On the Rate of Convergence in the Entropic Central Limit Theorem

Shiri Artstein [*], Keith Ball[†], Franck Barthe[‡] and Assaf Naor[§]

September 25, 2003

Abstract

We study the rate at which entropy is produced by linear combinations of independent random variables which satisfy a spectral gap condition.

1 Introduction

The Shannon entropy of a random variable X with density f is defined as $\mathrm{Ent}(X) = -\int_{\mathbb{R}} f \log f$, provided the positive part of the integral is finite. So $\mathrm{Ent}(X) \in [-\infty, +\infty)$. If X has variance 1, it is a classical fact that its entropy is well defined and bounded above by that of a standard Gaussian random variable G. By the Pinsker-Csiszar-Kullback inequality ([12], [9], [10]), G is a stable maximizer of entropy in the following strong sense:

$$d_{TV}(X - \mathbb{E}X, G)^2 \leq 2[\mathrm{Ent}(G) - \mathrm{Ent}(X)], \tag{1}$$

where d_{TV} is the total variation distance. It is therefore natural to track the convergence in the central limit theorem (CLT) in terms of entropy. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent copies of a random variable X. For notational convenience we assume that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$. Set

$$S_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}.$$

[*]Supported in part by the EU Grant HPMT-CT-2000-00037, The Minkowski center for Geometry and the Israel Science Foundation

[†]Supported in part by NSF Grant DMS-9796221

[‡]Supported in part by EPSRC Grant GR/R37210

[§]Supported in part by the BSF, Clore Foundation and EU Grant HPMT-CT-2000-00037

Linnik [11] used entropy to reprove convergence in distribution of S_n to G. He actually showed that $\mathrm{Ent}(S_n)$ converges to $\mathrm{Ent}(G)$ under some smoothness assumptions on the law of X. This is discussed in [4] where Barron established entropic convergence in the CLT in full generality: if the entropy of S_n ever becomes different from $-\infty$, then $\mathrm{Ent}(S_n)$ converges to $\mathrm{Ent}(G)$. This convergence is very strong, as (1) shows.

The fact that the sequence of entropies $\mathrm{Ent}(S_n)$ has a limit follows from the Shannon-Stam inequality ([13], [15]): for independent variables Y, Z and $\lambda \in (0,1)$ one has

$$\mathrm{Ent}(\sqrt{1-\lambda}\,Y + \sqrt{\lambda}Z) \geq (1-\lambda)\mathrm{Ent}(Y) + \lambda \mathrm{Ent}(Z), \tag{2}$$

therefore $(n\mathrm{Ent}(S_n))$ is super-additive and converges to its supremum. In particular $\mathrm{Ent}(S_{2^k})$ is non-decreasing. In [1] we confirmed an old conjecture: that the entire sequence $(\mathrm{Ent}(S_n))$ is non-decreasing.

Until recently, quantitative improvements on the Shannon-Stam inequality (2), have been somewhat elusive. A *qualitative* result in this direction, under pretty general hypotheses was obtained by Carlen and Soffer [8] (see also the works of Brown and Shimuzu [7] [14]). In [3] we obtained quantitative estimates for random variables X satisfying a Poincaré (or spectral gap) inequality with constant $c > 0$: i.e. so that for every smooth function s, one has

$$c\,\mathrm{Var}(s(X)) \leq \mathbb{E}[(s'(X))^2].$$

The main result in [3] asserts that independent copies X_1, X_2 of such a random variable X satisfy

$$\mathrm{Ent}\left(\frac{X_1+X_2}{\sqrt{2}}\right) - \mathrm{Ent}(X) \geq \frac{c}{2+2c}(\mathrm{Ent}(G) - \mathrm{Ent}(X)).$$

In this article we apply an extension of the variational method of [3] to estimate the entropy of sums (or linear combinations) of more than two copies of a random varaiable. Independently of [3] and almost simultaneously, Barron and Johnson proved an analogous but slightly weaker fact for the Fisher information. We have just learned that, by modifying their argument they are able to recover our results for entropy: the new argument appears in [5].

The principal result of this article is the following.

