Exact Classical Simulation of the GHZ Distribution

Gilles Brassard[1], Luc Devroye[2], and Claude Gravel[3]

1 Université de Montréal, CIFAR and ETH-ITS, brassard@iro.umontreal.ca
2 McGill University, lucdevroye@gmail.com
3 Université de Montréal, claude.gravel@bell.net

Abstract

John Bell has shown that the correlations entailed by quantum mechanics cannot be reproduced by a classical process involving non-communicating parties. But can they be simulated with the help of bounded communication? This problem has been studied for more than twenty years and it is now well understood in the case of bipartite entanglement. However, the issue was still widely open for multipartite entanglement, even for the simplest case, which is the tripartite Greenberger–Horne–Zeilinger (GHZ) state. We give an exact simulation of arbitrary independent von Neumann measurements on general n-partite GHZ states. Our protocol requires $O(n^2)$ bits of expected communication between the parties, and $O(n \log n)$ expected time is sufficient to carry it out in parallel. Furthermore, we need only an expectation of $O(n)$ independent unbiased random bits, with no need for the generation of continuous real random variables nor prior shared random variables. In the case of equatorial measurements, we improve earlier results with a protocol that needs only $O(n \log n)$ bits of communication and $O(\log^2 n)$ parallel time. At the cost of a slight increase in the number of bits communicated, these tasks can be accomplished with a constant expected number of rounds.

1998 ACM Subject Classification J.2 Physical Sciences and Engineering: Physics

Keywords and phrases Entanglement simulation, Greenberger–Horne–Zeilinger (GHZ) state, Multiparty entanglement, von Neumann's rejection algorithm, Knuth-Yao's sampling algorithm

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.7

1 Introduction

The issue of non-locality in quantum physics was raised in 1935 by Einstein, Podolsky and Rosen when they introduced the notion of entanglement [10]. Thirty years later, Bell proved that the correlations entailed by entanglement cannot be reproduced by classical local hidden variable theories between noncommunicating parties [2]. This momentous discovery led to the natural question of *quantifying* quantum non-locality.

A natural quantitative approach to the non-locality inherent in a given entangled quantum state is to study the amount of resources that would be required in a purely classical theory to reproduce exactly the probabilities corresponding to measuring this state. More formally, we consider the problem of *sampling* the joint discrete probability distribution of the outcomes obtained by people sharing this quantum state, on which each party applies locally some measurement on his share. Each party is given a description of his own measurement but not informed of the measurements assigned to the other parties. This task would be easy (for a theoretician!) if the parties were indeed given their share of the quantum state, but they are not. Instead, they must *simulate* the outcome of these measurements without any quantum resources, using as little *classical communication* as possible.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 7–23
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This conundrum was introduced by Maudlin in 1992 in the simplest case of linear polarization measurements at arbitrary angles on the two photons that form a Bell state [17]. Similar concepts were reinvented independently years later by other researchers [5, 20]. This led to a series of results, culminating with the protocol of Toner and Bacon to simulate arbitrary von Neumann measurements on a Bell state with a single bit of communication in the worst case [21]. Later, Regev and Toner extended this result by giving a simulation of the correlations entailed by arbitrary binary von Neumann measurements on arbitrary bipartite states of any dimension using two bits of communication, also in the worst case [19]. Inspired by Ref. [20], Cerf, Gisin and Massar showed that the effect of an arbitrary pair of positive-operator-valued measurements (POVMs) on a Bell state can also be simulated with a bounded amount of expected communication [8]. A more detailed early history of the simulation of quantum entanglement can be found in Ref. [4, Sect. 6].

All this prior work is concerned strictly with the simulation of *bipartite* entanglement. Much less is known when it comes to simulating multipartite entanglement with classical communication, a topic that is still teeming with major open problems. Consider the simplest case, which is the simulation of independent arbitrary von Neumann measurements on the tripartite GHZ state, named after Greenberger, Horne and Zeilinger [14], which we shall denote $|\Psi_3\rangle = \frac{1}{\sqrt{2}}|000\rangle + \frac{1}{\sqrt{2}}|111\rangle$, or more generally on its n-partite generalization $|\Psi_n\rangle = \frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$.

The easiest situation arises in the special case of *equatorial* measurements (defined in Section 2) on the GHZ state because all the marginal probability distributions obtained by tracing out one or more of the parties are uniform. Hence, it suffices in this case to simulate the n-partite correlation. Once this has been achieved, all the marginals can easily be made uniform [11]. Making the best of this observation, Bancal, Branciard and Gisin have given a protocol to simulate equatorial measurements on the tripartite and fourpartite GHZ states at an expected cost of 10 and 20 bits of communication, respectively [1]. Later on, Branciard and Gisin improved this in the tripartite case with a protocol using 3 bits of communication in the worst case [3]. The simulation of equatorial measurements on $|\Psi_n\rangle$ for $n \geq 5$ was handled subsequently by Brassard and Kaplan in a paper published in the 2012 edition of this Conference on Theory of Quantum Computation, Communication and Cryptography, with an expected cost of $O(n^2)$ bits of communication [6]. This was the best result obtained until now on this line of work.

Despite substantial effort, the case of *arbitrary* von Neumann measurements, even on the original tripartite GHZ state $|\Psi_3\rangle$, was still wide open. Here, we solve this problem in the general case of the simulation of the n-partite GHZ state $|\Psi_n\rangle$, for any n, under the *random bit model* introduced in 1976 by Knuth and Yao [16], in which the only source of randomness comes from the availability of independently distributed unbiased random bits. Furthermore, we have no needs for prior shared random variables between the parties. An expected number of $6n + 17$ perfect random bits suffices to carry out our simulation. The expected communication cost is $O(n^2)$ bits, but only $O(n \log n)$ *time* if we count one step for sending bits in parallel according to a realistic scenario in which no party has to send or receive more than one bit in any given step. Furthermore, in the case of equatorial measurements, we improve the earlier best result [6] with an expected communication cost of only $O(n \log n)$ bits and $O(\log^2 n)$ parallel time. At the cost of a slight increase in the number of bits communicated and the number of required random bits, these tasks can be accomplished with a constant expected number of rounds.

More formally, the quantum task that we want to simulate is as follows. Each party i holds one qubit from state $|\Psi_n\rangle = \frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$ and is given the description of a von Neumann measurement M_i. By local operations, they collectively perform $\otimes_{i=1}^{n} M_i$ on $|\Psi_n\rangle$, thus obtaining one outcome each, say $b_i \in \{-1,+1\}$, which is their output. The joint probability distribution $p(b)$ of the b_i's is defined by the joint set of measurements (see Section 2). Our purpose is to sample *exactly* this joint probability distribution by a purely classical process that involves no prior shared random variables and as little communication as possible. Our complete solution builds on four ingredients: (1) Gravel's decomposition of $p(b)$ as a convex combination of two sub-distributions [12, 13]; (2) Knuth and Yao's algorithm to sample exactly probability distributions assuming only a source of unbiased identically independently distributed (i.i.d.) bits [16]; (3) the *universal method of inversion* [9, for instance]; and (4) our own distributed version of the classic *von Neumann's rejection algorithm* [18].

