)} \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}}\, \kappa_\pi[c_{\mathbf{i}(1)}, c_{\mathbf{j}(1)}^*, \ldots, c_{\mathbf{i}(m)}, c_{\mathbf{j}(m)}^*].$$

Many of the above terms are in fact 0, since the $c_{i(k)_\ell}$ are $*$-free. Indeed, the mixed cumulant κ_π in the above sum is nonzero only when the indices of terms paired by π are all equal (and in this case it is 1). We record this with the function $\delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m))$ defined to equal 0 whenever π pairs any $c_{i(k)_\ell}$ with a $c_{j(k')_{\ell'}}^*$ with $i(k)_\ell \neq j(k')_{\ell'}$, and 1 if π always pairs like-indexed c's and c^*'s. Thus

$$\|T\|_{2m}^{2m} = \sum_{\pi \in NC^*(n,m)} \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}}\, \delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)).$$

Now, let us re-index the above sum. Denote the indices $\{i(1)_1, \ldots, i(m)_n\}$ by $p_1, \ldots, p_{nm}$, and let $\lambda(p_1, \ldots, p_{nm}) = \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)}$. Note, in any nonzero term in the above sum, the indices appearing in the product $\overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}}$ are exactly those paired to $p_1, \ldots, p_{nm}$ by π; identifying the pairing π with its corresponding permutation, we then have

$$\|T\|_{2m}^{2m} = \sum_{\pi \in NC_2^*(n,m)} \sum_{p_1, \ldots, p_{nm}} \lambda(p_1, \ldots, p_{nm}) \overline{\lambda(p_{\pi(1)}, \ldots, p_{\pi(nm)})}. \tag{3.11}$$

Applying the Cauchy-Schwarz inequality to the interior summation yields, for each π,

$$\sum_{p_1,\ldots,p_{nm}} \lambda(p_1,\ldots,p_{nm})\overline{\lambda(p_{\pi(1)},\ldots,p_{\pi(nm)})}$$
$$\leq \left[\sum_{p_1,\ldots,p_{nm}} |\lambda(p_1,\ldots,p_{nm})|^2\right]^{1/2} \cdot \left[\sum_{p_1,\ldots,p_{nm}} |\lambda(p_{\pi(1)},\ldots,p_{\pi(nm)})|^2\right]^{1/2}.$$

Since the sum is over all nm-tuples of indices and π is a permutation, the second term may be reordered to cancel the apparent π-dependence, yielding the same summation in both factors; i.e. the interior sum in Equation 3.11 is just

$$\sum_{p_1,\ldots,p_{nm}} |\lambda(p_1,\ldots,p_{nm})|^2.$$

Returning to our original indexing scheme, this becomes

$$\sum_{p_1,\ldots,p_{nm}} |\lambda(p_1,\ldots,p_{nm})|^2 = \sum_{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n} |\lambda_{\mathbf{i}(1)}\cdots\lambda_{\mathbf{i}(m)}|^2 = \left[\sum_{|\mathbf{i}|=n} |\lambda_{\mathbf{i}}|^2\right]^m,$$

and this last expression is $\|T\|_2^{2m}$ from Equation 3.10. Thus, Equation 3.11 and Corollary 3.2 together yield

$$\|T\|_{2m}^{2m} \leq \sum_{\pi\in NC_2^*(n,m)} \|T\|_2^{2m} = C_m^{(n)}\|T\|_2^{2m}.$$

Taking mth roots and letting $m\to\infty$, referring to the same limit calculated in Equation 3.7, we have thus proved the main theorem of this section:

Theorem 3.4. *Let c be a variance 1 circular, and let $T\in\mathcal{H}^{(n)}(c,I)$ for some countable index set I. Then*

$$\|T\| \leq \sqrt{e}\,\sqrt{n+1}\,\|T\|_2.$$

We note that this inequality (with the $\sqrt{n+1}$ factor) bears some resemblance to what Bożejko called *Nelson's inequality* in [Bo1]. The context of his inequality is different, however (his estimate is for the creation and annihilation operators on the full Fock space separately), and our result cannot be derived from his. We also note that the exact inequality yielded by the asymptotics of the Fuss-Catalan number is $\|T\| \leq \sqrt{(n+1)^{n+1}/n^n}\,\|T\|_2$, which is (quickly) asymptotic to to the form written above.

3.3. **The Haagerup inequality in $\mathcal{H}(u,I)$.** In Section 4, we will shortly demonstrate how to extend the above argument to general $\mathscr{R}$-diagonal elements; indeed, we will see that the partition sets $NC_2^*(n,m)$ still mediate the Haagerup constant in this case. The case of $\mathcal{H}(u,I)$ for a Haar unitary u will follow as a special case. However, the techniques of Section 4 do not produce the optimal constant ($\sqrt{e}$) multiplying the $O(n^{1/2})$-term in the inequality.

For $\mathcal{H}(u,I)$ (which is close to Haagerup's original context), we can actually approach the Haagerup constant through moments directly, without going over to cumulants as above. For $T\in\mathcal{H}^{(n)}(u,I)$, we have $T=\sum_{|\mathbf{i}|=n}\lambda_{\mathbf{i}}u_{\mathbf{i}}$ for some $\lambda_{\mathbf{i}}\in\mathbb{C}$. Thus

$$\begin{aligned}\|T\|_{2m}^{2m} &= \varphi[(TT^*)^m]\\ &= \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n\\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)}\cdots\lambda_{\mathbf{i}(m)}\overline{\lambda_{\mathbf{j}(1)}}\cdots\overline{\lambda_{\mathbf{j}(m)}}\,\varphi\left(u_{\mathbf{i}(1)}u_{\mathbf{j}(1)}^*\cdots u_{\mathbf{i}(m)}u_{\mathbf{j}(m)}^*\right).\end{aligned}$$

Now, since the u_k are $*$-free Haar unitaries, the term $\varphi(u_{\mathbf{i}(1)}u_{\mathbf{j}(1)}^*\cdots u_{\mathbf{i}(m)}u_{\mathbf{j}(m)}^*)$ is either 0 or 1; interpretting u_k as generators of a free group and $u_k^*=u_k^{-1}$, the term is 1 iff the word $u_{\mathbf{i}(1)}u_{\mathbf{j}(1)}^{-1}\cdots u_{\mathbf{i}(m)}u_{\mathbf{j}(m)}^{-1}$

reduces to the identity in the free group. (This follows immediately from the definition of the state φ_k on the free group factor $L(\mathbb{F}_k)$ in Section 2.2.) Let us record this with a delta function $\delta(\mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m))$, and so we have

$$\|T\|_{2m}^{2m} = \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}}\, \delta(\mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)). \tag{3.12}$$

Now, if $\delta(\mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)) = 1$ then the word $u_{\mathbf{i}(1)} u_{\mathbf{j}(1)}^{-1} \cdots u_{\mathbf{i}(m)} u_{\mathbf{j}(m)}^{-1}$ reduces to the identity in the free group. This means there is a reduction procedure: at some point in the word, there is a u_k adjacent to a u_k^{-1}, and once this pair has been removed, the reduced word has some $u_{k'}$ adjacent to a $u_{k'}^{-1}$, etc. We can produce a pair partition π of the set $\{1, \ldots, 2mn\}$ while reducing the word, by matching numbers based on the (starting) positions of terms in the word $u_{\mathbf{i}(1)} u_{\mathbf{j}(1)}^{-1} \cdots u_{\mathbf{i}(m)} u_{\mathbf{j}(m)}^{-1}$ as they are cancelled in the reduction to the identity. This is demonstrated in Figure 7.

$$u_1\ u_5\ u_2\ u_2^{-1} u_3^{-1} u_3^{-1} u_3\ u_2\ u_1\ u_1^{-1} u_2^{-1} u_1^{-1} u_1\ u_3\ u_3\ u_4^{-1} u_5^{-1} u_4^{-1} u_4\ u_5\ u_4\ u_3^{-1} u_5^{-1} u_1^{-1}$$

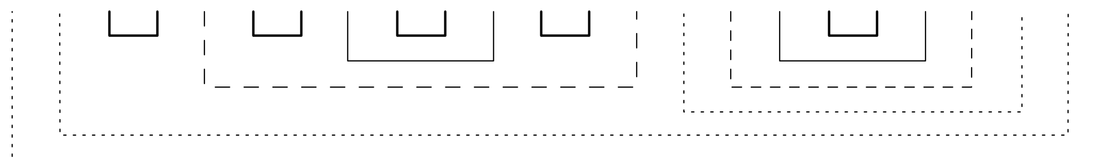


FIGURE 7. A reducible word in $\mathbb{F}_5$ resulting from an expansion in $\mathcal{H}(u, I)$, and a corresponding reduction partition in $NC_2^*(3, 4)$. (Heavier lines indicate earlier stages in the reduction procedure.)

