New Limits on Fault-Tolerant Quantum Computation

Harry Buhrman[*]
CWI and
U of Amsterdam
buhrman@cwi.nl

Richard Cleve[†]
U of Waterloo and
Perimeter Institute
cleve@cs.uwaterloo.ca

Monique Laurent [‡]
CWI, Amsterdam
M.Laurent@cwi.nl

Noah Linden[§]
U of Bristol
N.Linden@bristol.ac.uk

Alexander Schrijver
CWI and
U of Amsterdam
Lex.Schrijver@cwi.nl

Falk Unger[*]
CWI, Amsterdam
unger@cwi.nl

November 3, 2005

Abstract

We show that quantum circuits cannot be made fault-tolerant against a depolarizing noise level of $\theta = (6 - 2\sqrt{2})/7 \approx 45\%$, thereby improving on a previous bound of 1/2 (due to Razborov [14]). Our precise quantum circuit model enables *perfect* gates from the Clifford group (CNOT, Hadamard, S, X, Y, Z) and arbitrary additional one-qubit gates that are subject to depolarizing noise θ. We prove that this set of gates cannot be universal for arbitrary (even classical) computation, from which the upper bound on the noise threshold for fault-tolerant quantum computation follows.

1 Introduction

In the past decade, quantum computing has attracted much attention because of its ability to efficiently solve problems for which no efficient classical algorithms are known. Significant research efforts are dedicated to physically realizing quantum computers. A fundamental problem is to cope with noise, which creates major difficulties in storing and operating on quantum states reliably. A key advance was the realization that quantum error correcting codes [16, 18] exist and fault-tolerant quantum computation [15] is possible for a number of apparently reasonable error models. Subsequent

[*]Supported by the EU project RESQ IST-2001-37559 and the NWO vici project 2004-2009.
[†]Supported in part by Canada's NSERC, CIAR, MITACS, and the US ARO.
[‡]Supported by the Netherlands Organization for Scientific Research NWO 639.032.203.
[§]Supported by the EU project RESQ IST-2001-37559 and UK EPSRC IRC in Quantum Information Processing.

results have progressively improved on the first fault-tolerant schemes, proving better and better bounds on the noise tolerable in quantum computation (e.g. [4, 3]). Recent results suggest that fault-tolerant quantum computation is possible with gates that have as much as 3% of depolarizing errors [9].

In this paper we will concentrate on the opposite task of proving that, for certain noise levels, quantum computation is *impossible*. Our main result is as follows: Let CLIFFORD be the set of all (noiseless) Clifford gates

$$\mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \tag{1}$$

It is known that CLIFFORD together with *any* other one-qubit gate, not generated by the gates in CLIFFORD, form a universal set of gates for quantum computation [17, 11]. We show however, that such additional one-qubit gates should not be too noisy. More precisely, let CLIFFORD* be CLIFFORD augmented with arbitrary one-qubit gates with depolarizing error at least $\hat{\theta} = (6 - 2\sqrt{2})/7 \approx 45\%$. Then this set of gates is not capable anymore of computing arbitrary functions and thus is not universal. In other words, fault-tolerant quantum computation cannot be performed if there is this level of noise. Additionally we show in Remark 3 that, among all one-qubit gates that augment CLIFFORD, the so-called $\pi/8$-gate (see end of Section 3) is the type of gate that requires the most noise to render it incapable of universal quantum computation by our approach. That is, for other augmenting gates (e.g., $\pi/16$-gates), our approach will yield stronger bounds on the tolerable level. Our results also yield a simple proof that not all classical functions can be computed using Clifford gates (complementing results in [1]). In particular, in Corollary 1, we show that a boolean function which can be computed by Clifford circuits can be written as the parity of a subset of input bits.

The main idea of our approach is as follows. Assume we have a Clifford circuit C with n classical input bits $x = x_1, \dots x_n$ and one dedicated output qubit that, when measured in the computational basis, yields the output of the computation of C on x. Suppose now that the input is partitioned over two parties, Alice and Bob, such that Alice has k bits of x and Bob has $n - k$ bits. We first show how Alice, with the help of Bob, can compute the value of C on x with just a single classical bit of communication (Lemma 1). From this it follows that Clifford circuits can at the very best compute only those functions that require for any partion of the inputs a single bit of communication, and it is well known that many functions require more than one bit of communication. Next, we show in Lemma 2 how probabilistic mixtures of Clifford gates can be used to simulate any single qubit unitary gate, that has noise $\hat{\theta}(\approx 45\%)$. The proof of this Lemma relies on solving an optimization problem related to the Clifford polytope, defined as the convex hull of the set $\mathcal{C} \subseteq \mathbb{R}^{3\times 3}$ of Clifford rotation matrices in $\mathbb{R}^3$. Here, the matrices $\mathcal{C}$ are essentially the one-qubit Clifford gates in Bloch sphere representation.

Combining Lemmas 1 and 2, we get that all circuits with CLIFFORD*-gates and with respect to any distribution of the inputs can be computed by Alice and Bob with a single bit of communication (Lemma 3). Using the fact that there are functions which require communication more than one bit, we get our main result (Theorem 1): The set of gates in CLIFFORD* cannot be universal. We also generalize our result to the case that the inputs are quantum states, see Remark 2.

1.1 Related Work

There are only a few other results concerning the limits of fault-tolerant computation. These are not all strictly comparable to each other and our result; nevertheless, we review them and make some comparisons. See the introduction of [14] for some remarks that motivate the analysis of thresholds for fault-tolerant quantum computation.

The first results on upper bounds of the threshold decoherence rate were obtained by showing that quantum computers with faulty gates can be simulated efficiently on a classical computer. The first to prove one of these results were Aharonov and Ben-Or [2], with the value 97% for the noise. Later Harrow and Nielsen [8] showed that if 74% of depolarizing noise is applied to each output qubit of each gate, then (faulty) two-qubit gates cannot produce entanglement. They concluded that circuits containing only one- and two-qubit gates with depolarizing noise at least 74% can be simulated efficiently on a classical computer. Another result along these lines is due to Virmani *et al.* [19], however they use a different noise model (dephasing noise) that is not quantitatively comparable to depolarizing noise. Note that all these results do not exclude the possibility that quantum circuits with high noise can still do universal classical computations; our results imply this.

The only prior result of this latter type is due to Razborov [14], where a 50% upper bound on the noise threshold for depolarizing noise on qubits for circuits with two-qubit gates is obtained (and a weaker bound for k-qubit gates). The argument in [14] is essentially that, at this noise level, any superlogarithmic-depth quantum circuit (with constant error rate per qubit per time step) will be overwhelmed by the noise and produce a statistically meaningless outcome. (Thus the complexity theoretic assumption $BQP \neq QNC^1$ is invoked.) We note that it is shown in [5] that in fact log-depth quantum circuits can perform interesting feats, including efficient integer factorization (if combined by classical polynomial-time pre- and post-processing). Our error model is in most respects weaker than that of [14] (since our qubit errors are only occurring at the completion of non-CLIFFORD gates) and our bound of 45% is below 50%. In fairness, there is a sense in which the bound in [14] is stronger: it permits arbitrary (noisy) two-qubit gates; whereas, our only two-qubit gates are (perfect) CNOT gates.

2 Preliminaries and Notation

E_{ij} is the all-zero matrix, except for the entry i, j which is equal to 1. We also write $+$ for $+1$ and $-$ for -1. For matrices $A, B \in \mathbb{R}^{3\times 3}$ we define the inner product $\langle A, B\rangle$ as:

$$\langle A, B\rangle = tr(A^T B) = \sum_{i,j\in\{1,2,3\}} a_{ij} b_{ij}.$$

The following fact is used repeatedly: $\langle A, BC\rangle = \langle B^T A, C\rangle$ for $A, B, C \in \mathbb{R}^{3\times 3}$.

