Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 06-147

Stabilizer state breeding[1]

Erik Hostens, Jeroen Dehaene and Bart De Moor[2]

December 2006
Published in *Phys. Rev. A*

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/ehostens/reports/06-147.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. 32/16/32 86 65, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.be/scd*. E-mail: *erik.hostens@esat.kuleuven.be, jeroen.dehaene@esat.kuleuven.be, bart.demoor@esat.kuleuven.be*.

Abstract

We present a breeding protocol that distills pure copies of any stabilizer state from noisy copies and a pool of predistilled pure copies of the same state, by means of local Clifford operations, Pauli measurements and classical communication.

Stabilizer state breeding

Erik Hostens,* Jeroen Dehaene, and Bart De Moor
ESAT-SCD, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

We present a breeding protocol that distills pure copies of any stabilizer state from noisy copies and a pool of predistilled pure copies of the same state, by means of local Clifford operations, Pauli measurements and classical communication.

PACS numbers: 03.67.Mn

I. INTRODUCTION

In recent literature, much attention has been paid to the distillation of multipartite entanglement. Distillation is the recovery of entanglement that has been disrupted by the environment, by means of local operations and classical communication. Like quantum error correction codes, entanglement distillation protocols are indispensable procedures towards the realization of many quantum communication and quantum cryptography applications, requiring pure multipartite states that are shared by remote parties [1–9].

Many generalizations to a multipartite setting have been found, mainly based on protocols for bipartite entanglement distillation [10, 11]. They can be categorized according to asymptotic (hashing/breeding) [12–17] versus recurrence-like schemes [13–24], to whether they take noise in the recovering operations into account [13–15, 18, 24] or to the kind of quantum states they are designed for. The cited references are only suited for Calderbank-Shor-Steane (CSS) states or states that are locally equivalent [32] to CSS states (e.g. two-colorable graph states, GHZ states, cluster states), except Ref. [22] (for the W state) and the recent papers Refs. [15, 21, 24] (for arbitrary stabilizer states).

In this paper, we present a generalization of the breeding protocol for arbitrary stabilizer states, which is to a large extent similar to our previous paper Ref. [12] for CSS states. Again, we define a class of local Clifford operations that distribute the information content of multiple noisy copies of a stabilizer state without destroying the tensor product, i.e. they transform multiple copies of a pure stabilizer state into multiple copies of the same stabilizer state. Breeding, contrary to hashing [12], starts from k noisy copies of an n-qubit stabilizer state and a pool of $(1-\gamma)k$ predistilled pure copies of the same state, where k is considered large (asymptotic protocol). After the local Clifford operations, we locally measure the $(1-\gamma)k$ initially pure copies, yielding information on the global state. This information extraction reduces the entropy, resulting in k remaining copies that approach purity (zero entropy) and are suited for the application in mind. The measured copies are afterwards separable and can be discarded. The yield of the protocol is the net output of pure copies that is distilled for every noisy copy, and equals

$$\frac{\#(\text{pure output copies})-\#(\text{pure input copies})}{\#(\text{noisy input copies})} = \frac{k-(1-\gamma)k}{k} = \gamma.$$

Our result is very much inspired by the work of Glancy, Knill and Vasconcelos [21], in which (nonasymptotically) CSS-H states are purified by any stabilizer code, CSS states by a CSS code and arbitrary stabilizer states by a CSS-H code [33]. Contrary to Ref. [21], we use a permutation-based approach instead of a code-based approach [25]. But very similar cases can be distinguished. For CSS states, we use local Clifford operations that are built exclusively of Controlled-NOT (CNOT) operations, which correspond to CSS codes. In Ref. [12] however, more general local Clifford operations are possible for specific CSS states, but CNOTs do always satisfy. Arbitrary local Clifford operations can be used for states that satisfy certain orthogonality conditions [12]. It can be verified that these states are CSS-H states. And now we will show that local Clifford operations, built of CNOTs and represented by an orthogonal matrix, corresponding to CSS-H codes, are suited for the distillation of arbitrary stabilizer states.

This paper is organized as follows. In Sec. II A, we give a brief overview of the binary linear algebra framework in which we describe the stabilizer formalism. In the past, this "binary picture" has been proved useful in the context of distillation protocols [12, 25–27]. In Sec. II B, we summarize some properties of the strongly typical set that we will need to calculate γ. For the sake of readability, we prefer to make proofs not as rigorous as in Ref. [12], since they are highly similar. In Sec. III, we explain the breeding protocol and calculate γ. We illustrate this with a typical example in Sec. IV. We conclude in Sec. V.

II. PRELIMINARIES

A. The stabilizer formalism in the binary picture

In this section, we give a brief overview of the binary matrix description of Pauli operations, stabilizer states and Clifford operations. For a more elaborate discussion,

*Electronic address: erik.hostens@esat.kuleuven.be

we refer to Refs. [12, 28]. In the following, all addition and multiplication of binary objects is performed modulo 2.

1. Pauli operations

We use the following notation for Pauli matrices.

$$\begin{aligned}\sigma_{00} &= I_2 = \begin{bmatrix}1&0\\0&1\end{bmatrix},\\ \sigma_{01} &= \sigma_x = \begin{bmatrix}0&1\\1&0\end{bmatrix},\\ \sigma_{10} &= \sigma_z = \begin{bmatrix}1&0\\0&-1\end{bmatrix},\\ \sigma_{11} &= \sigma_y = \begin{bmatrix}0&-i\\i&0\end{bmatrix}.\end{aligned}$$

Let $v, w \in \mathbb{Z}_2^n$ and $a = \begin{bmatrix} v \\ w \end{bmatrix}$, then we denote

$$\sigma_a = \sigma_{v_1 w_1} \otimes \ldots \otimes \sigma_{v_n w_n}.$$

The Pauli group on n qubits is defined to contain all tensor products σ_a of Pauli matrices with an additional complex phase factor in $\{1, i, -1, -i\}$. In this paper, we only consider Hermitian Pauli operations, so we may exclude imaginary phase factors. It can be verified that Pauli operations satisfy the following commutation relation:

$$\sigma_a \sigma_b = (-1)^{a^T P b} \sigma_b \sigma_a, \text{ where } P = \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}. \quad (1)$$