As is clear from the figure, the pairing π resulting from a reduction P is non-crossing. (Indeed, the pairing is generated by matching two adjacent elements, removing this interval, and proceeding inductively; thus, the partition conforms to the recursive definition of non-crossing.) Also, since u's cancel only u^{-1}'s, such a pairing π is in $NC_2^*(n, m)$. Finally, since u_k must cancel only with u_k^{-1} (i.e. with the same index), π must match only terms with the same index.

We must be careful, since a given reducible word may have different reduction procedures, and therefore different associated pairings. Nevertheless, the above considerations show that we can simply overcount and conclude that

$$\delta(\mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)) \leq \sum_{\pi \in NC_2^*(n,m)} \delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)), \tag{3.13}$$

where $\delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m))$ has the same meaning as in Section 3.2: it is $\{0, 1\}$-valued, equalling 1 iff the pairing π only matches terms with the same index. Combining this with Equation 3.12, we have

$$\|T\|_{2m}^{2m} \leq \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}} \sum_{\pi \in NC_2^*(n,m)} \delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)).$$

We now proceed exactly as in Section 3.2: reindex $\mathbf{i}(1), \ldots, \mathbf{i}(m)$ as $p_1, \ldots, p_{mn}$ and denote by $\lambda(p_1, \ldots, p_{nm})$ the product of the relabeled terms $\lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)}$. Then, treating π as a permutation of the set $\{1, \ldots, 2mn\}$, the above sum becomes

$$\sum_{\pi \in NC_2^*(n,m)} \sum_{p_1, \ldots, p_{mn}} \lambda(p_1, \ldots, p_{mn}) \overline{\lambda(p_{\pi(1)}, \ldots, p_{\pi(mn)})}.$$

For each term in the internal sum, we apply the Cauchy-Schwarz inequality, and use the fact that π is a bijection to reindex as in the last section, achieving

$$\|T\|_{2m}^{2m} \leq \sum_{\pi \in NC_2^*(n,m)} \sum_{\mathbf{i}(1),\ldots,\mathbf{i}(m)} |\lambda_{\mathbf{i}(1)}|^2 \cdots |\lambda_{\mathbf{i}(m)}|^2 = |NC_2^*(n,m)| \left(\sum_{\mathbf{i}} |\lambda_{\mathbf{i}}|^2 \right)^m . \tag{3.14}$$

We may now complete the proof of Theorem 1.4.

Proof of Theorem 1.4. The generators of a free group $\mathbb{F}_k$ are $*$-free Haar unitary operators in $L(\mathbb{F}_k)$, and so any $f \in \mathbb{C}\mathbb{F}_k \subset L(\mathbb{F}_k)$ supported on words of length n in the generators and not their inverses is in $\mathcal{H}^{(n)}(u, I_k)$ where $I_k = \{1, \ldots, k\}$. Since the norm in $L(\mathbb{F}_k)$ restricts to the convolution norm on $\mathbb{C}\mathbb{F}_k$, it follows that $\|f\|_{2m} \to \|f\|_*$ as $m \to \infty$. Also, if $f(u_{\mathbf{i}}) = \lambda_{\mathbf{i}}$ (so f corresponds to T above), then $\|f\|_2^2 = \sum_{\mathbf{i}} |\lambda_{\mathbf{i}}|^2$. Hence, Equation 3.14 says

$$\|f\|_{2m} \leq |NC_2^*(n,m)|^{1/2m} \|f\|_2,$$

and so taking $m \to \infty$ and using Corollary 3.2 and the asymptotics of the Fuss-Catalan numbers from Equation 3.7, we have $\|f\|_* \leq \sqrt{(n+1)^{n+1}/n^n}\,\|f\|_2$. As $(n+1)^{n+1}/n^n \leq e(n+1)$, this proves the result for f with finite support. Since $\mathbb{C}\mathbb{F}_k$ is dense in $\ell^2(\mathbb{F}_k)$, this completes the proof. □

A few comments are in order. First, since any reducible word may have many different reductions to the identity, Equation 3.13 may, a priori, seem to be a massive over-estimate, and one wonders whether the resulting Haagerup inequality is close to optimal. In fact, the inequality

$$\|f\|_* \leq \sqrt{\frac{(n+1)^{n+1}}{n^n}}\|f\|_2$$

(for $f \in \ell^2(\mathbb{F}_k^+)$ supported on words of length n) is optimal, at least in the case $n = 1$: consider $\mathcal{H}(u, \mathbb{N})$. For $k > 1$ in $\mathbb{N}$, the element $T_k = u_1 + \cdots + u_k$ is in the 1-particle space, and satisfies $\|T_k\|_2 = \sqrt{k}$ and $\|T_k\| = 2\sqrt{k-1}$ (the latter was calculated in [AO]). Thus

$$\frac{\|T_k\|}{\|T_k\|_2} = 2 \cdot \sqrt{\frac{k-1}{k}}.$$

Since $\sqrt{(1+1)^{1+1}/1^1} = 2$, we have asymptotic agreement. Note that the case $n = 1$ presents, in some sense, the greatest ambiguity for reduction procedures.

Second, we note that a similar argument to the one above, applied to elements $f \in \ell^2(\mathbb{F}_k)$ supported on words of length n in the generators *and their inverses*, yields the original Haagerup inequality of Theorem 1.1; here, the relevant combinatorial structure whose enumeration determines the Haagerup constant is the one described immediately following Figure 3.

4. $\mathscr{R}$-diagonal Elements

In this section, we extend the techniques developed in Section 3 to all $\mathscr{R}$-diagonal elements. A similar reduction of the multidimensional case to the one-dimensional case is possible, but there is an obstruction: the main argument goes through only when the mixed cumulants are non-negative. We address this problem by replacing an $\mathscr{R}$-diagonal element with negative cumulants by a different $\mathscr{R}$-diagonal whose cumulants are positive and dominate the original's.

In Section 4.1, we calculate the 2-norm of an element T in the n-particle space, and develop the main estimate (which generalizes the proof of Theorem 3.4) of higher moments of TT^* in terms of the absolute values of the cumulants. Then, in Section 4.2, we show how to replace a given $\mathscr{R}$-diagonal element with a different one whose cumulants dominate the absolute values of the original's, and use this substitution to prove Theorem 1.3.

4.1. **Estimating moments for $T \in \mathcal{H}^{(n)}(a, I)$.** Let a be an $\mathscr{R}$-diagonal element in a C^*-probability space, and let $T \in \mathcal{H}^{(n)}(a, I)$. So, $T = \sum_{|\mathbf{i}|=n} \lambda_{\mathbf{i}} a_{\mathbf{i}}$ for some scalars $\lambda_{\mathbf{i}} \in \mathbb{C}$, where $\{a_i : i \in I\}$ are $*$-free $\mathscr{R}$-diagonal elements each with the same $*$-distribution as a. As in Equation 3.8 above, we have the following multinomial expansion for the mth moment of TT^*:

$$\begin{aligned}\|T\|_{2m}^{2m} &= \varphi[(TT^*)^m] \\ &= \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}} \, \varphi\left(a_{\mathbf{i}(1)} a^*_{\mathbf{j}(1)} \cdots a_{\mathbf{i}(m)} a^*_{\mathbf{j}(m)}\right). \qquad (4.1)\end{aligned}$$