By A^* we denote the conjugate transpose of matrix A.

An *n-qubit state* (or density matrix) ρ is a matrix $\rho \in \mathbb{C}^{2^n \times 2^n}$ with the properties $tr(\rho) = 1$ and $\rho = \rho^*$ (Hermiticity). An *n-qubit operation* (or *gate*) is a unitary matrix $U \in \mathbb{C}^{2^n \times 2^n}$, i.e., $U^*U = \mathbb{I}$. For such n-qubit state ρ and n-qubit operation U the application of U to ρ results in the state $U\rho U^*$.

2.1 Bloch-vector representation

In our further analysis it will be convenient to use the Bloch-sphere representation of 1-qubit states and 1-qubit operations, which we review now (see for example Chapters 4.2 and 8 in [12]).

For $\mathbf{r} \in \mathbb{R}^3$ define $\mathbf{r} \cdot \sigma = r_x X + r_y Y + r_z Z$, where $\sigma = (X, Y, Z)$ is the vector of Pauli-matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

Then, all 1-qubit density matrices ρ can be uniquely written in the form

$$\rho = \frac{\mathbb{I} + \mathbf{r} \cdot \sigma}{2} = \frac{\mathbb{I} + r_x X + r_y Y + r_z Z}{2},$$

where $\mathbf{r} \in \mathbb{R}^3$ and $||\mathbf{r}|| = \sqrt{r_x^2 + r_y^2 + r_z^2} \le 1$. We call $\mathbf{r}$ the *Bloch vector* of ρ.

For $\mathbf{n} \in \mathbb{R}^3$ with $||\mathbf{n}|| = 1$ and $\theta \in \mathbb{R}$ we define

$$U_{\mathbf{n}}(\theta) = \exp(-i\theta \mathbf{n} \cdot \sigma/2) = \cos(\theta/2)\mathbb{I} - i \sin(\theta/2)\mathbf{n} \cdot \sigma.$$

We first note that $U_{\mathbf{n}}(\theta)U_{\mathbf{n}}(\theta)^* = \mathbb{I}$, i.e., $U_{\mathbf{n}}(\theta)$ is unitary. Second, let the result of the quantum operation $U_{\mathbf{n}}(\theta)$ applied to state $\rho = \mathbb{I}/2 + \mathbf{r} \cdot \sigma/2$ be $\rho' = U_{\mathbf{n}}(\theta)^* \rho U_{\mathbf{n}}(\theta) = \mathbb{I}/2 + \mathbf{r}' \cdot \sigma/2$. Then $\mathbf{r}'$ is the image of rotating $\mathbf{r}$ around $\mathbf{n}$ by an angle θ. Third, all 1-qubit unitaries U can be written as

$$U = U_{\mathbf{n}}(\theta)$$

with $\mathbf{n} \in \mathbb{R}^3, \theta \in \mathbb{R}$ and $||\mathbf{n}|| = 1$ (ignoring an unimportant phase factor $\alpha \in \mathbf{C}$ with $|\alpha| = 1$).

Thus, one-qubit states and unitaries are isomorphic to vectors ,resp., rotations in $\mathbb{R}^3$. The set of all rotations in $\mathbb{R}^3$ is the group $SO(3)$.[1] We introduce some notation reflecting this isomorphism. For unitary $U \in \mathbb{C}^{2\times 2}$ we let $R_U \in SO(3)$ be the corresponding rotation matrix. We get a reverse operation (up to phase factors) by fixing one mapping $f : SO(3) \to \mathbb{C}^{2\times 2}$ with the property that for all unitary $U \in \mathbb{C}^{2\times 2}$ it holds $f(R_U) = \alpha U$ for some $\alpha \in \mathbb{C}$, $|\alpha| = 1$. We then write $U_R = f(R)$.

This can be extended to probabilistic mixtures of quantum operations. Let $\{p_i\}$ be a probability distribution, i.e., $\sum_i p_i = 1$ and $0 \le p_i$, and let $U_i \in \mathbb{C}^{2\times 2}$ be a 1-qubit unitary with corresponding Bloch representation $R_i \in \mathbb{R}^{3\times 3}$. Then the quantum operation E in which each U_i is applied with probability p_i has Bloch-representation $R_E = \sum_i p_i R_i$.

[1] This group will play a prominent role in the proof of Lemma 2, where some more notation can be found.

2.2 Noise

There are several models of noise considered in the literature. The most common one, which we consider too, is *depolarizing noise.* A 1-qubit state ρ to which depolarizing noise p is applied, becomes

$$(1-p)\rho + p\mathbb{I}/2.$$

Thus, with probability $1-p$ the state is not changed and with probability p the state is replaced with the completely mixed state.

It is not hard to see that applying depolarizing noise p to $\rho = \mathbb{I}/2 + \mathbf{r}\cdot\sigma/2$ yields $\rho' = \mathbb{I}/2 + \mathbf{r}'\cdot\sigma/2$, with $r' = (1-p)r$. So, this noise shrinks the Bloch vector of a state to $(1-p)$ of its original length.

We say that a 1-qubit gate implements the unitary U with noise p if it transforms states ρ into

$$(1-p)U^*\rho U + p\mathbb{I}/2. \tag{3}$$

This *quantum operation* can be seen as a two-stage process, in which first U and then depolarizing noise is applied. Let $R_U \in \mathbb{R}^{3\times 3}$ be the rotation matrix corresponding to the unitary U. Then this noisy quantum operation has Bloch-representation $(1-p)R_U$, i.e., it rotates a Bloch vector and scales it by a factor $1-p$.

For 1-qubit gates and depolarizing noise, the two representations are (up to unimportant global phase factors) equivalent. (See chapter 8.3 in [12] for more details.)

2.3 Clifford group

The (n-qubit) Clifford group contains all unitary operations that can be written as a product of tensor products of S, H and $CNOT$ (see eq. (1)). The Clifford group contains also all Pauli-operators X, Y, Z. We let Clifford be the set of all Clifford gates. Let Clifford* be the set of gates consisting of Clifford and arbitrary 1-qubit gates which have depolarizing noise at least $\hat{\theta} = (6-2\sqrt{2})/7$.

For a state with Bloch vector $\mathbf{r}$ we get:

$$S\left(\frac{1}{2}\mathbb{I} + \frac{r_x}{2}X + \frac{r_y}{2}Y + \frac{r_z}{2}Z\right)S^* = \frac{1}{2}\mathbb{I} - \frac{r_y}{2}X + \frac{r_x}{2}Y + \frac{r_z}{2}Z$$

Let R_S be the Bloch representation of S. Then R_S rotates Bloch vectors around the z-axis by $\pi/2$. In particular, the x-axis is mapped to $-y$ and y to x. For the Hadamard-gates we similarly have

$$H\left(\frac{1}{2}\mathbb{I} + \frac{r_x}{2}X + \frac{r_y}{2}Y + \frac{r_z}{2}Z\right)H^* = \frac{1}{2}\mathbb{I} + \frac{r_z}{2}X - \frac{r_y}{2}Y + \frac{r_x}{2}Z.$$

So the Bloch representation R_H of H negates the y-coordinate of a Bloch vector and swaps the x and z-coordinates, i.e., it is a rotation by π around the axis $1/\sqrt{2}(1,0,1)$.

We define $\mathcal{C}$ as the set of matrices which can be generated from R_S and R_H. A $C \in \mathcal{C}$ is called a *Clifford (rotation) matrix.* It is not hard to see that $\mathcal{C}$ contains exactly those rotations which map axes to axes (or their opposite). Those C have in each row and column exactly one non-zero entry, which must be either $+1$ or -1, and

$det(C) = 1$. Note that $\mathcal{C}$, being isomorphic to the 1-qubit Clifford group, is a group under matrix multiplication. Examples of Clifford matrices are

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

In Appendix A we need (and explain) more details about Clifford rotation matrices.

