Enumerating all bilocal Clifford distillation protocols through symmetry reduction

S. Jansen[1,2], K. Goodenough[3], S. de Bone[3,4], D. Gijswijt[1], and D. Elkouss[3,5]

[1]Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands.
[2]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands
[3]QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
[4]QuSoft, CWI, Science Park 123, 1098 XG Amsterdam, The Netherlands
[5]Networked Quantum Devices Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

Entanglement distillation is an essential building block in quantum communication protocols. Here, we study the class of near-term implementable distillation protocols that use bilocal Clifford operations followed by a single round of communication. We introduce tools to enumerate and optimise over all protocols for up to $n = 5$ (not necessarily equal) Bell-diagonal states using a commodity desktop computer. Furthermore, by exploiting the symmetries of the input states, we find all protocols for up to $n = 8$ copies of a Werner state. For the latter case, we present circuits that achieve the highest fidelity with perfect operations and no decoherence. These circuits have modest depth and number of two-qubit gates. Our results are based on a correspondence between distillation protocols and double cosets of the symplectic group, and improve on previously known protocols.

1 Introduction

Entanglement is an essential resource for a host of quantum communication tasks, including but not limited to secret-key generation [1], conference key agreement [2, 3], clock synchronisation [4], and distributed quantum computation [5, 6, 7]. Due to experimental limitations, entanglement is in practice always noisy, i.e. it has a non-unit fidelity with a target perfectly entangled state. A lower fidelity can lead to a lower rate at which one can perform certain tasks, or even yield their implementation impossible. Entanglement distillation is a set of procedures that increase the fidelity of the present entanglement by transforming multiple copies of a lower fidelity entangled state into (usually) a smaller number of copies with higher fidelity [8, 9, 10] (see [11] for a review).

Entanglement distillation has been studied in different settings. One such setting corresponds to the highly idealised scenario where one is given an asymptotic number of copies of a single state, and one can perform arbitrary local operations and classical communication (LOCC) [12, 13, 14]. Such protocols can in principle require an unbounded number of rounds of classical communication between Alice and Bob, rendering them infeasible in practice. A well-known explicit asymptotic protocol is the hashing protocol [8]. This protocol allows for the distillation to maximally entangled states at a finite rate, given that the given input state has a high enough input fidelity. Other scenarios include purification with the help of entanglement [15, 16, 17, 18], on a single copy only, known as filtering [19, 20], with environment assistance [21], of higher-dimensional states [22, 23] or with different classes of operations than LOCC [24, 25]. On the experimental side, entanglement distillation has been realised using photonic setups [26, 27, 28, 29, 30, 31] (where the distilled state is not stored in a memory afterwards), ions [32] (where the distillation was performed within a single node) and NV-centres in diamond [33] (where the distilled states were heralded and stored in memories for further use).

Our goal is to find good distillation protocols with modest requirements. In particular, protocols where Alice and Bob use a small number of entangled states [34, 35], and require only a single round of communication after performing

their local operations [34]. The above class of distillation protocols were first considered in [34], where they were called measure and exchange protocols. The semidefinite programming bounds found by Rozpedek et al. [34] allow to bound the optimal performance of measure and exchange protocols. Moreover, in some particular cases the existing protocols meet the bounds allowing to establish their optimality. Regarding the design of protocols, a heuristic procedure called the seesaw method allows to improve existing protocols [34]. More recently, Krastanov et al. investigated a genetic optimisation method for a subset of these protocols [36] and evaluated them including noisy operations.

Here, complementary to previous work, we find a systematic procedure to obtain good measure and exchange protocols.

To this end, we narrow down our investigation from general measure and exchange protocols to a practically relevant subset of protocols and states. We consider the distillation of Bell-diagonal states, where we use arbitrary noiseless *bilocal Clifford circuits* and measure out all but one of the qubit pairs. The measurement results are communicated between Alice and Bob, and the protocol is deemed successful if all pairs had correlated outcomes. We call this class of protocols *bilocal Clifford protocols* for short. This class of protocols includes a number of relevant protocols considered before in the literature [8, 10, 37, 38, 39, 40, 41, 42, 36].

The restriction to bilocal Clifford protocols and Bell-diagonal states allows us to reduce the finding of all bilocal Clifford protocols to enumerating all (double) cosets $\mathcal{D}_n \backslash \mathrm{Sp}(2n, \mathbb{F}_2) / \mathcal{K}_n$. Here, $\mathrm{Sp}(2n, \mathbb{F}_2)$ is the symplectic group over the field with two elements $\mathbb{F}_2$, $\mathcal{K}_n$ is the (possibly trivial) subgroup that preserves the input states and $\mathcal{D}_n$ is the distillation subgroup, which is the set of operations that leave both the success probability and fidelity invariant. One of our contributions in this work is to characterise this subgroup in terms of its generators and its order. We consider two cases for the input states - general input states (i.e. trivial symmetry group) and the n-fold tensor product of Werner states. For general input states, we find all protocols for up to $n = 5$ entangled pairs. For an n-fold tensor product of Werner states, we describe an algorithm that finds a complete set of double coset representatives. This allows us to optimise over all bilocal Clifford protocols when distilling an n-fold tensor product of a Werner state for n up to 8 pairs.

We find that for $n = 2, 3$ copies of a Werner state, the highest fidelity out of all bilocal Clifford protocols is achieved by protocols studied before in the literature. For $n = 4$ to 8, we find increased fidelities over previously considered distillation schemes. Furthermore, we find explicit circuits achieving the highest fidelity out of all bilocal Clifford protocols, see Appendix E. These circuits have comparable depth and number of two-qubit gates as previously studied protocols, highlighting also the practical feasibility of our findings.

This paper is structured as follows. In Section 2 we describe the preliminaries and notation needed throughout the paper. Section 3 explains bilocal Clifford distillation protocols and how the optimisation over such protocols can be rephrased as an optimisation over elements from the symplectic group $\mathrm{Sp}(2n, \mathbb{F}_2)$. In Section 4 we characterise the distillation subgroup $\mathcal{D}_n$. In Section 5 we prove a further reduction of our search space when the state to be distilled is an n-fold tensor product of a Werner state. In Section 6 we present our optimisation results. We end with conclusions and discussions in Section 7.

2 Preliminaries

We begin by setting some relevant notation. The field with two elements is denoted by $\mathbb{F}_2$. We use the notation U_i to denote a single-qubit operation on qubit i. The single-qubit operations that we use are the Pauli gates (I, X, Y and Z), the Hadamard gate (H) and the phase gate (S). Moreover, we denote by CNOT_{ij} a controlled-NOT operation with control qubit i and target qubit j, by CZ_{ij} a controlled-Z operation between qubits i and j and by SWAP_{ij} the operation that swaps qubits i and j.

2.1 Pauli group and Clifford group

The Pauli matrices are defined as

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{1}$$

The Pauli group with phases on n qubits $\overline{\mathcal{P}}_n$ consists of all $2^n \times 2^n$ matrices of the form $\lambda P_1 \otimes \cdots \otimes P_n$ with $\lambda \in \{\pm 1, \pm i\}$ and $P_i \in \{I, X, Y, Z\}$ for all $i \in \{1, \ldots, n\}$, together with standard matrix multiplication. Of particular interest to us is the Pauli group without any phase factors, $\mathcal{P}_n \cong \overline{\mathcal{P}}_n / \langle iI^{\otimes n} \rangle$. Here $\langle iI^{\otimes n} \rangle$ is the subgroup generated by $iI^{\otimes n}$. We will call this the Pauli group for short. An element of the group $\mathcal{P}_n$ is referred to as a Pauli string (of length n). The order of $\mathcal{P}_n$ equals $|\mathcal{P}_n| = 4^n$.

An important class of gates in quantum information theory are the so-called Clifford gates [43]. Circuits composed of Clifford gates are efficiently classically simulable, yet can be used to create complex quantum states, which are used for example in stabiliser error correction. The Clifford gates on n qubits form a group $\mathcal{C}_n$, and each $C \in \mathcal{C}_n$ induces an automorphism $f : \overline{\mathcal{P}}_n \to \overline{\mathcal{P}}_n$ on $\overline{\mathcal{P}}_n$ by conjugating each element with C, i.e. $f(P) = CPC^\dagger$. The Clifford group $\mathcal{C}_n$ is generated by Hadamard- (H_i) and phase (S_i) gates on each qubit ($1 \leq i \leq n$) and CNOT gates between every pair (i, j) of qubits. In matrix representation, these gates are given by

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, \tag{2}$$

$$\mathrm{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{3}$$

2.2 Binary representation of the Pauli group and the Clifford group

The elements of the Pauli group and the Clifford group can be described in terms of binary vectors and matrices, respectively. To see this, we first introduce the following notation for the Pauli matrices.

$$\tau_{00} = I, \quad \tau_{10} = X, \quad \tau_{11} = iY, \quad \tau_{01} = Z. \tag{4}$$

We extend this notation to tensor products of Pauli matrices as follows.

$$\tau_v \coloneqq \tau_{v_1 v_{n+1}} \otimes \cdots \otimes \tau_{v_n v_{2n}}, \quad v \in \mathbb{F}_2^{2n}, \tag{5}$$

As mentioned in Section 2.1, the global phase factors are not important in the context of this paper, so an element $\lambda\tau_v$, $\lambda \in \{\pm 1, \pm i\}$, of $\overline{\mathcal{P}}_n$ can be represented by the binary vector $v \in \mathbb{F}_2^{2n}$. The multiplication of the elements of $\overline{\mathcal{P}}_n$ corresponds then to the addition of the binary vectors.

For any $C \in \mathcal{C}_n$, the conjugation map f corresponds to a linear map on the set of binary vectors (and thus on $\overline{\mathcal{P}}_n$). The map f is an automorphism, and thus preserves the commutation relations of the elements of $\overline{\mathcal{P}}_n$. To see what this implies for the linear transformation in the binary picture, let $v, w \in \mathbb{F}_2^{2n}$. Then

$$\tau_v \tau_w = (-1)^{v^\mathsf{T} \Omega w} \tau_w \tau_v, \tag{6}$$

where $\Omega = \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}$. A proof of this formula can be found in Appendix A.

Let M denote the linear transformation corresponding to conjugation by C. It follows from equation (6) that

$$\tau_{Mv} \tau_{Mw} = (-1)^{(Mv)^\mathsf{T} \Omega M w} \tau_{Mw} \tau_{Mv}. \tag{7}$$

By equation (6), we know that τ_v and τ_w commute iff $v^\mathsf{T} \Omega w = 0$ and anti-commute iff $v^\mathsf{T} \Omega w = 1$. In order to preserve the commutation relations, it must then hold that $v^\mathsf{T} M^\mathsf{T} \Omega M w = v^\mathsf{T} \Omega w$ for all $v, w \in \mathbb{F}_2^{2n}$, so $M^\mathsf{T} \Omega M = \Omega$. The matrices M that satisfy this condition thus preserve the so-called symplectic inner product $\omega(v, w) \equiv v^\mathsf{T} \Omega w$ between any two $v, w \in \mathbb{F}_2^{2n}$. These matrices form a group known as the symplectic group over $\mathbb{F}_2$, denoted by $\mathrm{Sp}(2n, \mathbb{F}_2)$. The order of the symplectic group over $\mathbb{F}_2$ is well-known [44] to be equal to

$$|\mathrm{Sp}(2n, \mathbb{F}_2)| = 2^{n^2} \prod_{j=1}^{n} (4^j - 1). \tag{8}$$

The symplectic complement of a subspace V of $\mathbb{F}_2^{2n}$ is defined as the set of elements of $\mathbb{F}_2^{2n}$ that have zero symplectic inner product with all elements from V,

$$V^\perp = \{ v \in \mathbb{F}_2^{2n} \mid \omega(v, w) = 0 \ \forall w \in V \} . \tag{9}$$

The symplectic complement satisfies the following property,

$$\left(V^\perp\right)^\perp = V . \tag{10}$$

Calculations involving a symplectic matrix M can often be simplified by writing it as a

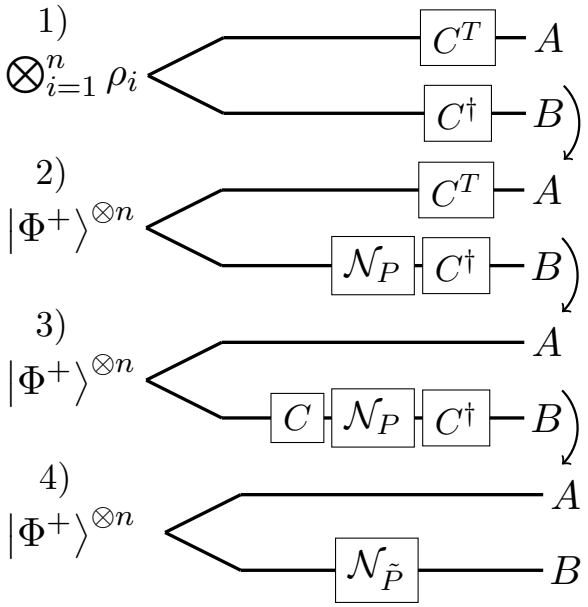


Figure 1: Schematic description of how bilocal Clifford circuits map n-qubit bipartite systems to n-qubit bipartite systems. From 1) to 2), we rewrite the state as $\otimes_{i=1}^{n}\rho_i = (I\otimes\mathcal{N})\left(|\Phi^+\rangle^{\otimes n}\right)$, where $\mathcal{N}(\cdot) = \sum_{P\in\mathcal{P}_n} p_P P(\cdot)P^\dagger$. In 3), we use the fact that $A^{\mathsf{T}}\otimes I|\Phi^+\rangle^{\otimes n} = I\otimes A|\Phi^+\rangle^{\otimes n}$ for any $2^n\times 2^n$ matrix A [46]. For 4), we use the fact the Cliffords act on the group of Pauli strings $\mathcal{P}_n$.

block matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, with $A, B, C, D \in M_{n\times n}(\mathbb{F}_2)$. From the condition $M^{\mathsf{T}}\Omega M = \Omega$ it follows that the blocks satisfy

$$\begin{aligned} B^{\mathsf{T}}D + D^{\mathsf{T}}B &= 0, \\ A^{\mathsf{T}}C + C^{\mathsf{T}}A &= 0, \\ A^{\mathsf{T}}D + C^{\mathsf{T}}B &= I_n. \end{aligned} \tag{11}$$

Moreover, the inverse of M is given by

$$M^{-1} = \begin{bmatrix} D^{\mathsf{T}} & B^{\mathsf{T}} \\ C^{\mathsf{T}} & A^{\mathsf{T}} \end{bmatrix}. \tag{12}$$

Let $\phi : \mathcal{C}_n \to \mathrm{Sp}(2n, \mathbb{F}_2)$ be the function that maps every Clifford gate to the corresponding symplectic matrix. This map is a surjective group homomorphism [45]. The symplectic group $\mathrm{Sp}(2n, \mathbb{F}_2)$ is thus generated by the images of a generating set of the Clifford group $\mathcal{C}_n$ under ϕ.