Theorem 1 *Let $X_1, \ldots, X_n$ be independent copies of a random variable X with density f. Assume that* $\mathrm{Var}(X) = 1$ *and that X satisfies the Poincaré inequality with constant c. Then for every $a \in \mathbb{R}^n$ with $\sum_{i=1}^n a_i^2 = 1$,*

$$\mathrm{Ent}(G) - \mathrm{Ent}\left(\sum_{i=1}^n a_i X_i\right) \leq \frac{\alpha(a)}{c/2 + (1-c/2)\alpha(a)}\left(\mathrm{Ent}(G) - \mathrm{Ent}(X)\right),$$

where $\alpha(a) = \sum_{i=1}^{n} a_i^4$.

Remark. In fact, we obtain a slightly stronger, more complicated inequality- see inequality (16).

Before passing to the proof, let us make a few comments. Note that when one of the a_i's tends to ± 1, there can be no increase of entropy. This is consistent with the fact that $\alpha(a)$ tends to 1 in this case. On the other hand, for equal coefficients $1/\sqrt{n}$ we obtain that the entropy distance from the Gaussian decays at rate $1/n$

$$\mathrm{Ent}(G) - \mathrm{Ent}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) \leq \frac{1}{1+\frac{c}{2}(n-1)}\left(\mathrm{Ent}(G) - \mathrm{Ent}(X)\right). \tag{3}$$

In conjunction with the Csiszar-Kulback-Pinsker inequality, the above theorem leads to a quantitative information-theoretic proof of the central limit theorem (under a spectral gap assumption). The rate of convergence obtained (in total variation distance) is $O(1/\sqrt{n})$, which is well known to be optimal. Moreover, we also obtain Berry-Esseen type estimates, but in total variation, for the rate of convergence to a Gaussian.

2 Proof of the main theorem

Assume as we may that the variables are centered, that is $\mathbb{E}X = \int x f(x)\, dx = 0$. The proof is divided into several steps:

2.1 Reduction to an inequality for Fisher information

Recall that the Fisher information of a random variable X with density f is

$$I(X) = I(f) = \int \frac{(f')^2}{f}.$$

According to Bruijn's identity (see e.g. [15]), Fisher information is the time derivative of entropy along the heat semi-group. A similar relation holds along the adjoint Ornstein-Uhlenbeck semigroup (see [2, 4]), where the evolute at time t of a variable X, denoted $X^{(t)}$, has the same law as $e^{-t}X + \sqrt{1-e^{-2t}}\, G$ where G is a normal Gaussian variable independent of X. More precisely if $\mathrm{Ent}(X)$ is finite then for every $t > 0$, $X^{(t)}$ has finite Fisher information and

$$\mathrm{Ent}(G) - \mathrm{Ent}(X) = \int_0^\infty \left(I(X^{(t)}) - 1\right)\, dt.$$

A very clear explanation of this appears in [8]. Also note that $I(G) = 1$. Since the evolute $(\sum a_i X_i)^{(t)}$ has the same distribution as $\sum a_i X_i^{(t)}$ it suffices to prove:

Theorem 2 *Let Y be a random variable with $\mathrm{Var}(Y) = 1$, finite entropy, and spectral gap at least $c > 0$. Let $t > 0$, $X = Y^{(t)}$ and $X_1, \dots, X_n$ be independent copies of X. Then for every $a \in \mathbb{R}^n$ with $\sum_{i=1}^n a_i^2 = 1$,*

$$I\left(\sum_{i=1}^n a_i X_i\right) - 1 \le \frac{\alpha(a)}{c/2 + (1 - c/2)\alpha(a)} \left(I(X) - 1\right).$$

Let us also point out that the hypotheses in the latter theorem are far from optimal; we could certainly get away with less smoothness for X than is guaranteed by its being an Ornstein-Uhlenbeck evolute. However the hypothesis is sufficient for our goal and ensures that the density f of X has an integrable second derivative (see e.g. the Appendix of [3]). The rest of the paper is devoted to the proof of Theorem 2 and keeps the same notation.