2.4 Communication Complexity

The setting for this is the following: Assume two separated parties, Alice and Bob, where Alice is given $x \in \{0,1\}^{m_A}$ and Bob $y \in \{0,1\}^{m_b}$, want to compute $f(x, y)$ for some fixed function $f : \{0,1\}^{m_A} \times \{0,1\}^{m_b} \to \{0,1\}$. We want that at least one party learns the result $f(x, y)$. In order to achieve this they can communicate bits, according to a predefined protocol. The *deterministic communication complexity* $C_f(n)$ of f is the smallest number c such that each protocol which always computes the correct result, needs at least c bits of communication for at least one input $x, y \in \{0,1\}^n$.

It is well-known that there are functions f where $C_f(n)$ is n, for example the inner product function (see [10]). The above can be extended to *randomized* communication, where the parties are additionally provided with a source which sends a sequence of random bits to Alice and the same sequence to Bob. The final result only has to be correct with some probability $1/2 + \epsilon$ for $\epsilon > 0$. However, also in this randomized setting there are "hard" functions. For example, it is known that the inner product function has randomized communication complexity $n - O(\log(1/\delta))$, if the outputs have to be correct with probability at least $1/2 - \delta$ (see also [10]).

3 The main result

We are now ready to prove the main lemma, which explains the idea of simulating Clifford circuits.

Lemma 1. *Let $f : \{0,1\}^{m_A} \times \{0,1\}^{m_B} \to \{0,1\}$ be a function that is computable with unbounded error by a quantum circuit C that uses only gates from* CLIFFORD *in the following sense. On input $(x, y) \in \{0,1\}^{m_A} \times \{0,1\}^{m_B}$, the inputs to C are $|x\rangle|y\rangle|0\ldots0\rangle$ and the output is the measured result of the first output qubit of C (measured in the computational basis). It suffices for the result to be correct with probability greater than* $1/2$ *(but it can be arbitrarily close to* $1/2$*). Then the deterministic communication complexity of f is one bit.*

Proof. We begin by noting that each qubit can be represented by two *shares*: a *classical share* consisting of two bits, and a *quantum share* consisting of one qubit and that logical Clifford operations can be performed on the shares themselves. When the classical share is ab, the logical qubit that the shares encode is that obtained by applying $X^a Z^b$ to the quantum share.

Assuming that a set of qubits is encoded in this manner, the operations H, S, and CNOT can be applied to the logical qubits by transversally performing operations on the shares that encode them (i.e., the logical qubits do not have to be reconstructed). For example, to apply H to a logical qubit, the two bits that make up its classical share are swapped and H is applied to its quantum share. This works correctly because

$$\begin{aligned} HX^aZ^b &= HX^aHHZ^bHH \\ &= Z^aX^bH \\ &= (-1)^{a\wedge b}X^bZ^aH, \end{aligned} \tag{4}$$

and $(-1)^{a\wedge b}$ is an irrelevant global phase.

To apply S to a logical qubit, a CNOT_2^1 is applied to the two bits of its classical share and S is applied to its quantum share. This case can be verified by noting that

$$\begin{aligned} SX^aZ^b &= i^aX^aSZ^aZ^b \\ &= i^aX^aZ^{a\oplus b}S, \end{aligned} \tag{5}$$

where we note that i^a is a global phase.

To simulate the application of CNOT gate on two logical qubits, with classical shares a_1b_1 and a_2b_2, CNOT_2^1 is applied to a_1a_2, CNOT_1^2 is applied to b_1b_2, and CNOT_2^1 is applied to the two quantum shares. In this case, we omit the details but note that the correctness can be verified using the identities

$$\begin{aligned} \mathrm{CNOT}_2^1(X\otimes I) &= (X\otimes X)\mathrm{CNOT}_2^1 \\ \mathrm{CNOT}_2^1(I\otimes X) &= (I\otimes X)\mathrm{CNOT}_2^1 \\ \mathrm{CNOT}_2^1(Z\otimes I) &= (Z\otimes I)\mathrm{CNOT}_2^1 \\ \mathrm{CNOT}_2^1(I\otimes Z) &= (Z\otimes Z)\mathrm{CNOT}_2^1. \end{aligned} \tag{6}$$

We first describe a *probabilistic* communication protocol for f. Alice operates on the classical shares while Bob operates on the quantum shares. The initial shares are easy to constuct: for each of Alice's input qubits $|x_j\rangle$, Alice sets her classical share to x_j0 and Bob sets his quantum share to $|0\rangle$; for each of Bob's input qubits $|y_j\rangle$, Alice sets her classical share to 00 and Bob sets his quantum share to $|y_j\rangle$. In this manner, Alice and Bob can simulate the execution of circuit C on input $|x\rangle|y\rangle|0\ldots0\rangle$ *without any communication* to obtain the shares of the output qubits of C. For Bob to obtain the measured output qubit, Alice sends the first bit of her classical share, a_1, to Bob, who applies X^{a_1} to his quantum share and measures it (Alice need not send b_1, the second bit of the classical share, since Bob is performing a measurement in the computational basis).

Finally, to obtain a *deterministic* communication protocol for f, we note that Bob need not actually manipulate quantum information; rather, he can simulate his quantum registers and his operations on them with a classical computer. Then, upon receipt of his classical bit from Alice, he can exactly determine the output probabilities of his measurement to determine which is more likely. □ □

We want to extend this Lemma 1, by replacing CLIFFORD with CLIFFORD*. To do that we first show how one can simulate arbitrary 1-qubit gates with depolarizing noise $\hat{\theta} = (6-2\sqrt{2})/7$ with Clifford operations.

Lemma 2. *Let U be a 1-qubit unitary and E_U be the following noisy version of it*

$$\rho \mapsto E_U(\rho) = (1-\hat{\theta})U\rho U^* + \hat{\theta}\mathbb{I}/2,$$

for any $\rho \in \mathbb{C}^{2\times 2}$. Then there is a probability distribution $\{p_C\}$ over $\mathcal{C}$ such that for all $\rho \in \mathbb{C}^{2\times 2}$ we have

$$E_U(\rho) = \sum_{C\in\mathcal{C}} p_C U_C \rho U_C^*.$$

Proof. Using Section 2.1 the lemma can be reformulated equivalently in Bloch representation: For any $S \in SO(3)$ there is a probability distribution $\{p_C\}$ over $\mathcal{C}$ such that

$$(1-\hat{\theta})S = \sum_{C\in\mathcal{C}} p_C C. \tag{7}$$

We will prove this latter statement. Define the *Clifford polytope*

$$P := \mathrm{conv}(\mathcal{C}) = \left\{ \sum_{C\in\mathcal{C}} p_C C \mid p_C \geq 0, \ \sum_{C\in\mathcal{C}} p_C = 1 \right\} \tag{8}$$

as the convex hull of the 24 Clifford rotation matrices in $\mathbb{R}^{3\times 3}$. We have to prove

$$(1-\hat{\theta})S \in P \ \text{ for any } S \in SO(3). \tag{9}$$

For this we use the fact that the Clifford polytope can be alternatively described by its facet description:

$$P = \left\{ S \in \mathbb{R}^{3\times 3} \mid \langle F, S\rangle \leq 1 \ \text{ for all } F \in \mathcal{F} \right\}, \tag{10}$$

where

$$\mathcal{F} := \left\{ C_1 B C_2 | C_1, C_2 \in \mathcal{C}, B \in \{B_1, B_1^T, B_2\} \right\}, \tag{11}$$

$$B_1 := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \ B_2 := \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

One can use the software from [7] for computing the facet description (10); we will give a direct proof in the Appendix. In view of (10), our claim (9) is equivalent to

$$(1-\hat{\theta})\langle F, S\rangle \leq 1 \ \text{ for all } S \in SO(3), \ F \in \mathcal{F}. \tag{12}$$