3 Bilocal Clifford protocols

This section covers the structure of the distillation protocols that are considered in this paper. We consider a system consisting of two parties, Alice and Bob, that share n entangled two-qubit states. We focus on states that are diagonal in the Bell basis. Bell-diagonal states naturally arise with realistic noise models such as dephasing and depolarizing. Moreover, any bipartite state can be *twirled* into a Bell-diagonal state while preserving the fidelity [9]. We note that the protocols found in our paper are also relevant to states that are not necessarily Bell diagonal — the performance of a protocol on a Bell-diagonal and a not necessarily Bell-diagonal state will be comparable as long as the two states in question are close in trace distance.

Bell-diagonal states can be written as

$$\begin{aligned} \rho = p_I \left|\Phi^+\right\rangle\left\langle\Phi^+\right| + p_X \left|\Psi^+\right\rangle\left\langle\Psi^+\right| \\ + p_Y \left|\Psi^-\right\rangle\left\langle\Psi^-\right| + p_Z \left|\Phi^-\right\rangle\left\langle\Phi^-\right|. \end{aligned} \tag{13}$$

The indices of the probabilities arise from the following correspondence between the Bell states and the Pauli matrices.

$$\begin{aligned} \left|\Phi^+\right\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = (I\otimes I)\left|\Phi^+\right\rangle, \\ \left|\Psi^+\right\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) = (I\otimes X)\left|\Phi^+\right\rangle, \\ \left|\Psi^-\right\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = (I\otimes -iY)\left|\Phi^+\right\rangle, \\ \left|\Phi^-\right\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) = (I\otimes Z)\left|\Phi^+\right\rangle. \end{aligned} \tag{14}$$

Equation (14) gives rise to a bijective mapping from the Bell states $\left|\Phi^+\right\rangle$, $\left|\Psi^+\right\rangle$, $\left|\Psi^-\right\rangle$ and $\left|\Phi^-\right\rangle$ to the Pauli matrices I, X, Y and Z, respectively. We denote a tensor product of n Bell-diagonal states by a tensor product of Pauli matrices, e.g. $\left|\Phi^+\right\rangle\otimes\left|\Psi^+\right\rangle\otimes\left|\Psi^-\right\rangle\otimes\left|\Phi^-\right\rangle$ is denoted by $I\otimes X\otimes Y\otimes Z$.

We generalise the notation of equation (13) and denote by p_P the probability that the system is in the state described by $P\in\mathcal{P}_n$. In the subscript we will not explicitly denote the tensor product, e.g. p_{XY} denotes the probability that the system is described by $X\otimes Y$. The initial state of the protocol consisting of n entangled two-qubit states can thus be fully described by the set of probabilities $\mathcal{Q} = \{p_{P_1P_2\ldots P_n} : P_i \in \{I, X, Y, Z\}\}$. We refer to such a system as an *n-qubit bipartite system*. When working in the binary representation, we write p_v instead of p_P,

where v is a binary vector and $p_v = p_P$ if v is the binary representation of P. It will be clear from the context which convention is used.

3.1 Bilocal Clifford circuits

The first step of the protocol is the performance of bilocal Clifford operations. That is, if Alice applies a Clifford operation $\tilde{C} \in \mathcal{C}_n$ to her qubits, then Bob applies $\tilde{C}^*$, the entry-wise complex conjugate of $\tilde{C}$, to his qubits (see Fig. 1). This leads to a permutation of the set $\mathcal{Q}$. In particular, each element p_P of $\mathcal{Q}$ is mapped [46] to $p_{\tilde{C}^\intercal P \tilde{C}^*}$, or equivalently, $p_{CPC^\dagger}$, where we defined $C = \tilde{C}^\intercal \in \mathcal{C}_n$. We denote the probabilities that describe the permuted state by $\tilde{p}_{P_1 P_2 ... P_n}$.

We note here that the most general permutation on $\mathcal{Q}$ by local unitaries consists of applying bilocal Cliffords followed by a Pauli string applied to either Alice or Bob's side [45]. These Pauli strings can be used to reorder locally the coefficients of the states.

Since (bilocal) Clifford operations form a group, the Clifford group has a group action on $\mathcal{Q}$. The (normal) subgroup of the Clifford group that fixes $\mathcal{Q}$ point-wise does not change any of the statistics, and is thus not of interest to us. This subgroup consists of all Pauli strings, and quotienting out the Cliffords by this subgroup leads to the symplectic group (over $\mathbb{F}_2$), $\mathrm{Sp}(2n, \mathbb{F}_2)$ [45, 47]. We can thus describe a bilocal Clifford operation by an element $M \in \mathrm{Sp}(2n, \mathbb{F}_2)$. To simplify notation, we sometimes slightly abuse the notation and denote by $C \in \mathrm{Sp}(2n, \mathbb{F}_2)$ the symplectic matrix corresponding to conjugation by $C \in \mathcal{C}_n$, but it should be kept in mind that always the symplectic matrix M is meant.

3.2 Measurements and postselection for bilocal Clifford protocols

In the second step, Alice and Bob perform measurements in the computational basis on $n-1$ of their qubits. Alice and Bob report their results to each other using classical communication. If the outcomes are equal, they keep the state that was not measured. In this case, the protocol is called *successful*. If the outcomes are not equal, they discard all states, and the protocol is not successful. The probability that a protocol is successful is equal to the probability that all measured states are either in the $|\Phi^+\rangle$ or in the $|\Phi^-\rangle$ state, which correspond to the I and Z Pauli matrix, respectively. The success probability of the protocol is thus equal to

$$p_{\mathrm{suc}} = \sum_{\substack{P_1 \in \{I,X,Y,Z\}, \\ Q_j \in \{I,Z\}}} \tilde{p}_{P_1 Q_2 ... Q_n}, \tag{15}$$

where we used the convention that the first two-qubit state is not measured. Moreover, the fidelity between the remaining state and the $|\Phi^+\rangle$ state is equal to

$$F_{\mathrm{out}} = \frac{\sum_{Q_j \in \{I,Z\}} \tilde{p}_{I_1 Q_2 ... Q_n}}{p_{\mathrm{suc}}}. \tag{16}$$

To simplify notation in the rest of this paper, we introduce the following two definitions.

Definition 3.1. The *base* of an n-qubit bipartite quantum system is given by

$$\mathscr{B} = \left\{ v \in \mathbb{F}_2^{2n} \mid v_i = 0 \ \forall i \in \{1, ..., n+1\} \right\}.$$

Note that the base vectors correspond to the Pauli strings $I_1 \otimes Q_2 \otimes ... \otimes Q_n \in \mathcal{P}_n$ with $Q_j \in \{I, Z\}$ for all $j \in \{2, ..., n\}$.

Definition 3.2. The *pillars* of an n-qubit bipartite quantum system are given by

$$\mathscr{P} = \left\{ v \in \mathbb{F}_2^{2n} \mid v_i = 0 \ \forall i \in \{2, ..., n\} \right\}.$$

The elements of the pillars correspond to the Pauli strings $P_1 \otimes Q_2 \otimes ... \otimes Q_n \in \mathcal{P}_n$ with $P_1 \in \{I, X, Y, Z\}$ and $Q_j \in \{I, Z\}$ for all $j \in \{2, ..., n\}$. The naming of the base and pillars is made clear when the probabilities p_P are ordered in an n-dimensional hypercube, where each dimension corresponds to a qubit pair, see Fig. 2.

Using these definitions, equation (15) can be rewritten as

$$p_{\mathrm{suc}} = \sum_{v \in \mathscr{P}} \tilde{p}_v, \tag{17}$$

and equation (16) as

$$F = \frac{\sum_{v \in \mathscr{B}} \tilde{p}_v}{\sum_{v \in \mathscr{P}} \tilde{p}_v}. \tag{18}$$

The fidelity as defined above corresponds to the I coefficient of the output state, and does not suffice to describe the output state completely.

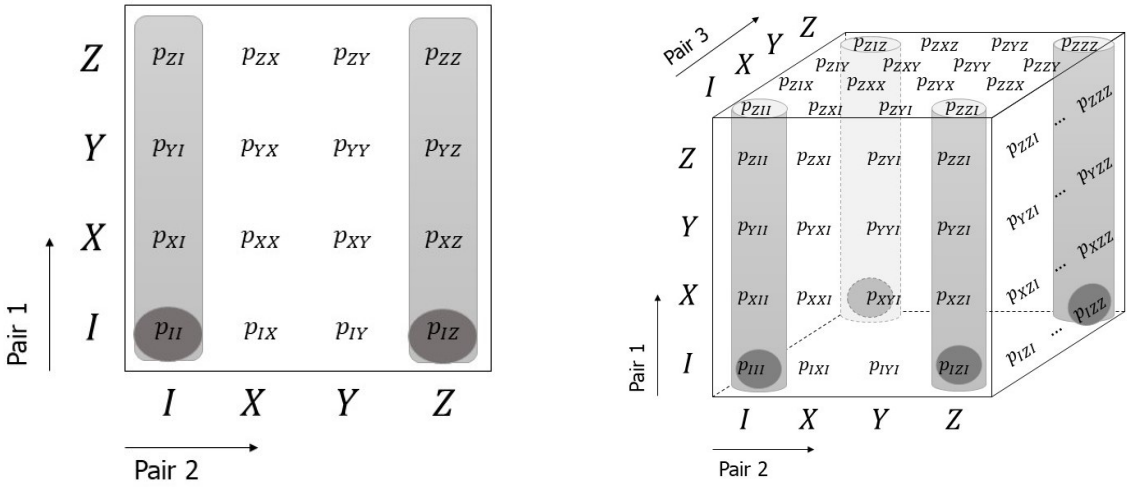


Figure 2: Probabilities that describe the state of a 2-pair system (left) and a 3-pair system (right). The light grey rectangles/cylinders highlight the probabilities that correspond to the pillars. The darker circles highlight the probabilities that correspond to the base. For the 3-pair system we have labelled here only the coefficients that are on the front, right and top face.

To describe the output state, we require the X, Y and Z coefficients as well. Similarly to equation (18), these coefficients F_i are described in terms of the probabilities by

$$F_i = \frac{\sum_{v \in \mathscr{B}+v_i} \tilde{p}_v}{\sum_{v \in \mathscr{P}} \tilde{p}_v}. \tag{19}$$

Here, v_i is $v_1 = e_1, v_2 = e_1 + e_{n+1}, v_3 = e_{n+1}$, corresponding to the X, Y and Z coefficients, respectively and we have used the standard basis vectors $\{e_i : i \in \{1, ..., 2n\}\}$ of $\mathbb{F}_2^{2n}$.

The fidelity, success probability and the three F_i coefficients are referred to as the *distillation statistics*. Importantly, we are not interested in permutations of the three F_i coefficients, since they can be permuted arbitrarily by local operations after the measurement step.

In the binary picture, the distillation statistics can be calculated using the inverse of the symplectic matrix, which can be efficiently calculated using (12). Let M be the symplectic matrix corresponding to a permutation $P \mapsto CPC^\dagger$, $C \in \mathcal{C}_n$. We wish to determine which binary vectors are mapped to the vectors corresponding to the base and the pillars by M. Since M permutes the binary vectors, this is equivalent to determining where the base and pillar vectors are mapped to by M^{-1}.

Finally, there is a direct analogy between our optimisation over bilocal Clifford protocols, and quantum error detection schemes of the form shown in Fig. 3. Such schemes will detect as errors the set of Pauli strings that do not end up in the pillars after applying the Clifford circuit C. We will not pursue this further in this paper, however.

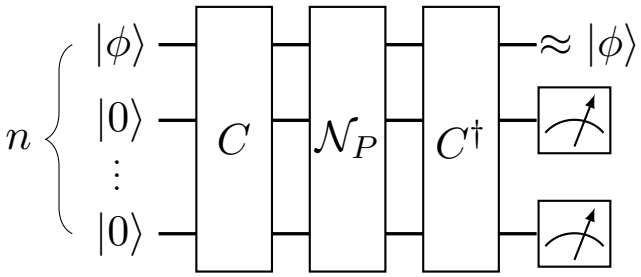


Figure 3: Equivalence between bilocal Clifford protocols and a subset of error detection schemes. This circuit detects as errors the set of Pauli strings that do not end up in the pillars after applying the Clifford circuit C.