We define precisely our problem in Section 2 and we formulate our convex decomposition of the GHZ distribution, which is the key to its simulation. Then, we explain how to sample according to a Bernoulli distribution even when only approximations to the distribution's parameter are available. We also explain how the classic von Neumann rejection algorithm can be used to sample in the sub-distributions defined by our convex decomposition. However, little attention is paid in Section 2 to the fact that the various parameters that define the joint distribution are not available in a single place. Section 3 is concerned with the communication complexity issues. It culminates with a complete protocol to solve our problem, as well as its complete analysis. This is followed by variations on the theme, in which we consider a parallel model of communication, an expected bounded-round solution, and improvements on the prior art for the simulation of equatorial measurements. We conclude with a discussion and open problems in Section 4.

2 Sampling exactly the GHZ distribution in the random bit model

Any von Neumann measurement on a single qubit can be conveniently represented by a point on the surface of a three-dimensional sphere, known as the Bloch sphere, whose spherical coordinates can be specified by an *azimuthal* angle $\theta \in [0, 2\pi)$ and an *elevation* angle $\varphi \in [-\pi/2, \pi/2]$. These parameters defines a Hermitian idempotent operator

$$M = x\,\sigma_1 + y\,\sigma_2 + z\,\sigma_3 = \begin{pmatrix} \sin\varphi & e^{-\imath\theta}\cos\varphi \\ e^{\imath\theta}\cos\varphi & -\sin\varphi \end{pmatrix},$$

where $x = \cos\theta\cos\varphi$, $y = \sin\theta\cos\varphi_j$, $z = \sin\varphi$, and σ_1, σ_2 and σ_3 are the Pauli operators. In turn, this operator defines a measurement in the usual way, which we shall also call M for convenience, whose outcome is one of its eigenvalues $+1$ or -1. The azimuthal angle θ represents the equatorial part of the measurement and the elevation angle φ represents its real part. A von Neumann measurement is said to be *equatorial* when its elevation angle $\varphi = 0$ vanishes and it is said to be *in the computational basis* when $\varphi = \pm\pi/2$.

Consider a set of n von Neumann single-qubit measurements M_j, represented by their parameters (θ_j, φ_j), $1 \le j \le n$. This set of operators defines a joint measurement $M = \otimes_{j=1}^{n} M_j$. In turn, this measurement defines a probability distribution p, which we shall call the *GHZ distribution*, on the set $\{-1,+1\}^n$. This distribution corresponds to the probability of all possible outcomes when the n-partite GHZ state $|\Psi_n\rangle = \frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$ is measured according to M.

It is shown in [12, 13], albeit in the usual computer science language in which von Neumann measurements are presented as a unitary transformation followed by a measurement in the computational basis, that the probability $p(b)$ of obtaining $b = (b_1, \ldots, b_n)$ in $\{-1,+1\}^n$ can be decomposed as

$$p(b) = \cos^2\left(\tfrac{\theta}{2}\right) p_1(b) + \sin^2\left(\tfrac{\theta}{2}\right) p_2(b)\,, \text{ where } \theta = \textstyle\sum_{j=1}^{n} \theta_j \text{ and} \tag{1}$$

$$p_1(b) = \frac{1}{2}\big(\mathrm{a}_1(b) + \mathrm{a}_2(b)\big)^2, \qquad p_2(b) = \frac{1}{2}\big(\mathrm{a}_1(b) - \mathrm{a}_2(b)\big)^2, \tag{2}$$

$$\mathrm{a}_1(b) = \prod_{j=1}^{n} \cos\left(\tfrac{1}{2}\left(\varphi_j - \tfrac{\pi}{2} b_j\right)\right), \qquad \mathrm{a}_2(b) = \prod_{j=1}^{n} -\sin\left(\tfrac{1}{2}\left(\varphi_j - \tfrac{\pi}{2} b_j\right)\right). \tag{3}$$

Hence, we see that distribution $p(b)$ is a convex combination of sub-distributions $p_1(b)$ and $p_2(b)$, in which the coefficients $\cos^2(\theta/2)$ and $\sin^2(\theta/2)$ depend only on the equatorial part of the measurements, whereas the sub-distributions depend only on their real part. Furthermore, the squares of a_1 and a_2 are themselves discrete probability distributions.

Sampling p is therefore a matter of sampling a Bernoulli distribution with defining parameter $\cos^2(\theta/2)$ before sampling either p_1 or p_2, whichever is the case. Notice that sampling p_2 reduces to sampling p_1 if, say, we replace φ_1 by $\varphi_1 + 2\pi$. As we shall see, full knowledge of the parameters is not required to sample p *exactly*. We shall see in subsection 2.1 how to sample a Bernoulli distribution with an arbitrary $p \in [0, 1]$ as parameter (not the same p as our probability distribution for GHZ) using a sequence of approximants converging to p and using an expected number of only five unbiased identically independently distributed (i.i.d.) random bits. Subsequently, we shall see in subsection 2.2 how to sample p_1 by modifying von Neumann's rejection algorithm in a way that it uses sequences of approximants and unbiased i.i.d. random bits. For simulating exactly the GHZ distribution, an expected number of $6n + 17$ perfect random bits is sufficient.

2.1 Sampling a Bernoulli distribution

Assume that only a random bit generator is available to sample a given probability distribution and that the parameters that specify this distribution are only accessible as follows: we can ask for any number of bits of each parameter, but will be charged one unit of cost per bit that is revealed. We shall also be charged for each random bit requested from the generator

To warm up to this conundrum, consider the problem of generating a Bernoulli random variable Y with parameter $p \in [0, 1]$. If $U = 0.U_1U_2\ldots$ is the binary expansion of a uniform $[0, 1)$ random variable, i.e. $U_1, U_2, \ldots$ is our source of unbiased independent random bits, and if $p = 0.p_1p_2\ldots$ is the binary expansion of p (in case $p = 1$ we can proceed as if it were $0.p_1p_2\ldots$ with each $p_i = 1$), we compare bits U_i and p_i for $i = 1, 2, \ldots$ until for the first time $U_i \neq p_i$. Then, if $U_i = 0 < p_i = 1$, we return $Y = 1$, and if $U_i = 1 > p_i = 0$, we return $Y = 0$. It is clear that $Y = 1$ if and only if $U < p$. Therefore, Y is Bernoulli(p). The expected number of bits required from p is precisely 2. The expected number of bits needed from our random bit source is also 2.

Now, suppose that the parameter p defining our Bernoulli distribution is given by $p = \cos^2(\theta/2)$, as in the case of our decomposition of the GHZ distribution. None of the parties can know θ precisely since it is distributed as a sum of θ_i's, each of which is known only by one individual party. If we could obtain as many physical bits of p as needed (although the expected number of required bits is as little as 2), we could use the idea given above in order to sample according to this Bernoulli distribution. However, it is not possible in general to know even the first bit of p given any fixed number of bits of the θ_i's. (For instance, if θ is arbitrarily close to $\pi/2$, we need arbitrarily many bits of precision about it before we can tell if the first bit in the binary expansion of $\cos^2(\theta/2)$ is 0 or 1). Nevertheless, we can use *approximations* of p, rather than *truncations*, which in turn can come from approximations of the θ_i's.