Let $F \in \mathcal{F}$ of the form $F = C_1 B C_2$ where $C_1, C_2 \in \mathcal{C}$. As $\langle F, S\rangle = \langle C_1^T S C_2^T, B\rangle$ and $C_1^T S C_2^T \in SO(3)$, (12) is equivalent to

$$\langle S, B\rangle \leq \frac{1}{1-\hat{\theta}} = 2\sqrt{2} - 1 \ \text{ for all } B \in \{B_1, B_2\}, \ S \in SO(3). \tag{13}$$

The case $B = B_1$ is easy to handle: For $S \in SO(3)$, $\langle S, B_1\rangle = \sum_{i=1}^3 S_{i1} \leq \sqrt{3} < 2\sqrt{2} - 1$. We now show (13) for $B = B_2$. Write $S \in \mathbb{R}^{3\times 3}$ as

$$S = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}. \tag{14}$$

Well-known necessary and sufficient conditions for $S \in SO(3)$ are

$$\mathbf{a}^T\mathbf{b} = 0,\ \mathbf{c} = \mathbf{a} \times \mathbf{b},\ \mathbf{a}^T\mathbf{a} = 1,\ \mathbf{b}^T\mathbf{b} = 1, \tag{15}$$

where $\times$ denotes the vector product, defined as

$$\mathbf{a} \times \mathbf{b} := (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)^T.$$

Recall that, for $\mathbf{a}, \mathbf{b}, \mathbf{c}$ as in (15), $\mathbf{a} = \mathbf{b} \times \mathbf{c}$ and $\mathbf{b} = \mathbf{c} \times \mathbf{a}$. Using $c_3 = a_1b_2 - a_2b_1$, we obtain $\langle B, S\rangle = a_1 - a_2 + b_1 + b_2 - a_1b_2 + a_2b_1$. Therefore our task is now to prove that the optimum value of the program

$$\begin{array}{ll} \max & f := a_1 - a_2 + b_1 + b_2 - a_1b_2 + a_2b_1 \\ \text{s.t.} & g_1 := a_1^2 + a_2^2 + a_3^2 = 1 \\ & g_2 := b_1^2 + b_2^2 + b_3^2 = 1 \\ & g_3 := a_1b_1 + a_2b_2 + a_3b_3 = 0 \end{array} \tag{16}$$

is at most $2\sqrt{2} - 1$; we in fact show that $\max f = 2\sqrt{2} - 1$. For this, consider a global maximizer (a, b) to the program (16). Then, the Karush-Kuhn-Tucker conditions have to be satisfied, since the gradient vectors $\{\nabla g_i(a, b) \mid i = 1, 2, 3\}$ are linearly independent; see, e.g., Theorem 12.1 in [13]. (Here, the gradient vector $\nabla g_i(a, b)$ consists of the partial derivatives with respect to the six variables $a_1, \dots, b_3$.) That is, there exist scalars $\lambda_1, \lambda_2, \lambda_3$ for which

$$\nabla f(a, b) + \sum_{i=1,2,3} \lambda_i \nabla g_i(a, b) = 0.$$

Equivalently, considering the partial derivatives first with respect to (a_1, a_2, a_3) and then with respect to (b_1, b_2, b_3)

$$\begin{array}{ccccccccc} \begin{pmatrix} 1 - b_2 \\ -1 + b_1 \\ 0 \end{pmatrix} & + & 2\lambda_1\mathbf{a} & & & + & \lambda_3\mathbf{b} & = & 0 \\ \begin{pmatrix} 1 + a_2 \\ 1 - a_1 \\ 0 \end{pmatrix} & & & + & 2\lambda_2\mathbf{b} & + & \lambda_3\mathbf{a} & = & 0. \end{array}$$

Multiplying the first and the second line by $\mathbf{c}^T = (\mathbf{a} \times \mathbf{b})^T$ (recall that $\mathbf{c} \perp \mathbf{a}, \mathbf{b}$) we get

$$\begin{aligned} 0 &= c_1(1 - b_2) + c_2(-1 + b_1) = c_1 - c_2 + a_3 \\ 0 &= c_1(1 + a_2) + c_2(1 - a_1) = c_1 + c_2 + b_3. \end{aligned}$$

Adding (resp. subtracting) these equations yields $2c_1 = -a_3 - b_3$ and $2c_2 = a_3 - b_3$. Squaring these two equations and then adding them gives $2a_3^2 + 2b_3^2 = 4c_1^2 + 4c_2^2$. Since the rows and columns in S are normalized, we get $2(1 - c_3^2) = 4(1 - c_3^2)$, from which we conclude $c_3^2 = 1$ and, therefore, $a_3 = b_3 = c_1 = c_2 = 0$. This implies $a_1^2 + b_1^2 = 1 = a_1^2 + a_2^2$ and thus $|b_1| = |a_2|$. Similarly one can establish $|a_1| = |b_2|$. On the basis of this observation we distinguish three cases.

1. $a_1 = b_2 = 0$. Then, $|a_2| = |b_1| = 1$ and $f = -a_2 + b_1 + a_2 b_1 \leq 1$.
2. $a_1 \neq 0$ and $a_1 = -b_2$. From $a^T b = 0$ we have $a_1(b_1 - a_2) = 0$, which gives $a_2 = b_1$. Then, $f = a_1 - a_2 + a_2 - a_1 + a_1^2 + a_2^2 = 1$.
3. $a_1 \neq 0$ and $a_1 = b_2$. From $a^T b = 0$ we have $a_1(b_1 + a_2) = 0$, which gives $a_2 = -b_1$. Then, $f = a_1 - a_2 - a_2 + a_1 - a_1^2 - a_2^2 = 2(a_1 - a_2) - 1$, which (under the condition $a_1^2 + a_2^2 = 1$) is clearly maximized by $a_1 = -a_2 = 1/\sqrt{2}$. Therefore, we find $\max f = 2\sqrt{2} - 1$.

Thus, we have shown that the optimum value of the program (16) is equal to $2\sqrt{2} - 1$, which concludes the proof. □ □

Lemma 3. *Let $f : \{0,1\}^{m_A} \times \{0,1\}^{m_B} \to \{0,1\}$ be a function and K a quantum circuit for f with correctness probability $c > 1/2$ which uses only gates from* Clifford* *and measurements in the computational basis. Then the* randomized *communication complexity of f is one bit.*

Proof. ¿From Lemma 1 we know how two parties, Alice and Bob, can simulate perfect Clifford gates. From Lemma 2 we know how they can jointly simulate the other noisy 1-qubit gates in Clifford*, where they use shared randomness to make sure that they always simulate the same Clifford gate.□ □

We can now prove an upper bound on the noise in fault-tolerant quantum computation.

Theorem 1. *The set of gates from* Clifford* *is not sufficient for arbitrary classical computation.*

Proof. The result follows by Lemma 3 and the fact that there are functions with communication complexity greater than 1 □. □

4 Discussion

4.1 The power of the stabilizer formalism

From Lemma 1 we also get that the set of functions computable in the stabilizer formalism is very limited and far from being universal.

Corollary 1. *Every function $f : \{0,1\}^n \to \{0,1\}$ which can be computed by a Clifford circuit, can be written in the form*

$$f(x_1 \ldots x_n) = c \oplus \bigoplus_{j \in S} x_j,$$

where $S \subseteq [n]$ is a subset of the input bits and $c \in \{0,1\}$.

Proof. The class of functions which can be computed by Clifford circuits with correctness probability greater than 1/2 is actually quite limited. Let $f : \{0,1\}^n \to \{0,1\}$ be a function which can be computed by a Clifford circuit C. Then we can simulate this circuit as in Lemma 1, where we give Alice the whole input, i.e., $m_A = n$ and $m_B = 0$.