2. Stabilizer states

A stabilizer state on n qubits is the simultaneous eigenvector, with eigenvalues 1, of n commuting Hermitian Pauli operations $(-1)^{b_i}\sigma_{s_i}$, where $s_i \in \mathbb{Z}_2^{2n}$ are linearly independent and $b_i \in \mathbb{Z}_2$, for $i = 1, \ldots, n$. These Pauli operations generate an Abelian subgroup of the Pauli group on n qubits, called the stabilizer $\mathcal{S}$. We will assemble the vectors s_i as the columns of a matrix $S \in \mathbb{Z}_2^{2n \times n}$ and the bits b_i in a vector $b \in \mathbb{Z}_2^n$. With (1), commutativity of the stabilizer is reflected by $S^T P S = 0$. The representation of $\mathcal{S}$ by S and b is not unique, as every other generating set of $\mathcal{S}$ yields an equivalent description. In the binary picture, a change from one generating set to another is represented by an invertible linear transformation $R \in \mathbb{Z}_2^{n \times n}$ acting on the right on S and acting appropriately on b. We have

$$\begin{aligned} S' &= SR \\ b' &= R^T b + d \end{aligned} \quad (2)$$

where $d \in \mathbb{Z}_2^n$ is a function of S and R but not of b [28]. In the context of distillation protocols, d can always be made zero [12].

Each S defines a total of 2^n orthogonal stabilizer states, one for each $b \in \mathbb{Z}_2^n$. As a consequence, all stabilizer states defined by S constitute a basis for $\mathcal{H}^{\otimes n}$, where $\mathcal{H}$ is the Hilbert space of one qubit. In the following, we will refer to this basis as the S-basis.

Let $|\psi_1\rangle$ and $|\psi_2\rangle$ be two stabilizer states represented by $S_1 = \begin{bmatrix} S_{1(z)} \\ S_{1(x)} \end{bmatrix}, b_1$ and $S_2 = \begin{bmatrix} S_{2(z)} \\ S_{2(x)} \end{bmatrix}, b_2$ respectively. Then $|\psi_1\rangle \otimes |\psi_2\rangle$ is a stabilizer state represented by

$$\begin{bmatrix} S_{1(z)} & 0 \\ 0 & S_{2(z)} \\ S_{1(x)} & 0 \\ 0 & S_{2(x)} \end{bmatrix}, \quad \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \quad (3)$$

If the phase factor vector b of a stabilizer state $|\psi\rangle$ represented by $S = \begin{bmatrix} S_z \\ S_x \end{bmatrix}$ is unknown, information on b by local measurements can be extracted as follows. From the definition of $|\psi\rangle$, it follows that $\sigma_{Sv}|\psi\rangle = (-1)^{v^T b + d}|\psi\rangle$ for every $v \in \mathbb{Z}_2^n$, where $d \in \mathbb{Z}_2$ is a function of S and v [28]. Consequently, we can find the inner product $v^T b$ by determining the eigenvalue of $|\psi\rangle$ for σ_{Sv}. We determine the eigenvalue for σ_g by locally measuring its nontrivial factors $\sigma_{g_i g_{n+i}}$ on qubits $i = 1, \ldots, n$, and taking the product of the outcomes. It follows that the eigenvalue for σ_g and σ_h can be determined both if and only if the factors $\sigma_{g_i g_{n+i}}$ and $\sigma_{h_i h_{n+i}}$ commute, for all i. In the binary picture: let $\mathcal{M}$ be a partition of $\{1, \ldots, n\}$ into three disjunct subsets M_z, M_x and M_y, and perform a σ_z, σ_x or σ_y measurement on qubits $i \in M_z$, M_x or M_y respectively. Note that there exist 3^n such partitions. Then from the outcomes of these measurements we can calculate $v^T b$ for all v that satisfy

$$\begin{aligned} \text{supp}\,(S_z v) \setminus \text{supp}\,(S_x v) &\subset M_z, \\ \text{supp}\,(S_x v) \setminus \text{supp}\,(S_z v) &\subset M_x, \\ \text{supp}\,(S_z v) \cap \text{supp}\,(S_x v) &\subset M_y, \end{aligned} \quad (4)$$

where the *support* of a vector $v \in \mathbb{Z}_2^n$ is defined as

$$\text{supp}\,(v) = \{i \in \{1, \ldots, n\} \mid v_i = 1\}.$$

All v satisfying (4) constitute a subspace $\mathcal{V}(\mathcal{M})$ of $\mathbb{Z}_2^n$. We denote $\dim[\mathcal{V}(\mathcal{M})]$ by $n(\mathcal{M})$.

3. Clifford operations

A Clifford operation Q, by definition, maps the Pauli group to itself under conjugation:

$$Q \sigma_a Q^\dagger = (-1)^\delta \sigma_b.$$

In the binary picture, a Clifford operation is represented by a matrix $C \in \mathbb{Z}_2^{2n \times 2n}$ and a vector $h \in \mathbb{Z}_2^{2n}$, where C is symplectic or $C^T P C = P$ [28]. The image of a Pauli operation σ_a under the action of a Clifford operation is then given by $(-1)^\epsilon \sigma_{Ca}$, where ϵ is a function of C, h and a.

If a stabilizer state $|\psi\rangle$, represented by S and b, is operated on by a Clifford operation Q, represented by C and h, $Q|\psi\rangle$ is a new stabilizer state whose stabilizer is given by $Q\mathcal{S}Q^\dagger$. As a result, this stabilizer is represented by

$$\begin{array}{rcl} S' & = & CS \\ b' & = & b + f \end{array} \tag{5}$$

where f is independent of b and can always be made zero, by performing an extra Pauli operation σ_g before the Clifford operation, where $S^T P g = f$ [12].