▶ **Definition 1.** A *k-bit approximation* of a quantity v is any $\hat{v}$ such that $|v - \hat{v}| \leq 2^{-k}$. A special case of k-bit approximation is the *k-bit truncation* $\hat{v} = \lfloor v2^k \rfloor / 2^k$. For convenience, we sometimes use the shorthands *k-approximation* and *k-truncation*. Note that the value of k corresponds to the number of bits in the fractional part, without limitation on the size of the integer part.

We postpone to Section 3.1 the detail of how these approximations can be obtained in a distributed setting. For the moment, assume that, for any k, we can obtain $p(k)$ so that $|p(k)-p| \leq 1/2^k$. Then, setting $U(k) = 0.U_1 \ldots U_k$, we have that $U \leq p$ if $U(k) \leq p(k) - 2/2^k$ whereas $U \geq p$ if $U(k) \geq p(k) + 1/2^k$. Thus, one can check if $U < p$ by generating only as many bits of U and increasingly good approximations of p as needed. These ideas are formalized in Algorithm 1. It is elementary to verify that the Y generated by this algorithm is Bernoulli(p) because $\mathbf{P}\{U < p\} = p$ if U is a continuous uniform random variable on $(0, 1)$.

Algorithm 1 Sampling a Bernoulli random variable with approximate defining parameter

```
Set $k \leftarrow 1$
Set $U(0) \leftarrow 0$
repeat forever
    Generate an i.i.d. unbiased bit $U_k$
    Compute $U(k) \leftarrow U(k-1) + U_k/2^k$ {hence $U(k) = 0.U_1 \ldots U_k$}
    Obtain $p(k)$ so that $|p(k) - p| \leq 1/2^k$
    if $U(k) \leq p(k) - 2/2^k$ then
        return $Y = 1$
    else if $U(k) \geq p(k) + 1/2^k$ then
        return $Y = 0$
    else
        $k \leftarrow k + 1$
    end if
end repeat
```

The number of iterations before Algorithm 1 returns a value, which is also its required number of independent unbiased random bits, is a random variable, say K. We have seen above that $\mathbf{E}\{K\}$, the expected value of K, would be exactly 2 if we could generate arbitrarily precise truncations of p. But since we can only obtain arbitrarily precise approximations instead, which is why we needed Algorithm 1 in the first place, we shall have to pay the price of a small increase in $\mathbf{E}\{K\}$.

$$\mathbf{P}\{K > k\} \leq \mathbf{P}\left\{|U(k) - p(k)| \leq \frac{2}{2^k}\right\} \leq \mathbf{P}\left\{|U - p| \leq \frac{4}{2^k}\right\} \leq \frac{8}{2^k}.$$

Therefore,

$$\mathbf{E}\{K\} = \sum_{k=0}^{\infty} \mathbf{P}\{K > k\} \leq \sum_{k=0}^{\infty} \min\left(1, \frac{8}{2^k}\right) = 5.$$

2.2 Sampling p_1 (or p_2) in the random bit model

As mentioned already, it suffices to concentrate on p_1 since one can sample p_2 in exactly the same way provided one of the angles φ_i is replaced by $\varphi_i + 2\pi$: this introduces the required minus sign in front of a_2 to transform p_1 into p_2. Let us define

$$\alpha_j = \cos\bigl(\tfrac{1}{2}\bigl(\varphi_j - \tfrac{\pi}{2}\bigr)\bigr) = \sin\bigl(\tfrac{1}{2}\bigl(\varphi_j + \tfrac{\pi}{2}\bigr)\bigr) \quad \text{and} \quad \beta_j = \cos\bigl(\tfrac{1}{2}\bigl(\varphi_j + \tfrac{\pi}{2}\bigr)\bigr) = -\sin\bigl(\tfrac{1}{2}\bigl(\varphi_j - \tfrac{\pi}{2}\bigr)\bigr)\,. \tag{4}$$

Consider n Rademacher[1] random variables B_j that take value -1 with probability β_j^2 and $+1$ with complementary probability α_j^2. The random vector with independent components given by $(B_1, \ldots, B_n)$ is distributed according to

$$q_1(b) \overset{\text{def}}{=} \prod_{j \in F_b} \beta_j^2 \prod_{j \in G_b} \alpha_j^2\,,$$

where $F_b = \{j \mid b_j = -1\}$ and $G_b = \{j \mid b_j = +1\}$ for all $b = (b_1, \ldots, b_n) \in \{-1, +1\}^n$. It is easy to verify that $q_1(b) = \mathrm{a}_1^2(b)$ for all b, where a_1 is given in Equation (3). Similarly, the random vector with independent components given by $(-B_1, \ldots, -B_n)$ is distributed according to

$$q_2(b) \overset{\text{def}}{=} \prod_{j \in F_b} \alpha_j^2 \prod_{j \in G_b} \beta_j^2 = \mathrm{a}_2^2(b)\,.$$

The key observation is that both q_1 and q_2 can be sampled without any needs for communication because each party j knows his own parameters α_j^2 and β_j^2, which is sufficient to draw independently according to local Rademacher random variable B_j or $-B_j$. Moreover, a single unbiased independent random bit s drawn by a designated party suffices to sample collectively from distribution $q = \frac{q_1 + q_2}{2}$, provided this bit is transmitted to all parties: everybody samples according to q_1 if $s = 0$ or to q_2 if $s = 1$. Now, It follows from Equation (2) that $p_1(b) + p_2(b) = \mathrm{a}_1^2(b) + \mathrm{a}_2^2(b) = q_1(b) + q_2(b)$ for all $b \in \{-1, +1\}^n$, and therefore $p_1(b) \le q_1(b) + q_2(b) = 2q(b)$.

The relevance of all these observations is that we can apply von Neumann's rejection algorithm [18] to sample $p_1(b)$ since it is bounded by a small constant (2) times an easy-to-draw probability distribution (q). For the moment, we assume once again the availability of a continuous uniform random generator, which we shall later replace by a source of unbiased independent random bits. We also assume for the moment that we can compute the α_i's, $p_1(b)$, $q_1(b)$ and $q_2(b)$ exactly. This gives rise to Algorithm 2.

Algorithm 2 Sampling p_1 using von Neumann's rejection algorithm

```
repeat
    Generate $U$ uniformly on $[0, 1)$
    Generate independent Rademacher random variables $B_1, \ldots, B_n$
        with parameters $\alpha_1^2, \ldots, \alpha_n^2$
    Generate an unbiased independent random bit $S$
    if $S = 1$ then
        set $B \leftarrow (B_1, \ldots, B_n)$
    else
        set $B \leftarrow (-B_1, \ldots, -B_n)$
    end if
until $(q_1(B) + q_2(B))\, U \le p_1(B)$
```

By the general principle of von Neumann's rejection algorithm, probability distribution p_1 is successfully sampled after an expected number of 2 iterations round the loop because

[1] A Rademacher random variable is like a Bernoulli, except that it takes value ±1 rather than 0 or 1.

$p_1(b) \le 2q(b)$ for all $b \in \{-1,+1\}^n$. Within one iteration, 2 expected independent unbiased random bits suffice to generate each of the n Rademacher random variables by a process similar to what is explained in the second paragraph of Section 2.1. Hence an expected total of $2n+1$ random bits are needed each time round the loop for an expected grand total of $4n+2$ bits to sample p_1. But of course, this does not take account of the (apparent) need to generate continuous uniform $[0,1)$ random variable U. It follows that the expected total amount of work required by Algorithm 2 is $O(n)$, provided we count infinite real arithmetic at unit cost. Furthermore, the time taken by this algorithm, divided by n, is stochastically smaller than a geometric random variable with constant mean, so its tail is exponentially decreasing.