4 Preservation of distillation statistics

The distillation statistics from equations (17), (18) and (19) are the relevant parameters for quantifying an entanglement distillation protocol. Furthermore, there exist non-identical bilocal Clifford circuits which result in the same distillation statistics. To find all bilocal Clifford protocols, it is thus sufficient to find a representative bilocal Clifford protocol for each unique collection of distillation statistics. In this section we characterise these representatives for general input states.

First, we specify the set of bilocal Clifford operations that preserve the distillation statistics. We denote this set by $\mathcal{D}_n$. Now observe that $\mathcal{D}_n$ is a subgroup of $\mathrm{Sp}(2n, \mathbb{F}_2)$. Moreover, let $M \in \mathrm{Sp}(2n, \mathbb{F}_2)$ and consider the corresponding distillation protocol. We can freely add or remove elements from $\mathcal{D}_n$ at the end of this protocol, without changing the distillation statistics. That is, all elements in the right coset $\mathcal{D}_n M = \{DM : D \in \mathcal{D}_n\}$ yield the same distillation statistics. Instead of optimising over all possible Clifford circuits it thus suffices to optimise over the right cosets of $\mathcal{D}_n$ in $\mathrm{Sp}(2n, \mathbb{F}_2)$.

4.1 Characterising the subgroup that preserves distillation statistics using base and pillars

In this section we explain the relation between the base and the pillars, which were introduced in definitions 3.1 and 3.2, respectively, and the distillation subgroup $\mathcal{D}_n$.

From equations (17), (18) and (19) it can be observed that for a general initial state, the operations that preserve the distillation statistics are precisely those operations that leave simultaneously both the base and pillars invariant, and permute the three affine subspaces $\mathscr{B} + e_1, \mathscr{B} + e_{n+1}$ and $\mathscr{B} + e_1 + e_{n+1}$. In the following lemma it is first proven that invariance of the base implies invariance of the pillars, and vice versa.

Lemma 4.1. *Let $\mathcal{Q}$ be an n-qubit bipartite quantum system with base $\mathscr{B} \subseteq \mathbb{F}_2^{2n}$ and pillars $\mathscr{P} \subseteq \mathbb{F}_2^{2n}$. Let $\pi : \mathbb{F}_2^{2n} \to \mathbb{F}_2^{2n}$, $\pi(v) = Mv$, with $M \in \mathrm{Sp}(2n, \mathbb{F}_2)$. Then $\pi[\mathscr{B}] = \mathscr{B} \iff \pi[\mathscr{P}] = \mathscr{P}$.*

Proof. We first prove $\pi[\mathscr{B}] = \mathscr{B} \implies \pi[\mathscr{P}] = \mathscr{P}$. For this, we first show that the pillars form the symplectic complement of the base, i.e. $\mathscr{B}^\perp = \mathscr{P}$ (see equation (9)). Recall from Definition 3.1 that $v \in \mathscr{B}$ if and only if $v_i = 0$ for $i = 1, ..., n+1$. Note that $\mathscr{B}$ is a subspace of $\mathbb{F}_2^{2n}$. The symplectic inner product between v and $w \in \mathbb{F}_2^{2n}$, is equal to $\omega(v, w) = v^T \Omega w$. This is equal to zero for all $v \in \mathscr{B}$ if and only if $w_i = 0$ for all $i \in \{2, ..., n\}$, so iff $w \in \mathscr{P}$.

Let $v \in \mathscr{B}$ and $w \in \mathscr{P}$. Then $\omega(v, w) = 0$, and since $M \in \mathrm{Sp}(2n, \mathbb{F}_2^{2n})$, we have that $\omega(\pi(v), \pi(w)) = 0$ as well. Since by assumption $\pi(v) \in \mathscr{B}$, it follows that $\pi(w) \in \mathscr{P}$. Finally, since π is an automorphism, we know that it is bijective and thus $\pi[\mathscr{P}] = \mathscr{P}$.

For the other direction, we use the fact that $\mathscr{P}^\perp = \mathscr{B}$, see equation (10). Then, the above argument can be repeated with $\mathscr{B}$ and $\mathscr{P}$ interchanged to conclude that $\pi[\mathscr{B}] = \mathscr{B} \iff \pi[\mathscr{P}] = \mathscr{P}$. □

We will now show that not only does preservation of the base imply preservation of the fidelity and success probability, but that it also implies that the other coefficients of the output state are preserved, up to a permutation.

Lemma 4.2. *Let $\pi : \mathbb{F}_2^{2n} \to \mathbb{F}_2^{2n}$, $\pi(v) = Mv$, with $M \in \mathrm{Sp}(2n, \mathbb{F}_2)$ such that $\pi[\mathscr{B}] = \mathscr{B}$. Then the coefficients F_1, F_2, F_3 are permuted amongst each other after applying π. In other words, the three coefficients of the output state are stabilised as a set after application of M.*

Proof. We have the following decomposition of the linear space $\mathscr{P}$ into four cosets of the subspace $\mathscr{B}$:

$$\mathscr{P} = \mathscr{B} \cup (\mathscr{B} + e_1) \cup (\mathscr{B} + e_1 + e_{n+1}) \cup (\mathscr{B} + e_{n+1}) .$$

Since $\pi[\mathscr{B}] = \mathscr{B}$, it follows from Lemma 4.1 that $\pi[\mathscr{P}] = \mathscr{P}$. So π permutes the three cosets $v_i + \mathscr{B}$, $v_i = e_1, e_1 + e_{n+1}, e_{n+1}$.

It follows by equation (19) that π permutes the coefficients F_1, F_2, and F_3. □

From Lemma 4.1 and 4.2 we conclude that the operations that preserve the distillation statistics for arbitrary input states are precisely the operations that leave the base invariant. We use this observation to characterise the subgroup $\mathcal{D}_n$ that preserves the distillation statistics. In the trivial case that $n = 1$, we have $\mathcal{D}_1 = \mathrm{Sp}(2, \mathbb{F}_2)$. In this case, the only base element is the identity I, which is always mapped to itself under an automorphism. For all $n > 1$, however, $\mathcal{D}_n$ is a proper subgroup of $\mathrm{Sp}(2n, \mathbb{F}_2)$. Consider for instance the Hadamard gate on the second qubit, which is an element of $\mathrm{Sp}(2n, \mathbb{F}_2)$. This gate induces the swap of X_2 and Z_2 and hereby changes the base.

4.2 Generators of the subgroup that preserves distillation statistics

The goal of this section is to characterise the distillation subgroup $\mathcal{D}_n$. In particular, we find the distillation subgroup in terms of a generating set T_n.

Lemma 4.3. *The distillation subgroup is generated by the set T_n, i.e. $\mathcal{D}_n = \langle T_n \rangle$, where*

$$T_n = \{H_1, S_1, \ldots, S_n\} \cup \{\mathrm{CNOT}_{ij} \mid 1 \leq j < i \leq n\} \cup \{\mathrm{CNOT}_{ij} \mid 2 \leq i < j \leq n\}. \quad (20)$$

Proof. By inspection, each element of T_n preserves the base, so by Lemmas 4.1 and 4.2 we have that $\langle T_n \rangle \subseteq \mathcal{D}_n$. For the other inclusion, let $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathcal{D}_n$. We show that we can reduce such an arbitrary M to the identity matrix by left-multiplication by elements in $\langle T_n \rangle$. An overview of the basic matrix operations corresponding to multiplication by elements of $\mathrm{Sp}(2n, \mathbb{F}_2^{2n})$ is given in Appendix B.

First, note that if $M \in \mathcal{D}_n$, then by definition $M[\mathscr{B}] \subseteq \mathscr{B}$ and $M[\mathscr{P}] \subseteq \mathscr{P}$. In the binary picture this implies that

$$B_{ij} = 0 \text{ if } (i,j) \neq (1,1), \qquad D_{12} = \cdots = D_{1n} = 0.$$

Since M has full rank, we cannot have $B_{11} = D_{11} = 0$. Hence, by multiplying M from the left by I, H_1 or $H_1 S_1$, we may assume that $B_{11} = 0$ (such that $B = 0$) and $D_{11} = 1$.

That M has full rank implies that the last n columns of M are linearly independent. By using CNOT gates from $\{\mathrm{CNOT}_{ij} \mid 1 \leq j < i \leq n\} \subseteq T_n$ and $\{\mathrm{CNOT}_{ij} \mid 2 \leq i < j \leq n\} \subseteq T_n$, the D submatrix can be reduced to the identity matrix.

Since $D = I$ and $B = 0$, it follows from (11) that $A = I$ and C is symmetric. For $1 \leq j < i$ denote $S_{ij} := (S_j \mathrm{CNOT}_{ij})^2 \in \langle T_n \rangle$. Left-multiplication by S_{ij} corresponds to adding row i to row $n+j$ and adding rows i and j to row $n+i$. Note that this preserves the fact that $A = D = I$ and $B = 0$.

For $j = 1, \ldots, n-1$ (in this order), we can multiply M from the left by elements from $\{S_j\} \cup \{S_{ij} \mid i > j\} \subseteq \langle T_n \rangle$ to ensure that $C_{ji} = 0$ for all $1 \leq j \leq i \leq n$. This implies that C is strictly lower triangular. But if C is strictly lower triangular and symmetric, $C = 0$. This implies that $M = I$. □

4.3 Order of the subgroup that preserves distillation statistics

As noted before, for general input states it is sufficient to only consider the right cosets of $\mathcal{D}_n$ in $\mathrm{Sp}(2n, \mathbb{F}_2)$. To see how much looking at cosets of $\mathcal{D}_n$ in $\mathrm{Sp}(2n, \mathbb{F}_2)$ limits the search space of protocols, in this section a formula for the order of $\mathcal{D}_n$ is presented and proved. As mentioned earlier, in the trivial case that $n = 1$ we have $\mathcal{D}_1 = \mathrm{Sp}(2, \mathbb{F}_2)$, and thus $|\mathcal{D}_1| = |\mathrm{Sp}(2, \mathbb{F}_2)| = 6$. For $n \geq 2$ the order of $\mathcal{D}_n$ is given in Theorem 4.4.

Theorem 4.4. *For an n-to-1 distillation protocol, with $n > 1$, the order of $\mathcal{D}_n$ is given by*

$$|\mathcal{D}_n| = 6 \cdot 2^{n^2-1} \prod_{j=1}^{n-1} (2^j - 1).$$

Proof. First note that $D \in \mathcal{D}_n$ is fully determined by how it maps each of the standard basis vectors $\{e_i : i \in \{1, ..., 2n\}\}$ of $\mathbb{F}_2^{2n}$. We count how many transformations of the standard basis vectors are possible.

Let us start by looking at e_{2n}. This is a base element, thus it must again be transformed to a base element, because D preserves the distillation statistics. There are 2^{n-1} base elements, but the identity element, the zero vector, is always mapped to itself by D. Thus there are $2^{n-1} - 1$ possibilities for the transformation of e_{2n}. That all transformations are indeed possible, is proved by giving a construction. Suppose that e_{2n} is mapped to a base element $v \equiv De_{2n} \in \mathscr{B}$. We show that v can be transformed to e_{2n} through left-multiplication by elements of $\mathcal{D}_n$, see Appendix B. The transformation from e_{2n} to v can then be obtained by taking the product of the inverses of these generators in reverse order.

Note that $v_1, ..., v_{n+1} = 0$. The vector v can be transformed to e_{2n} by taking the following steps.

1. If $v_{2n} = 0$, apply a $\mathrm{CNOT}_{ni} \in \mathcal{D}_n$ gate with i chosen such that $v_{n+i} = 1$. Note that there always is a i such that this is possible, because otherwise v is the zero vector, which corresponds to the identity element $I^{\otimes n}$.
2. For all $i \in \{2, ..., n\}$ with $v_{n+i} = 1$, apply a $\mathrm{CNOT}_{in} \in \mathcal{D}_n$ gate.

Steps 1 and 2 are visually summarised below.

$$v = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline 0 \\ \cdot \\ \cdot \end{array}\right] \xrightarrow{1} \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline 0 \\ \cdot \\ 1 \end{array}\right] \xrightarrow{2} \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 1 \end{array}\right] = e_{2n}$$

Given the transformation of e_{2n} by D, we now wish to determine the number of possible transformations for e_n. We know that left-multiplication by D preserves the symplectic inner product. Hence, since $\omega(e_n, e_{2n}) = 1$, it must hold that $\omega(De_n, De_{2n}) = 1$. Observe that for every non-zero element $u \in \mathbb{F}_2^{2n}$, exactly for half of the elements of $\mathbb{F}_2^{2n}$ the symplectic inner product with u is equal to one[1]. Thus there are $\frac{|\mathbb{F}_2^{2n}|}{2} = \frac{4^n}{2} = 2^{2n-1}$ possibilities for the transformation of e_n.

We show that each of those transformations can indeed be achieved. Suppose that D has mapped e_n to a vector $w \equiv De_n \in \mathbb{F}_2^{2n}$. Because $\omega(De_n, De_{2n}) = 1$ and De_{2n} is a base vector, we know that there is at least one $i \in \{2, \ldots, n\}$ such that $w_i = 1$. Since we can always apply a $\mathrm{CNOT}_{in} \in \mathcal{D}_n$ gate, which does not affect the vector e_{2n}, we can assume without loss of generality that $w_n = 1$. Now w can be transformed to e_n without affecting e_{2n} by taking the following steps.

3. For all i with $w_i = 1$ apply a $\mathrm{CNOT}_{ni} \in \mathcal{D}_n$ gate.
4. For all $i \neq n$ with $w_{n+i} = 1$, apply a $\mathrm{CZ}_{in} \in \mathcal{D}_n$ gate. This operation results in the addition of row i to row $2n$ and the addition of row n to row $n+i$. Note that this operation leaves the base invariant, so indeed $\mathrm{CZ}_{in} \in \mathcal{D}_n$.
5. If $w_{2n} = 1$, apply the gate $S_n \in \mathcal{D}_n$ on qubit n.