Inspecting the proof of Lemma 1 we see that in each step Alice always updates her a_i's and b_i's by computing the parity of two bits. So, the final bit she sends over, say a_i, is just the parity of some of the input bits. Thus we can write $a_i = \bigoplus_{j \in S} x_j$, for some $S \subseteq [n]$. Bob initializes all his quantum bits to $|0\rangle$, so he starts with the state $|\psi^0\rangle = |0 \dots 0\rangle$. Further, Bob just applies the circuit C to his state and measures the i-th qubit of $X^{a_i} C|\psi^0\rangle$ in the computational basis.

It is known that the probability for measuring 1 in a Clifford circuits is either 0, 1/2 or 1 (see [12] page 463). It cannot be 1/2 in our case, because that would mean that the circuit does not compute a result of function f. So, measuring the i-th bit of $C|\psi^0\rangle$ yields a bit $c \in \{0,1\}$ with certainty. But this means that $f(x) = c \oplus a_i = c \oplus \bigoplus_{j \in S} x_j$.□. □

We mention that Aaronson and Gottesman proved [1] that there is a log-space machine which on any input a Clifford circuit[2] C a classical circuit consisting only of CNOT and NOT gates, which outputs 1 iff C on input of the all zero state $|0\rangle^{\otimes n}$ accepts. Our corollary extends this slightly: For every Clifford circuit C computing a boolean function, there is an equivalent (for classical inputs) classical circuit which uses only NOT- and CNOT-gates. Using the result from [1] we see that we can compute the bit c in the proof of Corollary 1 in log-space and it is also clear that the circuit Alice uses to compute a_i can be computed in log-space.

4.2 Extensions and Implications

Remark 1. Our threshold theorem says the following. Let f be a function such that it requires more than one bit of communication in order to compute it, when the input bits are partitioned over Alice and Bob. There is no quantum circuit consisting out of *perfect* Clifford operations and single qubit gates with noise $\hat{\theta}(45\%)$ that can computes f. We can strenthen this result to allow a small amount of perfect single qubits as well. Assume that f requires m bits of communication to be computed. There is no quantum circuit that uses perfect Clifford operations, s perfect single qubit gates, and single qubit gates with noise $\hat{\theta}$that computes f, for $2s + 1 < m$. The reason we get this strengthening is because in our simulation, Lemmas 1 and 2, Alice sends to Bob whenever he wants to perform a perfect single qubit gate on some qubit, her classical share a and b of that specific qubit. Bob can now perform the perfect qubit gate on that qubit and they proceed as in Lemma 1 and 2. By the end of the simulation Alice has sent $2s + 1$ bits to Bob and he will be able to compute f, contradicting that the communication complexity of f is at least $m > 2s + 1$.

Remark 2. Theorem 1 actually also holds when Alice and Bob get quantum states as inputs and they are provided with entanglement. It is no problem for Bob to start with a quantum state as an input. For Alice we do the following. We let her teleport her

[2]which they call Stabilizer circuits

quantum input to Bob bit by bit, using the standard scheme for teleportation (see e.g. [12]). When Alice teleports a qubit, which corresponds to the i-th input qubit of the simulated circuit C, she measures two classcial bits. Now, if she does not send these to Bob, but rather initializes her a_i, b_i with these bits, Alice and Bob obtain the correct representation for qubits of C as in Lemma 1. Since the inner product function has communication complexity $\Omega(n)$ even in the presence of entanglement [6] we see that Theorem 1 is also true for quantum inputs.

Note however, that now Bob can no longer compute the result of the circuit with certainty, because he does not even know his input state. The correctness probability will only be the same as that of the simulated circuit C itself.

Remark 3. From the proof of Lemma 2 we see that the rotation matrix S which achieves the optimal value, is

$$\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}. \tag{17}$$

If the (3,3)-entry is 1, this matrix performs a rotation of $\pi/4$ around one of the axes. The $\pi/8$-gate

$$T = \begin{pmatrix} \exp(-i\pi/8) & 0 \\ 0 & \exp(i\pi/8) \end{pmatrix}$$

performs a rotation of $\pi/4$ around the z-axis. So, the $\pi/8$-gate is the one which is "hardest" to simulate by gates from CLIFFORD.

Remark 4. It is straightforward to extend the results to functions with m output bits, if the communication complexity of the function is also higher than m.

5 Acknowledgements

We thank Nebojša Gvozdenović and Hartwig Bosse and also Troy Lee and Stephanie Wehner for useful discussions and Komei Fukuda for providing the software `cdd+` [7]. We also thank Scott Aaronson and Daniel Gottesman for discussions about their work.

References

[1] S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. Technical report, arXiv:quant-ph/0406196 v4, 2004.

[2] Dorit Aharonov and Michael Ben-Or. Polynomial simulations of decohered quantum computers. *Proceedings of the 37th IEEE FOCS*, page 46?55, 1996.

[3] Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error. In *STOC*, pages 176–188, 1997.

[4] A. R. Calderbank and Peter Shor. Good quantum error-correcting codes exist. *Phys. Rev. A 52, 2493*, 1996.

[5] R. Cleve and J. Watrous. Fast parallel circuits for the quantum fourier transform. In *FOCS*, pages 526–536, 2000.

[6] Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp. Quantum entanglement and the communication complexity of the inner product function. *Selected papers from the First NASA International Conference on Quantum Computing and Quantum Communications*, pages 61–74, 1998.

[7] Komei Fukuda. *Software* `cdd+`. available from http://www.ifor.math.ethz.ch/ fukuda.

[8] Aram Harrow and Michael Nielsen. How robust is a quantum gate in the presence of noise? *Phys. Rev. A 68, 012308*, 2003.

[9] E. Knill. Quantum computing with realistically noisy devices. *Nature 434, 39-44*, 2005.

[10] Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, 1997.

[11] G. Nebe, E. M. Rains, and N. J. A. Sloane. The invariants of the clifford groups. *Designs, Codes and Cryptography*, 24(99), 2001. math.CO/0001038.

[12] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[13] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer Press, New York, 1999.

[14] Alexander Razborov. An upper bound on the threshold quantum decoherence rate. *Quantum Information and Computation*, 4(3):222–228, 2004.

[15] P. Shor. Fault-tolerant quantum computation. In *Proc. 37th Annual Symposium on Foundations of Computer Science*, pages 56–65. IEEE Computer Society Press, 1996.

[16] Peter Shor. Scheme for reducing decoherence in quantum memory. *Phys. Rev. A 52, 2493*, 1995.

[17] R. Solovay. talk at mathematical sciences research institute. "".

[18] Andrew Steane. Error correcting codes in quantum theory. *Phys. Rev. Lett. 77, 793*, 1996.

[19] S. Virmani, S. Huelga, and M. Plenio. Classical simulatability, entanglement breaking, and quantum computation thresholds. Technical report, quant-ph/0408076, 2004.

A Computing the facets of the Clifford polytope

We give here the facet description for the Clifford polytope P defined in (8) as the convex hull of the set $\mathcal{C}$ consisting of the 24 Clifford rotations of the 3-space. Define the polyhedron

$$Q := \{S \in \mathbb{R}^{3\times 3} \mid \langle F, S\rangle \leq 1 \ \text{ for all } F \in \mathcal{F}\},$$

where $\mathcal{F}$ is as in (11). Our objective is to show the equality $P = Q$.

To start with, let us prove the easy inclusion $P \subseteq Q$. For this, let $C \in \mathcal{C}$ and $F \in \mathcal{F}$ of the form $F = C_1 B C_2$ with $C_1, C_2 \in \mathcal{C}$ and $B \in \{B_1, B_1^T, B_2\}$. Then, $\langle F, C\rangle = \langle B, C_1^T C C_2^T\rangle$. As $C_1^T C C_2^T \in \mathcal{C}$, it suffices to verify that $\langle B, C\rangle \leq 1$ for any $C \in \mathcal{C}$ and $B = B_1, B_2$. (We have used here the fact that $\mathcal{C}$ is a group which is closed under transposing matrices.) For $C \in \mathcal{C}$, the inequality $\langle B_1, C\rangle \leq 1$ is obvious and the inequality $\langle B_2, C\rangle \leq 1$ can be checked by direct inspection.