Now, we modify and adapt this algorithm to eliminate the need for the continuous uniform U (and hence its generation), which is not allowed in the random *bit* model. Furthermore, we eliminate the need for infinite real arithmetic and for the exact values of $q_1(B)$, $q_2(B)$ and $p_1(B)$, which would be impossible to obtain in our distributed setting since the parameters needed to compute these values are scattered among all parties, and replace them with approximations—we postpone to Section 3.2 the issue of how these approximations can be computed. (On the other hand, arbitrarily precise values of the α_i's *are* available to generate independent Rademacher random variables with these parameters because each party will be individually responsible to generate his own Rademacher.)

In each iteration of Algorithm 2, we generated a pair (U,B). However, we did not really need U: we merely needed to generate a Bernoulli random variable Y for which

$$\mathbf{P}\{Y=1\} = \mathbf{P}\left\{(q_1(B)+q_2(B))\,U \le p_1(B)\right\}.$$

For this, we adapt the method developed for Algorithm 1. Again, we denote by $U(k)$ the k-bit truncation of U, so that $U(k) < U < U(k)+2^{-k}$, except with probability 0. Furthermore, we use L_k (L for *left*) and R_k (R for *right*) to denote k-bit approximations of $q_1(B)+q_2(B)$ and $p_1(B)$, respectively, so that $|L_k - \big(q_1(B)+q_2(B)\big)| \le 2^{-k}$ and $|R_k - p_1(B)| \le 2^{-k}$. Then using ε_k to denote arbitrary real numbers in the interval $(-1,1)$,

$$\begin{aligned} |U(k)L_k - U(q_1(B)+q_2(B))| &= \left|U(k)L_k - U\Big(L_k + \frac{\varepsilon_k}{2^k}\Big)\right| \\ &= \left|(U(k)-U)L_k - \frac{U\varepsilon_k}{2^k}\right| \;\le\; \frac{L_k}{2^k} + \frac{1}{2^k} \;\le\; \frac{3}{2^k}\,. \end{aligned}$$

Similarly, $|R_k - p_1(B)| \le \dfrac{1}{2^k}\,.$

Thus, we know that $Y=1$ whenever $U(k)L_k + 3/2^k < R_k - 1/2^k$, whereas $Y=0$ whenever $U(k)L_k - 3/2^k > R_k + 1/2^k$. Otherwise, we are in the uncertainty zone and we need more bits of U, $q_1(B)+q_2(B)$ and $p_1(B)$ before we can decide on the value of Y. This is formalized in Algorithm 3 (on next page).

It follows from the above discussion that this algorithm can be used to sample random variable Y, which is used as terminating condition in Algorithm 2, in order to eliminate the need for the generation of a continuous uniform random variable $U \in [0,1)$ and for the precise values of $q_1(B)$, $q_2(B)$ and $p_1(B)$. Since $L_k \to q_1(B)+q_2(B)$ and $R_k \to p_1(B)$ as $k \to \infty$, Algorithm 3 halts with probability 1. Let K be a random variable corresponding to the value of k upon exiting from the **repeat forever** loop in the algorithm, which is the number of times round the loop and hence the number of bits needed from U and the precision in $q_1(B)+q_2(B)$ and $p_1(B)$ required in order to sample correctly Bernoulli random variable Y. Next, we calculate an upper-bound on $\mathbf{E}\{K\}$, the expected value of K.

Algorithm 3 Generator for the stopping condition in Algorithm 2

```
Note: $B \in \{-1,+1\}^n$ is given to the algorithm, generated according to $\frac{q_1+q_2}{2}$
Set $k \leftarrow 1$
Set $U(0) \leftarrow 0$
repeat forever
    Generate an i.i.d. unbiased bit $U_k$
    Compute $U(k) \leftarrow U(k-1) + U_k/2^k$ {hence $U(k) = 0.U_1 \ldots U_k$}
    Compute $L_k$ and $R_k$ from $B$
    if $U(k)\,L_k - R_k < -\frac{4}{2^k}$ then
        return $Y = 1$
    else if $U(k)\,L_k - R_k > \frac{4}{2^k}$ then
        return $Y = 0$
    else
        $k \leftarrow k+1$
    end if
end repeat
```

If the algorithm has not yet halted after having processed $U(k)$, L_k and R_k, then we know that

$$\begin{aligned}
&|U(q_1(B)+q_2(B)) - p_1(B)| \\
&\quad = \Big|\big(U(q_1(B)+q_2(B)) - U(k)L_k\big) + \big(R_k - p_1(B)\big) + \big(-R_k + U(k)L_k\big)\Big| \\
&\quad \leq |U(q_1(B)+q_2(B)) - U(k)L_k| + |R_k - p_1(B)| + |R_k - U(k)L_k| \\
&\quad \leq \frac{3}{2^k} + \frac{1}{2^k} + \frac{4}{2^k} \quad = \quad \frac{8}{2^k}\,.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbf{P}\{K > k \mid B\} &\leq \mathbf{P}\{|U(q_1(B)+q_2(B)) - p_1(B)| \leq 8/2^k \mid B\} \\
&= \mathbf{P}\left\{U \in \left(\frac{p_1(B)}{2q(B)} - \frac{1}{2}\frac{8}{2^k}\frac{1}{q(B)}\ ,\ \frac{p_1(B)}{2q(B)} + \frac{1}{2}\frac{8}{2^k}\frac{1}{q(B)}\right)\right\} \\
&\leq \frac{8}{2^k}\frac{1}{q(B)}\,.
\end{aligned}$$

Thus, using k_0 to denote $\left\lceil 3 + \log_2\left(\frac{1}{q(B)}\right)\right\rceil$,

$$\begin{aligned}
\mathbf{E}\{K \mid B\} &= \sum_{k=0}^{\infty} \mathbf{P}\{K > k \mid B\} \\
&\leq \sum_{k=0}^{\infty} \min\left(1, \frac{8}{2^k q(B)}\right) \\
&\leq \sum_{k<k_0} 1 + \sum_{k \geq k_0} \frac{8}{2^k q(B)} \\
&\leq 5 + \log_2\left(\frac{1}{q(B)}\right) \qquad \text{(this step requires a messy calculation).}
\end{aligned}$$

Now, we uncondition in order to conclude:

$$\begin{aligned}\mathbf{E}\{K\} &\leq 5+\sum_{b\in\{-1,+1\}^n} q(b)\log_2\left(\frac{1}{q(b)}\right)\\ &= H(q)+5 \qquad (5)\\ &\leq n+5\,, \qquad (6)\end{aligned}$$

where $H(q)$ denote the entropy of distribution $q=\frac{q_1+q_2}{2}$.

3 Communication complexity of sampling

In this section, we consider the case in which the sampler of the previous section no longer has full knowledge of the GHZ distribution to be simulated. The sampler, whom we call *the leader* in a distributed setting, has to communicate through classical channels in order to obtain partial knowledge of the parameters belonging to the other parties. Partial knowledge results in approximation of the parameters involved in sampling the GHZ distribution, but, as we saw in the previous section, we know how to sample *exactly* in the random bit model using such approximations.