Steps 3 to 5 are visually summarised below.

$$w = \begin{bmatrix} \cdot \\ \cdot \\ 1 \\ \hline \cdot \\ \cdot \\ \cdot \end{bmatrix} \xrightarrow{3} \begin{bmatrix} 0 \\ 0 \\ 1 \\ \hline \cdot \\ \cdot \\ \cdot \end{bmatrix} \xrightarrow{4} \begin{bmatrix} 0 \\ 0 \\ 1 \\ \hline 0 \\ 0 \\ \cdot \end{bmatrix} \xrightarrow{5} \begin{bmatrix} 0 \\ 0 \\ 1 \\ \hline 0 \\ 0 \\ 0 \end{bmatrix} = e_n$$

Thus indeed, given the transformation of e_{2n}, there are 2^{2n-1} possible transformations of e_n. Combining this with the number of transformations of e_{2n}, we find that there are $2^{2n-1}(2^{n-1}-1)$ possible transformations for e_n and e_{2n} together.

[1] Let $u \in \mathbb{F}_2^{2n}$, such that $u_k = 1$. Then the vectors $u' \in \mathbb{F}_2^{2n}$ satisfying $\omega(u, u') = 1$ can be constructed by choosing $u'_j \in \{0, 1\}$ randomly for $j \in \{1, \ldots, 2n\} \setminus \{n+k\}$ and then choosing $u'_{n+k} \in \{0, 1\}$ such that $\omega(u, u') = 1$.

The elements of $\mathcal{D}_n$ that leave e_n and e_{2n} invariant form a subgroup that is isomorphic to $\mathcal{D}_{n-1}$, with the number of cosets in $\mathcal{D}_n$ equal to $2^{2n-1}(2^{n-1}-1)$. Thus

$$|\mathcal{D}_n| = 2^{2n-1}(2^{n-1}-1)|\mathcal{D}_{n-1}|.$$

By induction on n it follows that

$$\begin{aligned} |\mathcal{D}_n| &= |\mathcal{D}_1| \prod_{j=2}^{n} 2^{2j-1}(2^{j-1}-1) \\ &= 6 \cdot 2^{\sum_{j=2}^{n}(2j-1)} \prod_{j=2}^{n}(2^{j-1}-1) \\ &= 6 \cdot 2^{n^2-1} \prod_{j=1}^{n-1}(2^j - 1). \end{aligned}$$

□

The following corollary is a direct consequence of Theorem 4.4.

Corollary 4.5. *The index of $\mathcal{D}_n$ in $\mathrm{Sp}(2n, \mathbb{F}_2)$ is given by*

$$[\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n] = \frac{1}{3}(2^n - 1) \prod_{j=1}^{n}(2^j + 1).$$

Proof. Recall that $|\mathrm{Sp}(2n, \mathbb{F}_2)| = 2^{n^2} \prod_{j=1}^{n}(4^j - 1)$. As a result,

$$\begin{aligned} [\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n] &= \frac{|\mathrm{Sp}(2n, \mathbb{F}_2)|}{|\mathcal{D}_n|} \\ &= \frac{2^{n^2} \prod_{j=1}^{n}(4^j - 1)}{6 \cdot 2^{n^2-1} \prod_{j=1}^{n-1}(2^j - 1)} \\ &= \frac{\prod_{j=1}^{n}(2^j - 1)(2^j + 1)}{3 \prod_{j=1}^{n-1}(2^j - 1)} \\ &= \frac{1}{3}(2^n - 1) \prod_{j=1}^{n}(2^j + 1). \end{aligned}$$

□

For comparison, we list the values of $|\mathrm{Sp}(2n, \mathbb{F}_2)|$ and $[\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n]$ in Table 1 for $n = 2, 3, 4, 5$.

4.4 Finding a transversal

In this section, we briefly describe a way to find a transversal for the cosets of $\mathcal{D}_n$ in $\mathrm{Sp}(2n, \mathbb{F}_2)$. A transversal is a set that contains exactly one element for each of the cosets. Once this

	2	3	4	5
$\lvert \mathrm{Sp}(2n, \mathbb{F}_2)\rvert$	720	1451520	47377612800	24815256521932800
$[\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n]$	15	315	11475	782595

Table 1: Values of $\lvert \mathrm{Sp}(2n, \mathbb{F}_2)\rvert$ and $[\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n]$ for $n = 2, 3, 4, 5$.

transversal is found, it can be applied to an arbitrary n-qubit input state to calculate all possible distillation statistics that can be achieved using bilocal Clifford circuits. From this set of distillation statistics, the optimal protocol based on any optimality criterion can be selected.

In order to find a transversal, random elements from the symplectic group $\mathrm{Sp}(2n, \mathbb{F}_2)$ are sampled. A sampled element is added to the set of representatives if the corresponding coset is not yet represented in this set. Recall that two elements belong to the same coset if they result in the same distillation statistics (for a general input state). This is the case if and only if the same Pauli strings are mapped to the base. More formally, consider an n-qubit pairs bipartite system with base $\mathscr{B}$ in the binary picture. Let $M_1,\ M_2 \in \mathrm{Sp}(2n, \mathbb{F}_2)$. Let $\mathscr{V}$ denote the set of binary vectors that are mapped to the base by M_1 and let $\mathscr{W}$ denote the set of binary vectors that are mapped to the base by M_2. Then M_1 and M_2 belong to the same coset if and only if $\mathscr{V} = \mathscr{W}$. Because M_1 and M_2 permute the binary Pauli vectors, this is equivalent to $M_1^{-1}[\mathscr{B}] = M_2^{-1}[\mathscr{B}]$.

The sampling is continued until the set of representatives has size $[\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n]$. Note that finding a transversal in the way described in this section is equivalent to the coupon collector's problem. Hence, it has expected running time $\mathcal{O}([\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n] \log[\mathrm{Sp}(2n, \mathbb{F}_2) : \mathcal{D}_n])$.

5 Reduction for n-fold tensor products of Werner states

Here we describe our reduction of the search space when the input state is an n-fold tensor product of Werner states. A Werner state has coefficients $p_I = F_{\mathrm{in}},\ p_X = p_Y = p_Z = \frac{1-F_{\mathrm{in}}}{3}$, and its n-fold tensor product is highly symmetric — it is left invariant under any element of $\mathcal{K}_n$, where

$$\mathcal{K}_n = \langle\, \{\mathrm{SWAP}_{ij}\}_{1\le i<j\le n} \cup \{H_i\}_{i=1}^n \cup \{S_i\}_{i=1}^n \,\rangle.$$

We leverage this symmetry by noting that the distinct distillation protocols correspond to the double cosets $\mathcal{D}_n \backslash \mathrm{Sp}(n, \mathbb{F}_2)/\mathcal{K}_n$, similar to our argument before for right cosets for general input states. In this section, we describe how one can rewrite an arbitrary symplectic matrix M to another symplectic matrix M' of a specific form, which is in the same double coset as M. Such a representative M' of the double coset has a smaller number of free parameters, reducing the search space significantly.

Recall that an overview of the basic matrix operations corresponding to elements of $\mathrm{Sp}(2n, \mathbb{F}_2^{2n})$ is given in Appendix B.

Lemma 5.1. *Let $M \in \mathrm{Sp}(2n, \mathbb{F}_2)$. There exists M' in the double coset $\mathcal{D}_n M \mathcal{K}_n$ that is of the form*

$$M' = \begin{bmatrix} A & B \\ 0 & A^{\mathsf{T}} \end{bmatrix},\ A = \left[\begin{array}{c|c} 1 & 0 \\ \hline a & I_{n-1} \end{array}\right],$$

$$B = \left[\begin{array}{c|c} 0 & b^{\mathsf{T}} \\ \hline b & E + ba^{\mathsf{T}} \end{array}\right],$$

where $a, b \in \mathbb{F}_2^{n-1}$ and $E \in \mathbb{F}_2^{(n-1)\times(n-1)}$ is symmetric with zeroes on the diagonal.

Proof. Let $M' \in \mathcal{D}_n M \mathcal{K}_n$ be such that

$$M'_{ij} = \delta_{ij} \quad \text{for } (i,j) \in [n] \times \{2, \dots, k\} \tag{21}$$

with $1 \le k \le n$ as large as possible. Note that for $k = 1$ the condition is trivially fulfilled.

Claim: $k = n$.
Proof of claim. Suppose that $k < n$. Then $M'_{k+1,k+1} = 0$, otherwise we can use row operations on M' (left-multiplication by matrices $\mathrm{CNOT}_{k+1,i} \in \mathcal{D}_n$) to obtain $M'_{i,k+1} = \delta_{i,k+1}$ for all $i \in [n]$ while keeping (21), contradicting the maximality of k.

Note that the above condition $M'_{k+1,k+1} = 0$ needs to hold after applying operations to M' that preserve the form in equation (21).

Thus, by permuting rows in $\{k+1,\dots,n\}$ (left-multiplication by matrices $\mathrm{SWAP}_{ij} \in \mathcal{D}_n$) or permuting columns in $\{1,k+1,\dots,n\}$ (right-multiplication by matrices $\mathrm{SWAP}_{ij} \in \mathcal{K}_n$) we deduce that $M'_{ij} = 0$ for $(i,j) \in \{k+1,\dots,n\} \times \{1,k+1,\dots,n\}$. Since we can swap column i and $n+i$ by multiplying from the right with $H_i \in \mathcal{K}_n$, we also have $M'_{ij} = 0$ for $(i,j) \in \{k+1,\dots,n\} \times \{n+1,n+k+1,\dots,2n\}$. To summarise, we have

$$\begin{aligned} M'_{ij} &= 0 \quad \text{for } i \in \{k+1,\dots,n\} \\ &\qquad \text{and } j \in [2n] \setminus \{n+2,\dots,n+k\} \\ M'_{ij} &= \delta_{ij} \quad \text{for } i \in \{2,\dots,k\} \text{ and } j \in \{2,\dots,k\}. \end{aligned}$$

Since rows $k+1,\dots,n$ must have zero symplectic inner product with rows $2,\dots,k$, it follows that rows $k+1,\dots,n$ must in fact be equal to zero. Since M' has full rank, this implies that $k = n$. ■

Consider the first row of M'. We have $M'_{1,j} = 0$ for $j = 2,\dots,n$. If $M'_{1,1} = M'_{1,n+1} = 0$, then the fact that this row has zero symplectic inner product with rows $2,\dots,n$ implies that the first row is equal to zero, which is not possible as M' has full rank. So by multiplying from the right by I, H_1, or S_1H_1 which are in $\mathcal{K}_n$, we may assume that $M'_{1,1} = 1$ and $M'_{1,n+1} = 0$.

Writing $M' = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$, we see that A and B have the following form:

$$A = \left[\begin{array}{c|c} 1 & 0 \\ \hline a & I_{n-1} \end{array}\right], \quad B = \left[\begin{array}{c|c} 0 & d^{\mathsf{T}} \\ \hline b & E' \end{array}\right].$$

Since row 1 has zero symplectic inner product with rows $2,\dots,n$, it follows that $d = b$. Note that for $1 \le i < j \le n-1$ the symplectic inner product of rows $i+1$ and $j+1$ is equal to $a_ib_j + E'_{ji} + a_jb_i + E'_{ij}$. Since this inner product is zero, the matrix $E := E' + ba^{\mathsf{T}}$ is symmetric. By multiplying from the right by $H_iS_iH_i \in \mathcal{K}_n$ $(i = 2,\dots,n)$ if necessary, we may set the diagonal elements of E' such that the diagonal elements of E are zero.

Recall from Lemma 4.3 that $S_{ij} := (S_j\mathrm{CNOT}_{ij})^2 \in \mathcal{D}_n$ for $1 \le j < i$. Recall furthermore that left-multiplication of M' by S_{ij} amounts to adding row i to row $n+j$ and adding rows i and j to row $n+i$. By left-multiplication by elements $S_{ij} \in \mathcal{D}_n$ and $S_i \in \mathcal{D}_n$ we may (without changing the first n rows of M') assume that C is a strictly upper triangular matrix. Since the first n columns of M' must have pairwise zero symplectic inner product, this implies that in fact $C = 0$. Since $A^{\mathsf{T}}D + C^{\mathsf{T}}B = I_n$, it follows that $D = (A^{\mathsf{T}})^{-1} = A^{\mathsf{T}}$, where we have used that A is self-inverse.

□

Note that for any permutation $\pi \in S_{n-1}$, we can replace a, b, E by $\pi(a), \pi(b), \pi(E)$ (permuting both rows and columns) by multiplying M' simultaneously from the left and the right by elements SWAP_{ij}, since SWAP_{ij} is an element of $\mathcal{D}_n$ and $\mathcal{K}_n$ for $2 \le i < j \ne n$. Also, we can replace (a,b) by (b,a) or $(a,a+b)$ by multiplication from the left and right by elements from $\{S_1, H_1\}$. Hence, to cover all cases, it suffices to enumerate over the triples (a,b,E) where $a \le b \le a+b$ and E runs over the adjacency matrices of graphs on $n-1$ nodes (up to isomorphism).

6 Optimisation results

In the previous sections we have outlined our methods for finding all possible bilocal Clifford protocols, which were described in Section 3. In the following we report our findings, first for up to $n = 5$ general Bell-diagonal input states, second for up to $n = 8$ identical Werner states.

6.1 Achieved distillation statistics for general input states

In Fig. 4 we show the achievable $(p_{\mathrm{suc}},\ F_{\mathrm{out}})$ pairs for $n = 2,3,4,5$ copies of a state with coefficients $p_I = 0.7$, $p_X = 0.15$, $p_Y = 0.10$, $p_Z = 0.05$. We also plot the envelope, indicating the best performing schemes. Moreover, our results for $n = 5$ clearly show that picking an arbitrary coset does not give a good protocol in general.