The reverse inclusion $Q \subseteq P$ follows from the following result.

Theorem 2. *Any facet of the polytope P is defined by an inequality of the form $\langle F, S\rangle \leq 1$ where $F \in \mathcal{F}$.*

The rest of the Appendix is devoted to the proof of this result. We first need to go in more detail into the structure of the Clifford matrices.

A.1 Preliminaries about the Clifford matrices

Each matrix $C \in \mathcal{C}$ corresponds to a "signed permutation" (σ, s), where $\sigma \in \mathrm{Sym}(3)$ and $s \in \{\pm 1\}^3$. Namely, C has nonzero entries precisely at the $(\sigma(i), i)$-positions with $C_{\sigma(i),i} = s_i$ for $i = 1, 2, 3$; we then also denote C as $C_{\sigma,s}$. The condition $\det(C) = 1$ translates into $s_1 s_2 s_3 = \mathrm{sign}(\sigma)$; that is, $s_1 s_2 s_3 = 1$ if σ is an even permutation (i.e., one of $\sigma_1 := (1,2,3)$, $\sigma_2 := (2,3,1)$, $\sigma_3 := (3,1,2)$) and $s_1 s_2 s_3 = -1$ if σ is an odd permutation (i.e., one of $\sigma_4 := (1,3,2)$, $\sigma_5 := (2,1,3)$, $\sigma_6 := (3,2,1)$). Thus the set $\mathcal{C}$ of Clifford matrices is naturally partitioned into six subclasses

$$\mathcal{C} = \bigcup_{\sigma \in \mathrm{Sym}(3)} \mathcal{C}_\sigma, \ \text{where } \mathcal{C}_\sigma := \{C_{\sigma,s} \mid s \in \{\pm 1\}^3, \ s_1 s_2 s_3 = \mathrm{sign}(\sigma)\}$$

with $|\mathcal{C}_\sigma| = 4$. For convenience we display in the table below the six subclasses $\mathcal{C}_\sigma$; the nonzero entries are indicated by $*$.

Even permutations	Odd permutations
$\sigma_1 = (1,2,3) \quad \mathcal{C}_{\sigma_1} : \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{pmatrix}$	$\sigma_4 = (1,3,2) \quad \mathcal{C}_{\sigma_4} : \begin{pmatrix} * & 0 & 0 \\ 0 & 0 & * \\ 0 & * & 0 \end{pmatrix}$
$\sigma_2 = (2,3,1) \quad \mathcal{C}_{\sigma_2} : \begin{pmatrix} 0 & 0 & * \\ * & 0 & 0 \\ 0 & * & 0 \end{pmatrix}$	$\sigma_5 = (2,1,3) \quad \mathcal{C}_{\sigma_5} : \begin{pmatrix} 0 & * & 0 \\ * & 0 & 0 \\ 0 & 0 & * \end{pmatrix}$
$\sigma_3 = (3,1,2) \quad \mathcal{C}_{\sigma_3} : \begin{pmatrix} 0 & * & 0 \\ 0 & 0 & * \\ * & 0 & 0 \end{pmatrix}$	$\sigma_6 = (3,2,1) \quad \mathcal{C}_{\sigma_6} : \begin{pmatrix} 0 & 0 & * \\ 0 & * & 0 \\ * & 0 & 0 \end{pmatrix}$

Table 1

The following observation can be directly verified and will be useful for the proof.

Observation 1. *Let $\sigma \in \mathrm{Sym}(3)$. Then, $\sum_{C\in\mathcal{C}_\sigma} C = 0$. Moreover, for any position $(\sigma(i), i)$ corresponding to a nonzero entry for matrices in $\mathcal{C}_\sigma$ and for $d \in \{\pm 1\}$, there exist $C, C' \in \mathcal{C}_\sigma$ with $C + C' = 2dE_{\sigma(i),i}$, which implies $dE_{\sigma(i),i} \in P$. Thus, $\pm E_{i,j} \in P$ for any $i, j = 1, 2, 3$.*

We now proceed with the proof of Theorem 2. Let $\langle F, S\rangle \le b$ be an inequality defining a facet of P, where $F \in \mathbb{R}^{3\times 3}$ and $b \in \mathbb{R}$. That is, the inequality $\langle F, S\rangle \le b$ is valid for P, which means that $\langle F, S\rangle \le b$ holds for any $S \in P$, and the set

$$\mathcal{R}_F := \{C \in \mathcal{C} \mid \langle F, C\rangle = b\}$$

contains nine affinely independent matrices. Without loss of generality, we may assume that $b = 1$. Indeed, $b \ge 0$ since $0 \in P$. Moreover, $b > 0$ for, otherwise, we would have $F_{ij} = 0$ for all $i, j = 1, 2, 3$, implying $F = 0$, in view of Observation 1. Thus, by rescaling, we can now assume that the facet is of the form $\langle F, S\rangle \le 1$. We sometimes speak of the "facet F" for short. Our objective is to show that $F = C_1BC_2$ for some $C_1, C_2 \in \mathcal{C}$, $B \in \{B_1, B_1^T, B_2\}$.

Call $F, F' \in \mathbb{R}^{3\times 3}$ *equivalent* if $F' = C_1FC_2$ for some $C_1, C_2 \in \mathcal{C}$. Then, as $\mathcal{C}$ is a group, $\langle F', S\rangle \le 1$ defines a facet of P if and only if $\langle F, C\rangle \le 1$ does. Moreover, $\mathcal{R}_{F'} = C_1\mathcal{R}_FC_2 = \{C_1CC_2 \mid C \in \mathcal{R}_F\}$. This property will be used repeatedly throughout the proof as it permits to exploit symmetry and to reduce the number of case checking.

The proof is based on a detailed inspection of the structure of the set $\mathcal{R}_F$. We begin with collecting several properties of the matrix F and the set $\mathcal{R}_F$.

Observation 2. $|\mathcal{R}_F \cap \mathcal{C}_\sigma| \le 3$ *for any* $\sigma \in \mathrm{Sym}(3)$.

Proof. If $\mathcal{C}_\sigma \subseteq \mathcal{R}_F$, then $\langle F, C\rangle = 1$ for any $C \in \mathcal{C}_\sigma$, which implies $4 = \sum_{C\in\mathcal{C}_\sigma}\langle F, C\rangle$, contradicting the fact that $\sum_{C\in\mathcal{C}_\sigma} C = 0$ by Observation 1. ☐ ☐

Observation 3. *If $F_{ij} = d \in \{-1, 1\}$, then all $C \in \mathcal{C}$ with $C_{ij} = d$ belong to R_F.*

Proof. Let $C \in \mathcal{C}$ with $C_{ij} = d$. There exists $C' \in \mathcal{C}$ with $C + C' = 2dE_{ij}$. Summing $\langle F, C\rangle \le 1$ and $\langle F, C'\rangle \le 1$ yields $\langle F, C + C'\rangle \le 2$. As $\langle F, C + C'\rangle = 2dF_{ij} = 2$, we have the equalities $\langle F, C\rangle = \langle F, C'\rangle = 1$, which implies $C \in \mathcal{R}_F$. ☐ ☐

Observation 4. *Let $C \ne C' \in \mathcal{R}_F \cap \mathcal{C}_\sigma$ (for some $\sigma \in \mathrm{Sym}(3)$) and assume that $C_{\sigma(i),i} = C'_{\sigma(i),i} = d \in \{\pm 1\}$ for some $i \in \{1, 2, 3\}$. Then, $F_{\sigma(i),i} = d$ and $F_{\sigma(j),j} + \mathrm{sign}(\sigma)dF_{\sigma(k),k} = 0$ with $\{j, k\} = \{1, 2, 3\} \setminus \{i\}$.*

Proof. Equality $F_{\sigma(i),i} = d$ follows from the fact that $C + C' = 2dE_{\sigma(i),i}$. Then, $1 = \langle F, C\rangle$ implies $0 = F_{\sigma(j),j}C_{\sigma(j),j} + F_{\sigma(k),k}C_{\sigma(k),k}$. Using $C_{\sigma(i),i}C_{\sigma(j),j}C_{\sigma(k),k} = \mathrm{sign}(\sigma)$, we find $F_{\sigma(j),j} + \mathrm{sign}(\sigma)dF_{\sigma(k),k} = 0$. ☐ ☐

Our last observation is an easy corollary of the former two observations.