3.1 Sampling a Bernoulli distribution whose parameter is distributed

In order to sample the GHZ distribution, we know from Section 2 that we must first sample the Bernoulli distribution with parameter $\cos^2(\theta/2)$, where $\theta=\sum_{j=1}^n\theta_j$. Let us say that the leader is party number 1. Since he knows only θ_1, he must communicate with the other parties to obtain partial knowledge about θ_i for $i\geq 2$. The problem of sampling a Bernoulli distribution with probability $\cos^2(\theta/2)$ reduces to learning the sum θ with sufficient precision in order to use Algorithm 1.

The problem of computing a k-bit approximation of $\cos^2(\theta/2)=\cos^2\bigl(\sum_{i=1}^n\theta_i/2\bigr)$ is relatively easy. Define $\vartheta=\theta/2$ and $\vartheta_i=\theta_i/2$ for each i. If the leader obtains an ℓ-bit approximation $\hat{\vartheta}_i$ of each ϑ_i, $i\geq 2$, and if we define $\hat{\vartheta}=\sum_{i=1}^n\hat{\vartheta}_i$, we need to find the value of ℓ for which $\cos^2(\hat{\vartheta})$ is a k-bit approximation of $\cos^2(\vartheta)$. It is an elementary exercise in Taylor series expansion to verify that $|\cos^2(\vartheta)-\cos^2(\hat{\vartheta})|\leq n/2^\ell$. Hence, it suffices to choose $\ell=k+\lceil\log_2 n\rceil$ in order to conclude as required that $|\cos^2(\vartheta)-\cos^2(\hat{\vartheta})|\leq 2^{-k}$. Taking into account the integer part of each ϑ_i, which must also be communicated, and remembering that $0\leq\vartheta_i\leq 2\pi$ since it is an angle [2], the required number of communicated bits in the sequential model is therefore $(n-1)(\ell+3)=(n-1)\bigl(3+k+\lceil\log_2 n\rceil\bigr)$, which is $O(kn+n\log n)$. In our case, the expected value of k is bounded by 5 (see the analysis of the Bernoulli sampling Section 2.1), so that this operation requires an expected communication of $O(n\log n)$ bits.

3.2 Approximating a product of bounded numbers

Once the leader has produced a bit Z with probability $\cos^2(\theta/2)$, he samples either p_1 or p_2, depending on whether he got $Z=0$ or $Z=1$. The problem of sampling p_2 reduces to sampling p_1 if the leader replaces his own φ_1 with $\varphi_1+2\pi$; thus we concentrate on sampling p_1. Of course, the leader does not know φ_i for $i\geq 2$. This problem reduces

[2] Actually, $0\leq\vartheta_i\leq\pi$ since ϑ_i is a *half* angle and one fewer bit is needed to communicate its integer part, but we prefer to consider here the more general case of approximating the cosine square of a sum of arbitrary angles.

to learning with sufficient precision the products $\mathrm{a}_1(B) = \prod_{j=1}^{n} \cos\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$ and $\mathrm{a}_2(B) = \prod_{j=1}^{n} -\sin\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$, given that the B_j's are independent Rademacher distributions with parameters α_j^2, $1 \leq i \leq n$ defined in Equation (4). Once these products are known with $k+2$ bits of precision, the left and right k-bit approximations L_k and R_k are easily computed, which allows us to run the modified von Neumann's rejection algorithm from Section 2.2.

In this section, we explain how to compute a k-bit approximation to $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$ at an expected communication cost of $O(kn + n\log n)$ bits. For our specific application of simulating the GHZ distribution, we proved at the end of Section 2.2 (Equation 6) that the expected value of k is bounded by $n+5$. It follows that an expected cost of $O(n^2)$ bits suffices to carry out the simulation.

Given $B = (B_1, \ldots, B_n)$ with the B_i's distributed according to non-identical independent Rademachers with parameter $\cos^2\left(\frac{1}{2}\left(\varphi_i - \frac{\pi}{2}\right)\right)$ or $\cos^2\left(\frac{1}{2}\left(\varphi_i + \frac{\pi}{2}\right)\right)$, we need to compute k-bit approximations of $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$. We use c_j and s_j to denote $\cos\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$ and $-\sin\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$, respectively, as well as $\hat{c}_j$ and $\hat{s}_j$ to denote their respective ℓ-truncations. We need to determine ℓ such that the products $\prod_{j=1}^{n} \hat{c}_j$ and $\prod_{j=1}^{n} \hat{s}_j$ are k-approximations of $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$, respectively. *Notice that each party knows exactly his own c_j and s_j, and hence $\hat{c}_j$ and $\hat{s}_j$ can be transmitted directly to the leader, rather than approximations of the φ_i's.* For each c_j, there exists $\varepsilon_j \in [-1,1]$ such that $c_j = \hat{c}_j + \frac{\varepsilon_j}{2^\ell}$; thus, using I to denote $\{1,2,\ldots,n\}$, we have

$$\prod_{j=1}^{n} c_j = \sum_{A\in\mathcal{P}(I)} \prod_{j\in A} \hat{c}_j \prod_{j\notin A} \frac{\varepsilon_j}{2^\ell} = \prod_{j=1}^{n} \hat{c}_j + \sum_{A\in\mathcal{P}(I)\setminus I} \prod_{j\in A} \hat{c}_j \prod_{j\notin A} \frac{\varepsilon_j}{2^\ell}$$

and hence we can bound the error as follows:

$$\left|\prod_{j=1}^{n} c_j - \prod_{j=1}^{n} \hat{c}_j\right| \leq \sum_{j=1}^{n}\left(\binom{n}{j}\frac{1}{2^{j\ell}}\right) - 1 = \left(1 + \frac{1}{2^\ell}\right)^n - 1\,.$$

Setting $\ell = \left\lceil -\log_2\left(\left(1+2^{-k}\right)^{1/n} - 1\right)\right\rceil \leq k + \lceil \log_2 n \rceil + 2$, we have

$$\left|\prod_{j=1}^{n} c_j - \prod_{j=1}^{n} \hat{c}_j\right| \leq \frac{1}{2^k}\,.$$

Taking account of the need to transmit the ℓ-truncations to both c_j and s_j, which consists of the sign of these numbers in addition to the first ℓ bits of their binary expansion, the expected communication cost is $2(n-1)(\ell+1)$ bits, which indeed is $O(kn + n\log n)$.

3.3 Protocol for sampling the GHZ distribution

We are finally ready to glue all the pieces together into Algorithm 4 (on next page), which samples exactly the GHZ distribution under arbitrary von Neumann measurements, thus solving our conundrum. Its correctness is proven below, and it is shown that the expected amount of randomness used in this process is upper-bounded by $6n + 17$ bits and an expected $O(n^2)$ bits of communication suffice to complete the task. Variations are discussed subsequently.

