Noise in quantum and classical computation & non-locality

Unger, F.P.

Citation for published version (APA):
Chapter 5

Perfect stabilizer operations and noisy 1-qubit unitaries

This chapter is based on the paper


In this chapter we prove another noise bound for certain interesting classes of gates. We show that quantum circuits using perfect Clifford operations (CNOT, Hadamard, S, X, Y, Z and measurements in the computational basis) and noisy 1-qubit unitaries cannot be made fault-tolerant if the depolarizing noise on the 1-qubit gates is at least \( \hat{\theta} = (6 - 2\sqrt{2})/7 \approx 45\% \). We further show that if we additionally allow noise-free measurements in the computational basis, perfect classical control and perfect classical side processing, then above the noise rate of \( \hat{\theta} \) the circuits become efficiently simulatable on a classical computer. This last result is tight, since at lower noise rates this gate set becomes quantum universal, that is, it is possible to simulate any other quantum circuit efficiently. A corollary of our approach is that circuits consisting only of gates from the Clifford group cannot be universal for classical computation.

5.1 Introduction

In the previous chapters we have proved very general upper bounds on the noise rates for which fault-tolerant computing is possible. They were very general, because we allowed all possible gates (Chapter 3) respectively all unitary gates (Chapter 4) with bounded fan-in. All multi-qubit gates had the same amount of noise.
Restricted gate sets

However, this might not reflect the actual properties of a physical system. It might be hard in practice to physically perform all possible quantum operations. And among those which can be implemented it might be that certain quantum operations, including storing qubits, might have lower noise rates or even no noise at all. This particular situation was studied by Bravyi and Kitaev in [21], where they argue that some realistic proposals for topological quantum computation have exactly these properties. They consider an example, where perfect stabilizer operations, perfect classical control and perfect classical side-computation is allowed (see definitions in Section 5.2). Then they prove that if it is additionally possible to create certain pure 1-qubit states, then universal fault-tolerant quantum computation becomes possible. Because of this these states were called “magic” states. However, they also show that it is not necessary to be able to prepare these magic states perfectly. It is possible to distill better magic states from several copies of noisy magic states. Thus, perfect stabilizer operations and noisy magic states are enough for universal quantum computation. We will follow a similar path in this chapter.

Model and its computational power

We will assume that we can perfectly implement the set of all stabilizer operations \( \text{Stab} \), which includes Clifford gates \( \text{Clifford} \) (see Section 5.2.2), preparation of computational basis states and measurements in the computational basis (see Section 5.2.3). The Gottesman-Knill Theorem says that this set of gates can be efficiently simulated classically (see also [2]), and therefore it seems plausible that it is not sufficient for universal quantum computation. We prove this intuition later rigorously in Corollary 5.3.2. On the other hand, it is known that \( \text{Clifford} \) alone together with any other 1-qubit unitary gate, not generated by the gates in \( \text{Clifford} \), form a universal set of gates for quantum computation [89, 66]. We show, however, that such additional 1-qubit gates should not be too noisy.

Main results

More precisely, let \( \text{Clifford}^* \) be \( \text{Clifford} \) augmented with arbitrary 1-qubit unitary gates with depolarizing error at least \( \hat{\theta} = (6 - 2\sqrt{2})/7 \approx 45\% \). Then this set of gates is not capable of computing arbitrary functions and therefore is not even classically universal, which is proved in Theorem 5.4.3. In particular, fault-tolerant (quantum) computation cannot be performed if there is at least this level of noise. Our second result in Theorem 5.4.4 states that circuits with arbitrary classical control and that use gates from \( \text{Stab} \) and 1-qubit unitaries with noise at least \( \hat{\theta} \) can be simulated efficiently on classical computers. This last result is tight, as we explain in Section 5.5, based on results from [83, 21]. Their results imply that at noise rates less than \( \hat{\theta} \approx 45.3\% \) it is possible to do efficient
universal quantum computation if perfect stabilizer operations, perfect classical control and perfect classical side computation are available.

On the way, we give a characterization of the convex closure of all 1-qubit Clifford operations (Lemma 5.4.1).

Outline of proof ideas

We first show in Section 5.3 that the set of all Clifford operations is not universal, i.e., that it is is impossible to compute every function with bounded error. In particular, in Corollary 5.3.2, 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 (complementing results in [2]). The argument uses results from communication complexity.

We then show in Lemma 5.4.1 that all 1-qubit unitaries with noise at least $\hat{\theta}$ can be seen as probabilistic mixtures of 1-qubit Clifford operations. In the proof we first compute the smallest polytope $P$ that contains all 1-qubit Clifford gates. Then we show that any 1-qubit unitary with noise at least $\hat{\theta}$ lies inside $P$. Together with the fact that Clifford operations alone are not universal this establishes the first result Theorem 5.4.3.

The same Lemma together with the Gottesman-Knill theorem implies our second result Theorem 5.4.4.