2.2 Variational representations of the Fisher information

Our approach relies on the following result, which we proved in [1]

Theorem 3 *Let $w : \mathbb{R}^n \to (0, \infty)$ be a continuously twice differentiable density on $\mathbb{R}^n$ with*

$$\int \frac{\|\nabla w\|^2}{w}, \int \|Hess(w)\| < \infty.$$

Let a be a unit vector in $\mathbb{R}^n$ and h the marginal density in direction a defined by

$$h(t) = \int_{ta + a^\perp} w.$$

Then the Fisher information of the density h satisfies

$$J(h) \le \int \frac{(\mathrm{div}(pw))^2}{w},$$

for any continuously differentiable vector field $p : \mathbb{R}^n \to \mathbb{R}^n$ with the property that for every x, $\langle p(x), a \rangle = 1$ (and, say, $\int \|p\| w < \infty$). If w satisfies $\int \|x\|^2 w(x) < \infty$, then there is equality for some suitable vector field p.

The previous formula is very short but it leads to tedious calculations when applied to specific p. For this reason we provide another formulation of it, which is easier to use in the present situation.

Lemma 4 *Let w be as in Theorem 3 and let $p : \mathbb{R}^n \to \mathbb{R}^n$ be twice continuously differentiable, such that each coordinate function p_i is bounded and has bounded derivatives of first and second order. Then*

$$\int \frac{(\mathrm{div}(pw))^2}{w} = \int w \left[\mathrm{Tr}((Dp)^2) + \langle \mathrm{Hess}(-\log w)p, p\rangle \right].$$

Proof of the Lemma: Distributing the divergence and the square gives

$$\int \frac{(\mathrm{div}(pw))^2}{w} = \int \left(w(\mathrm{div}(p))^2 + \frac{\langle \nabla w, p\rangle^2}{w} + 2\mathrm{div}(p)\langle \nabla w, p\rangle \right).$$

On the other hand

$$\int w \left[\mathrm{Tr}((Dp)^2) + \langle \mathrm{Hess}(-\log w)p, p\rangle \right] = \int \left(w\mathrm{Tr}((Dp)^2) + \frac{\langle \nabla w, p\rangle^2}{w} - \langle \mathrm{Hess}(w)p, p\rangle \right).$$

So the identity we want amounts to

$$\int \left(w(\mathrm{div}(p))^2 + 2\mathrm{div}(p)\langle \nabla w, p\rangle - w\mathrm{Tr}((Dp)^2) + \langle \mathrm{Hess}(w)p, p\rangle \right) = 0.$$

This can be checked by integration by parts, once for the second term and twice for the fourth term, in order to get rid of derivatives of w. ■

Let X with density f be as in Theorem 2 and set $w(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n)$. Then the function $h(t) = \int_{ta+a^\perp} w$ is the density of $\sum_{i=1}^n a_i X_i$. The latter results yield

$$I\left(\sum_{i=1}^n a_i X_i \right) \le \int w \left[Tr(Dp)^2 + \langle \mathrm{Hess}(-\log w)p, p\rangle \right], \tag{4}$$

for any sufficiently integrable vector field $p : \mathbb{R}^n \to \mathbb{R}^n$ with the property that for every x, $\langle p(x), a\rangle = 1$. Note that the best choice of p would give equality in (4), whereas the simplest choice ($p(x) = a$ for all x) recovers $I\left(\sum_{i=1}^n a_i X_i\right) \le I(X)$, a consequence of the Blachman-Stam inequality [6, 15]. As in our previous works, our strategy is to optimize over a workable sub-family of test functions p.

2.3 Choice of particular test functions

In order to exploit the product structure of w, in expression (4), we restrict ourselves to functions p_i which are linear combinations of the constant function and functions depending only on one variable. It seems to be sufficient to choose a single one-variable function:

so our test function has the form $p_i(x_1,\dots,x_n)=c_{i0}+\sum_{j=1}^n c_{ij}r(x_j)$ and the only restriction on p is that $\sum_{i=1}^n a_ip_i\equiv 1$. We would like to choose $p=(a_1r(x_1),a_2r(x_2),\dots,a_nr(x_n))$ but since this does not satisfy the condition we project it orthogonally onto the hyperplane $\{x:\langle a,x\rangle=1\}$. So we choose

$$p_i(x_1,\dots,x_n)=a_i+a_i(1-a_i^2)r(x_i)-a_i\sum_{j\neq i}a_j^2r(x_j),$$

where r is a real-valued function on $\mathbb{R}$.