Let Q_1 and Q_2 be two Clifford operations represented by $\begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix}$ and $\begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix}$ respectively, where all blocks are in $\mathbb{Z}_2^{n\times n}$. Then $Q_1 \otimes Q_2$ is a Clifford operation represented by

$$\begin{bmatrix} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{bmatrix}. \tag{6}$$

The breeding protocol that will be introduced in Sec. III, considers $\bar{k}$ stabilizer states on n qubits that are shared by n remote parties, each holding corresponding qubits of all $\bar{k}$ states. Each stabilizer state is represented by the same S. According to (3), the overall state is then represented by

$$\begin{bmatrix} I_{\bar{k}} \otimes S_z \\ I_{\bar{k}} \otimes S_x \end{bmatrix},\ \bar{b}' = \begin{bmatrix} b_1 \\ \vdots \\ b_{\bar{k}} \end{bmatrix}.$$

Since it is more convenient to arrange all qubits per party, we rewrite the stabilizer matrix by permuting rows and columns as follows:

$$\begin{bmatrix} S_z \otimes I_{\bar{k}} \\ S_x \otimes I_{\bar{k}} \end{bmatrix} = S \otimes I_{\bar{k}},\ \bar{b} = \begin{bmatrix} \bar{b}_1 \\ \vdots \\ \bar{b}_n \end{bmatrix}, \tag{7}$$

where the entries of $\bar{b}'$ are permuted appropriately into $\bar{b} \in \mathbb{Z}_2^{n\bar{k}}$. All parties perform the same local Clifford operation, built only of CNOTs. It is shown in Ref. [28] that a Clifford operation, composed of CNOTs, is represented by

$$C = \begin{bmatrix} A & 0 \\ 0 & A^{-T} \end{bmatrix},$$

where A is invertible. With (6), the overall Clifford operation is then represented by

$$\begin{bmatrix} I_n \otimes A & 0 \\ 0 & I_n \otimes A^{-T} \end{bmatrix}, \tag{8}$$

Furthermore, we will demand that $A \in \mathbb{Z}_2^{\bar{k}\times\bar{k}}$ is orthogonal, i.e. $A^{-T} = A$. In that case, the representation of the overall state after the local Clifford operation can be transformed back into the original form of (7) by multiplication with $R = I_n \otimes A^T$ on the right, since

$$(I_{2n} \otimes A)(S \otimes I_{\bar{k}})(I_n \otimes A^T) = S \otimes I_{\bar{k}}.$$

Using (2) and (5), $\bar{b}$ is then transformed as follows:

$$\bar{b} \to (I_n \otimes A)\bar{b}. \tag{9}$$

B. Strongly typical set

In this section, we introduce the information-theoretical notion of a *strongly typical set*. For proofs, we refer to Ref. [12]. Good introductory material on information theory can be found in Ref. [29]. We use the compact notation $x \approx y$ for $x = y \pm \epsilon$, where $\epsilon \to 0$ for $k \to \infty$.

Let $X = (X_1, \ldots, X_k)$ be a sequence of independent and identically distributed discrete random variables, each having event set Ω with probability function $p : \Omega \to [0,1] : a \to p(a)$. The strongly typical set $\mathcal{T}_\epsilon^{(k)}$ is defined as the set of sequences $x = (x_1, \ldots, x_k) \in \Omega^k$ for which the sample frequencies $f_a(x) = |\{x_i \mid x_i = a\}|/k$ are close to the true values $p(a)$, or:

$$x \in \mathcal{T}_\epsilon^{(k)} \ \Leftrightarrow\ |f_a(x) - p(a)| < \epsilon,\ \forall a \in \Omega.$$

It can be verified that $p(\mathcal{T}_\epsilon^{(k)}) \geq 1 - \delta$, where $\delta = O(k^{-1}\epsilon^{-2})$. Therefore, $p(\mathcal{T}_\epsilon^{(k)}) \approx 1$ for $k \to \infty$. In words, a random sequence x will almost certainly be contained in the strongly typical set.

Let Ω be partitioned into subsets Ω_j $(j = 1, \ldots, t)$. We define the function

$$y(x) = (\Omega_{j_1}, \ldots, \Omega_{j_k}),\ \text{where } x_i \in \Omega_{j_i},\ \text{for } i = 1, \ldots, k.$$

In Sec. III C, we will encounter the following problem. Given some $u \in \mathcal{T}_\epsilon^{(k)}$, calculate the number $|\mathcal{N}_u|$ of sequences $v \in \mathcal{T}_\epsilon^{(k)}$ that satisfy $y(v) = y(u)$, or

$$\mathcal{N}_u = \{v \in \mathcal{T}_\epsilon^{(k)} \mid y(v) = y(u)\}.$$

It can be verified that

$$\frac{\log_2 |\mathcal{N}_u|}{k} \approx H(X) - H(Y),$$

where

$$\begin{array}{rcl} H(X) & = & -\sum\limits_{a\in\Omega} p(a)\log_2 p(a) \\ \text{and}\ \ H(Y) & = & -\sum\limits_{j=1}^{t} p(\Omega_j)\log_2 p(\Omega_j) \end{array}$$

are the entropies of X and $y(X)$ respectively.

III. BREEDING

In this section, we show how the protocol works and calculate the yield. The standard information-theoretical interpretation of hashing/breeding protocols is as follows. We have k copies of a mixed stabilizer state ρ, that we wish to purify. The global state of the copies can be regarded as a classical ensemble of pure states or, equivalently, as an unknown pure state of which we know the a-priori probabilities. With vanishing error probability, this unknown state can then be assumed to be contained in the strongly typical set $\mathcal{T}^{(k)}$. The protocol consists of local Clifford operations, followed by local measurements. These measurements yield information on the unknown state, eliminating all elements $\mathcal{T}^{(k)}$ that do not match the outcomes. The protocol ends when all elements but one are eliminated from $\mathcal{T}^{(k)}$ and we are left with a known pure state.

This section is organized as follows. In Sec. III A, we go into more detail on the information-theoretical interpretation as explained above. In Sec. III B, we show which measurements maximize the amount of extracted information or, equivalently, the probability that an element of $\mathcal{T}^{(k)}$ is eliminated. Finally, in Sec. III C, we calculate the minimal number of measurements necessary to reduce $\mathcal{T}^{(k)}$ to a singleton with probability approaching unity. The remaining element is the formerly unknown pure state.