The term $\varphi(a_{\mathbf{i}(1)} a^*_{\mathbf{j}(1)} \cdots a_{\mathbf{i}(m)} a^*_{\mathbf{j}(m)})$ can be calculated, via Equation 2.5, as

$$\varphi\left(a_{\mathbf{i}(1)} a^*_{\mathbf{j}(1)} \cdots a_{\mathbf{i}(m)} a^*_{\mathbf{j}(m)}\right) = \sum_{\pi \in NC(2mn)} \kappa_\pi[a_{\mathbf{i}(1)}, a^*_{\mathbf{j}(1)}, \ldots, a_{\mathbf{i}(m)}, a^*_{\mathbf{j}(m)}].$$

Since the a_i are $*$-free, the above mixed cumulant is nonzero only when the indices of terms connected by π are all equal. We record this with the function $\delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m))$ defined above, which equals 0 whenever π connects two differently-indexed elements, and 1 if all connected elements have like-indices. It is, then, true that

$$\varphi\left(a_{\mathbf{i}(1)} a^*_{\mathbf{j}(1)} \cdots a_{\mathbf{i}(m)} a^*_{\mathbf{j}(m)}\right) = \sum_{\pi \in NC(2mn)} \kappa_\pi[a_{\mathbf{i}(1)}, a^*_{\mathbf{j}(1)}, \ldots, a_{\mathbf{i}(m)}, a^*_{\mathbf{j}(m)}] \delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)).$$

In the special case $m = 1$, this reduces to

$$\varphi\left(a_{\mathbf{i}} a^*_{\mathbf{j}}\right) = \sum_{\pi \in NC(2n)} \kappa_\pi[a_{\mathbf{i}}, a^*_{\mathbf{j}}] \delta(\pi, \mathbf{i}, \mathbf{j}). \qquad (4.2)$$

Now, let π be a partition with $\delta(\pi, \mathbf{i}, \mathbf{j}) = 1$. Thus, each block of π connects only terms with a single index i. Since a_i is $\mathscr{R}$-diagonal, its only nonzero $*$-cumulants are $\kappa_{2n}[a_i, a_i^*, \ldots, a_i, a_i^*]$ and $\kappa_{2n}[a_i^*, a_i, \ldots, a_i^*, a_i]$. Hence, π still contributes a zero in Equation 4.2 unless, in each block of π, the a_i's and a_i^*'s alternate. But in this case ($m = 1$), all the a_i^*'s are to the right of all the a_i's, and hence alternating sequences have length at most 2. So π contributes only if it is a pair partition. Since the cumulants $\kappa_2[a_j, a_j] = \kappa_2[a_j^*, a_j^*] = 0$ for each j, such a π only pairs $*$'s to non-$*$'s, and so π is actually a $*$-pairing: $\pi \in NC_2^*(n, 1)$. As shown in Section 3.2, the only element of $NC_2^*(n, 1)$ is ϖ. So the sum in Equation 4.2 reduces to at most a single term,

$$\varphi\left(a_{\mathbf{i}} a^*_{\mathbf{j}}\right) = \kappa_\varpi[a_{\mathbf{i}}, a^*_{\mathbf{j}}] \delta(\varpi, \mathbf{i}, \mathbf{j}).$$

Since $a_{\mathbf{i}} = a_{i_1} \cdots a_{i_n}$ and $a^*_{\mathbf{j}} = a^*_{j_n} \cdots a^*_{j_1}$, $\delta(\varpi, \mathbf{i}, \mathbf{j}) = 1$ iff $\mathbf{i} = \mathbf{j}$, and in this case, $\kappa_\varpi[a_{\mathbf{i}}, a^*_{\mathbf{i}}]$ is equal to the product $\kappa_2[a_{i_1}, a^*_{i_1}] \cdots \kappa_2[a_{i_n}, a^*_{i_n}]$ which (since the a_i are identically distributed) equals $\kappa_2[a, a^*]^n$. So Equation 4.1 yields

$$\|T\|_2^2 = \sum_{|\mathbf{i}|=|\mathbf{j}|=n} \lambda_{\mathbf{i}} \overline{\lambda_{\mathbf{j}}} \varphi\left(a_{\mathbf{i}} a^*_{\mathbf{j}}\right) = \sum_{|\mathbf{i}|=n} |\lambda_{\mathbf{i}}|^2 \kappa_2[a, a^*]^n.$$

Finally, we note that the second cumulant of a centred random variable is equal to its second moment, and since $\mathscr{R}$-diagonal elements have vanishing first moment, it follows that

$$\|T\|_2^2 = \sum_{|\mathbf{i}|=n} |\lambda_{\mathbf{i}}|^2 \|a\|_2^{2n}. \qquad (4.3)$$

Similar considerations are not enough to explicitly calculate higher moments, since alternating sequences can have greater length (e.g. in $\|T\|_4^4$, terms corresponding to partitions with blocks of sizes 2 and 4 may contribute), and calculations become unwieldy very quickly. Nevertheless, we can estimate the higher moments using only pair partitions, to great effect. In general, from Equation 4.1 we have

$$\|T\|_{2m}^{2m} = \sum_{\pi \in NC(2mn)} \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}} \cdot \vartheta[\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)],$$

where

$$\vartheta[\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)] = \kappa_\pi[a_{\mathbf{i}(1)}, a^*_{\mathbf{j}(1)}, \ldots, a_{\mathbf{i}(m)}, a^*_{\mathbf{j}(m)}] \delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)).$$

Now, in any term where $\delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)) = 1$, each block of π connects only a_i's and a_i^*'s for a single index i. Since a_i is $\mathscr{R}$-diagonal, its only nonvanishing $*$-cumulants are alternating, and so the term is zero unless a's and a^*'s alternate within each block of π. This is an important set of non-crossing partitions; we call it $NC^*(n,m)$ (so $NC_2^*(n,m)$ is the subset of $NC^*(n,m)$ consisting of only pair partitions). It is important to note that, as per our definition of alternating, the size of each block of a partition in $NC^*(n,m)$ must be *even*. (The sequence $a, a^*, \ldots, a, a^*, a$ is not alternating in our sense, since an $\mathscr{R}$-diagonal element still has vanishing cumulants for this list.)

Using this notation, the above summation becomes

$$\|T\|_{2m}^{2m} = \sum_{\pi \in NC^*(n,m)} \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}} \cdot \vartheta[\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)].$$

Fix $\mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)$, and let $\pi \in NC^*(n,m)$ be such that $\delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)) = 1$. Let $\{V_1, \ldots, V_k\}$ be the blocks of π. Since all indices of elements in a single block V_j are equal (to, say, i), and since a_i has the same distribution as a, we have that $\varphi_{V_j}[a_{\mathbf{i}(1)}, a^*_{\mathbf{j}(1)}, \ldots, a_{\mathbf{i}(m)}, a^*_{\mathbf{j}(m)}] = \varphi_{V_j}[a_{n,m}]$, where

$$a_{n,m} = \overbrace{\underbrace{a,\ \ldots, a,}_{n}\ \underbrace{a^*, \ldots, a^*}_{n}, \ldots, \underbrace{a,\ \ldots, a,}_{n}\ \underbrace{a^*, \ldots, a^*}_{n}}^{2m \text{ groups}}$$

is independent of the indices. Consequently, we have (for π with $\delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)) = 1$)

$$\kappa_\pi[a_{\mathbf{i}(1)}, a^*_{\mathbf{j}(1)}, \ldots, a_{\mathbf{i}(m)}, a^*_{\mathbf{j}(m)}] = \kappa_\pi[a_{n,m}]. \tag{4.4}$$

Thus, for $\pi \in NC^*(n,m)$, we have

$$\vartheta[\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)] = \kappa_\pi[a_{n,m}] \delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)),$$

and so

$$\|T\|_{2m}^{2m} = \sum_{\pi \in NC^*(n,m)} \kappa_\pi[a_{n,m}] \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} \lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \overline{\lambda_{\mathbf{j}(1)}} \cdots \overline{\lambda_{\mathbf{j}(m)}} \delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)).$$

We now estimate this sum by associating to each $\pi \in NC^*(n,m)$ a refinement $\pi_r \in NC_2^*(n,m)$ as follows: for each block $V = \{k_1 < k_2 < \cdots < k_{2\ell}\}$ in π, the pairings $k_1 \sim k_2$, $k_3 \sim k_4$, $\ldots$, $k_{2\ell-1} \sim k_{2\ell}$ are in π_r.