Now for smooth r with compact support inequality (4) is valid. Denoting $\phi=-\log f$, the two integrands on the right hand side of (4) are

$$\mathrm{Tr}\Big((Dp)^2\Big)=\sum_{1\le i,j\le n}\frac{\partial p_i}{\partial x_j}\frac{\partial p_j}{\partial x_i}=\sum_{i=1}^n a_i^2(1-a_i^2)^2r'(x_i)^2+\sum_{i\neq j}a_i^3a_j^3r'(x_i)r'(x_j),$$

and

$$\begin{aligned}\langle\mathrm{Hess}(-\log w)p,p\rangle &= \sum_{i=1}^n\phi''(x_i)a_i^2\left(1+(1-a_i^2)r(x_i)-\sum_{j\neq i}a_j^2r(x_j)\right)^2\\ &= \sum_{i=1}^n\phi''(x_i)a_i^2\Bigg[1+(1-a_i^2)^2r(x_i)^2+\sum_{j\neq i}a_j^4r(x_j)^2-2\sum_{j\neq i}a_j^2r(x_j)\\ &\qquad +2(1-a_i^2)r(x_i)-2\sum_{j\neq i}(1-a_i^2)a_j^2r(x_j)r(x_i)+\\ &\qquad \sum_{i\neq j\neq k\neq i}a_j^2a_k^2r(x_j)r(x_k)\Bigg].\end{aligned}$$

Integrating against the product density $w(x)$ and using the fact that $\sum a_i^2=1$, this identity reduces to

$$\begin{aligned}&\int w\left[Tr(Dp)^2+\langle\mathrm{Hess}(-\log w)p,p\rangle\right]\\ &= J+W\left(\int f(r')^2+\int f\phi''r^2\right)+JV\int fr^2+J(W-V)\left(\int fr\right)^2\\ &\qquad +2U\left(\int f\phi''r-J\int fr\right)-2W\left(\int fr\right)\left(\int f\phi''r\right)+M\left(\int fr'\right)^2,\end{aligned}$$

where:

$$J=I(X),\quad U=1-\sum_{i=1}^n a_i^4,\quad V=\sum_{i=1}^n a_i^4-\sum_{i=1}^n a_i^6,\quad W=U-V$$

$$\text{and } M=\left(\sum_{i=1}^n a_i^3\right)^2-\sum_{i=1}^n a_i^6.$$

Our aim is to choose r to make the estimate small. This is a standard variational problem except for the term $M\left(\int fr'\right)^2$. If the random variable X is symmetric, so that f is an even function, then we will choose an even function r and so the term in question will vanish. This gives a slightly better result for symmetric random variables. In the the general case we can control the term because

$$\left(\int fr'\right)^2 = \left(\int f'r\right)^2 \leq \left(\int \frac{(f')^2}{f}\right)\left(\int fr^2\right) = J\left(\int fr^2\right). \quad (5)$$

However, our choice of vector field p is invariant under the addition of a constant to r and this property simplifies considerably the calculations. In order to maintain this invariance we improve the estimate (5) slightly as follows. Setting $m = \int fr$ we have

$$\left(\int fr'\right)^2 = \left(\int f'(r-m)\right)^2 \leq \left(\int \frac{(f')^2}{f}\right)\left(\int f(r-m)^2\right) = J\left(\int fr^2 - m^2\right). \quad (6)$$

Lastly, the coefficient M is at most V by Hölder's inequality.

Hence we obtain the estimate

$$I\left(\sum_{i=1}^{n} a_i X_i\right) - J \leq T(r)$$

where

$$\begin{aligned} T(r) \;=\;& W\left(\int f(r')^2 + \int f\phi'' r^2\right) + 2JV\int fr^2 + J(W-2V)\left(\int fr\right)^2 \\ & +2U\left(\int f\phi'' r - J\int fr\right) - 2W\left(\int fr\right)\left(\int f\phi'' r\right). \end{aligned}$$

Our goal is to find r that makes this quantity as negative as possible.