A. Protocol

As noted in Sec. II A, all 2^n stabilizer states represented by the same $S \in \mathbb{Z}_2^{2n\times n}$ constitute a basis for $\mathcal{H}^{\otimes n}$, which we call the S-basis. The protocol starts with k identical copies of a mixed state ρ that is diagonal in this basis and $(1-\gamma)k$ copies of the pure stabilizer state represented by S and $b = 0$. The copies of ρ could for instance result from distributing k pure copies via imperfect quantum channels. If ρ is not diagonal in the S-basis, it can always be made that way by performing a local POVM measurement [13]. We have

$$\rho = \sum_{b\in\mathbb{Z}_2^n} p(b)|\psi_b\rangle\langle\psi_b|,$$

where $|\psi_b\rangle$ stands for the stabilizer state represented by S and b. The mixed state ρ can be regarded as a statistical ensemble of pure states $|\psi_b\rangle$ with probabilities $p(b)$. Consequently, $\rho^{\otimes k}$ is an ensemble of pure states represented by $S \otimes I_k$ and $\tilde{b}$, with probabilities

$$p(\tilde{b}) = p(\tilde{b}') = \prod_{i=1}^{k} p(b_i).$$

Again, the entries of $\tilde{b}$ correspond to the nk phase factors ordered per party instead of per copy like $\tilde{b}'$. We define $\bar{k} = (2-\gamma)k$ and $\bar{b}$ from $\tilde{b}$ as follows:

$$\bar{b}_i = \begin{bmatrix} \tilde{b}_i \\ 0 \end{bmatrix}, \text{ for } i = 1,\ldots,n,$$

and we have that $\rho^{\otimes k} \otimes (|\psi_0\rangle\langle\psi_0|)^{\otimes(1-\gamma)k}$ is represented by (7).

The ensemble can be interpreted as an unknown pure state. The probability that this state is represented by $\tilde{b}$ is then equal to $p(\tilde{b})$. Let this unknown pure state be represented by $\tilde{u}$. With probability approaching unity, $\tilde{u}$ is contained in the set $\mathcal{T}^{(k)}$, defined as in Sec. II B. Here, Ω is the set of all $b \in \mathbb{Z}_2^n$. So with negligible error probability, we may assume that $\tilde{u} \in \mathcal{T}^{(k)}$.

The protocol consists of the following steps:

1. Each party applies local Clifford operations (8) with orthogonal A, that, with (9), result in the transformation $\bar{u}_i \to A\bar{u}_i$, for $i = 1,\ldots,n$.
2. The resulting last $(1-\gamma)k$ copies (which were the initial pure states) are measured locally, yielding information on $\tilde{u}$. Afterwards, these copies are in a separable state and can be discarded.

After each measurement, we eliminate every $\tilde{b} \in \mathcal{T}^{(k)}$ that is inconsistent with the measurement outcome. The protocol ends when all $\tilde{b} \neq \tilde{u}$ are eliminated from $\mathcal{T}^{(k)}$ and only $\tilde{u}$ is left.

B. Maximal information extraction

The local Clifford operations (8) transform each $\bar{b}_i$ to $A\bar{b}_i$, for $i = 1,\ldots,n$. As the last $(1-\gamma)k$ entries of $\bar{b}_i$ are zero, and the last $(1-\gamma)k$ copies are the ones measured, the only relevant part of A is the lower left $(1-\gamma)k \times k$ part. We define Q as the transpose of this part. We show in Appendix A, for arbitrary full rank Q, how to construct an orthogonal matrix A with lower left $(1-\gamma)k \times k$ part ${Q'}^T$, where Q' is either equal to Q or equivalent to Q for the protocol.

We calculate the probability that some $\tilde{b} \neq \tilde{u}$ is not eliminated after the local measurement of one of the $(1-\gamma)k$ ancillary n-qubit states (take the ith). Let $\mathcal{M}$ be the partition according to which this state is measured and q the ith column of Q . If we organize $\tilde{b}$ in the following matrix:

$$\tilde{B} = \begin{bmatrix} b_1 & \cdots & b_k \end{bmatrix} = \begin{bmatrix} \tilde{b}_1^T \\ \vdots \\ \tilde{b}_n^T \end{bmatrix},$$

then, by the local Clifford operations, b_{k+i} is transformed into $\tilde{B}q$. Defining $\tilde{U}$ in the same way, u_{k+i} is transformed into $\tilde{U}q$. We know from Sec. II A that the measurement reveals $V^T u_{k+i} = V^T\tilde{U}q$, where $\mathrm{col}\,(V)$, the column space of $V \in \mathbb{Z}_2^{n\times n(\mathcal{M})}$, equals $\mathcal{V}(\mathcal{M})$. The information contained by $V^T\tilde{U}q$ is maximal when q is uniformly distributed over $\mathbb{Z}_2^k$. Indeed, as $V^T\tilde{U}q$ is a linear

function of q, then $V^T\tilde{U}q$ will be distributed uniformly too (over the range of $V^T\tilde{U}$).

If and only if $V^T(\tilde{B}+\tilde{U})q = 0$, then $\tilde{b}$ is not eliminated from $\mathcal{T}^{(k)}$ by the measurement. Indeed, in that case $\tilde{b}$ would have the same measurement outcome $V^T\tilde{B}q$ as $\tilde{u}$. Let $\Delta\tilde{b} = \tilde{b}+\tilde{u}$ and $d(\mathcal{M}, \Delta\tilde{b})$ the rank of $V^T\Delta\tilde{B}$, which is at most $n(\mathcal{M})$. Then the probability of $V^T\Delta\tilde{B}q = 0$ is equal to $2^{-d(\mathcal{M},\Delta\tilde{b})}$, as this is the inverse of the number of elements in the range of $V^T\Delta\tilde{B}$. The same reasoning can be applied to all measurements. Consequently, the probability that some $\tilde{b}$ will not be eliminated after all measurements is equal to

$$2^{-k\sum_{\mathcal{M}} m(\mathcal{M})d(\mathcal{M},\Delta\tilde{b})}, \tag{10}$$

where $km(\mathcal{M})$ is the number of measurements corresponding to partition $\mathcal{M}$. Note that $\sum_{\mathcal{M}} m(\mathcal{M}) = 1-\gamma$.