Correctness of the protocol: The part occurring before the first "repeat" (line 5) samples a Bernoulli with parameter $\cos^2\left(\sum_{i=1}^{n} \theta_i/2\right)$, which allows the leader to decide whether to

Algorithm 4 Complete protocol for sampling the GHZ distribution in the sequential model

```
The leader, who is party number 1, communicates with the other parties in order to
obtain increasingly precise approximations of $\theta = \sum_{i=1}^{n} \theta_i$ until he can sample random
bit $Z$ according to *exact* Bernoulli random distribution with parameter $\cos^2(\theta/2)$
if $Z = 1$ then
    The leader adds $2\pi$ to his own $\varphi$-parameter i.e. $\varphi_1 \leftarrow \varphi_1 + 2\pi$
    {to sample $p_2$ rather than $p_1$}
end if
{Now entering the modified von Neumann's "distributed" sampler for $p_1$}
repeat
    The leader generates a fair random bit $S$ and broadcasts it to the other parties
    {The bit $S$ determines whether to sample $q_1$ or $q_2$}
    Locally, each party $j$ generates a random $B_j \in \{-1,+1\}$ according to an independent
    Rademacher distribution so that $B_j = +1$ with probability $\cos^2\bigl(\frac{1}{2}\bigl(\varphi_j - \frac{\pi}{2}\bigr)\bigr)$
    {Random variable $B = (B_1, \ldots, B_n)$ is now sampled according to $q_1$}
    if $S = 1$ then
        Each party does $B_j \leftarrow -B_j$
        {In this case, random variable $B = (B_1, \ldots, B_n)$ is now sampled according to $q_2$}
    end if
    {Random variable $B = (B_1, \ldots, B_n)$ is sampled according to $q = \frac{q_1+q_2}{2}$}
    {The leader starts talking with the other parties to decide whether to accept $B$}
    Each party computes $c_j = \cos\bigl(\frac{1}{2}\bigl(\varphi_j - \frac{\pi}{2}B_j\bigr)\bigr)$ and $s_j = -\sin\bigl(\frac{1}{2}\bigl(\varphi_j - \frac{\pi}{2}B_j\bigr)\bigr)$
    The leader sets $k \leftarrow 1$
    The leader sets $U(0) \leftarrow 0$
    repeat forever
        The leader generates an i.i.d. unbiased bit $U_k$
        The leader computes $U(k) \leftarrow U(k-1) + U_k/2^k$ {hence $U(k) = 0.U_1 \ldots U_k$}
        The leader requests $(k + 3 + \lceil \log_2 n \rceil)$-approx. of $c_j$ and $s_j$ from each party $j \geq 2$
        The leader uses this information to compute $(k+2)$-approximations of $\mathrm{a}_1(B)$ and
        $\mathrm{a}_2(B)$, which are used to compute $k$-bit approximations $L_k$ of $\mathrm{a}_1^2(B) + \mathrm{a}_2^2(B)$ and
        $R_k$ of $p_1(B)$
        if $U(k)L_k - R_k < -\frac{4}{2^k}$ then
            Set $Y \leftarrow 1$ and break from the repeat forever loop. {Vector $B$ is accepted}
        else if $U(k)L_k - R_k > \frac{4}{2^k}$ then
            Set $Y \leftarrow 0$ and break from the repeat forever loop. {Vector $B$ is rejected}
        else
            Set $k \leftarrow k + 1$ and continue the repeat forever loop
            {The leader does not yet have enough information to decide whether to accept or
            reject $B$. Therefore, he needs to compute the next bit of $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$.
            For this, he needs more information from all the other parties.}
        end if
    end repeat
until $Y = 1$ {accepting}
The leader informs all the other parties that the simulation is complete and, therefore,
that the time has come for each party $j$ (including the leader himself) to output his
current value of $B_j$
```

sample B according to p_1 (by leaving his φ_1 unchanged) or according to p_2 (by adding 2π to his φ_1). Notice that the leader does not have to inform the other parties of this decision since they do not need to know if the sampling will be done according to p_1 or p_2. In Section 3.1, we showed how to sample exactly a Bernoulli with parameter $\cos^2\bigl(\sum_{i=1}^n \theta_i/2\bigr)$ when the θ_i's are not known to the leader for $i \geq 2$.

The part within the outer "repeat" loop (lines 5 to 27) is essentially von Neumann's rejection algorithm, which has been adapted and modified to work in a distributed scenario. The leader must first know which of q_1 or q_2 to sample. For this purpose, he generates an unbiased random bit S and broadcasts it to the other parties. Sampling either q_1 or q_2 can now be done locally and independently by each party j, yielding a tentative $B_j \in \{-1,+1\}$. The parties will output these B_j's only at the end, provided this round is not rejected. Now, each party uses his B_j to compute locally $c_j = \cos\bigl(\frac{1}{2}\bigl(\varphi_j - \frac{\pi}{2}B_j\bigr)\bigr)$ and $s_j = -\sin\bigl(\frac{1}{2}\bigl(\varphi_j - \frac{\pi}{2}B_j\bigr)\bigr)$, which will be sent bit by bit to the leader upon request, thus allowing him to compute increasingly precise approximations L_k and R_k of $q_1(B)+q_2(B)$ and $p_1(B)$, respectively. These values are used to determine whether a decision can be made to accept or reject this particular B, or whether more information is needed to make this decision. As shown at the end of Section 2.2 (Equation 6), the expected number of bits needed in L_k and R_k before we can break out of the "repeat forever" loop is $k \leq n+5$. At that point, flag Y tells the leader whether or not this was a successful run of von Neumann's rejection algorithm. If $Y=0$, the entire process has to be restarted from scratch, except for the initial Bernoulli sampling, at line 6. On the other hand, once the leader gets $Y=1$, he can finally tell the other parties that they can output their B_j's because, according to von Neumann's rejection algorithm, this signals that the vector $(B_1,\ldots,B_n)$ is distributed according to p_1 (or p_2, depending on the initial Bernoulli). Also according to von Neumann's rejection algorithm, we have an expectation of $C=2$ rounds of the outer "repeat" loop before we can thus conclude successfully.

Expected communication cost and number of random coins: The expected amount of randomness used in this process is upper-bounded by $6n+17$ bits. This is calculated as follows: the expected number of bits for sampling Bernoulli Z is bounded by 5. This is followed by an expectation of $C=2$ rounds of von Neumann's rejection algorithm (the outer "repeat" loop). In each of these rounds, we need 1 bit for S and expect 2 bits for each of the B_j's (hence $2n$ in total), before entering the "repeat forever" loop. The expected number of times round this loop is bounded by $n+5$, and one more random bit U_k is needed each time. Putting it all together, the expected number of random bits is bounded by $5+2(1+2n+(n+5)) = 6n+17$.

The expected amount of communication is dominated by the leader's need to obtain increasingly accurate approximations of c_j and s_j from all other parties at line 17 in order to compute increasingly accurate approximations of L_k and R_k, which he needs in order to decide whether or not to break from the "repeat forever" loop and, in such case, whether or not to accept B as final output. On the k-th time round the loop, the leader needs $k+3+\lceil \log_2 n \rceil$ bits of precision plus one bit of sign about each c_j and s_j, $j \geq 2$ (in addition to having full knowledge about his own c_1 and s_1, of course). This would be very expensive if all those bits had to be resent each time round the loop, with increasing values of k. Fortunately, this process works well if the parties send *truncations* of these values to the leader, because each truncation simply adds one bit of precision to the previous one. Hence, it suffices for the leader to request $2(5+\lceil \log_2 n \rceil)$ bits from each other party at the onset, when $k=1$, and only two additional bits per party are needed afterwards for each subsequent trip round the loop

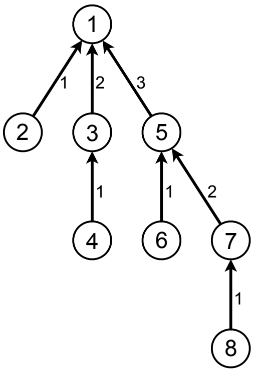

Figure 1 Binomial tree structure defining the parallel model.