Furthermore, we consider the $n = 2$ scenario where the two input states are equal, i.e. both have equal values for the p_I, p_X, p_Y, p_Z parameters. By comparing all analytic expressions of the output fidelity as a function of p_I, p_X, p_Y, p_Z, we find that the DEJMPS protocol achieves the highest fidelity out of all bilocal Clifford protocols (see [48] for the details).

While we do not explore this direction, the results can also be applied to less symmetric cases, i.e. when the n pairs are not the same. This situation is, for example, relevant when states

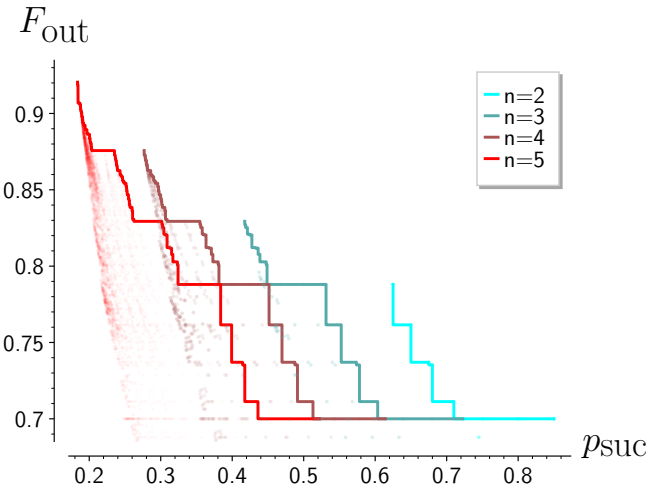


Figure 4: Achievable $(p_{\text{suc}},\ F_{\text{out}})$ pairs for $n = 2$ to 5 copies of a state with coefficients $p_I = 0.7$, $p_X = 0.15$, $p_Y = 0.10$, $p_Z = 0.05$. The highest achievable pairs are indicated by a solid line for each number of copies. Not included in the plot are those distillation protocols with fidelity smaller than $F = 0.68$.

arrive at different times, and thus experience different amounts of decoherence.

6.2 Achieved distillation statistics for n-fold Werner states

Here we show our results for the case of an n-fold tensor product of a Werner state. First, we list the number of cases to check (i.e. the number of triples (a, b, E), see Section 5) and the number of distinct distillation protocols for this scenario in Table 2.

The number of cases and distinct protocols still rapidly grow with n, but our reduction allowed to consider all possible distillation protocols for $n = 8$ in about a day of computer run-time. This should be compared with a naive optimisation over all elements of the symplectic group — for $n = 8$ the ratio between the order of the symplectic group and the number of cases to check is approximately $2 \cdot 10^{34}$. For $n = 9$ and higher however, our current method becomes infeasible, requiring too many cases to consider. One could imagine improving on Lemma V.1 as to reduce the gap between the number of cases to check and the number of distinct distillation protocols. This could potentially only allow for an enumeration up to $n \approx 11$. This can be made more quantitative by considering that a lower bound on the number of double cosets is given by $\frac{|\mathrm{Sp}(2n,\mathbb{F}_2)|}{|\mathcal{D}_n||\mathcal{K}_n|} = \mathcal{O}\left(\frac{2^{\frac{n^2}{2}}}{6^n n!}\right)$, since the size of a double coset can be at most $|\mathcal{D}_n| \cdot |\mathcal{K}_n|$. We note here that we are approximating the number of distinct distillation statistics by the number of double cosets. With such a lower bound, it is clear that a full enumeration becomes infeasible for $n \gtrapprox 11$, even in the best-case scenario.

We note here that the large number of distinct protocols only means that a full enumeration of the distillation protocols becomes infeasible for larger n. In particular, it does not necessarily preclude an optimisation of the distillation protocols for a given metric, such as the output fidelity. In this paper, however, we only consider an optimisation by first fully enumerating all distinct protocols.

In order to gauge the advantage of the optimal protocols that we find for Werner states, we compare them with the class of protocols we call *concatenated DEJMPS protocols*. These are bilocal Clifford protocols that are built from multiple iterations of the DEJMPS protocol [10], see Appendix C for more information. The concatenated DEJMPS protocols form a natural generalisation of the (nested) entanglement pumping protocols [11].

We first investigate the increase in fidelity $F_{\text{out}} - F_{\text{in}}$ conditioned on the success of the distillation protocol. We plot the increase in fidelity as a function of the input fidelity F_{in} for $n = 2, 3, \ldots, 8$ in Fig. 5. The dotted lines correspond to the concatenated DEJMPS protocols, the solid lines correspond to the protocols that achieve the highest output fidelity found with our optimisation. For completeness, we show the success probabilities and fidelities for the optimised protocols for $n = 2, 3, \ldots, 8$ in Tables 5, 6, 7 and 8 in Appendix E.

Let us now discuss Fig. 5. First we observe that for $n = 2, 3$, the optimal protocols correspond to the original DEJMPS [10] and double-selection [37] protocols. However, for $n > 3$, we find distillation protocols that outperform the concatenated DEJMPS protocols.

We find that the optimal protocol for $n = 4$ achieves the same fidelity as the concatenated DEJMPS protocol for $n = 5$, and can be executed with a circuit of the same depth as the concatenated DEJMPS protocol. This protocol achieves the same distillation statistics as the protocol found with different means in the recent work from [49].

	2	3	4	5	6	7	8
Cases	2	10	60	561	6358	111540	2917980
Distinct protocols	2	5	13	34	108	379	1736

Table 2: Number of cases to check and number of distinct distillation protocols for n identical Werner states.

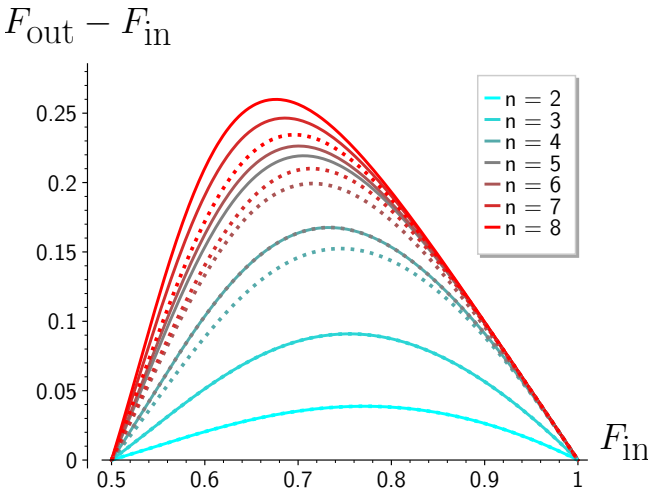


Figure 5: Comparison between the increase in fidelity $F_{\text{out}} - F_{\text{in}}$ with our optimisation (solid) and concatenated DEJMPS protocol (dotted), for $n = 2$ to 8 identical Werner states with fidelity F_{in}. Note how the $n = 5$ concatenated DEJMPS protocol overlaps with an optimised $n = 4$ protocol.

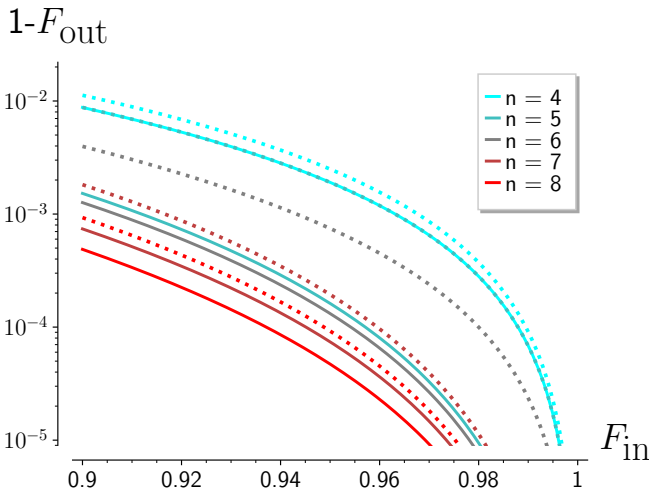


Figure 6: Comparison between the achieved infidelities $1 - F_{\text{out}}$ with our optimisation (solid) and concatenated DEJMPS protocol (dotted), for $n = 4$ to 8 identical Werner states with fidelity F_{in}.

For $n = 5$ there is a large gap between the optimised protocols and the concatenated DEJMPS protocol. We make this now more quantitative by expanding the F_{out} for high input fidelity $F_{\text{in}} \approx 1$. For $n = 5$, the concatenated DEJMPS protocol has quadratic scaling in the infidelity,

$$1 - \frac{2}{3}(1-F)^2 + \mathcal{O}\left((1-F)^3\right) , \tag{22}$$

while the optimised protocol has a cubic scaling in the infidelity

$$1 - \frac{10}{9}(1-F)^3 + \mathcal{O}\left((1-F)^4\right) . \tag{23}$$

This is particularly surprising, since previous protocols with five or less pairs [36] have a scaling that is at most quadratic in the infidelity. We list the scaling of the found protocols in Table. 9.

Next, we investigate the behavior of the protocols for high fidelities $F_{\text{in}} \approx 1$. In Fig. 6 we plot the infidelity $1 - F_{\text{out}}$ as a function of the input fidelity F_{in}. We observe that it is possible to reach fidelities of around 0.999 using six copies of Werner states with fidelity $F_{\text{in}} = 0.9$. We do not plot the results for $n = 2, 3$ since we find no improvements with respect to previous protocols.

We have seen that the optimised distillation protocols are capable of achieving a higher fidelity than the concatenated DEJMPS protocols. However, the optimised distillation protocols also have a lower success probability. This motivates us to investigate three metrics, each of which combines the success probability and the quality of the resultant state. As the first metric, we use the distillable entanglement rate which we approximate by combining the distillation protocol together with a *hashing* protocol [8]. That is, given n entangled pairs, we first perform an n-to-1 distillation protocol and then use the output as input for the hashing protocol. The rate r at which this procedure produces maximally entangled state is given by

$$r = \frac{(1 - H(p)) \cdot p_{\text{suc}}}{n} , \tag{24}$$

where $H(p)$ is the entropy in bits of the probability distribution $p = (p_{\text{I}}, p_X, p_Y, p_Z)$ correspond-

ing to the output state. This metric has been used previously and is sometimes called the hashing yield [36].

We compare the achieved rate of all found distillation protocols and the concatenated DEJMPS protocols in Fig. 7. We show both achieved rates and the ratio between them. We find that for $n > 3$ and fidelities $\lesssim 0.76$ the optimal bilocal Clifford protocols achieve up to rates three times greater than concatenated DEJMPS protocols. Conversely, for high fidelities it suffices to use concatenated DEJMPS protocols if one is interested in maximising the asymptotic rate.

For the second metric, we consider the more practical application of achieving a certain threshold fidelity F_{tar} using (at most) one round of distillation. While there exist a number of applications that require a minimum fidelity, we consider here device-independent quantum key distribution. If one assumes only depolarising noise, a bound on the minimum fidelity is given [50] by $F_{\text{tar}} = 0.930025$ to perform device-independent quantum key distribution. In what follows, we will assume that it is necessary to achieve F_{tar}, and that the other coefficients of the output state are equal.

In Fig. 8 we show the average rate at which entanglement can be distilled to F_{tar} using a single round of distillation. We find a similar behaviour as in Fig. 7, where for higher fidelities (≈ 0.78) it suffices to perform concatenated DEJMPS protocols, while for lower fidelities the optimised protocols achieve a larger rate.

Finally, we consider the metric of the success probability times the *relative entropy of entanglement* [51, 52, 53]. The relative entropy of entanglement (REE) is an upper bound on the distillable entanglement [53], and can be computed exactly for Bell-diagonal states [51]. The success probability times the REE is a quantity that has been used to capture how well a protocol concentrates the present entanglement, and has even been used to show optimality of some distillation protocols, see [34]. We note that we have found that the success probability times the REE decreases as the number of input states increases. We thus consider maximising the above quantity for fixed values of n. To this end, we plot the difference between the success probability and the REE for all optimised protocols and concatenated DEJMPS protocols

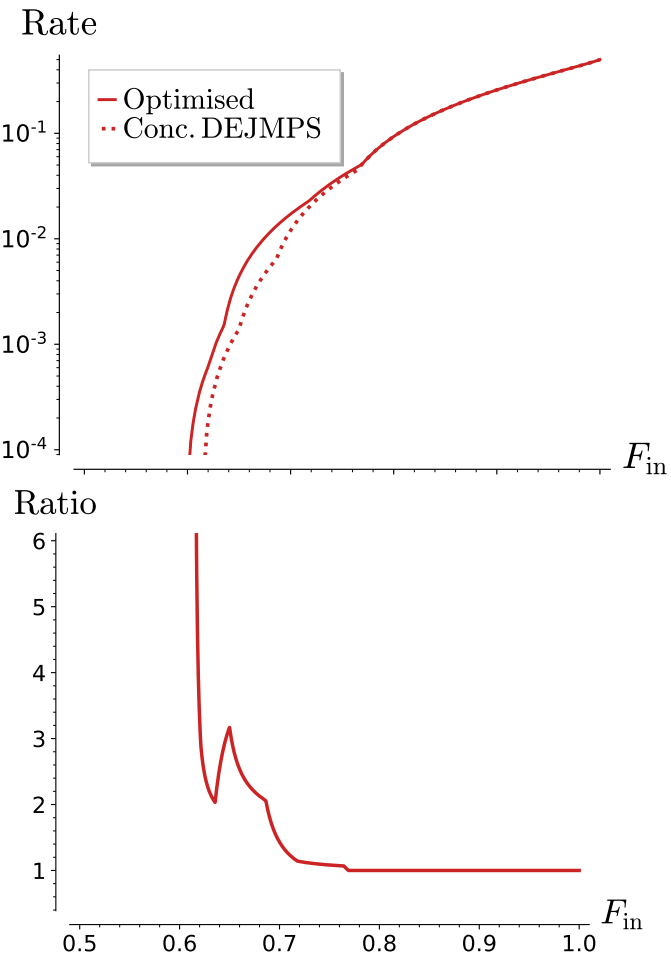


Figure 7: Comparison between the achieved rates after distilling and then hashing with our optimisation (solid) and the concatenated DEJMPS (dotted) protocol. For both cases, we take the envelope of all protocols on $n = 2$ to 8 identical Werner states with fidelity F_{in}. Top) achieved rates with our optimisation and concatenated DEJMPS protocols. Bottom) Ratio between the rate with our optimisation and concatenated DEJMPS protocols.

in Fig. 9 for $n = 2, \ldots, 8$. We find for all $n > 3$ that the full optimisation over bilocal Clifford protocols allows for a larger value of the success probability times the REE, in particular for higher fidelities. This should be contrasted with the results from Figs. 7 and 8, where the full optimisation only showed an improvement over concatenated DEJMPS protocols for lower fidelities.