2.4 Optimization in the test function r

In this subsection we apply variational methods to find the optimum value of $T(r)$. Their statement in the appropriate Sobolev spaces, as well as the questions of existence and regularity are discussed in detail in the appendix of [3]. So, in the following we shall ignore technical issues.

In order to deal with the quadratic terms $J(W-2V)\left(\int fr\right)^2$ and $-2W\left(\int fr\right)\left(\int f\phi'' r\right)$ we introduce a Lagrange multiplier λ and choose r to be λr_0 where r_0 is the minimiser of

$$\int f(r')^2 + \int f\phi'' r^2 + J\beta\int fr^2 + 2\left(\int f\phi'' r - J\int fr\right)$$

and $\beta = 2V/W$. The Euler-Lagrange equation for this minimization problem is

$$-(fr_0')' + fr_0\phi'' + J\beta fr_0 = Jf - f\phi''. \tag{7}$$

By integrating this equation against r_0 and 1 respectively, we obtain that any solution r_0 of (7) automatically satisfies the following relations

$$\int f(r_0')^2 + \int f\phi'' r_0^2 + J\beta \int fr_0^2 = J\int fr_0 - \int f\phi'' r_0, \tag{8}$$

$$\int f\phi'' r_0 + J\beta \int fr_0 = 0. \tag{9}$$

Set $A = \int fr_0$ and $B = \int f\phi'' r_0$. Using equations (8) and (9) we obtain

$$T(\lambda r) = AJ\lambda(1+\beta)\left(\lambda(1+A)W - 2U\right).$$

The optimal choice of λ is $U/W(1+A)$ and the estimate is then

$$T(\lambda r_0) = -\frac{JU^2(1+\beta)}{W}\frac{A}{A+1}. \tag{10}$$

To finish the proof we need to estimate $A = \int fr_0$ from below: we want to know that if U and W are close to 1 and β is small then A is large so that the quantity above captures almost all of J. From now on, we suppress the subscript and use r instead of r_0. The equation satisfied by r reads

$$-(fr')' + f\phi'' r + J\beta fr = f(J - \phi''). \tag{11}$$

Writing $f' = -f\phi'$ and dividing by f we can rewrite it as

$$-r'' + \phi' r' + \phi'' r + J\beta r = J - \phi''.$$

If we set $r = s'$ then we can integrate the equation to get

$$-s'' + \phi' s' + J\beta s = Jx - \phi' \tag{12}$$

where the constant of integration is chosen so that $\int fs = 0$ (assuming that our random variable is centred so that $\int fx = 0$). Multiplying (12) by f we get

$$-(fs')' + J\beta fs = Jfx - f\phi' \tag{13}$$

Now equation (13) is the Euler-Lagrange equation for a different minimisation problem in which the objective function is

$$Q(s) = \frac{1}{2}\int f(s')^2 + \frac{J\beta}{2}\int f\left(s - \frac{x}{\beta} + \frac{\phi'}{J\beta}\right)^2. \tag{14}$$

Observe that in our new notation $A = \int fs'$. Integrating equation (13) against s and x respectively we get

$$\int f(s')^2 + J\beta \int fs^2 = J \int fsx - \int fs'$$

and

$$\int fs' + J\beta \int fsx = J \int fx^2 - \int f = J - 1$$

since f is the density of a random variable with variance 1.

As before these equations yield an expression of $Q(s)$ in terms of $A = \int fs'$ only. One obtains that

$$Q(s) = \frac{1+\beta}{2\beta} A. \tag{15}$$

So in order to bound A from below it suffices to bound the functional Q. Note that the value of Q on the zero function is

$$Q(0) = \frac{J}{2\beta} \int f \left(x - \frac{\phi'}{J} \right)^2 = \frac{J-1}{2\beta}.$$