Since π_r is a refinement of π, if π only connects like-indexed elements then π_r does as well, and so $\delta(\pi, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)) \leq \delta(\pi_r, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m))$. Hence, we may estimate (by taking absolute values)

$$\|T\|_{2m}^{2m} \leq \sum_{\pi \in NC^*(n,m)} |\kappa_\pi[a_{n,m}]| \sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} |\lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \cdot \lambda_{\mathbf{j}(1)} \cdots \lambda_{\mathbf{j}(m)}| \delta(\pi_r, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)).$$

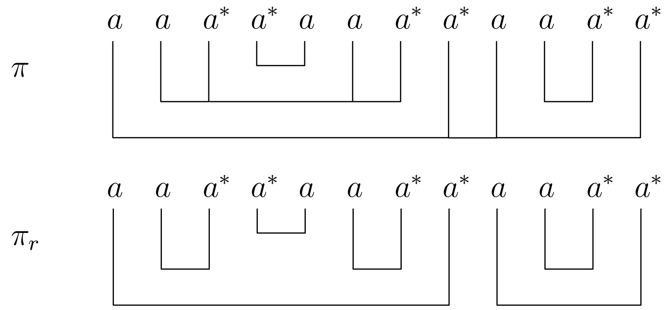


FIGURE 8. A partition $\pi \in NC^*(2,3)$, and the corresponding $\pi_r \in NC_2^*(2,3)$.

We can now reindex the interior sum the same way we did in Section 3.2: denote the indices $\{i(1)_1, \ldots, i(m)_n\}$ by $p_1, \ldots, p_{nm}$, and this time let $\lambda(p_1, \ldots, p_{nm}) = |\lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)}|$. Then allowing π_r to refer both to the pair-partition and the associated permutation, we have

$$\begin{aligned}
&\sum_{\substack{|\mathbf{i}(1)|=\cdots=|\mathbf{i}(m)|=n \\ |\mathbf{j}(1)|=\cdots=|\mathbf{j}(m)|=n}} |\lambda_{\mathbf{i}(1)} \cdots \lambda_{\mathbf{i}(m)} \cdot \lambda_{\mathbf{j}(1)} \cdots \lambda_{\mathbf{j}(m)}| \delta(\pi_r, \mathbf{i}(1), \mathbf{j}(1), \ldots, \mathbf{i}(m), \mathbf{j}(m)) \\
&= \sum_{p_1, \ldots, p_{nm}} \lambda(p_1, \ldots, p_{nm}) \lambda(p_{\pi_r(1)}, \ldots, p_{\pi_r(nm)}) \\
&\leq \left[\sum_{p_1, \ldots, p_{nm}} \lambda(p_1, \ldots, p_{nm})^2 \right]^{1/2} \cdot \left[\sum_{p_1, \ldots, p_{nm}} \lambda(p_{\pi_r(1)}, \ldots, p_{\pi_r(nm)})^2 \right]^{1/2},
\end{aligned}$$

where we have applied the Cauchy-Schwarz inequality. Since the sum is over all indices $p_1, \ldots, p_{nm}$ and since π_r is a permutation, the second term above can be reindexed to yield the first term, and hence the interior sum is

$$\leq \sum_{p_1, \ldots, p_{nm}} \lambda(p_1, \ldots, p_{nm})^2 = \left[\sum_{|\mathbf{i}|=n} |\lambda_{\mathbf{i}}|^2 \right]^m.$$

Combining this with Equation 4.3 yields the following estimate, which is the main lemma of this section.

Lemma 4.1. *Let $T \in \mathcal{H}^{(n)}(a, I)$ for a $\mathscr{R}$-diagonal. Then for $m \geq 1$,*

$$\|T\|_{2m} \leq \left[\sum_{\pi \in NC^*(n,m)} |\kappa_\pi[a_{n,m}]| \right]^{1/2m} \frac{1}{\|a\|_2^n} \|T\|_2.$$

If the cumulants of a are all non-negative, then $\kappa_\pi[a_{n,m}] \geq 0$ as well, and the above summation reduces to a one-dimensional calculation.

Corollary 4.2. *If the cumulants of a are non-negative, then $\|T\| \leq \dfrac{\|a^n\|}{\|a\|_2^n} \|T\|_2$.*

Proof. By Equation 2.5,

$$\|a^n\|_{2m}^{2m} = \varphi([a^n(a^*)^n]^m) = \sum_{\pi \in NC(2nm)} \kappa_\pi[a_{n,m}].$$

As explained above, since a is $\mathscr{R}$-diagonal, $\kappa_\pi[a_{n,m}] = 0$ unless $\pi \in NC^*(n,m)$. Thus, from Lemma 4.1, we have

$$\|T\|_{2m}^{2m} \le \left[\sum_{\pi \in NC^*(n,m)} |\kappa_\pi[a_{n,m}]| \right] \frac{1}{\|a\|_2^{2nm}} \|T\|_2^{2m}$$

$$= \left[\sum_{\pi \in NC(2nm)} \kappa_\pi[a_{n,m}] \right] \frac{1}{\|a\|_2^{2nm}} \|T\|_2^{2m} = \frac{\|a^n\|_{2m}^{2m}}{\|a\|_2^{2nm}} \|T\|_2^{2m}.$$

The result now follows by taking $2m$th roots, and letting m tend to ∞. □

Hence, in this case, the question of Haagerup's inequality is reduced to determining the growth-rate of $\|a^n\|/\|a\|_2^n$, which was addressed in [Lar] (and will be discussed in the next section). However, if some cumulants of a are negative, we must work harder to make such an estimate.

4.2. **Strong Haagerup inequalities.** To reduce the calculation in Section 4.1 to the one-dimensional case when a can have negative cumulants, our strategy is to replace a with a different $\mathscr{R}$-diagonal element b whose cumulants are positive and dominate the absolute values of a's cumulants. We will do this in a way that allows close control of both $\|b\|$ and $\|b\|_2$.

To begin, we bound the growth of the nonvanishing cumulants of a.

Lemma 4.3. *Let a be an $\mathscr{R}$-diagonal element in a C^*-probability space. Then the nonvanishing cumulants of a satisfy*

$$|\alpha_n[a]|, |\beta_n[a]| \le \frac{1}{4}(2^4\|a\|)^{2n},$$

where $\alpha_n[a]$ and $\beta_n[a]$ are the determining sequences of a from Equation 2.6.

Proof. From Equation 2.4, we have

$$\alpha_n[a] = \kappa_{2n}[a, a^*, \dots, a, a^*] = \sum_{\sigma \in NC(2n)} \varphi_\sigma[a, a^*, \dots, a, a^*]\mu(\sigma, 1_{2n}).$$

(The sum is over all of $NC(2n)$ since all σ are less than 1_{2n}, the largest element.) Therefore, from Equation 2.3 we have

$$|\alpha_n[a]| \le \sum_{\sigma \in NC(2n)} |\varphi_\sigma[a, a^*, \dots, a, a^*]| 4^{2n-1} = 4^{2n-1} \sum_{\sigma \in NC(2n)} \prod_{V \in \sigma} |\varphi_V[a, a^*, \dots, a, a^*]|.$$

Let $V_1, \dots, V_r$ be the blocks of a given $\sigma \in NC(2n)$; so $|V_1| + \cdots + |V_r| = 2n$. Well, $\varphi_{V_j}[a, a^*, \dots, a, a^*] = \varphi(a^{\epsilon_1} \cdots a^{\epsilon_{|V_j|}})$ where $\epsilon_i \in \{1, *\}$. Since φ is a state on a C^*-algebra, this gives

$$|\varphi_{V_j}[a, a^*, \dots, a, a^*]| \le \|a^{\epsilon_1} \cdots a^{\epsilon_{|V_j|}}\| \le \|a\|^{|V_j|}.$$

Hence, $|\varphi_\sigma[a, a^*, \dots, a, a^*]| = \prod_{j=1}^r |\varphi_{V_j}[a, a^*, \dots, a, a^*]| \le \prod_{j=1}^r \|a\|^{|V_j|} = \|a\|^{2n}$, and so

$$|\alpha_n[a]| \le 4^{2n-1} \sum_{\sigma \in NC(2n)} \|a\|^{2n} = 4^{2n-1} C_{2n} \|a\|^{2n}.$$

The result for $\alpha_n[a]$ now follows from the fact that $C_{2n} \le 4^{2n}$. The argument for $\beta_n[a]$ is identical. □

Thus, we need only construct an $\mathscr{R}$-diagonal element whose determining sequences are positive and bounded below by $\frac{1}{4}(2^4\|a\|)^{2n}$.