(one for c_j and one for s_j). All counted, a total of $2(n-1)(k+5+\lceil\log_2 n\rceil)$ bits will have been requested from all other parties by the time we have gone through the "repeat forever" loop k times. Since the expected value of k upon exiting this loop is bounded by $n+5$, the expected number of bits that have to be communicated to the leader to complete von Neumann's rejection algorithm (lines 5 to 27) is bounded by $2(n-1)((n+5)+5+\lceil\log_2 n\rceil)$. This is $O(n^2)$ expected bits of communication. The additional amount of communication required to sample Bernoulli Z at step 1 (which is $(n-1)(5+\log_2 n)$ bits) and for the leader to broadcast to all parties the value of S, as well as synchronization bits by which he needs to inform the other parties of success or failure each time round the loop is negligible. All counted, Algorithm 4 needs $O(n)$ bits of randomness and $O(n^2)$ bits of communication in order to sample exactly the GHZ distribution under arbitrary von Neumann measurements.

Using Equation (5) rather than Equation (6), we shall show in the final journal version of this work that Algorithm 4 needs only $O(n\log n)$ bits of communication in order to sample exactly the GHZ distribution under computational-basis von Neumann measurements. Of course, $O(n)$ bits of communication would suffice, even in the worst case, if we knew ahead of time that all measurements are in the computational basis, but our protocol works seamlessly with $O(n\log n)$ expected bits of communication even if the measurements are not *exactly* in the computational basis, and if up to $O(\log n)$ of the measurements are arbitrary.

3.4 Variations on the theme

We can modify Algorithm 4 in a variety of ways to improve different parameters at the expense of others. Here, we mention briefly three of these variations: the parallel model, bounding the number of rounds, and the simulation of equatorial measurements.

The parallel model: Until now, we have considered only a *sequential model* of communication, in which the leader has a direct channel with everyone else. In this model, communication takes place sequentially because the leader cannot listen to everyone at the same time. However, it is legitimate to consider a *parallel model*, in which arbitrary many pairs of parties can communicate simultaneously. In this model, any number of bits can be sent and received in the same time step, provided no party has to send or receive more than one bit at any given time. If we make the parties communicate with one another following the binomial tree structure shown in Fig. 1, with the leader at the root, we shall show in the final journal version of this work that the exact simulation of the GHZ distribution under arbitrary independent von Neumann measurements can be accomplished within $O(n \log n)$ expected parallel time. The expected total number of bits communicated with this approach is slightly greater than with Algorithm 4, but it remains $O(n^2)$.

Reducing the number of rounds: Algorithm 4 is efficient in terms of the number of bits of randomness as well as the number of bits of communication, but it requires an expected $O(n)$ rounds, in which the leader and all other parties take turn at sending messages. This could be prohibitive if they are far apart and their purpose is to try to convince examiners that they are actually using true entanglement and quantum processes to produce their joint outputs, because it would prevent them from responding quickly enough to be credible. We leave it for the reader as an exercise to verify that if we change line 24 of Algorithm 4 from "$k \leftarrow k+1$" to "$k \leftarrow 2k$", the expected number of rounds is decreased from $O(n)$ to $O(\log n)$. If in addition we start with "$k \leftarrow n$" instead of "$k \leftarrow 1$" at line 12, the expected number of rounds becomes a constant. (Alternatively, we could start with "$k \leftarrow n$" at line 12 and step with "$k \leftarrow k+n$" at line 24.)

Equatorial measurements: Recall that equatorial measurements are those for which $\varphi_j = 0$ for each party j. In this case, the leader can sample according to p_1 or p_2, without any help or communication from the other parties, since he has complete knowledge of their vanished elevation angles. Therefore, he can run steps 5 to 27 of Algorithm 4 all by himself! However, he needs to communicate in step 1 of Algorithm 4 in order to know from which of p_1 or p_2 to sample. The only remaining need for communication occurs in step 28, which has to be modified from "The leader informs all the other parties that the simulation is complete" to "The leader informs all the other parties of which value of $B_j \in \{-1,+1\}$ he has chosen for them".

Only step 1 requires significant communication since the new step 28 needs only the transmission of $n-1$ bits. We have already seen at the end of Section 3.1 that step 1, which is a distributed version of Algorithm 1, requires an expected communication of $O(n \log n)$ bits in the sequential model. This is therefore the complexity of our simulation, which is an improvement over the previously best technique known to simulate the GHZ distribution under arbitrary equatorial von Neumann measurements [6], which required an expectation of $O(n^2)$ bits of communication.

A more elegant protocol can be obtained if we adapt Equations (1), (2) and (3), which were given at the beginning of Section 2 to define the GHZ probability distribution $p(b)$ for $b \in \{-1,+1\}^n$, to the special case of equatorial measurements. Because all the elevation angles φ_j vanish, these formulas reduce to

$$p(b) = \begin{cases} 2^{1-n}\cos^2\left(\frac{\theta}{2}\right) & \text{if } b \in X \\ 2^{1-n}\sin^2\left(\frac{\theta}{2}\right) & \text{if } b \notin X \end{cases} \qquad \text{where } X = \left\{ b \in \{-1,+1\}^n \;\middle|\; \prod_{j=1}^{n} b_j = +1 \right\}.$$

Now, each party j other than the leader can simply choose an independent unbiased Rademacher $b_j \in \{-1,+1\}$ as final output, without any consideration of his own input θ_j nor communication with anyone else, and inform the leader of this choice. It simply remains for the leader to choose his own b_1 in order to make $\prod_{j=1}^{n} b_j$ equal to $+1$ with probability $\cos^2(\theta/2)$ or -1 with probability $\sin^2(\theta/2)$. For this, we still need step 1 from Algorithm 4, which requires an expected communication of $O(n \log n)$ bits. We shall show in the final journal version of this work that this process can be achieved with only $O(\log^2 n)$ expected time steps in the parallel model of communication.

4 Discussion and open problems

We have addressed the problem of simulating the effect of arbitrary independent von Neumann measurements on the qubits forming the general GHZ state $\frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$ distributed among n parties. Rather than doing the actual quantum measurements, the parties must sample the exact GHZ probability distribution by purely classical means, which necessarily requires communication in view of Bell's theorem. Our main objective was to find a protocol that solves this conundrum with a finite amount of expected communication, which had only been known previously to be possible when the von Neumann measurements are restricted to being equatorial (a severe limitation indeed). Our solution needs only $O(n^2)$ bits of communication, which can be dispatched in $O(n \log n)$ time if bits can be sent in parallel according to a realistic scenario in which nobody has to send or receive more than one bit in any given step. We also improved on the former art in the case of equatorial measurements, with $O(n \log n)$ bits of communication and $O(\log^2 n)$ parallel time.