Let us conclude with an investigation of circuits that achieve the highest fidelity for $n = 4$ to 8. Interestingly, these protocols can be implemented with low-depth circuits. We performed a search over circuits of the form described in Appendix D, to find circuits that achieve the highest fidelity.

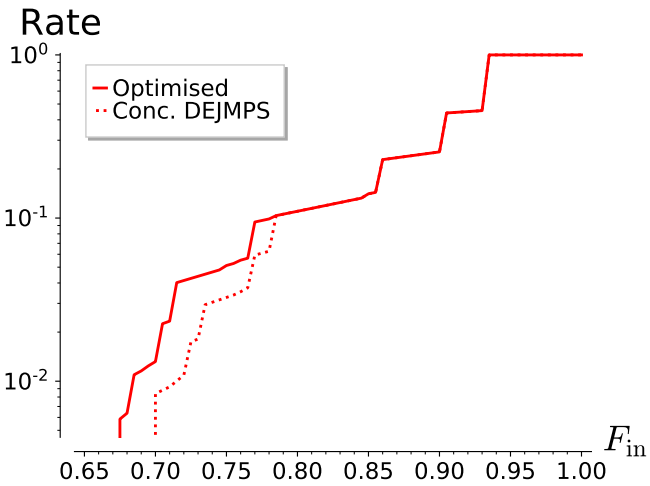


Figure 8: Average rate at which entanglement can be distilled to a minimum fidelity of $F_{\text{tar}} = 0.930025$, using both our optimisation (solid) and concatenated DEJMPS (dotted) protocols.

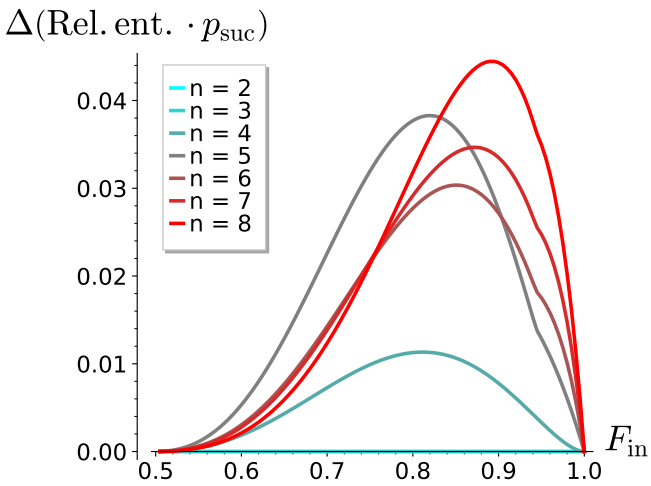


Figure 9: Difference between the product of the success probability and relative entropy of entanglement of the resultant state, optimised over all bilocal Clifford protocols and all concatenated DEJMPS protocols.

We report these circuits in Appendix D. For $n = 4$ to 8, we find a total number of two-qubit gates of 4, 7, 8, 11 and 13. Furthermore, the corresponding circuit depths are 3, 5, 6, 6 and 7, respectively. For comparison, the circuit from [49] for $n = 4$ pairs has 4 two-qubit gates and depth 5. This protocol can be converted to our optimal $n = 4$ protocol by left-multiplication with elements in $\mathcal{D}_n$ and right-multiplication with elements in $\mathcal{K}_n$. Therefore, both protocols achieve the same distillation statistics. The protocol from [54] for $n = 5$ pairs, which achieves the same fidelity and success probability as the concatenated DEJMPS protocol, has 4 two-qubit gates and depth 4. We note here that the fidelity and success probability do not necessarily need to correspond to a unique specific distillation protocol. As an example, for $n = 8$ there are four distinct protocols that achieve the highest fidelity. These protocols have the same fidelity and success probability, but differ in their F_i components.

7 Conclusions and discussions

Our goal in this paper was to find good distillation protocols requiring modest resources. For this, we introduced the class of bilocal Clifford protocols which generalises many existing protocols. The protocols in this class require only a single round of communication between the end parties and the implementation of Clifford gates. Within this class, we leveraged group theoretic tools to find all distillation protocols for up to $n = 5$ pairs for general Bell-diagonal states and up to $n = 8$ pairs for the n-fold tensor product of a Werner state.

Some of the protocols that we found strongly improve upon the fidelities and rates of previous protocols. Moreover, we give explicit circuits for the optimal protocols for the n-fold Werner state case, with $n = 2$ to $n = 8$. These circuits have comparable depth and number of two-qubit gates as previous protocols, indicating the experimental feasibility of the new protocols. If the improved performance holds with noisy operations, then it will translate in improved forecasts for the performance of near-term quantum networks [54, 55] or distributed quantum computation [7]. Finally, since we have enumerated all bilocal Clifford protocols up to $n = 5$, it is possible to pick and choose the protocol that maximises any figure of merit for any particular set of input states. Our software can be found at [48].

In this work we considered distilling one entangled pair out of n pairs. The results here could be extended to n to m distillation protocols by generalising Lemma 5.1 and the characterisation of the distillation subgroup to the case of n to m distillation. Such distillation protocols would allow for a more refined trade-off between the fidelity and the success probability/rate. Another interesting avenue would be to generalise the tools to higher dimensional entangled states.

Acknowledgements

This work was supported by the QIA project (funded by European Union's Horizon 2020, Grant Agreement No. 820445) and by the Netherlands Organization for Scientific Research (NWO/OCW), as part of the Quantum Software Consortium program (Project No. 024.003.037/3368). The authors thank T. Coopmans for initial discussions, and F. Rozpędek and S. Bäuml for valuable feedback on the manuscript.

References

[1] Charles H. Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". Theoretical Computer Science **560**, 7–11 (2014).

[2] Kai Chen and Hoi-Kwong Lo. "Conference key agreement and quantum sharing of classical secrets with noisy ghz states". In Proceedings. International Symposium on Information Theory, 2005. ISIT 2005. Pages 1607–1611. IEEE (2005).

[3] Jérémy Ribeiro, Gláucia Murta, and Stephanie Wehner. "Fully device-independent conference key agreement". Physical Review A **97**, 022307 (2018).

[4] Richard Jozsa, Daniel S. Abrams, Jonathan P. Dowling, and Colin P. Williams. "Quantum clock synchronization based on shared prior entanglement". Physical Review Letters **85**, 2010 (2000).

[5] Lov K. Grover. "Quantum telecomputation" (1997). arXiv:quant-ph/9704012.

[6] J. Ignacio Cirac, Artur Ekert, Susana F. Huelga, and Chiara Macchiavello. "Distributed quantum computation over noisy channels". Physical Review A **59**, 4249 (1999).

[7] Naomi H. Nickerson, Ying Li, and Simon C. Benjamin. "Topological quantum computing with a very noisy network and local error rates approaching one percent". Nature communications **4**, 1–5 (2013).

[8] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters. "Purification of noisy entanglement and faithful teleportation via noisy channels". Physical Review Letters **76**, 722 (1996).

[9] Charles H. Bennett, David P. Divincenzo, John A. Smolin, and William K. Wootters. "Mixed-state entanglement and quantum error correction". Physical Review A **54**, 3824–3851 (1996).

[10] David Deutsch, Artur Ekert, Richard Jozsa, Chiara Macchiavello, Sandu Popescu, and Anna Sanpera. "Quantum privacy amplification and the security of quantum cryptography over noisy channels". Physical Review Letters **77**, 2818–2821 (1996).

[11] Wolfgang Dür and Hans J. Briegel. "Entanglement purification and quantum error correction". Reports on Progress in Physics **70**, 1381 (2007).

[12] Igor Devetak and Andreas Winter. "Distillation of secret key and entanglement from quantum states". Proceedings of the Royal Society A: Mathematical, Physical and engineering sciences **461**, 207–235 (2005).

[13] Francesco Buscemi and Nilanjana Datta. "Distilling entanglement from arbitrary resources". Journal of Mathematical Physics **51**, 102201 (2010).

[14] Felix Leditzky, Nilanjana Datta, and Graeme Smith. "Useful states and entanglement distillation". IEEE Transactions on Information Theory **64**, 4689–4708 (2017).

[15] Vlatko Vedral. "On bound entanglement assisted distillation". Physics Letters A **262**, 121–124 (1999).

[16] Satoshi Ishizaka. "Bound entanglement provides convertibility of pure entangled states". Physical Review Letters **93**, 190501 (2004).

[17] Ferran Riera-Sàbat, Pavel Sekatski, Alexander Pirker, and Wolfgang Dür. "Entanglement purification by counting and locating errors with entangling measurements". Physical Review A **104**, 012419 (2021).

[18] Ferran Riera-Sàbat, Pavel Sekatski, Alexander Pirker, and Wolfgang Dür. "Entanglement-assisted entanglement purification". Physical Review Letters **127**, 040502 (2021).

[19] Nicolas Gisin. "Hidden quantum nonlocality revealed by local filters". Physics Letters A **210**, 151–156 (1996).

[20] Frank Verstraete, Jeroen Dehaene, and Bart De Moor. "Local filtering operations

on two qubits". Physical Review A **64**, 010101 (2001).

[21] Francesco Buscemi and Nilanjana Datta. "General theory of environment-assisted entanglement distillation". IEEE transactions on information theory **59**, 1940–1954 (2012).

[22] Miguel A. Martín-Delgado and Miguel Navascués. "Single-step distillation protocol with generalized beam splitters". Physical Review A **68**, 012322 (2003).

[23] Hector Bombin and Miguel A. Martín-Delgado. "Entanglement distillation protocols and number theory". Physical Review A **72**, 032313 (2005).

[24] Fernando G. S. L. Brandao and Nilanjana Datta. "One-shot rates for entanglement manipulation under non-entangling maps". IEEE Transactions on Information Theory **57**, 1754–1760 (2011).

[25] Bartosz Regula, Kun Fang, Xin Wang, and Mile Gu. "One-shot entanglement distillation beyond local operations and classical communication". New Journal of Physics **21**, 103017 (2019).

[26] Paul G. Kwiat, Salvador Barraza-Lopez, Andre Stefanov, and Nicolas Gisin. "Experimental entanglement distillation and 'hidden' non-locality". Nature **409**, 1014–1017 (2001).

[27] Jian-Wei Pan, Sara Gasparoni, Rupert Ursin, Gregor Weihs, and Anton Zeilinger. "Experimental entanglement purification of arbitrary unknown states". Nature **423**, 417–422 (2003).

[28] Takashi Yamamoto, Masato Koashi, Şahin Kaya Özdemir, and Nobuyuki Imoto. "Experimental extraction of an entangled photon pair from two identically decohered pairs". Nature **421**, 343–346 (2003).

[29] Philip Walther, Kevin J. Resch, Časlav Brukner, Aephraim M. Steinberg, Jian-Wei Pan, and Anton Zeilinger. "Quantum nonlocality obtained from local states by entanglement purification". Physical Review Letters **94**, 040504 (2005).

[30] Luo-Kan Chen, Hai-Lin Yong, Ping Xu, Xing-Can Yao, Tong Xiang, Zheng-Da Li, Chang Liu, He Lu, Nai-Le Liu, Li Li, et al. "Experimental nested purification for a linear optical quantum repeater". Nature Photonics **11**, 695–699 (2017).

[31] Sebastian Ecker, Philipp Sohr, Lukas Bulla, Marcus Huber, Martin Bohmann, and Rupert Ursin. "Experimental single-copy entanglement distillation". Physical Review Letters **127**, 040506 (2021).

[32] Rainer Reichle, Dietrich Leibfried, Emanuel Knill, Joseph W. Britton, R. Bradford Blakestad, John D. Jost, Christopher Langer, Roee Ozeri, Signe Seidelin, and David J. Wineland. "Experimental purification of two-atom entanglement". Nature **443**, 838–841 (2006).

[33] Norbert Kalb, Andreas A. Reiserer, Peter C. Humphreys, Jacob J. W. Bakermans, Sten J. Kamerling, Naomi H. Nickerson, Simon C. Benjamin, Daniel J. Twitchen, Matthew Markham, and Ronald Hanson. "Entanglement distillation between solid-state quantum network nodes". Science **356**, 928–932 (2017).

[34] Filip Rozpędek, Thomas Schiet, Le Phuc Thinh, David Elkouss, Andrew C. Doherty, and Stephanie Wehner. "Optimizing practical entanglement distillation". Physical Review A **97**, 062333 (2018).