Now we use the fact that f satisfies a Poincaré inequality with constant c. This will force the minimum of Q to be of the order of $Q(0)$. Since $\int fs = 0$ the Poincaré inequality gives

$$c \int fs^2 \leq \int fs'^2.$$

Thus we get

$$\begin{aligned} Q(s) &= \frac{1}{2} \int fs'^2 + \frac{J\beta}{2} \int f \left(s - \frac{x}{\beta} + \frac{\phi'}{J\beta} \right)^2 \\ &\geq \frac{1}{2} \int f \left(cs^2 + J\beta \left(s - \frac{x}{\beta} + \frac{\phi'}{J\beta} \right)^2 \right) \\ &\geq \frac{1}{2} \cdot \frac{cJ\beta}{c + J\beta} \int f \left(-\frac{x}{\beta} + \frac{\phi'}{J\beta} \right)^2 \\ &= \frac{c}{2(c + J\beta)} Q(0) = \frac{c(J-1)}{2\beta(c + \beta J)}. \end{aligned}$$

Substituting into the expression (15) for $Q(s)$ in terms of A, we obtain the lower bound

$$A \geq \frac{c(J-1)}{(1+\beta)(c+J\beta)}.$$

Substituting this into (10) rearranging and using the facts that $W = U - V$ and $\beta = 2V/W$ gives an estimate slightly better than

$$I\left(\sum_{i=1}^{n} a_i X_i \right) - J \leq -\frac{c(J-1)U^2}{cU + 2V - cV}, \tag{16}$$

where we have used the fact that $J \geq 1$, which is a consequence of the Cramer-Rao inequality. It is easy to check that $V \leq U - U^2$ and this implies the desired estimate

$$I\left(\sum a_i X_i\right) - 1 \leq (J-1)\,\frac{1-U}{cU/2+1-U} = (J-1)\,\frac{\alpha(a)}{c(1-\alpha(a))/2+\alpha(a)}.$$

References

[1] S. Artstein, K. Ball, F. Barthe, and A. Naor. Solution of Shannon's Problem on the Monotonicity of Entropy. *Submitted*, 2002.

[2] D. Bakry and M. Emery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX*, number 1123 in Lect. Notes in Math., pages 179–206. Springer, 1985.

[3] K. Ball, F. Barthe, and A. Naor. Entropy jumps in the presence of a spectral gap. *Duke Math. J.* 119, No. 1, 41-63, 2003.

[4] A. R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14:336–342, 1986.

[5] A. R. Barron and O. Johnson. Fisher information inequalities and the central limit theorem. Preprint, ArXiv:math.PR/0111020

[6] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Info. Theory*, 2:267–271, 1965.

[7] L. D. Brown. A proof of the central limit theorem motivated by the Cramer-Rao inequality. In Kalliampur et al., editor, *Statistics and Probability: Essays in Honor of C. R. Rao*, pages 314–328, Amsterdam, 1982. North-Holland.

[8] E. A. Carlen and A. Soffer. Entropy production by block variable summation and central limit theorem. *Commun. Math. Phys.*, 140(2):339–371, 1991.

[9] I. Csiszar. Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen. *Publications of the Mathematical Institute, Hungarian Academy of Sciences*, VII, Series A:137–157, 1962.

[10] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Info. Theory*, 4:126–127, 1967.

[11] Ju. V. Linnik. An information theoretic proof of the central limit theorem with lindeberg conditions. *Theory Probab. Appl.*, 4:288–299, 1959.

[12] M.Š. Pinsker. Information and information stability of random variables and processes. Holden-Day, San Francisco, 1964.

[13] C. E. Shannon and W. Weaver. *The mathematical theory of communication.* University of Illinois Press, Urbana, IL, 1949.

[14] R. Shimizu. On Fisher's amount of information for location family. In Patil et al., editor, *A modern course on statistical distributions in scientific work*, Boston, MA, 1974. D. Reidel.

[15] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Info. Control*, 2:101–112, 1959.

2000 Mathematics Subjects Classification: 94A17, 60F05.

Shiri Artstein

School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel; E-mail artst@post.tau.ac.il

Keith M. Ball,

Department of Mathematics, University College London, Gower Street, London WC1 6BT, United Kingdom; E-mail: kmb@math.ucl.ac.uk

Franck Barthe,

Laboratoire d'Analyse et de Mathématiques Appliquées, CNRS UMR8050, Université de Marne-la-Vallée, Cité Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée, Cedex 2, France; E-mail: barthe@math.univ.mlv.fr

Assaf Naor,

Theory Group, Microsoft Research, One Microsoft Way, Redmond WA 98052-6399, USA; E-mail: anaor@microsoft.com