Knuth and Yao [16] initiated the study of the complexity of generating random integers (or bit strings) with a given probability distribution $p(b)$, assuming only the availability of a source of unbiased identically independently distributed random bits. They showed that any sampling algorithm must use an expected number of bits at least equal to the entropy $\sum_b p(b) \log_2(1/p(b))$ of the distribution, and that the best algorithm does not need more than two additional bits. For further results on the bit model in random variate generation, see Ref. [9, Chap. XIV].

The GHZ distribution has an entropy no larger than n, and therefore Knuth and Yao have shown that it could be sampled with no more than $n+2$ expected random bits if all the parameters were concentrated in a single place [16]. Even though we have studied the problem of sampling this distribution in a setting in which the defining parameters (here the description of the von Neumann measurements) are distributed among n parties, and despite the fact that our main purpose was to minimize communication between these parties, we were able to succeed with $6n+17$ expected random bits, which is just above six times the bound of Knuth and Yao. The amount of randomness required by our protocols does not depend significantly on the actual measurements they have to simulate. However, some sets of measurements entail a probability distribution $p(B)$ whose entropy $H(p)$ is much smaller than n. In the extreme case of having all measurements in the computational basis, $H(p)$ is a single bit! Can there be protocols that succeed with as few as $H(p)+2$ expected random bits, thus meeting the bound of Knuth and Yao, or failing this as few as $O(H(p))$ expected random bits, no matter how small $H(p)$ is for the given set of von Neumann measurements? Notice that all the protocols presented here require $\Omega(n)$ random bits since they ask each party to sample independently at least once a Rademacher random variable, a hurdle that can only be alleviated in the case of measurements in the computational basis.

Are our protocols optimal in terms of the required amount of communication? Could we simulate arbitrary von Neumann measurements as efficiently as the case of equatorial measurements, i.e. with $O(n \log n)$ bits of communication? We leave this as open question, but point out that Broadbent, Chouha and Tapp have proved an $\Omega(n \log n)$ lower bound on the *worst case* communication complexity of simulating measurements on n-partite GHZ states [7], a result that holds even for equatorial measurements, and even under the promise that $\cos \sum_{i=1}^{n} \theta_i = \pm 1$ [15].

As a recent development, which we shall formalize in the final journal version of this work, we have discovered how to simulate more general multipartite states than the GHZ state. For instance, we know how to simulate the so-called W state $\frac{1}{\sqrt{3}}|100\rangle + \frac{1}{\sqrt{3}}|010\rangle + \frac{1}{\sqrt{3}}|001\rangle$

and more generally

$$W_n = \frac{1}{\sqrt{n}} \left(|10^{n-1}\rangle + |010^{n-2}\rangle + |0010^{n-3}\rangle + \cdots + |0^{n-1}1\rangle \right)$$

with $O(n^3)$ expected bits of communication and the need of only $O(n^2)$ expected unbiased independent random bits. However, we leave for further research the problem of simulating arbitrary positive-operator-valued measurements (POVMs) on the single-qubit shares of GHZ states (or on more general multipartite states), as well as the problem of simulating multipartite entanglement other than equatorial von Neumann measurements on the tripartite GHZ state [3] with *worst-case* bounded classical communication.

Acknowledgements. We wish to thank Marc Kaplan and Nicolas Gisin for stimulating discussions about the simulation of entanglement. Furthermore, Marc has carefully read Ref. [12], in which the decomposition of the GHZ distribution as a convex combination of two sub-distributions was first accomplished, and he has pointed out that the lower bound from Ref. [7] applies even in the case of equatorial measurements. Alain Tapp pointed out that the entropy of the GHZ distribution can be as small as one bit in the case of measurements in the computational basis. We are also grateful for the useful feedback provided by the TQC reviewers.

G. B. is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chair program, the Canadian Institute for Advanced Research (CIFAR), and the Institute for Theoretical Studies (ITS) at the ETH Zürich. L. D. is supported in part by NSERC and Fonds de recherche du Québec – Nature et technologies (FRQNT).

References

1 J.-D. Bancal, C. Branciard and N. Gisin, "Simulation of equatorial von Neumann measurements on GHZ states using nonlocal resources", *Advances in Mathematical Physics* **2010**:293245, 2010.

2 J. S. Bell, "On the Einstein-Podolsky-Rosen paradox", *Physics* **1**:195–200, 1964.

3 C. Branciard and N. Gisin, "Quantifying the nonlocality of Greenberger-Horne-Zeilinger quantum correlations by a bounded communication simulation protocol", *Physical Review Letters* **107**:020401, 2011.

4 G. Brassard. "Quantum communication complexity", *Foundations of Physics* **33**(11): 1593–1616, 2003.

5 G. Brassard, R. Cleve and A. Tapp, "Cost of exactly simulating quantum entanglement with classical communication", *Physical Review Letters* **83**:1874–1877, 1999.

6 G. Brassard and M. Kaplan, "Simulating equatorial measurements on GHZ states with finite expected communication cost", *Proceedings of 7th Conference on Theory of Quantum Computation, Communication, and Cryptography (TQC)*, pp. 65–73, 2012.

7 A. Broadbent, P. R. Chouha and A. Tapp, "The GHZ state in secret sharing and entanglement simulation", *Proceedings of the Third International Conference on Quantum, Nano and Micro Technologies*, pp. 59–62, 2009.

8 N. Cerf, N. Gisin and S. Massar, "Classical teleportation of a quantum bit", *Physical Review Letters* **84**(11):2521–2524, 2000.

9 L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

10 A. Einstein, B. Podolsky and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?", *Physical Review* **47**:777–780, 1935.

11 N. Gisin, personal communication, 2010.

12 C. Gravel, "Structure de la distribution de probabilité de l'état GHZ sous l'action de mesures de von Neumann locales", M.Sc. thesis, Department of Computer Science and Operations Research, Université de Montréal: `https://papyrus.bib.umontreal.ca/jspui/handle/1866/5511`, 2011.

13 C. Gravel, "Structure of the probability distribution for the GHZ quantum state under local von Neumann measurements", *Quantum Physics Letters* **1**(3):87–96, 2012.

14 D. M. Greenberger, M. A. Horne and A. Zeilinger, "Going beyond Bell's theorem", in *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafatos (Kluwer Academic, Dordrecht), pp. 69–72, 1989.

15 M. Kaplan, personal communication, 2013.

16 D. E. Knuth and A. C. Yao, "The complexity of nonuniform random number generation", in: *Algorithms and Complexity*, edited by J. E. Traub, pp. 357–428, Academic Press, New York, 1976.

17 T. Maudlin, "Bell's inequality, information transmission, and prism models", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 404–417, 1992.

18 J. von Neumann, "Various techniques used in connection with random digits. Monte Carlo methods", *National Bureau Standards* **12**:36–38, 1951. Reprinted in *Collected Works*, **5**:768–770, Pergamon Press, 1963.

19 O. Regev and B. Toner, "Simulating quantum correlations with finite communication", *SIAM Journal on Computing* **39**(4):1562–1580, 2009.

20 M. Steiner, "Towards quantifying non-local information transfer: finite-bit non-locality", *Physics Letters A* **270**:239–244, 2000.

21 B. Toner and D. Bacon, "Communication cost of simulating Bell correlations", *Physical Review Letters* **91**:187904, 2003.