C. Minimal number of measurements

So far we have given an information-theoretical interpretation of the protocol: we start with an unknown pure state (represented by $\tilde{u}$), which is almost certainly contained in $\mathcal{T}^{(k)}$. Consecutive measurements rule out all inconsistent $\tilde{b} \in \mathcal{T}^{(k)}$. The probability that a particular $\tilde{b} \neq \tilde{u}$ survives this process equals (10). Consequently, the probability that any $\tilde{b} \neq \tilde{u}$ survives the process is at most

$$\sum_{f\not\equiv 0} N_f^* 2^{-k\sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M})} \tag{11}$$

where the sum runs over all functions $f : f(\mathcal{M}) \in \{0, 1, \ldots, n(\mathcal{M})\}$ that are not identical to zero. N_f^* is the number of $\tilde{b} \in \mathcal{T}^{(k)}$ for which $d(\mathcal{M}, \Delta\tilde{b}) = \dim[\Delta\tilde{B}^T\mathcal{V}(\mathcal{M})] = f(\mathcal{M})$. Let $N_f^* = 2^{k\alpha_f^*}$. We will show later that α_f^* (and α_f, defined below) is independent of k. Then (11) vanishes in the limit as k goes to infinity, if and only if the following inequalities hold:

$$\sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M}) > \alpha_f^*, \text{ for all } f \not\equiv 0. \tag{12}$$

Let $N_f = 2^{k\alpha_f}$ be the number of $\tilde{b} \in \mathcal{T}^{(k)}$ for which $d(\mathcal{M}, \Delta\tilde{b}) \leq f(\mathcal{M})$. Evidently,

$$N_f = \sum_{f'\leq f} N_{f'}^*, \tag{13}$$

where $f' \leq f$ stands for: $f'(\mathcal{M}) \leq f(\mathcal{M})$ for all $\mathcal{M}$. The inequalities

$$\sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M}) > \alpha_f, \text{ for all } f \not\equiv 0, \tag{14}$$

are equivalent to (12). Indeed, it follows from (13) that $\alpha_f \approx \alpha_{f'} \approx \alpha_{f'}^*$ for some $f' \leq f$ [34]. Since $\sum_{\mathcal{M}} m(\mathcal{M})f'(\mathcal{M}) > \alpha_{f'}^* \approx \alpha_{f'} \approx \alpha_f$ implies $\sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M}) > \alpha_f$, a solution to (14) is also a solution to (12) and vice versa.

This leaves us to calculate α_f. Let $\mathcal{G}_f(\mathcal{M})$ be an $n(\mathcal{M}) - f(\mathcal{M})$ dimensional subspace of $\mathcal{V}(\mathcal{M})$. For a space $\mathcal{G}$, we define

$$\mathcal{L}(\mathcal{G}) = \{\Delta\tilde{b} \in \mathbb{Z}_2^{nk} \mid G^T\Delta\tilde{B} = 0, \text{ where col}\,(G) = \mathcal{G}\}.$$

It follows that $d(\mathcal{M}, \Delta\tilde{b}) \leq f(\mathcal{M})$ for all $\Delta\tilde{b} \in \mathcal{L}[\mathcal{G}_f(\mathcal{M})]$. We then have

$$N_f = |\bigcup_{\mathcal{G}_f} \mathcal{L}\left(\sum_{\mathcal{M}} \mathcal{G}_f(\mathcal{M})\right) \cap \mathcal{T}^{(k)}|$$

where the union runs through all functions $\mathcal{G}_f : \mathcal{M} \to \mathcal{G}_f(\mathcal{M})$, where every $\mathcal{G}_f(\mathcal{M})$ is a subspace of $\mathcal{V}(\mathcal{M})$ with dimension $n(\mathcal{M}) - f(\mathcal{M})$. It follows that

$$N_f = r \max_{\mathcal{G}_f} |\mathcal{L}\left(\sum_{\mathcal{M}} \mathcal{G}_f(\mathcal{M})\right) \cap \mathcal{T}^{(k)}|,$$

where $1 \leq r \leq$ the total number of functions $\mathcal{G}_f$, which is independent of k. Therefore,

$$\alpha_f \approx \log_2(\max_{\mathcal{G}_f} |\mathcal{L}\left(\sum_{\mathcal{M}} \mathcal{G}_f(\mathcal{M})\right) \cap \mathcal{T}^{(k)}|)/k.$$

We now calculate $\log_2 |\mathcal{L}(\mathcal{G}) \cap \mathcal{T}^{(k)}|/k$. Note that $\Delta\tilde{b} \in \mathcal{L}(\mathcal{G})$ if and only if $G^T\Delta b_i = 0$, for $i = 1, \ldots, k$. The cosets Ω_j $(j = 1, \ldots, t)$ of the space $\mathcal{G}^\perp = \{v \in \mathbb{Z}_2^n | G^T v = 0\}$ constitute a partition of $\mathbb{Z}_2^n$. Note that $t = 2^{\dim(\mathcal{G})}$. We want to know the logarithm, divided by k, of the number of $\tilde{b} \in \mathcal{T}^{(k)}$ for which b_i is in the same coset as u_i, for all $i = 1, \ldots, k$. We know from Sec. II B that this is approximately

$$\begin{aligned} H - C(\mathcal{G}) \text{ where } \quad H &= -\sum_{b\in\mathbb{Z}_2^n} p(b)\log_2 p(b) \\ \text{and } \quad C(\mathcal{G}) &= -\sum_{j=1}^{t} p(\Omega_j)\log_2 p(\Omega_j). \end{aligned}$$

Let

$$H_f = \min_{\mathcal{G}_f} C\left(\sum_{\mathcal{M}} \mathcal{G}_f(\mathcal{M})\right).$$

It follows that

$$\alpha_f \approx H - H_f. \tag{15}$$

The yield γ is maximized by minimizing the total number of measurements. With (14) and (15), this results in the following linear programming (LP) problem:

$$\begin{aligned} &\text{minimize } \sum_{\mathcal{M}} m(\mathcal{M}) \\ &\text{subject to } \sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M}) > H - H_f, \text{ for all } f \not\equiv 0. \end{aligned}$$