Best gate

It is interesting to point out that among all 1-qubit unitary gates, the so-called $\pi/8$-gate (see Section 5.6) is the gate that requires the most noise to render it incapable of universal quantum computation by our approach. That is, augmenting the Clifford gates CLIFFORD with other gates (e.g., $\pi/16$-gates), our approach will yield stronger bounds on the tolerable noise level.

5.1.1 Organization

This chapter is organized as follows: In the beginning of Section 5.2 we introduce some notation and review some standard facts about Bloch-sphere representations from Section 2.5.1 and explain how depolarizing noise acts on the Bloch sphere. We then introduce the Clifford group and stabilizer operations in Section 5.2.2. Section 5.3 contains the result that the gate set CLIFFORD cannot be universal and only allows to compute parity functions, see Corollary 5.3.2. The proof uses a reduction to communication complexity (introduced in Section 2.4) and the fact that there are functions with non-trivial communication complexity. Section 5.4 shows that gates from CLIFFORD*, together with all stabilizer operations and perfect classical control are classically simulatable and thus probably not quantum-universal. It can be read independently of the preceding section. Section
Chapter 5. Perfect stabilizer operations and noisy 1-qubit unitaries

5.5 shows how results from [83, 21] imply a lower bound on $\hat{\theta}$. In Section 5.6 we then discuss some possible extensions, including different noise models and show that the $\pi/8$-gate is in some sense the most fault-tolerant gate.

5.2 Preliminaries and notation

Recall from Chapter 2 that $E_{ij}$ is the all-zero matrix, except for the entry $i, j$ which is equal to 1. For matrices $A, B \in \mathbb{R}^{3 \times 3}$ we define as before the inner product $\langle A, B \rangle$ as:

$$\langle A, B \rangle = \text{Tr}(A^TB) = \sum_{i,j \in \{1,2,3\}} a_{ij}b_{ij}.$$ The following fact is used repeatedly: $\langle A, BC \rangle = \langle B^T A, C \rangle$ for $A, B, C \in \mathbb{R}^{3 \times 3}$.

Recall Section 2.5 where we saw that 1-qubit states $\rho \in \mathbb{C}^{2 \times 2}$ are isomorphic to vectors $r \in \mathbb{R}^3$ via $\rho = I_2 + r \cdot \sigma = I_2 + r_x X + r_y Y + r_z Z$, and 1-qubit unitary operations $U \in \mathbb{C}^{2 \times 2}$ are isomorphic to rotations $R \in SO(3)$ via $U_n(\theta) = \exp \left( -i\theta n \cdot \sigma \right) = I_2 \cos \frac{\theta}{2} - i n \cdot \sigma \sin \frac{\theta}{2}$, where $n \in \mathbb{R}^3$ with $||n|| = 1$ is the axis and $\theta \in \mathbb{R}$ the angle of the rotation $R$. 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 $f : SO(3) \rightarrow \mathbb{C}^{2 \times 2}$ with the property that for all unitary $U \in \mathbb{C}^{2 \times 2}$ it holds that $f(R_U) = \alpha U$ 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 $0 \leq p_i$, 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 $R_E = \sum_i p_i R_i$.

5.2.1 Noise

The noise model we consider is again depolarizing noise. We repeat its definition from page 34. A 1-qubit state $\rho$ to which depolarizing noise $p$ is applied, becomes

$$\rho \mapsto (1 - p)\rho + pI/2.$$ Thus, with probability $1 - p$ the state is not changed, and with probability $p$ the state is replaced with the completely mixed state.
5.2. Preliminaries and notation

It is not hard to see that applying depolarizing noise \( p \) to \( \rho = \frac{I}{2} + r \cdot \sigma / 2 \) yields \( \rho' = \frac{I}{2} + r' \cdot \sigma / 2 \), 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 operation \( U \) with noise \( p \) if it transforms states \( \rho \) into

\[
(1 - p)U\rho U^\dagger + pI/2.
\]

This quantum operation can be seen as a two-stage process, in which first \( U \) and then depolarizing noise is applied. Let \( R_U \in \mathbb{R}^{3\times3} \) 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 Section 8.3 in [68] for more details.

### 5.2.2 Clifford group

The \((n\text{-qubit})\) Clifford group contains all unitary operations that can be written as a product of tensor products of \( S, H \) and CNOT\(^1_2 \) (see equation (5.2)).

\[
\text{CNOT}^1_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \quad (5.2)
\]

We denote the set of all operations which can be generated in this way by \textbf{CLIFFORD}. In particular, the Clifford group contains also the Pauli group \((S^2 = Z, HS^2H = X \text{ and } ZX = iY)\), which is the tensor product of all Pauli operators \( I, X, Y, Z \).

Let \textbf{CLIFFORD}\(^*\) be the set of gates consisting of \textbf{CLIFFORD} and arbitrary 1-qubit gates followed by depolarizing noise at least \( \hat{\theta} = (6 - 2\sqrt{2})/7 \).