[35] Kun Fang, Xin Wang, Marco Tomamichel, and Runyao Duan. "Non-asymptotic entanglement distillation". IEEE Transactions on Information Theory **65**, 6454–6465 (2019).

[36] Stefan Krastanov, Victor V. Albert, and Liang Jiang. "Optimized entanglement purification". Quantum **3**, 123 (2019).

[37] Keisuke Fujii and Katsuji Yamamoto. "Entanglement purification with double selection". Physical Review A **80**, 042308 (2009).

[38] Hans J. Briegel, Wolfgang Dür, J. Ignacio Cirac, and Peter Zoller. "Quantum repeaters: the role of imperfect local operations in quantum communication". Physical Review Letters **81**, 5932 (1998).

[39] Wolfgang Dür and Hans J. Briegel. "Entanglement purification for quantum computation". Physical Review Letters **90**, 067901 (2003).

[40] Wolfgang Dür, Hans J. Briegel, J. Ignacio Cirac, and Peter Zoller. "Quantum repeaters based on entanglement purification". Physical Review A **59**, 169 (1999).

[41] Liangzhong Ruan, Wenhan Dai, and Moe Z. Win. "Adaptive recurrence quantum entanglement distillation for two-kraus-

operator channels". Physical Review A **97**, 052332 (2018).

[42] Karl Gerd H. Vollbrecht and Frank Verstraete. "Interpolation of recurrence and hashing entanglement distillation protocols". Physical Review A **71**, 062325 (2005).

[43] Daniel Gottesman. "Theory of fault-tolerant quantum computation". Physical Review A **57**, 127 (1998).

[44] Emil Artin. "Geometric algebra". Chapter 6, pages 143–147. Interscience Publishers New York. (1957). 1 edition.

[45] Jeroen Dehaene, Maarten Van den Nest, Bart De Moor, and Frank Verstraete. "Local permutations of products of Bell states and entanglement distillation". Physical Review A **67**, 022310 (2003).

[46] Mark M. Wilde. "Quantum information theory". Cambridge University Press. (2013). arXiv:1106.1445.

[47] Jeroen Dehaene and Bart De Moor. "Clifford group, stabilizer states, and linear and quadratic operations over gf (2)". Physical Review A **68**, 042318 (2003).

[48] "Enumerating distillation protocols code". DOI: 10.4121/15082515. Accessed: 2021-02-04.

[49] Xuanqiang Zhao, Benchi Zhao, Zihe Wang, Zhixin Song, and Xin Wang. "Practical distributed quantum information processing with LOCCNet". npj Quantum Information **7**, 159 (2021).

[50] Ernest Y.-Z. Tan, Pavel Sekatski, Jean-Daniel Bancal, René Schwonnek, Renato Renner, Nicolas Sangouard, and Charles C.-W. Lim. "Improved DIQKD protocols with finite-size analysis" (2020). arXiv:2012.08714.

[51] Vlatko Vedral, Martin B. Plenio, Michael A. Rippin, and Peter L. Knight. "Quantifying entanglement". Physical Review Letters **78**, 2275 (1997).

[52] Vlatko Vedral and Martin B. Plenio. "Entanglement measures and purification procedures". Physical Review A **57**, 1619 (1998).

[53] Vlatko Vedral. "The role of relative entropy in quantum information theory". Reviews of Modern Physics **74**, 197 (2002).

[54] Takaaki Matsuo, Clément Durand, and Rodney Van Meter. "Quantum link bootstrapping using a RuleSet-based communication protocol". Physical Review A **100**, 052320 (2019).

[55] Michelle Victora, Stefan Krastanov, Alexander Sanchez de la Cerda, Steven Willis, and Prineha Narang. "Purification and entanglement routing on quantum networks" (2020). arXiv:2011.11644.

[56] Dmitri Maslov and Martin Roetteler. "Shorter stabilizer circuits via bruhat decomposition and quantum circuit transformations". IEEE Transactions on Information Theory **64**, 4729–4738 (2018).

[57] Sergey Bravyi and Dmitri Maslov. "Hadamard-free circuits expose the structure of the Clifford group". IEEE Transactions on Information Theory **67**, 4546–4563 (2021).

A Background on binary picture

For completeness, we give here more background and derivations on the binary picture used in this work. Firstly, we give a derivation of equation (1). Suppose that we have two elements $\tau_v, \tau_w \in \overline{\mathcal{P}}_n$, with $v, w \in \mathbb{F}_2^{2n}$. Then

$$\begin{aligned}\tau_v\tau_w &= (\tau_{v_1 v_{n+1}} \otimes \cdots \otimes \tau_{v_n v_{2n}})(\tau_{w_1 w_{n+1}} \otimes \cdots \otimes \tau_{w_n w_{2n}}) \\ &= \bigotimes_{k=1}^{n} \tau_{v_k v_{n+k}} \tau_{w_k w_{n+k}}.\end{aligned} \tag{25}$$

For all $k \in \{1, \ldots, n\}$, we have

$$\begin{aligned}\tau_{v_k v_{n+k}} \tau_{w_k w_{n+k}} &= X^{v_k} Z^{v_{n+k}} X^{w_k} Z^{w_{n+k}} \\ &= X^{v_k} (-1)^{v_{n+k} w_k} X^{w_k} Z^{v_{n+k}} Z^{w_{n+k}} \\ &= (-1)^{v_{n+k} w_k} X^{v_k + w_k} Z^{v_{n+k} + w_{n+k}} \\ &= (-1)^{v_{n+k} w_k} \tau_{v_k + w_k, v_{n+k} + w_{n+k}}.\end{aligned} \tag{26}$$

As a result,

$$\begin{aligned}\tau_v\tau_w &= \bigotimes_{k=1}^{n} (-1)^{v_{n+k} w_k} \tau_{v_k + w_k, v_{n+k} + w_{n+k}} \\ &= (-1)^{\sum_{k=1}^{n} v_{n+k} w_k} \ \tau_{v+w}\end{aligned} \tag{27}$$

We can rewrite $\sum_{k=1}^{n} v_{n+k} w_k$ in terms of the vectors v and w:

$$\sum_{k=1}^{n} v_{n+k} w_k = w^{\mathsf{T}} \Xi v, \qquad \Xi = \begin{bmatrix} 0 & I_n \\ 0 & 0 \end{bmatrix}. \tag{28}$$

Hence, the product of τ_v and τ_w is given by

$$\tau_v\tau_w = (-1)^{w^{\mathsf{T}} \Xi v} \tau_{v+w}. \tag{29}$$

Combining equation (29) for $\tau_v\tau_w$ and $\tau_v\tau_w$, we finally obtain

$$\tau_v\tau_w = (-1)^{w^{\mathsf{T}} \Xi v + v^{\mathsf{T}} \Xi w} \tau_w\tau_v = (-1)^{v^{\mathsf{T}} \Xi^{\mathsf{T}} w + v^{\mathsf{T}} \Xi w} \tau_w\tau_v = (-1)^{v^{\mathsf{T}} \Omega w} \tau_w\tau_v, \quad \Omega = \Xi + \Xi^{\mathsf{T}} = \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}. \tag{30}$$

Let $C \in \mathcal{C}_n$ be a Clifford operation and $f : \overline{\mathcal{P}}_n \to \overline{\mathcal{P}}_n$, $f(P) = CPC^\dagger$ be the corresponding automorphism. Let $\pi : \mathbb{F}_2^{2n} \to \mathbb{F}_2^{2n}$ be the representation of f in the binary picture. Let $v, w \in \mathbb{F}_2^{2n}$. Then we know that $C\tau_{v+w}C^\dagger = (-1)^{w^{\mathsf{T}} \Xi v} C\tau_v\tau_w C^\dagger = (-1)^{w^{\mathsf{T}} \Xi v} C\tau_v C^\dagger C\tau_w C^\dagger$. In the binary representation, the prefactor $(-1)^{w^{\mathsf{T}} \Xi v}$ does not make a difference. Thus, $\pi(v + w) = \pi(v) + \pi(w)$, so π is a linear map, and there exists a binary $2n \times 2n$ matrix M such that $\pi(v) = Mv$ for all $v \in \mathbb{F}_2^{2n}$.

B Basic row/column operations corresponding to symplectic matrices

In this section we give an overview of the basic row and column operations used in the proofs of Lemma 4.3, Theorem 4.4 and Lemma 5.1 and the corresponding elements of $\mathrm{Sp}(2n, \mathbb{F}_2)$. Let $M \in \mathrm{Mat}_{2n \times k}(\mathbb{F}_2)$ be a binary $(2n \times k)$-matrix with $k \geq 1$. Multiplying M from the left by an element of $\mathrm{Sp}(2n, \mathbb{F}_2)$ results in basic row operations on the rows of M. The basic row operations corresponding to left multiplication by elements of $\mathrm{Sp}(2n, \mathbb{F}_2)$ are summarized in Table 3 [56].

Similarly, multiplying $M \in \mathrm{Mat}_{k \times 2n}(\mathbb{F}_2)$ from the right by an element of $\mathrm{Sp}(2n, \mathbb{F}_2)$ results in basic column operations on the columns of M. These column operations are summarized in Table 4.

Note that we are particularly interested in the cases $k = 2n$ (Lemma 4.3, Lemma 5.1) and $k = 1$ (Theorem 4.4), although Table 3 and Table 4 hold true for any $k \geq 1$.

Element of $\mathrm{Sp}(2n, \mathbb{F}_2)$	Row operation
H_i	Swapping rows i and $n+i$
S_i	Adding row i to row $n+i$
CNOT_{ij}	Adding row i to row j and adding row $n+j$ to row $n+i$
$S_{ij} = (S_j \mathrm{CNOT}_{ij})^2$	Adding row i to row $n+j$ and adding rows i and j to row $n+i$
CZ_{ij}	Adding row i to row $n+j$ and adding row j to row $n+i$
SWAP_{ij}	Swapping rows i and j and swapping rows $n+i$ and $n+j$

Table 3: Basic row operations corresponding to left multiplication of $M \in \mathrm{Mat}_{2n\times k}(\mathbb{F}_2)$ by the elements of $\mathrm{Sp}(2n, \mathbb{F}_2)$ in the first column.

Element of $\mathrm{Sp}(2n, \mathbb{F}_2)$	Column operation
H_i	Swapping columns i and $n+i$
S_i	Adding column $n+i$ to column i
CNOT_{ij}	Adding column j to column i and adding column $n+i$ to column $n+j$
$S_{ij} = (S_j \mathrm{CNOT}_{ij})^2$	Adding columns $n+i$ and $n+j$ to column i and adding column $n+i$ to column j
CZ_{ij}	Adding column $n+i$ to column j and adding column $n+j$ to column i
SWAP_{ij}	Swapping columns i and j and swapping columns $n+i$ and $n+j$

Table 4: Basic column operations corresponding to right multiplication of $M \in \mathrm{Mat}_{k\times 2n}(\mathbb{F}_2)$ by the elements of $\mathrm{Sp}(2n, \mathbb{F}_2)$ in the first column.

C Concatenated DEJMPS protocols

Here we describe the distillation protocols which we compare our results with. These are all based on the so-called DEJMPS protocol [10]. The DEJMPS protocol takes two pairs of Bell-diagonal states, and outputs one state. It performs bilocal single-qubit rotations on both pairs, then a bilocal CNOT, and finally a measurement on one of the pairs where a success is achieved only when correlated outcomes are observed. It is clear that the DEJMPS protocol is an example of a bilocal Clifford protocol. The DEJMPS protocol can be generalised to a number of pairs $n > 2$ by applying the DEJMPS protocol multiple times.

Since the DEJMPS protocol corresponds to 2-1 distillation, the possible ways of combining the different pairs correspond to the number of binary trees on n unlabeled nodes for an n-fold tensor product of input states. Furthermore, for each of the performed DEJMPS protocols (corresponding to each parent of the binary tree), we consider all possible single-qubit rotations. The *concatenated DEJMPS protocols* are then all protocols that arise in this fashion. Note that this class includes well known variants of DEJMPS such as (nested) entanglement pumping protocols [11, 40, 38] or double selection [37].

D Distillation circuits

In this section we are concerned with finding circuits that achieve the highest fidelity for $n = 4$ to 8 for an n-fold tensor product of a Werner state[2].

We first note that one could use techniques for general Clifford circuit decompositions to decompose the symplectic matrices of the form in 5.1. However, we found that this would in general lead to circuits with high depths. Instead, we first find that any distillation protocol has a circuit in a given form. Then, we randomly generate circuits of that form, until we find circuits that achieve the highest fidelity, and have small depth.

D.1 Reducing the circuit search space

We use the Bruhat decomposition from [57, 56], which allows to write any Clifford circuit C in the form $C = FWF'$, with F and, F' elements of the so-called Borel subgroup[3] and W a layer of Hadamard gates followed by a permutation $\sigma \in \mathcal{S}_n$. The Borel subgroup $\mathcal{B}_n$ is generated by X_i, S_i for $1 \leq i \leq n$ and CNOT_{ij} with $1 \leq j < i \leq n$. For convenience, we denote such CNOT gates as $\mathrm{CNOT}^{\uparrow}$ gates. Now note that the Borel subgroup $\mathcal{B}_n$ is a subgroup of the distillation subgroup $\mathcal{D}_n$. This implies that the F part of any circuit in the form $C = FWF'$ does not change the distillation statistics. Thus, any distillation protocol has a corresponding circuit of the form WF'. Furthermore, since the distillation subgroup $\mathcal{D}_n$ contains elements that arbitrarily permute qubits 2 to n, we can restrict to permutations that are either the identity, or exchange qubit 1 with j. In practice, we have found that it is sufficient to only consider $W = H_2 H_3 \dots H_n$.