IV. ILLUSTRATION WITH A THREE-COLORABLE GRAPH STATE

We illustrate our protocol with an example. Any stabilizer state is locally equivalent to a graph state, i.e. any stabilizer state can be reversibly transformed into a graph state by means of one-qubit Clifford operations [30, 31]. In order not to be covered by our previous result [12], the given example should be a state that is not equivalent to a CSS state or a two-colorable graph state. The 5-qubit ring state, which is an example of a three-colorable graph state [15], is such a state. We calculate the yield for the following mixture:

$$\rho = F|\psi_0\rangle\langle\psi_0| + \frac{1-F}{2^5-1}\sum_{b\in\mathbb{Z}_2^5\setminus\{0\}}|\psi_b\rangle\langle\psi_b|,$$

where $|\psi_b\rangle$ is binary represented by S and b, and

$$S = \begin{bmatrix} \theta \\ I_5 \end{bmatrix}, \text{ with } \theta = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

We will only consider the following partitions:

$$\begin{array}{ll} \mathcal{M}_1: & M_z=\{3,4,5\},\ M_x=\{1,2\},\ M_y=\emptyset; \\ \mathcal{M}_2: & M_z=\{1,4,5\},\ M_x=\{2,3\},\ M_y=\emptyset; \\ \mathcal{M}_3: & M_z=\{1,2,5\},\ M_x=\{3,4\},\ M_y=\emptyset; \\ \mathcal{M}_4: & M_z=\{1,2,3\},\ M_x=\{4,5\},\ M_y=\emptyset; \\ \mathcal{M}_5: & M_z=\{2,3,4\},\ M_x=\{1,5\},\ M_y=\emptyset. \end{array}$$

By restricting to these partitions, we risk finding a sub-optimal solution. However, we will show below that the solution found is in fact optimal. It can be verified that these partitions satisfy $n(\mathcal{M}_i)=2$, and for no other partition $n(\mathcal{M})>2$. For symmetry reasons, we may assume that the optimal $m(\mathcal{M}_i)$ will be equal for all $\mathcal{M}_i$. We have

$$\mathcal{V}(\mathcal{M}_1) = \mathrm{col}\,(V) \text{ where } V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

and $f(\mathcal{M}_1)$ can be either 0, 1 or 2:

$\rightarrow$ if $f(\mathcal{M}_1)=0$ then $\mathcal{G}_f(\mathcal{M}_1)=\mathcal{V}(\mathcal{M}_1)$,

$\rightarrow$ if $f(\mathcal{M}_1)=1$ then $\mathcal{G}_f(\mathcal{M}_1)=\mathrm{col}\,(e_1)$ or $\mathrm{col}\,(e_2)$,

$\rightarrow$ if $f(\mathcal{M}_1)=2$ then $\mathcal{G}_f(\mathcal{M}_1)=\{0\}$,

where $e_i\in\mathbb{Z}_2^5$ is a vector with all zeros except on position i. Analogous derivations hold for the other $\mathcal{M}_i$. For this highly symmetric example, we can follow the next intuitive train of thoughts: $H-H_f$ larger $\Leftarrow$ H_f smaller $\Leftarrow$ $C(\mathcal{G})$ smaller [where $\mathcal{G}=\sum_{i=1}^5\mathcal{G}_f(\mathcal{M}_i)$] $\Leftarrow$ t smaller $\Leftarrow$ $\dim(\mathcal{G})$ smaller. So we have to choose $\mathcal{G}_f(\mathcal{M}_i)$ for different i as overlapping as possible to find the highest lower bounds in the LP problem formulation. For this example, there is a one-to-one relationship between H_f and $\dim(\mathcal{G})$:

$\dim(\mathcal{G})$	$H_f[\dim(\mathcal{G})]$
0	0
1	$-[F+\frac{15}{31}(1-F)]\log_2[F+\frac{15}{31}(1-F)]$ $-[\frac{16}{31}(1-F)]\log_2[\frac{16}{31}(1-F)]$
2	$-[F+\frac{7}{31}(1-F)]\log_2[F+\frac{7}{31}(1-F)]$ $-3[\frac{8}{31}(1-F)]\log_2[\frac{8}{31}(1-F)]$
3	$-[F+\frac{3}{31}(1-F)]\log_2[F+\frac{3}{31}(1-F)]$ $-7[\frac{4}{31}(1-F)]\log_2[\frac{4}{31}(1-F)]$
4	$-[F+\frac{1}{31}(1-F)]\log_2[F+\frac{1}{31}(1-F)]$ $-15[\frac{2}{31}(1-F)]\log_2[\frac{2}{31}(1-F)]$
5	$-F\log_2 F-31\frac{1-F}{31}\log_2(\frac{1-F}{31})=H$

It can then be verified that the yield γ equals $1-\sum_{i=1}^5 m(\mathcal{M}_i)=1-5m$ where m is the solution to the following LP problem:

$$\begin{array}{lrcl} \text{minimize} & m & & \\ \text{subject to} & 10m & > & H-H_f(0)=H \\ & 8m & > & H-H_f(1) \\ & 6m & > & H-H_f(2) \\ & 4m & > & H-H_f(3) \\ & 2m & > & H-H_f(4) \end{array}$$

Numerical calculation shows that the first inequality is the most binding. Therefore,

$$\gamma = 1-\frac{H}{2},$$

which corresponds to our intuition that every measurement yields $n(\mathcal{M}_i)=2$ bits of information. However, this will not be true for a less symmetric mixed state ρ. Note that this solution is optimal. Indeed, we cannot gain more than 2 bits per measurement, as $n(\mathcal{M})\leq 2$ for all $\mathcal{M}$. We have plotted γ as a function of F in Fig. 1.

V. CONCLUSION

We have presented a breeding protocol that works for any multipartite stabilizer state. Starting with k noisy copies of a stabilizer state and a pool of $(1-\gamma)k$ predistilled pure copies of the same state, the protocol consists of local Clifford operations on noisy and pure copies, followed by local Pauli measurements on the initially pure copies to extract information on the global state. The yield γ is calculated as the solution of a linear programming problem. We have illustrated this with a typical example.