Observation 5. *If $F_{\sigma(i),i} = d \in \{\pm 1\}$ (for some $\sigma \in \mathrm{Sym}(3)$), then $F_{\sigma(j),j} + \mathrm{sign}(\sigma)dF_{\sigma(k),k} = 0$, where $\{i, j, k\} = \{1, 2, 3\}$.*

One can verify that, for $F = B_1, B_1^T$, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 2$ for all $\sigma \in \mathrm{Sym}(3)$ while, for $F = B_2$, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for $\sigma = \sigma_1, \sigma_5$. Based on this observation we now distinguish two cases: Either, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| \leq 2$ for all $\sigma \in \mathrm{Sym}(3)$ (in which case we show that F is equivalent to B_1 or B_1^T), or $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for some $\sigma \in \mathrm{Sym}(3)$ (in which case we show that F is equivalent to B_2).

A.2 The case $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for some $\sigma \in \mathrm{Sym}(3)$

Using symmetry, we may assume that $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$ for the (odd) permutation $\sigma = \sigma_4$. We prove this in detail to show how this kind of symmetry argument works. Define the matrices

$$C_1 = \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & - \\ 0 & - & 0 \end{pmatrix},\ C_2 = \begin{pmatrix} 0 & + & 0 \\ 0 & 0 & + \\ + & 0 & 0 \end{pmatrix},\ C_3 = \begin{pmatrix} 0 & 0 & + \\ + & 0 & 0 \\ 0 & + & 0 \end{pmatrix} \tag{18}$$

lying, resp., in $\mathcal{C}_{\sigma_4}$, $\mathcal{C}_{\sigma_3}$, $\mathcal{C}_{\sigma_2}$. Recall that $+, -$ stand for $1, -1$, respectively. Our assumption is that $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_i}| = 3$ for some $i = 1, \ldots, 6$; we show that one can replace F by another equivalent facet F' in such a way that $i = 4$ holds. For this, suppose first $i = 2, 3$. As the mapping $X \mapsto XC_i$ maps $\mathcal{C}_{\sigma_i}$ to $\mathcal{C}_{\sigma_1}$, we can replace the facet F by $F' := FC_i$ and then we find $|\mathcal{R}_{F'} \cap \mathcal{C}_{\sigma_1}| = 3$ since $\mathcal{R}_{F'} = \mathcal{R}_F C_i$. Thus we may assume $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_1}| = 3$. As the mapping $X \mapsto XC_1$ maps $\mathcal{C}_{\sigma_1}$ to $\mathcal{C}_{\sigma_4}$, replacing the facet F by $F' := FC_1$, we find $|\mathcal{R}_{F'} \cap \mathcal{C}_{\sigma_4}| = 3$. Thus we can now assume $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_i}| = 3$ for some $i = 4, 5, 6$. If $i = 5$, as the mapping $X \mapsto XC_3$ maps $\mathcal{C}_{\sigma_5}$ to $\mathcal{C}_{\sigma_4}$, replace F by $F' := FC_3$; if $i = 6$, the mapping $X \mapsto XC_2$ maps $\mathcal{C}_{\sigma_6}$ to $\mathcal{C}_{\sigma_4}$ and one can replace F by $F' := FC_2$; in both cases we get back to the case when $|\mathcal{R}_{F'} \cap \mathcal{C}_{\sigma_4}| = 3$.

Thus we now assume $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_4}| = 3$. Moreover, we may assume that the following matrices from $\mathcal{C}_{\sigma_4}$

$$\begin{pmatrix} + & 0 & 0 \\ 0 & 0 & - \\ 0 & + & 0 \end{pmatrix},\ \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & + \\ 0 & + & 0 \end{pmatrix},\ \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & - \\ 0 & - & 0 \end{pmatrix} \tag{19}$$

belong to $\mathcal{R}_F$. (To see this, replace if necessary F by FC, where $C \in \mathcal{C}_{\sigma_1}$.) Using Observation 4, we obtain $F_{11} = -1, F_{23} = -1, F_{32} = 1$. From this we get by Observation 3 that also the matrices

$$\begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & + \end{pmatrix},\ \begin{pmatrix} - & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & - \end{pmatrix} \in \mathcal{C}_{\sigma_1}, \tag{20}$$

$$\begin{pmatrix} 0 & 0 & + \\ + & 0 & 0 \\ 0 & + & 0 \end{pmatrix},\ \begin{pmatrix} 0 & 0 & - \\ - & 0 & 0 \\ 0 & + & 0 \end{pmatrix} \in \mathcal{C}_{\sigma_2}, \tag{21}$$

$$\begin{pmatrix} 0 & + & 0 \\ 0 & 0 & - \\ - & 0 & 0 \end{pmatrix},\ \begin{pmatrix} 0 & - & 0 \\ 0 & 0 & - \\ + & 0 & 0 \end{pmatrix} \in \mathcal{C}_{\sigma_3} \tag{22}$$

$$\tag{23}$$

belong to $\mathcal{R}_F$. By Observation 4, we also obtain $F_{22} = F_{33}$, $F_{12} = F_{31}$ and $F_{13} = -F_{21}$.

Claim 1. *There exists also an even permutation $\sigma \in \mathrm{Sym}(3)$ for which $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 3$.*

Proof. Assume for contradiction that, for $i = 1, 2, 3$, the set $\mathcal{R}_F \cap \mathcal{C}_{\sigma_i}$ contains only the respective two matrices from (20)-(22). Choose a subset $\mathcal{B} \subseteq \mathcal{R}_F$ consisting of nine affinely independent matrices and such that $\mathcal{R}_F \cap \mathcal{C}_{\sigma_4} \subseteq \mathcal{B}$. We have $|\mathcal{B} \cap \mathcal{C}_{\sigma_1}| \leq 1$, since the two matrices in (20) are affinely dependent with the last two matrices in (19). Similarly, $|\mathcal{B} \cap \mathcal{C}_{\sigma_2}| \leq 1$, $|\mathcal{B} \cap \mathcal{C}_{\sigma_3}| \leq 1$. As $|\mathcal{B}| = 9$, we deduce that $|\mathcal{B} \cap \mathcal{C}_{\sigma_5}| \geq 2$ or $|\mathcal{B} \cap \mathcal{C}_{\sigma_6}| \geq 2$. Assume first that $|\mathcal{B} \cap \mathcal{C}_{\sigma_5}| \geq 2$. Say, $C \neq C' \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_5}$. Then C and C' have the same nonzero entry $d \in \{-1, 1\}$ in some position (k, l). By Observation 4 this yields $F_{kl} = d$. Now, there is also an even permutation σ for which $k = \sigma(l)$. By Observation 3 we then deduce that at least two matrices from $\mathcal{C}_\sigma$ must be in $\mathcal{R}_F$, which contradicts our assumption. The other case $|\mathcal{B} \cap \mathcal{C}_{\sigma_6}| \geq 2$ goes analogously.□ □

It is sufficient to consider the case $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_1}| = 3$. Indeed, if $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_2}| = 3$, then one may replace F by $C_3 F C_4$ with C_3 as in (18) and

$$C_4 := \begin{pmatrix} 0 & 0 & + \\ - & 0 & 0 \\ 0 & - & 0 \end{pmatrix},$$

since the mapping $X \mapsto C_3 X C_4$ maps $\mathcal{C}_{\sigma_2}$ to $\mathcal{C}_{\sigma_1}$ and preserves the set of three matrices from (19), as well as the set of 6 matrices from (20)-(22) (namely, (20) $\to$ (22) $\to$ (21) $\to$ (20)). One can handle the case when $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_3}| = 3$ in the same way.