Lemma 4.4. *Let $(\mathscr{A}, \varphi)$ be a C^*-probability space, and let γ and λ be positive constants. There exists an $\mathscr{R}$-diagonal element $b = b_{\gamma,\lambda} \in \mathscr{A}$ with $\alpha_n[b] = \beta_n[b] = \gamma \cdot \lambda^{2n}$.*

Proof. As shown in [NS3] (and also in [S]), there is a *free Poisson* element $p = p_{\frac{\gamma}{2},\lambda}$ which is self-adjoint and satisfies $\kappa_n[p,\ldots,p] = \frac{1}{2}\gamma\cdot\lambda^n$. Let p_1, p_2 be free copies of this Poisson element, and let $q = p_1 - p_2$. As κ_n is a multilinear functional, and as p_1 and $-p_2$ are free (so their mixed cumulants vanish), we have

$$\kappa_n[q,\ldots,q] = \kappa_n[p_1,\ldots,p_1] + \kappa_n[-p_2,\ldots,-p_2] = (1+(-1)^n)\kappa_n[p,\ldots,p] = \begin{cases}\gamma\cdot\lambda^n, & n \text{ even,}\\ 0, & n \text{ odd}\end{cases}.$$

Now, let u be a Haar unitary $*$-free from q. By Theorem 4.2(2) in [NS2], $b = qu$ is $\mathscr{R}$-diagonal. (The conditions of the theorem require the C^*-probability space to be tracial; however, we may simply restrict φ to the unital C^* algebra generated by the normal elements q and u, where it is always a trace.) Since b is $\mathscr{R}$-diagonal, we can compute its determining sequences by

$$\alpha_n[b] = \beta_n[b] = \sum_{\pi\in NC^*(n,1)} \varphi_\pi[b,b^*,\ldots,b,b^*]\mu(\pi,1_{2n}).$$

Well, since $\pi \in NC^*(n,1)$, all blocks in π are of even size and alternately connect b's and b^*'s. Hence, for each block V in π,

$$\varphi_V[b,b^*,\ldots,b,b^*] = \varphi[(bb^*)^{|V|/2}] = \varphi[(quu^*q)^{|V|/2}] = \varphi[q^{|V|}] = \varphi_V[q,\ldots,q], \tag{4.5}$$

and thus $\varphi_\pi[b,b^*,\ldots,b,b^*] = \varphi_\pi[q,\ldots,q]$ for $\pi\in NC^*(n,1)$.

Now, suppose σ is a partition in $NC(2n)\setminus NC^*(n,1)$ – i.e. σ contains a block $V = \{k_1,\ldots,k_r\}$ with two successive elements $k_\ell < k_{\ell+1}$ of the same parity. (Indeed, $NC^*(n,1)$ consists of non-crossing partitons whose blocks always successively pair b's and b^*'s in the pattern $[b,b^*,\ldots,b,b^*]$ – i.e. the blocks must alternately pair even and odd numbers in $\{1,\ldots,2n\}$.) But then there is an odd number of elements between k_ℓ and $k_{\ell+1}$, and so some block in σ must be of odd size. Since q is an even element, it follows that $\varphi_\sigma[q,\ldots,q] = 0$. Hence, we also have $\kappa_{2n}[q,\ldots,q] = \sum_{\pi\in NC^*(n,1)}\varphi_\pi[q,\ldots,q]\mu_{2n}(\pi,1_{2n})$, and so from Equation 4.5,

$$\begin{aligned}\alpha_n[b] &= \sum_{\pi\in NC^*(n,1)} \varphi_\pi[b,b^*,\ldots,b,b^*]\mu_{2n}(\pi,1_{2n})\\ &= \sum_{\pi\in NC^*(n,1)} \varphi_\pi[q,\ldots,q]\mu_{2n}(\pi,1_{2n}) = \kappa_{2n}[q,\ldots,q] = \gamma\cdot\lambda^{2n}.\end{aligned}$$

□

Following the argument of Corollary 4.2, we see that if we choose an $\mathscr{R}$-diagonal element b which satisfies $\alpha_n[b] \ge |\alpha_n[a]|$ for all n then letting $b_{n,m}$ be the list corresponding to $[b^n(b^*)^n]^m$, we have $\kappa_\pi[b_{n,m}] \ge |\kappa_\pi[a_{n,m}]|$, and so

$$\|b^n\|_{2m}^{2m} = \sum_{\pi\in NC^*(n,m)} \kappa_\pi[b_{n,m}] \ge \sum_{\pi\in NC^*(n,m)} |\kappa_\pi[a_{n,m}]|.$$

Hence, from Lemma 4.1, we have

$$\|T\|_{2m} \le \frac{\|b^n\|_{2m}}{\|a\|_2^n}\|T\|_2. \tag{4.6}$$

In order for this to yield useful information, we must choose b in such a way that its variance and norm are well-controlled by those of a. In the following lemma, we choose $b = b_{\gamma,\lambda}$ as in Lemma 4.4 to optimally bound the ratio $\|b^n\|/\|a\|_2^n$.

Lemma 4.5. *Let a be $\mathscr{R}$-diagonal, and define $\lambda = 2^8\|a\|^2/\|a\|_2$ and $\gamma = \|a\|_2^2\lambda^{-2}$. Set $b = b_{\gamma,\lambda}$, as in Lemma 4.4. Then $\|b\|_2 = \|a\|_2$, and*

$$\frac{\|b^n\|}{\|a\|_2^n} \le 515\,\sqrt{e}\,\sqrt{n}\,\frac{\|a\|^2}{\|a\|_2^2}.$$

Proof. For $\mathscr{R}$-diagonal b, Corollary 3.2 in [Lar] says that $\|b^n\| \le \sqrt{e}\,\sqrt{n}\,\|b\|\,\|b\|_2^{n-1}$. Note that, since b is centred, $\|b\|_2^2 = \kappa_2[b,b^*]$ which, from Lemma 4.4, equals $\gamma\cdot\lambda^2 = \|a\|_2^2$. Hence,

$$\frac{\|b^n\|}{\|a\|_2^n} = \frac{\|b^n\|}{\|b\|_2^n} \le \sqrt{e}\,\sqrt{n}\,\frac{\|b\|}{\|b\|_2} = \sqrt{e}\,\sqrt{n}\,\frac{\|b\|}{\|a\|_2}. \tag{4.7}$$

For the norm $\|b\|$, we have $b = qu$ where u is unitary, and so $\|b\| = \|q\| = \|p_1 - p_2\| \le 2\|p_1\|$. The norm of a free Poisson was calculated in [VDN]; the result is $\|p_1\| = \lambda(1+\sqrt{\gamma/2})^2$, so

$$\sqrt{\gamma/2} = 2^{-1/2}\cdot\|a\|_2\lambda^{-1} = 2^{-8.5}\frac{\|a\|_2^2}{\|a\|^2} \le 2^{-8.5},$$

and so

$$\|b\| \le 2\cdot 2^8\frac{\|a\|^2}{\|a\|_2}\cdot(1+2^{-8.5})^2 \le 515\frac{\|a\|^2}{\|a\|_2},$$

yielding the result. □

We now stand ready to prove the main result of this paper.