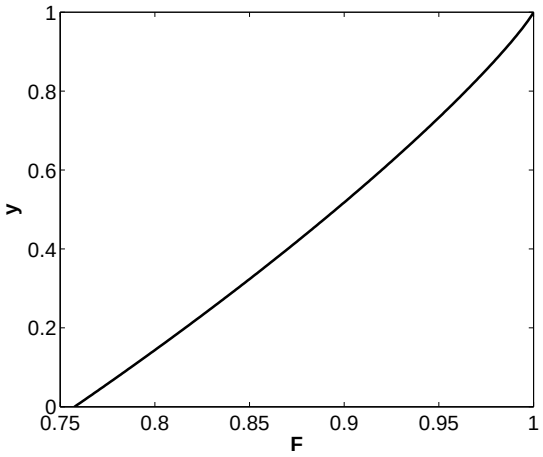

FIG. 1: the yield γ of the protocol for the state $\rho = F|\psi_0\rangle\langle\psi_0| + (1-F)/31(\mathbb{1} - |\psi_0\rangle\langle\psi_0|)$ as a function of F.

Acknowledgments

Research funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). Research supported by Research Council KUL: GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0197.02 (power islands), G.0141.03 (identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), G.0452.04 (new quantum algorithms), G.0499.04 (statistics), G.0211.05 (nonlinear), G.0226.06 (cooperative systems and optimization), G.0321.06 (tensors), G.0302.07 (SVM/Kernel), research communities (ICCoS, ANMMM, MLDM); IWT: PhD Grants, McKnow-E, Eureka-Flite2; Belgian Federal Science Policy Office: IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling', 2002-2006); EU: ERNSI.

APPENDIX A: CONSTRUCTION OF ORTHOGONAL A

Before we show how A is constructed, we prove the following theorems.

Theorem 1 *Any symmetric matrix $W \in \mathbb{Z}_2^{n\times n}$ of rank r can be factorized as follows:*

$$W = RDR^T,$$

where R is invertible and

(i) $D = \begin{bmatrix} I_{r/2} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \\ & 0 \end{bmatrix}$ *if W has zero diagonal,*

(ii) $D = \begin{bmatrix} I_r & \\ & 0 \end{bmatrix}$ *if W has nonzero diagonal.*

Proof: We prove that if the theorem is true for all $n \leq N$, it also holds for $n = N+1$. Note that the theorem is trivial for zero matrices, as $0 = R0R^T$, and matrices of zero dimension.

(i) Without loss of generality, we may consider (nonzero) $W \in \mathbb{Z}_2^{(N+1)\times(N+1)}$ of the following form:

$$W = \begin{bmatrix} 0 & 1 & a^T \\ 1 & 0 & b^T \\ a & b & W_2 \end{bmatrix},$$

where a, b, W_2 have appropriate dimensions and W_2 has zero diagonal. Indeed, note that identical permutations of rows and columns of W are for free, as they can be absorbed into R as follows:

$$PWP^T = RDR^T \Rightarrow W = (P^TR)D(P^TR)^T.$$

Since $W_2 + ab^T + ba^T$ has zero diagonal and is an $(N-1)\times(N-1)$ matrix, we can write:

$$W_2 + ab^T + ba^T = R_2D_2R_2^T.$$

It follows that $W = RDR^T =$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & a & R_2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & D_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & b^T \\ 0 & 1 & a^T \\ 0 & 0 & R_2^T \end{bmatrix}.$$

By construction, R is invertible because R_2 is.

(ii) Again, without loss of generality, we may consider W of the form:

$$W = \begin{bmatrix} 1 & a^T \\ a & W_2 \end{bmatrix}.$$

We can write:

$$W_2 + aa^T = R_2D_2R_2^T,$$

where D_2 is either (i) or (ii). It follows that

$$W = RDR^T = \begin{bmatrix} 1 & 0 \\ a & R_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} 1 & a^T \\ 0 & R_2^T \end{bmatrix}.$$

If D_2 is of the form (i), it can be brought to (ii), by using the identity

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = VV^T, \quad \text{with } V = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

This ends the proof. □

Corollary 1 *If and only if a given symmetric matrix $W \in \mathbb{Z}_2^{n\times n}$ is not both full rank and zero-diagonal, we can find square M such that $W = M^T M$.*

Proof: Using Theorem 1, we have $W = RDR^T$. If D is of the form (ii), we take M equal to R^T with the rightmost $n-r$ columns set to zero. If D is of the form (i) and not full rank, we can find U such that $D = U^T U$, by using the identity

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = V^T V, \quad \text{with } V = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Then take M equal to UR^T with the rightmost $n-r$ columns set to zero.

Finally, we show that if W is full rank and zero-diagonal, there is no M satisfying $M^T M = W$. An equivalent statement is that there exists no square M such that $M^T M = D = I \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. As $x^T D x = 0$ for all x, $M^T M = D$ implies that $y^T y = 0$ for all $y = Mx$. Consequently, M cannot be full rank. But then $M^T M = D$ cannot be true, as D is full rank. This ends the proof. □

Theorem 2 *A matrix $W \in \mathbb{Z}_2^{n\times r}$ can be extended to an orthogonal matrix $A \in \mathbb{Z}_2^{n\times n}$ by adding columns, if and only if*

- $W^T W = I_r$,
- $e \notin \mathrm{col}\,(W)$, *where $e \in \mathbb{Z}_2^n$ is the all-ones vector.*

Proof: Define a full rank matrix $Y \in \mathbb{Z}_2^{n\times(n-r)}$ such that $W^T Y = 0$. By Theorem 1, we can find R and D such that $Y^T Y = RDR^T$. For now, we assume that $Y^T Y$ is full rank. As $e \notin \mathrm{col}\,(W)$, we know that $D = I_{n-r}$. Otherwise $Y^T Y$ has a zero diagonal, or $y_i^T y_i = 0$ for all columns y_i of Y. Equivalently, we have $y_i^T e = 0$ for all i, or $Y^T e = 0$, which contradicts $e \notin \mathrm{col}\,(W)$. Let $Z = YR^{-T}$, then $A = \begin{bmatrix} W & Z \end{bmatrix}$ is orthogonal. Indeed, $Z^T W = R^{-1} Y^T W = 0$ and $Z^T Z = R^{-1} Y^T Y R^{-T} = D = I$.