The set $\mathcal{R}_F$ already contains the matrices

$$D_1 := \begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & + \end{pmatrix}, \; D_2 := \begin{pmatrix} - & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & - \end{pmatrix}$$

from $\mathcal{C}_{\sigma_1}$ (displayed in (20)). The remaining two matrices of $\mathcal{C}_{\sigma_1}$ are

$$D_3 := \begin{pmatrix} + & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & + \end{pmatrix}, \; D_4 := \begin{pmatrix} + & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & - \end{pmatrix}.$$

If $D_4 \in \mathcal{R}_F$, one may replace the facet F by $F' := D_2 F D_1$ to obtain that $D_1, D_2, D_3 \in \mathcal{R}_{F'}$, since the mapping $X \mapsto D_2 X D_1$ maps $\{D_1, D_2, D_4\}$ to $\{D_1, D_2, D_3\}$ and leaves the set of 3 matrices from (19) invariant as well as the set of 6 matrices from (20)-(22). Thus we may assume that $D_3 \in \mathcal{R}_F$.

By Observation 4, we find that $F_{33} = F_{22} = 1$. As $F_{22} = 1$, Observation 5 implies that $F_{31} = F_{13}$. Similarly, $F_{33} = 1$ implies that $F_{12} = F_{21}$. Putting all equations together we obtain $F_{12} = F_{21} = -F_{13} = -F_{31} = -F_{12}$, implying they are all zero. Thus

$$F = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} B_2 \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{24}$$

A.3 The case $|\mathcal{R}_F \cap \mathcal{C}_\sigma| \leq 2$ for all $\sigma \in \mathrm{Sym}(3)$

Let again $\mathcal{B} \subseteq \mathcal{R}_F$ consist of nine affinely independent matrices. As $|\mathcal{R}_F| \geq 9$, $|\mathcal{R}_F \cap \mathcal{C}_\sigma| = 2$ for at least three permutations σ. W.l.o.g. we can assume that two of those permutations are odd permutations and that they are equal, say, to σ_4 and σ_6 (replacing if necessary F by an equivalent facet). Further we may assume $\mathcal{R}_F$ contains the following two matrices of $\mathcal{C}_{\sigma_4}$:

$$\begin{pmatrix} - & 0 & 0 \\ 0 & 0 & - \\ 0 & - & 0 \end{pmatrix}, \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & + \\ 0 & + & 0 \end{pmatrix} \in \mathcal{R}_F. \tag{25}$$

This can be seen using the following two mappings $X \mapsto C_2XC_2$ (with C_2 defined as in (18)) and $X \mapsto CX$ (with $C \in \mathcal{C}_{\sigma_1}$) which permit to map any subset of size 2 of $\mathcal{C}_{\sigma_4}$ to any other such subset and which preserve $\mathcal{C}_{\sigma_6}$ as well. We choose the basis $\mathcal{B}$ containing the two matrices of (25). ¿From Observation 4 we find $F_{11} = -1$ and $F_{23} = -F_{32} \neq \pm 1$; the latter inequality follows froim the fact that $|\mathcal{R}_F \cap \mathcal{C}_{\sigma_4}| = 2$ combined with Observation 3. As $F_{11} = -1$, by Observation 3,

$$\begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & + \end{pmatrix}, \begin{pmatrix} - & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & - \end{pmatrix} \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_1} \tag{26}$$

and Observation 4 implies $F_{22} = F_{33} \neq \pm 1$. At most one of the two matrices in (26) belongs to $\mathcal{B}$ since they are affinely dependent with the matrices in (25). Say, $|\mathcal{B} \cap \mathcal{C}_{\sigma_1}| = 1$.

Let us now examine which two matrices of $\mathcal{C}_{\sigma_6}$ belong to $\mathcal{R}_F$. Set

$$C_5 := \begin{pmatrix} + & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & - \end{pmatrix} \in \mathcal{C}_{\sigma_1},\ X_1 := \begin{pmatrix} 0 & 0 & - \\ 0 & - & 0 \\ - & 0 & 0 \end{pmatrix} \in \mathcal{C}_{\sigma_6}.$$

The two mappings $X \mapsto XC_5$ and $X \mapsto C_5X$ preserve the set of matrices in (25) and permit to map any other matrix of $\mathcal{C}_{\sigma_6}$ to the matrix X_1. Therefore we can assume w.l.o.g. that $X_1 \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$. The second matrix of $\mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$ does not have entry -1 at the position $(2,2)$ since, otherwise, $F_{22} = -1$ contradicting an earlier claim. Hence the second matrix in $\mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$ is

$$X_2 := \begin{pmatrix} 0 & 0 & + \\ 0 & + & 0 \\ - & 0 & 0 \end{pmatrix}, \quad \text{or } X_3 := \begin{pmatrix} 0 & 0 & - \\ 0 & + & 0 \\ + & 0 & 0 \end{pmatrix}.$$

1. Consider first the case when $X_2 \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$. Then, $F_{31} = -1$ and $F_{22} = -F_{13} \neq \pm 1$. As $F_{31} = -1$, we have

$$\begin{pmatrix} 0 & + & 0 \\ 0 & 0 & - \\ - & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & - & 0 \\ 0 & 0 & + \\ - & 0 & 0 \end{pmatrix} \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_3} \tag{27}$$

and $F_{12} = F_{23} \neq \pm 1$. As $\mathcal{B}$ contains at most three of the matrices X_1, X_2 and in (27), we must have $|\mathcal{B} \cap \mathcal{C}_{\sigma_2}| = 2$ or $|\mathcal{B} \cap \mathcal{C}_{\sigma_5}| = 2$. We obtained eralier that $F_{33} = F_{22} = -F_{13} \neq \pm 1$ and $F_{12} = F_{23} = -F_{32} \neq \pm 1$. In other words, the second and third columns of F contain no entry ± 1. On the other hand, the two matrices from $B \cap \mathcal{C}_{\sigma_i}$ $(i = 2, 5)$ have one common nonzero entry which therefore is located in the first column, at the position $(2, 1)$. This implies $F_{21} = \pm 1$.

(a) If $F_{21} = 1$, then Observation 5 implies $F_{12} = F_{33}$ and $F_{13} = -F_{32}$. Combining with the former relations on entries of F, we find

$$F = \begin{pmatrix} - & 0 & 0 \\ + & 0 & 0 \\ - & 0 & 0 \end{pmatrix}. \tag{28}$$

(b) If $F_{21} = -1$, then in the same way we find

$$F = \begin{pmatrix} - & 0 & 0 \\ - & 0 & 0 \\ - & 0 & 0 \end{pmatrix}. \tag{29}$$

In both cases we find that F is equivalent to B_1.

2. Consider now the case when $X_3 \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_6}$. Then, $F_{13} = -1$, $F_{22} = -F_{31} \neq \pm 1$,

$$\begin{pmatrix} 0 & 0 & - \\ - & 0 & 0 \\ 0 & + & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & - \\ + & 0 & 0 \\ 0 & - & 0 \end{pmatrix} \in \mathcal{R}_F \cap \mathcal{C}_{\sigma_2} \tag{30}$$

and $F_{21} = F_{32} \neq \pm 1$. In the same way as in the first case one finds that F is equivalent to B_1^T.

This concludes the proof of Theorem 2.