Proof of Theorem 1.3. We will check that the element $b = b_{\gamma,\lambda}$ with coefficients chosen as in Lemma 4.5 has all positive cumulants which dominate the absolute values of the cumulants of a. First, we have (as used above) $\alpha_1[b] = |\alpha_1[a]|$. For higher cumulants, using Lemma 4.4,

$$\alpha_n[b] = \gamma\cdot\lambda^{2n} = \|a\|_2^2\left(2^8\frac{\|a\|^2}{\|a\|_2}\right)^{2n-2} = \frac{1}{4}(2^4\|a\|)^{2n}\cdot\left(\frac{\|a\|}{\|a\|_2}\right)^{2n-4}2^{8n-14},$$

and since $n \ge 2$ and $\|a\|_2 \le \|a\|$, this is $\ge \frac{1}{4}(2^4\|a\|)^{2n}$ which is, by Lemma 4.3, $\ge |\alpha_n[a]|$. Having shown that $\alpha_n[b] \ge |\alpha_n[a]|$ for all n, we may now use Equation 4.6. We have (taking the limit as $m\to\infty$)

$$\|T\| \le \frac{\|b^n\|}{\|a\|_2^n}\|T\|_2,$$

and from Lemma 4.5 this yields the result:

$$\|T\| \le 515\,\sqrt{e}\,\frac{\|a\|^2}{\|a\|_2^2}\,\sqrt{n}\,\|T\|_2.$$

If the cumulants of a are all non-negative, then Equation 4.6 holds with $b = a$, and then Equation 4.7 yields the tighter estimate. □

We conclude this section with a discussion of Brown measure.

Theorem 4.6. *Let a be an $\mathscr{R}$-diagonal element which is not a scalar multiple of a Haar unitary, and let ν_a be its Brown measure. There are constants $C_1(n), C_2(n) \asymp \sqrt{n}$ such that*

$$C_1(n)\|z^n\|_2 \le \|z^n\|_\infty \le C_2(n)\|z^n\|_2.$$

To be clear, the norms in the statement of the theorem are the usual measure-theoretic norms,

$$\|z^n\|_\infty = \sup_{z\in\operatorname{supp}\nu_a}|z^n|\,,\quad \|z^n\|_2 = \left[\int |z^n|^2\,d\nu_a(z,\bar{z})\right]^{1/2}.$$

Proof. Fist note from Theorem 2.1, there is a function $f\colon [0,\|a\|_2] \to \mathbb{R}_+$ which is continuous and satisfies $f(\|a\|_2) > 0$, such that $d\nu_a = f(r)dr\, d\theta$ with supp ν_a equal to an annulus whose outer radius is $\|a\|_2$. Of course, this means that $\sup_{\text{supp }\nu_a} |z^n| = \|a\|_2^n$. For the 2-norm, let M be the supremum of f on $[0,\|a\|_2]$; then

$$\int_{\text{supp }\nu_a} |z^n|^2\, d\nu_a(z,\bar{z}) = \int_0^{2\pi}\int_0^{\|a\|_2} r^{2n}\, f(r)\, dr\, d\theta \le \frac{2\pi M}{2n+1}\|a\|_2^{2n+1} = \frac{2\pi M \|a\|_2}{2n+1}\|z^n\|_\infty^2,$$

and this shows that $\|z^n\|_\infty/\|z^n\|_2 \gtrsim \sqrt{n}$. For the reverse inequality, since f is continuous and $f(\|a\|_2) > 0$, there are $\epsilon, m > 0$ such that $f(r) \ge m > 0$ for $r \in [\|a\|_2 - \epsilon, \|a\|_2]$, and so since $f \ge 0$ everywhere,

$$\int_0^{2\pi}\int_0^{\|a\|_2} r^{2n}\, f(r)\, dr\, d\theta \ge 2\pi m \int_{\|a\|_2-\epsilon}^{\|a\|_2} r^{2n}\, dr = \frac{2\pi m\|a\|_2}{2n+1}\left(1-(1-\epsilon/\|a\|_2)^{2n+1}\right)\|a\|_2^{2n} \gtrsim \frac{\|z^n\|_\infty^2}{n}.$$

□

As discussed in Section 2.1, the Brown measure of a non-normal element a (as most $\mathscr{R}$-diagonal elements are) does not respect mixed moments; that is, $\varphi(a^*a) \neq \int |z|^2\, d\nu_a(z,\bar{z})$ in general, and so forth. Nevertheless, as we see in Theorem 4.6, a Haagerup inequality with the same $O(n^{1/2})$-behaviour holds in the space $\mathcal{H}L^2(\nu_a)$ of holomorphic L^2 functions with respect to the Brown measure of any $\mathscr{R}$-diagonal element. $\mathcal{H}L^2(\nu_a)$ is, in some sense, the commutative model for our spaces $\mathcal{H}(a,I)$ (at least in the case where $|I|=1$), and so we see that the Brown measure does retain some information about mixed moments.

5. Strong Ultracontractivity

In this final section, we apply our strong Haagerup inequality (Theorem 1.3) to give strong ultracontractive bounds for the Ornstein-Uhlenbeck semigroup on $\mathcal{H}(a,I)$. In Section 5.1 we define the O-U semigroup in this general context, and show that it is a natural generalization of the free O-U semigroup considered in [Bi2]. In Section 5.2, we prove optimal ultracontractive bounds, and discuss applications to free groups.

5.1. **Ornstein-Uhlenbeck semigroups.** Let a be $\mathscr{R}$-diagonal. Consider the operator N_{fin}, defined on the algebraic direct sum $\bigoplus_{n=0}^\infty \mathcal{H}^{(n)}(a,I)$ (which is, of course, dense in $L^2(\mathcal{H}(a,I),\varphi)$) as the linear extension of $N_{\text{fin}}(h_n) = n\, h_n$ for $h_n \in \mathcal{H}^{(n)}(a,I)$. Since $h_n \perp h_m$ for $n \neq m$ (this follows from the $*$-freeness of the a_i), the operator N_{fin} is symmetric and lower-semi-bounded by 0. Thus, by the Friedrich's extension theorem, N_{fin} extends to a densely-defined (unbounded) self-adjoint operator N on $L^2(\mathcal{H}(a,I),\varphi)$, and this operator is postive semidefinite. We will refer to N as the **number operator** affiliated with $\mathcal{H}(a,I)$.

Proposition 5.1. *The number operator N affiliated with $\mathcal{H}(a,I)$ generates a $\mathscr{C}_0$ contraction semigroup e^{-tN} on $L^2(\mathcal{H}(a,I),\varphi)$.*

Proof. Since the spaces $\mathcal{H}^{(n)}(a,I)$ reduce N, we see easily that e^{-tN} must act via

$$e^{-tN}\sum_{n=0}^\infty h_n = \sum_{n=0}^\infty e^{-nt} h_n.$$

It is then immediately verified that e^{-tN} is a contraction semigroup, since $e^{-nt} \le 1$ for all $t \ge 0$. To prove that is it $\mathscr{C}_0$, it suffices to show that $w\text{-}\lim_{t\downarrow 0} e^{-tN}h = h$ for each $h \in L^2(\mathcal{H}(a,I),\varphi)$. Let $h = \sum h_n$ and $g = \sum g_n$; since $h_n \perp g_m$ for $n \neq m$,

$$\langle e^{-tN}h, g\rangle = \left\langle \sum_{n=0}^\infty e^{-nt}h_n, \sum_{m=0}^\infty g_m \right\rangle = \sum_{n=0}^\infty e^{-nt}\langle h_n, g_n\rangle.$$

As both h and g are in L^2, the sequence $\langle h_n, g_n\rangle$ is in ℓ^1, and since $e^{-nt} \le 1$, it follows from the dominated convergence theorem that

$$\lim_{t\downarrow 0}\sum_{n=0}^{\infty} e^{-nt}\langle h_n, g_n\rangle = \sum_{n=0}^{\infty}\langle h_n, g_n\rangle = \langle h, g\rangle.$$

□

An important example of this number operator is given in the case of a circular element $a = c$. In this case, $\mathcal{H}(c, I)$ is naturally isomorphic to the holomorphic space $\mathcal{H}_0(\mathscr{H})$ over a Hilbert space $\mathscr{H}$ of dimension $|I|$, as defined in the first author's paper [Ke], and the number operator N above is just the free Ornstein-Uhlenbeck (number) operator N_0 considered in that paper. N_0 is the restriction to the holomorphic space $\mathcal{H}_0(\mathscr{H})$ of the free Ornstein-Uhlenbeck operator defined in [Bi2] on the free group factor $L(\mathbb{F}_{|I|})$, which coincides with the 0-Gaussian factor $\Gamma_0(\mathscr{H})$ introduced in [Vo] and further developed in [BoS, BKS]. There is a family of such spaces $\Gamma_q(\mathscr{H})$ for $-1 \le q \le 1$ (with $q = 1$ corresponding to the classical theory of Gaussian random variables, and $q = -1$ the hyperfinite II_1-factor), and Biane introduced number operators N_q affiliated to each of them. We should also note that, in [Bi1], Biane introduced a space $\mathscr{C}_{hol}(\mathscr{H})$ isomorphic to $\mathcal{H}(c, I)$, but did not consider the action of a number operator on it.