This leaves us to prove that $Y^T Y$ is full rank. If not, then there exists some $x \neq 0$ that satisfies $Y^T Y x = 0$. By the definition of Y, it follows that $Yx \in \mathrm{col}\,(W)$ or $Yx = Wz$ for some $z \neq 0$. But then $W^T W z = W^T Y x = 0$, which contradicts $W^T W = I$. This ends the proof. □

We now show, for a given full rank $k \times (1-\gamma)k$ matrix Q, how to construct an orthogonal $\bar{k} \times \bar{k}$ matrix A with lower left part equal to Q'^T, where Q' is either equal to Q or equivalent to Q for the protocol. This is the problem adressed in Sec. III B. We perform the following steps:

1. find square matrix M such that $M^T M = I + Q^T Q$;
2. create orthogonal A^T from $W = \begin{bmatrix} Q \\ M \end{bmatrix}$ by adding columns.

By Corollary 1, step 1 is possible provided that $I + Q^T Q$ is not both full rank and zero diagonal. In that case, this can be solved by adding just one column to Q, as the resulting matrix Q' then has an odd number of columns. Consequently, we have one measurement more, but as k is large, this will not influence the yield. By Theorem 2, Step 2 is possible provided that $e \notin \mathrm{col}\,(W)$. Let $e \in \mathrm{col}\,(W)$. Then there exists some $x \neq 0$ that satisfies $Qx = e$ and $Mx = e$. Without loss of generality, we may assume that $x_1 = 1$. If we add any column (take the second) of Q to the first, yielding Q', and repeat step 1, e will be no longer in $\mathrm{col}\,(W')$. This is shown as follows. Let $e \in \mathrm{col}\,(W')$, then there exists some y satisfying $Q'y = e$ and $M'y = e$. From the construction of Q' and the fact that Q is full rank, we have $y_i = x_i$ for all $i \neq 2$ and $y_2 = 1 + x_2$. Consequently, $y^T y \neq x^T x$. This is contradicted by $x^T x + y^T y = x^T (M^T M + Q^T Q) x + y^T (M'^T M' + Q'^T Q') y = 0$. Finally, note that Q' is equivalent to Q for the protocol, as the outcomes of the first two measurements in the latter case can be calculated from the corresponding outcomes in the former case.

[1] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W.K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[2] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Phys. Rev. Lett. **77**, 2818 (1996).

[3] A.K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[4] A. Karlsson, M. Koashi, and N. Imoto, Phys. Rev. A **59**, 162 (1999).

[5] M. Hillery, V. Buzek, and A. Berthiaume, Phys. Rev. A **59**, 1829 (1999).

[6] R. Cleve, D. Gottesman, and H.-K. Lo, Phys. Rev. Lett. **83**, 648 (1999).

[7] C. Crépeau, D. Gottesman, and A. Smith, in *Proc. STOC* (2002), quant-ph/0206138.

[8] W. Dür, J. Calsamiglia, and H.-J. Briegel, Phys. Rev. A **71**, 042336 (2005).

[9] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest, and H.-J. Briegel, in *Proceedings of the International School of Physics "Enrico Fermi" on "Quantum computers, Algorithms and Chaos"* (2005), quant-ph/0602096.

[10] C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J.A. Smolin, and W.K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).

[11] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, and W.K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[12] E. Hostens, J. Dehaene, and B. De Moor, Phys. Rev. A

73, 042316 (2006).
[13] H. Aschauer, W. Dür, and H.-J. Briegel, Phys. Rev. A **71**, 012319 (2005).
[14] C. Kruszynska, S. Anders, W. Dür, and H.-J. Briegel, Phys. Rev. A **73**, 062328 (2006).
[15] C. Kruszynska, A. Miyake, H.-J. Briegel, and W. Dür, *Entanglement purification protocols for all graph states*, quant-ph/0606090.
[16] E. Maneva and J.A. Smolin, *Improved two-party and multi-party purification protocols*, in *Quantum Computation and Quantum information*, edited by J. Samuel and J. Lomonaco, Vol. **305** of AMS Contemporary Mathematics (American Mathematical Society, Providence, RI, 2002); e-print quant-ph/0003099.
[17] K. Chen and H.-K. Lo, *Multi-partite quantum cryptographic protocols with noisy GHZ states*, quant-ph/0404133.
[18] W. Dür, H. Aschauer, and H.-J. Briegel, Phys. Rev. Lett. **91**, 107903 (2003).
[19] M. Murao, M.B. Plenio, S. Popescu, V. Vedral, and P.L. Knight, Phys. Rev. A **57**, R4075 (1998).
[20] A. Acin, E. Jane, W. Dür, and G. Vidal, Phys. Rev. Lett. **85**, 4811 (2000).
[21] S. Glancy, E. Knill, and H.M. Vasconcelos, *Entanglement purification of any stabilizer state*, quant-ph/0606125.
[22] A. Miyake and H.-J. Briegel, Phys. Rev. Lett. **95**, 220501 (2005).
[23] K. Goyal, A. McCauley, and R. Raussendorf, *Purification of large bi-colorable graph states*, quant-ph/0605228.
[24] A. Kay and J.K. Pachos, *Optimality of purification protocols and upper-bounds to fault-tolerance*, quant-ph/0608080.
[25] E. Hostens, J. Dehaene, and B. De Moor, *The equivalence of two approaches to the design of entanglement distillation protocols*, quant-ph/0406017.
[26] J. Dehaene, M. Van den Nest, B. De Moor, and F. Verstraete, Phys. Rev. A **67**, 022310 (2003).
[27] E. Hostens, J. Dehaene, and B. De Moor, Phys. Rev. A **73**, 062337 (2006).
[28] J. Dehaene and B. De Moor, Phys. Rev. A **68**, 042318 (2003).
[29] T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley & Sons, Inc., 1991).
[30] M. Van den Nest, J. Dehaene, and B. De Moor, Phys. Rev. A **69**, 022316 (2004).
[31] D. Schlingemann, Quant. Inf. Comp. **2**, 307 (2002).
[32] Two states are called locally equivalent if they can be transformed into one another by local operations.
[33] A CSS-H code (state) is a CSS code (state) that is invariant under the action of a local Hadamard operation on every qubit. Furthermore, it should be possible to define logical $\sigma_x^{(L)}$ and $\sigma_z^{(L)}$ that are transformed into one another by a local Hadamard operation on every qubit.
[34] We make use of the fact that if $2^{k\alpha} = \sum_{i=1}^{g} 2^{k\alpha_i}$ for some natural number g (independent of k) and $k \to \infty$, then $\alpha \approx \alpha_i$ for some i.