By the results from [57], any element F' from $\mathcal{B}_n$ can be written as a layer of $\mathrm{CNOT}^{\uparrow}$ gates, a layer of CZ gates, a layer of phase gates and a layer of Pauli gates. Firstly, the Pauli gates are in the kernel of ϕ, and thus can be left out. Secondly, the layer of S gates can be moved to the beginning. To see this, first note that phase gates commute with CZ gates. Then, since $\mathrm{CNOT}_{ij} S_i = S_i \mathrm{CNOT}_{ij}$ and $S_j \mathrm{CNOT}_{ij} = \mathrm{CNOT}_{ij} \mathrm{CZ}_{ij} S_i S_j$, the layer of S gates can be moved to the beginning. Since Werner states are invariant under S, the layer of S gates can be removed without changing the distillation statistics. In the above process of moving the S gates to the beginning, the layer that only had $\mathrm{CNOT}^{\uparrow}$ gates will now have CZ gates as well. In the binary picture we have the following identities,

$$\mathrm{CNOT}_{ij} \mathrm{CZ}_{kl} = \mathrm{CZ}_{kl} \mathrm{CNOT}_{ij} \ , \tag{31}$$

$$\mathrm{CNOT}_{ij} \mathrm{CZ}_{ij} = \mathrm{CZ}_{ij} \mathrm{CNOT}_{ij} \ , \tag{32}$$

$$\mathrm{CNOT}_{ik} \mathrm{CZ}_{ij} = \mathrm{CZ}_{ij} \mathrm{CNOT}_{ik} \ , \tag{33}$$

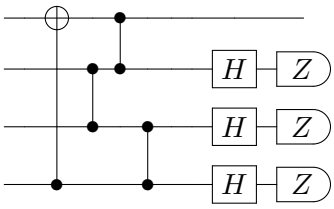
$$\mathrm{CNOT}_{ik} \mathrm{CZ}_{jk} = \mathrm{CZ}_{ij} \mathrm{CZ}_{jk} \mathrm{CNOT}_{ik} \ , \tag{34}$$

where the i, j, k, l are assumed to be distinct, and can be verified using Tables 3 and 4. By using the above identities, the CZ gates can be moved through to the original layer of CZ gates.

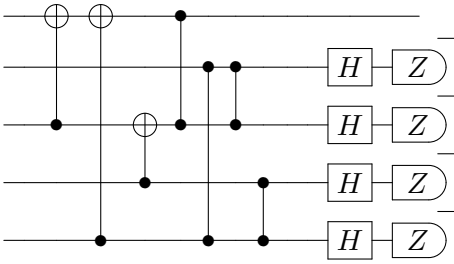
It is thus sufficient to consider only elements F' that consist of a layer of $\mathrm{CNOT}^{\uparrow}$ gates and a layer of CZ gates. Now, to find circuits we randomly generate circuits consisting of a layer of $\mathrm{CNOT}^{\uparrow}$ gates, a layer of CZ gates, and a Hadamard gate on all qubits except the first. We found several circuits that achieved the largest fidelity, and choose the one with smallest depth. We report the found circuits in Fig. 10.

[2] We are not interested in the cases $n = 2$ and $n = 3$, since for those cases the concatenated DEJMPS protocols are optimal.

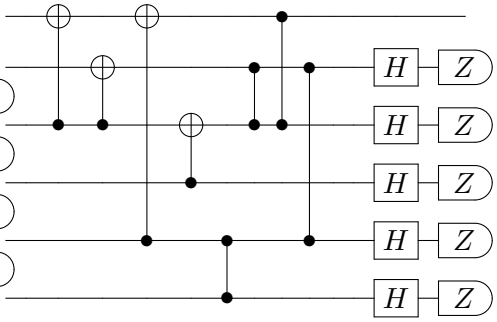
[3] We use a different convention from [57, 56], where the target index for the CNOT gates in the Borel subgroup is larger than the control index.



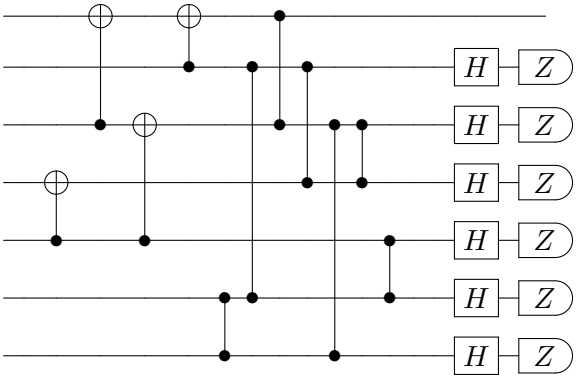
(a) $n = 4$, #2-qubit gates $= 4$, depth $= 3$.



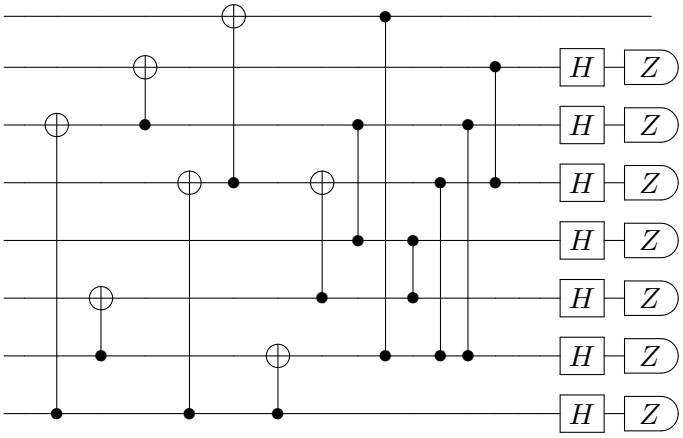
(b) $n = 5$, #2-qubit gates $= 7$, depth $= 5$.



(c) $n = 6$, #2-qubit gates $= 8$, depth $= 6$



(d) $n = 7$, #2-qubit gates $= 11$, depth $= 6$.



(e) $n = 8$, #2-qubit gates $= 13$, depth $= 7$.

Figure 10: Circuits that achieve the maximum fidelity for n. These circuits are applied by both Alice and Bob, after which they measure the last $n - 1$ qubits, and communicate their outcomes to each other. When the outcomes for all individual qubit pairs are correlated, the distillation protocol was deemed successful.

E Analytical expressions

We report here the analytical expressions of the fidelity and success probability that correspond to the found optimal schemes. The input state is an n-fold tensor product of a Werner state with fidelity F. For completeness, we report here as well the distillation statistics expressed in the infidelity $\epsilon \equiv 1 - F$, and the scaling of the output fidelity as a function of the infidelity.

n	p_{suc}
2	$\frac{8}{9}F^2 - \frac{4}{9}F + \frac{5}{9}$
3	$\frac{32}{27}F^3 - \frac{4}{9}F^2 + \frac{7}{27}$
4	$\frac{32}{27}F^4 - \frac{4}{9}F^2 + \frac{4}{27}F + \frac{1}{9}$
5	$\frac{80}{27}F^4 - \frac{80}{27}F^3 + \frac{10}{9}F^2 - \frac{5}{27}F + \frac{2}{27}$
6	$\frac{128}{243}F^6 + \frac{320}{243}F^5 - \frac{256}{243}F^4 + \frac{16}{243}F^3 + \frac{40}{243}F^2 - \frac{14}{243}F + \frac{1}{27}$
7	$\frac{2048}{2187}F^7 - \frac{128}{2187}F^6 + \frac{320}{729}F^5 - \frac{796}{2187}F^4 - \frac{44}{2187}F^3 + \frac{49}{729}F^2 - \frac{37}{2187}F + \frac{37}{2187}$
8	$\frac{6656}{6561}F^8 - \frac{1024}{6561}F^7 + \frac{1664}{6561}F^6 - \frac{64}{6561}F^5 - \frac{1120}{6561}F^4 + \frac{416}{6561}F^3 - \frac{4}{6561}F^2 - \frac{16}{6561}F + \frac{53}{6561}$

Table 5: Success probability for the protocols with the highest output fidelity for $n = 2$ to 8.

n	$p_{\text{suc}} \cdot F_{\text{out}}$
2	$\frac{10}{9}F^2 - \frac{2}{9}F + \frac{1}{9}$
3	$\frac{28}{27}F^3 - \frac{1}{9}F + \frac{2}{27}$
4	$\frac{8}{9}F^4 + \frac{8}{27}F^3 - \frac{2}{9}F^2 + \frac{1}{27}$
5	$\frac{32}{27}F^5 - \frac{20}{27}F^4 + \frac{10}{9}F^3 - \frac{20}{27}F^2 + \frac{5}{27}F$
6	$\frac{32}{27}F^6 - \frac{112}{243}F^5 + \frac{80}{243}F^4 + \frac{8}{243}F^3 - \frac{32}{243}F^2 + \frac{10}{243}F + \frac{1}{243}$
7	$\frac{2368}{2187}F^7 - \frac{592}{2187}F^6 + \frac{196}{729}F^5 - \frac{44}{2187}F^4 - \frac{199}{2187}F^3 + \frac{20}{729}F^2 - \frac{2}{2187}F + \frac{8}{2187}$
8	$\frac{6784}{6561}F^8 - \frac{51}{6561}F^7 - \frac{32}{6561}F^6 + \frac{832}{6561}F^5 - \frac{560}{6561}F^4 - \frac{8}{6561}F^3 + \frac{52}{6561}F^2 - \frac{8}{6561}F + \frac{13}{6561}$

Table 6: Product of the success probability and the output fidelity for the protocols with the highest output fidelity for $n = 2$ to 8.

n	p_{suc}
2	$1-\frac{4}{3}\epsilon+\frac{8}{9}\epsilon^2$
3	$1-2\epsilon+\frac{4}{3}\epsilon^2$
4	$1-2\epsilon+\frac{4}{3}\epsilon^2-\frac{8}{27}\epsilon^3$
5	$1-\frac{14}{3}\epsilon+\frac{28}{3}\epsilon^2-\frac{256}{27}\epsilon^3+\frac{400}{81}\epsilon^4-\frac{256}{243}\epsilon^5$
6	$1-5\epsilon+\frac{32}{3}\epsilon^2-12\epsilon^3+\frac{608}{81}\epsilon^4-\frac{608}{243}\epsilon^5+\frac{256}{729}\epsilon^6$
7	$1-7\epsilon+\frac{190}{9}\epsilon^2-\frac{944}{27}\epsilon^3+\frac{928}{27}\epsilon^4-\frac{544}{27}\epsilon^5+\frac{1600}{243}\epsilon^6-\frac{2048}{2187}\epsilon^7$
8	$1-\frac{23}{3}\epsilon+\frac{244}{9}\epsilon^2-\frac{1540}{27}\epsilon^3+\frac{6280}{81}\epsilon^4-\frac{16832}{243}\epsilon^5+\frac{28768}{729}\epsilon^6-\frac{9472}{729}\epsilon^7+\frac{4096}{2187}\epsilon^8$

Table 7: Success probability for the protocols with the highest output fidelity for $n = 2$ to 8, expressed in the infidelity $\epsilon \equiv 1 - F$.

n	$p_{\mathrm{suc}} \cdot F_{\mathrm{out}}$
2	$1-2\epsilon+\frac{10}{9}\epsilon^2$
3	$1-3\epsilon+\frac{10}{3}\epsilon^2-\frac{4}{3}\epsilon^3$
4	$1-3\epsilon+\frac{10}{3}\epsilon^2-\frac{44}{27}\epsilon^3+\frac{8}{27}^4$
5	$1-5\epsilon+\frac{92}{9}\epsilon^2-\frac{284}{27}\epsilon^3+\frac{440}{81}\epsilon^4-\frac{272}{243}\epsilon^5$
6	$1-\frac{17}{3}\epsilon+\frac{122}{9}\epsilon^2-\frac{466}{27}\epsilon^3+\frac{992}{81}\epsilon^4-\frac{1112}{243}\epsilon^5+\frac{512}{729}\epsilon^6$
7	$1-7\epsilon+\frac{190}{9}\epsilon^2-\frac{320}{9}\epsilon^3+\frac{2936}{81}\epsilon^4-\frac{1816}{81}\epsilon^5+\frac{5680}{729}\epsilon^6-\frac{2560}{2187}\epsilon^7$
8	$1-8\epsilon+\frac{259}{9}\epsilon^2-\frac{544}{9}\epsilon^3+\frac{2180}{27}\epsilon^4-\frac{17000}{243}\epsilon^5+\frac{27872}{729}\epsilon^6-\frac{2912}{243}\epsilon^7+\frac{3584}{2187}\epsilon^8$

Table 8: Product of the success probability and the output fidelity for the protocols with the highest output fidelity for $n = 2$ to 8, expressed in the infidelity $\epsilon \equiv 1 - F$.

n	F_{out}
2	$1-\frac{2}{3}\epsilon-\mathcal{O}\left(\epsilon^2\right)$
3	$1-\frac{1}{3}\epsilon-\mathcal{O}\left(\epsilon^2\right)$
4	$1-\frac{2}{3}\epsilon^2-\mathcal{O}\left(\epsilon^3\right)$
5	$1-\frac{10}{9}\epsilon^3-\mathcal{O}\left(\epsilon^4\right)$
6	$1-\frac{8}{9}\epsilon^3-\mathcal{O}\left(\epsilon^4\right)$
7	$1-\frac{13}{27}\epsilon^3-\mathcal{O}\left(\epsilon^4\right)$
8	$1-\frac{8}{27}\epsilon^3-\mathcal{O}\left(\epsilon^4\right)$

Table 9: Scaling of the output fidelity around $\epsilon \approx 0$ for the protocols with the highest output fidelity for $n = 2$ to 8, where $\epsilon \equiv 1 - F$.