The main theorem of [Ke] shows as a special case (the case $q = 0$) that the semigroup e^{-tN} affiliated with $\mathcal{H}(c, I)$ is not only a contraction semigroup on $L^2(\mathcal{H}(c, I), \varphi)$ (for tracial φ), but is in fact *strongly hypercontractive*:

Theorem 5.2 (Theorem 4 in [Ke])**.** *Let $r > 2$ be an even integer, and let $t_J(2, r) = \frac{1}{2}\log\frac{r}{2}$. Then for $t \ge t_J(2, r)$, e^{-tN} is a contraction from $L^2(\mathcal{H}(c, I), \varphi)$ to $L^r(\mathcal{H}(c, I), \varphi)$.*

This *strong hypercontractivity* theorem is the precise analogue of the same theorem in the context of the spaces $\mathcal{H}L^r(\mathbb{C}^n, \gamma)$ (where γ is Gauss measure) proved by Janson in [Ja]. (We should note, however, that Janson's theorem holds from $L^p \to L^r$ for $0 < p \le r < \infty$, not just the discrete values in [Ke].) The time t_J is shorter than the least time to contraction t_N in the real spaces $L^r(\mathbb{R}^n, \gamma)$, where the hypercontractivity inequalities were first proved and studied by Nelson in [N]. The main theorem of [Bi2] is the generalization of Nelson's hypercontractivity theorem to the q-Gaussian factors.

5.2. **Ultracontractivity.** In the classical holomorphic case studied by Janson, while the semigroup e^{-tN} is a contractive map from $\mathcal{H}L^2$ to $\mathcal{H}L^r$ for any $r > 2$ once t is large enough, it is also unbounded for $t < t_J(2, r)$. As a result, the semigroup e^{-tN} does not map $\mathcal{H}L^2$ into the algebra of bounded functions for any time. Of course, in the classical context, the algebra of bounded functions contains no holomorphic functions save constants; even in the full real spaces, the same effect holds. This is essentially due to the fact that the kernel of the semigroup e^{-tN} in these cases, the *Mehler kernel*, is not a bounded function.

A semigroup is called *ultracontractive* if it maps L^2 into L^∞ for all $t > 0$. The Ornstein-Uhlenbeck semigroups studied by Nelson and Janson (and many others) fail to be ultracontractive. Nevertheless, the non-commutative counterpart e^{-tN_0} on the free group factor *is* ultracontractive, as shown in [Bi2] and essentially in [Bo1].

Proposition 5.3 (Corollary 3 in [Bi2])**.** *The free Ornstein-Uhlenbeck semigroup e^{-tN_0} is ultracontractive; there is $C > 0$ with*

$$\|e^{-tN_0}\|_{L^2(\Gamma_0)\to L^\infty(\Gamma_0)} \le C\, t^{-3/2}, \quad 0 < t < 1.$$

(In general the function $t \mapsto \|e^{-tN_0}X\|_r$ is decreasing for any X and r, hence it is only small-time behaviour which is interesting.) Bożejko later generalized this theorem to all the Γ_q factors with $-1 < q < 1$; see [Bo2].

The generators of the algebra Γ_0 (the free group factor) are $*$-free semicircular elements. Thus, the $*$-algebra generated by $\mathcal{H}(c, I)$ is contained in Γ_0, and the ultracontractive $O(t^{-3/2})$-bound of Proposition 5.3 also holds for the semigroup e^{-tN} affiliated with $\mathcal{H}(c, I)$ defined above. Using our main theorem, Theorem 1.3, we may essentially follow Biane's argument and prove a stronger form of Proposition 5.3 not only for the algebra $\mathcal{H}(c, I) \cong \mathcal{H}_0(\mathscr{H})$, but in fact for all $\mathcal{H}(a, I)$ with a $\mathscr{R}$-diagonal. Indeed, we find that the short-time behaviour in the $\mathscr{R}$-diagonal case is $O(t^{-1})$.

Theorem 5.4. *Let a be $\mathscr{R}$-diagonal, and let N be the number operator affiliated with $\mathcal{H}(a, I)$. Then e^{-tN} is ultracontractive; for each $h \in L^2(\mathcal{H}(a, I), \varphi)$, $e^{-tN}h \in \mathcal{H}(a, I)$ for $t > 0$, and moreover*

$$\|e^{-tN}h\| \le \frac{1}{2} C_a\, t^{-1}\, \|h\|_2 \quad t > 0. \tag{5.1}$$

(Here C_a is the same constant as in Theorem 1.3.) We refer to Theorem 5.4 as *strong ultracontractivity*, as it is a stronger version of the inequality in Proposition 5.3 which holds when the semigroup is restricted to a holomorphic subspace. This is similar in spirit to the stronger form of hypercontractivity [Ja] which holds in the holomorphic version of Nelson's setup in [N]. We emphasize, again, that ultracontractivity is a *strictly* non-commutative effect in this case, since the semigroup is unbounded from $L^2 \to L^\infty$ in the classical (real and holomorphic) contexts. Theorem 5.4 is thus an essentially non-commutative result which highlights the interesting phenomenon that many functional inequalities improve in the holomorphic category.

Proof. Let $h = \sum_{n=0}^\infty h_n$ with $h_n \in \mathcal{H}^{(n)}(a, I)$. We estimate

$$\|e^{-tN}h\| = \left\| \sum_{n=0}^\infty e^{-nt} h_n \right\| \le \sum_{n=0}^\infty e^{-nt} \|h_n\|.$$

We now employ Theorem 1.3, which implies that $h_n \in \mathcal{H}(a, I)$ and $\|h_n\| \le C_a\, \sqrt{n}\, \|h_n\|_2$. Thus,

$$\|e^{-tN}h\| \le C_a \sum_{n=0}^\infty \sqrt{n}\, e^{-nt} \|h_n\|_2 \le C_a \left[\sum_{n=0}^\infty n\, e^{-2nt} \right]^{1/2} \cdot \left[\sum_{n=0}^\infty \|h_n\|_2^2 \right]^{1/2},$$

where we have used the Cauchy-Schwarz inequality. The second factor is just $\|h\|_2$. The first factor is the derivative of $-\frac{1}{2}\sum_{n=0}^\infty e^{-2nt} = -\frac{1}{2}\frac{1}{1-e^{-2t}}$. The reader may readily verify that we thus have

$$\|e^{-tN}h\| \le C_a \frac{e^{-t}}{1 - e^{-2t}}\, \|h\|_2$$

for all $t > 0$. This shows that $e^{-tN}h \in \mathcal{H}(a, I)$. Moreover, the function $t \mapsto \frac{te^{-t}}{1-e^{-2t}}$ is decreasing on $\mathbb{R}_+$ and has limit $1/2$ at $t = 0$. This proves Equation 5.1. $\square$

It is typical to prove, from a bound like Equation 5.1, a Sobolev inequality of the form $\|h\|_p \le C\langle Nh, h\rangle$, $h \in \mathscr{D}(N)$ for an appropriate $p > 2$; indeed, if e^{-tN} in Theorem 5.4 were a classical sub-Markovian semigroup defined on L^2 of a Radon measure, we could use the standard techniques in, for example, [CSV], to prove a strong Sobolev imbedding theorem (for *any* $p < \infty$) in this case. However, the techniques necessary to implement such a proof use the Marcinkewicz interpolation theorem in a fundamental way. As pointed out in [Ke], holomorphic spaces like $\mathcal{H}(a, I)$ (in particular in the case $a = c$) tend not to be complex interpolation scale (at least in the $|I| = \infty$ case). Thus, we are unable to prove a Sobolev inequality for $\mathcal{H}(a, I)$ using known-techniques.

We finally remark that one interesting new application of this theorem is to the discrete O-U semigroup on the free group $\mathbb{F}_k$ (or rather its restriction to $\mathbb{F}_k^+$). As noted above, the algebra

$\mathcal{H}(u, I_k)$ with u a Haar unitary and $|I_k| = k$ is isomorphic to the convolution-norm closure of $\mathbb{F}_k^+$ in $L(\mathbb{F}_k)$, and thus $L^2(\mathcal{H}(u, I_k), \varphi_k) \cong \ell^2(\mathbb{F}_k^+)$, where the number operator N acts by $Nw = nw$ on a word w of length n. The same semigroup e^{-tN} defined on all of $\mathbb{F}_k$ was essentially introduced in [H], and has been studied in [JLX, JX] with a view towards L^p-contraction bounds; to the authors' knowledge, Theorem 5.4 yields the first ultracontractive bound in that context.

Acknowledgement. The authors kindly thank Drew Armstrong and Heather Armstrong for their combinatorial insights.

References

[AO] Akemann, C.; Ostrand, P:. *Computing norms in group C^*-algebras.* Amer. J. Math. **98**, 1015–1047 (1976)

[Ba] Bargmann, V.: *On a Hilbert space of analytic functions and an associated integral transform.* Comm. Pure Appl. Math. **14**, 187-214 (1961)

[BJ] Bisch, D.; Jones, V.: *Algebras associated to intermediate subfactors.* Invent. Math. **128**, 89-157 (1997)

[Bi1] Biane, P.: *Segal-Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems.* J. Funct. Anal. **144**, 232–286 (1997)

[Bi2] Biane, P.: *Free hypercontractivity.* Commun. Math. Phys. **184**, 457-474 (1997)

[BiS] Biane, P.; Speicher, R.: *Stochastic calculus with respect to free Brownian motion and analysis on Wigner space.* Probab. Theory Related Fields **112**, 373-409 (1998)

[BKS] Bozejko, M., Kümmerer, B., Speicher, R.: *q-Gaussian processes: non-commutative and classical aspects.* Commun. Math. Phys. **185**, 129-154 (1997)

[Bo1] Bożejko, M.: *A q-deformed probability, Nelson's inequality and central limit theorems.* Nonlinear fields: classical, random, semiclassical (Karpacz, 1991), 312–335, World Sci. Publishing, River Edge, NJ, 1991

[Bo2] Bożejko, M.: *Ultracontractivity and strong Sobolev inequality for q-Ornstein-Uhlenbeck semigroup* $(-1 < q < 1)$. Infin. Dimens. Anal. Quantum Probab. Relat. Top. **2** 204-220 (1999)

[BoS] Bozejko, M., Speicher, R.: *An example of a generalized Brownian motion.* Comm. Math. Phys. **137**, 519–531 (1991)

[Br] Brown, L.: *Lidskiĭ's theorem in the type* II *case.* Geometric methods in operator algebras (Kyoto, 1983), 1–35, Pitman Res. Notes Math. Ser., **123**, Longman Sci. Tech., Harlow, 1986.

[CPS] Chatterji, I.; Pittet, Ch.; Saloff-Coste, L.: *Connected Lie Group and Property RD.* To appear.

[CSV] Coulhon, T.; Saloff-Coste, L.; Varopoulos, N.Th.: *Analysis and geometry on groups.* Cambridge Tracts in Mathematics, 100. Cambridge University Press, Cambridge, 1992.

[E] Edelman, P.: *Chain enumeratation and non-crossing partitions.* Discrete Math. **31**, 171-180 (1980)

[G] Gross, L.: *Hypercontractivity over complex manifolds.* Acta. Math. **182**, 159-206 (1999)

[H] Haagerup, U.: *An example of a nonnuclear C^*-algebra, which has the metric approximation property.* Invent. Math. **50** 279-293 (1978/79)

[HL] Haagerup, U.; Larsen, F.: *Brown's spectral distribution measure for $\mathscr{R}$-diagonal elements in finite von Neumann algebras.* J. Funct. Anal. **176**, 331-367 (2000)

[HP] Hiai, F.; Petz, D.: *Properties of free entropy related to polar decomposition.* Commun. Math. Phys. **202**, 421-444 (1999)

[Ja] Janson, S.: *On hypercontractivity for multipliers on orthogonal polynomials.* Ark. Math. **21**, 97-110 (1983)

[JLX] Junge, M.; Le Merdy, C.; Xu, Q.: *Calcul fonctionnel et fonctions carrées dans les espaces L^p non commutatifs.* C. R. Math. Acad. Sci. Paris **337** 93-98 (2003)

[JX] Junge, M.; Xu, Q.: *Théorèmes ergodiques maximaux dans les espaces L_p non commutatifs.* C. R. Math. Acad. Sci. Paris **334** 773-778 (2002)

[Ke] Kemp, T.: *Hypercontractivity in non-commutative holomorphic spaces.* Commun. Math. Phys. **259**, 615-637 (2005)

[Kr] Kreweras, G.: *Sur les partitions non-croisses d'un cycle.* Discrete Math. **1**, 333-350 (1972)

[Laf1] Lafforgue, Vincent *Une démonstration de la conjecture de Baum-Connes pour les groupes réductifs sur un corps p-adique et pour certains groupes discrets possédant la propriété (T).* C. R. Acad. Sci. Paris Sér. I Math. **327** 439-444 (1998)

[Laf2] Lafforgue, V.: *A proof of property (RD) for cocompact lattices of* $\mathrm{SL}(3, \mathbb{R})$ *and* $\mathrm{SL}(3, \mathbb{C})$. J. Lie Theory **10**, 255-267 (2000)

[Lar] Larsen, F.: *Powers of $\mathscr{R}$-diagonal elements.* J. Operator Theory **47**, 197-212 (2002)

[N] Nelson, E.: *The free Markov field.* J. Funct. Anal. **12**, 211-227 (1973)

[NS1] Nica, A.; Speicher, R.: *$\mathscr{R}$-diagonal pairs—a common approach to Haar unitaries and circular elements.* Fields Inst. Commun., **12**, 149-188 (1997)

[NS2] Nica, A.; Speicher, R.: *Commutators of free random variables.* Duke Math. J. **92**, 553-592 (1998)

[NS3] Nica, A.; Speicher, R.: *Lectures on the Combinatorics of Free Probability.* London Mathematical Society Lecture Note Series, no. 335, Cambridge University Press, 2006

[NSS] Nica, A.; Shlyakhtenko, D.; Speicher, R.: *Maximality of the microstates free entropy for $\mathscr{R}$-diagonal elements.* Pac. J. Math. **187**, 333-347 (1999)
[O] Oravecz, F.: *On the powers of Voiculescu's circular element.* Studia Math. **145** 85-95 (2001)
[R] Rudin, W.: *Real and complex analysis.* Third edition. McGraw-Hill Book Co., New York, 1987.
[S] Speicher, R.: *Multiplicative functions on the lattice of noncrossing partitions and free convolution.* Math. Ann. **298**, 611–628 (1994)
[Va1] Valette, A.: *On the Haagerup inequality and groups acting on $\widetilde{A}_n$-buildings.* Ann. Inst. Fourier (Grenoble) **47** 1195-1208 (1997)
[Va2] Valette, A.: *Introduction to the Baum-Connes conjecture.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2002.
[Vo] Voiculescu, D.V.: *Symmetries of some reduced free product C^* algebras.* In: *Operator Algebras and their Connection with Topology and Ergodic Theory, Lecture Notes in Mathematics,* Vol. **1132**, Berlin-Heidelberg-New York: Springer, 1985, pp. 566-588
[VDN] Voiculescu, D.V.; Dykema, K.; Nica, A.: *Free random variables.* CRM Monograph Series, **1**. American Mathematical Society, Providence, RI, 1992.

$*$ DEPARTMENT OF MATHEMATICS, MIT, CAMBRIDGE, MA, 02139
E-mail address: tkemp@math.mit.edu

$**$ DEPARTMENT OF MATHEMATICS AND STATISTICS, QUEEN'S UNIVERSITY, KINGSTON, ON, K7L 3N6
E-mail address: speicher@mast.queensu.ca