**Bloch-vector representation of Clifford operations** For a state with Bloch vector \( r \) we get:

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

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-gate we similarly have

\[
H \left( \frac{1}{2}I + \frac{r_x}{2}X + \frac{r_y}{2}Y + \frac{r_z}{2}Z \right) H^* = \frac{1}{2}I + \frac{r_z}{2}X - \frac{r_y}{2}Y + \frac{r_x}{2}Z.
\]
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 $\pi$ around the axis $(1, 0, 1)/\sqrt{2}$.

We define $C$ as the set of matrices which can be generated from $R_S$ and $R_H$. A $C \in C$ is called a Clifford (rotation) matrix. It is not hard to see that $C$ 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 $C$, 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 B we need (and explain) more details about Clifford rotation matrices, which are only necessary for one technical result in Lemma 5.4.1, which can alternatively be obtained by computer software [46].

5.2.3 Stabilizer operations and the Gottesman-Knill Theorem

We conclude this section by defining a few more terms. The set of stabilizer operations, denoted by $\text{Stab}$, contains all operations generated by the Clifford group $\text{CLIFFORD}$ and additionally preparations of computational basis states and measurements in the computational basis. The Gottesman-Knill Theorem says that this set of gates can be efficiently simulated classically, see also [2].

Classical control (see page 35) means that later gates may arbitrarily depend on earlier measurement outcomes. In particular, this means that arbitrary classical side computation is allowed. In [2] it is shown that quantum circuits using only operations from $\text{Stab}$ in which perfect classical control is allowed are also efficiently simulatable on a classical computer, i.e., the simulation can be done with at most a polynomial overhead over the number of quantum operations and the number of (classical) operations needed for the classical control.

5.3 The power of Clifford circuits

The main idea of this section is as follows. Assume we have a Clifford circuit $C$ (i.e. a circuit composed of the gates in (5.2)) with $n$ classical input bits $x = x_1, \ldots, x_n$ 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 bits $S \subseteq \{1, \ldots, n\}$ of $x$ and Bob has bits $\{1, \ldots, n\}\setminus S$. We first show how Alice, with the help of Bob, can compute the value of $C$ on $x$ with just a single
5.3. The power of Clifford circuits

classical bit of communication (Lemma 5.3.1) for any partition $S$. Recall that in Section 2.4 we defined the worst-case partition communication complexity of $f : \{0, 1\}^n \rightarrow \{0, 1\}$ as $D_{\text{worst}}(f) = \max_{S \subseteq \{1, \ldots, n\}} D^S(f)$, where $D^S(f)$ is the (deterministic) communication complexity of $f$ when the bits in $S$ are given to Alice and all others to Bob. Hence, Clifford circuits can at the very best compute only those functions that require a single bit of communication for any partition of the inputs; it is well known that most functions require more than one bit of communication, see Section 2.4.

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

5.3.1. Lemma. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a function that is computable with unbounded error\(^1\) by a quantum circuit $C$ that uses only gates from CLIFFORD, ancillas initialized to $|0\rangle$ and one single-qubit measurement in the computational basis, which determines the output. Then the deterministic worst-case partition communication complexity of $f$ is at most one bit.

Proof: In the simulation of the circuit $C$ we represent the $j$-th qubit by two shares: a classical share consisting of two bits $a_j, b_j$, and a quantum share consisting of 1-qubit. A state $|\psi_C\rangle$ of $C$ will be encoded by

$$|\psi_C\rangle := \bigotimes_j X_j^{a_j} Z_j^{b_j} |\psi\rangle,$$

where $|\psi\rangle$ is the state of all quantum shares and the indices $j$ denote the qubits on which the operators $X$ and $Z$ act. We call $|\psi_C\rangle$ the logical state (of $C$).

Assuming that the set of qubits of $C$ is encoded in this manner, the operations $H, S$, and CNOT can be applied to the logical qubits by separately performing operations on the shares that encode them (i.e., the logical qubits do not have to be reconstructed). The reason why this works is because for any Clifford operation $C = H, S, \text{CNOT}_1$ and any tensor product of Pauli operators $P$ there is a tensor product of Pauli operators $P_2$ with $CP_1 = P_2C$. For example, to apply $H$ to the logical qubit $i$, the two bits that make up its classical share are swapped and $H$ is applied to its quantum share. This works correctly because

$$H_j X_j^{a_j} Z_j^{b_j} |\psi\rangle = H_j X_j^{a_j} H_j H_j Z_j^{b_j} H_j |\psi\rangle = Z_j^{a_j} X_j^{b_j} H_j |\psi\rangle = (-1)^{a_j \wedge b_j} X_j^{b_j} Z_j^{a_j} H_j |\psi\rangle,$$

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

\(^1\)That means, that the output is only correct with probability greater than $1/2$, but can go arbitrarily close to $1/2$. 

Chapter 5. Perfect stabilizer operations and noisy 1-qubit unitaries

To apply $S$ to a logical qubit, the $b$-part of the classical share is updated to $b := a \oplus b$ and $S$ is applied to its quantum share. This case can be verified by noting that

$$S_j X^a_j Z^b_j |\psi\rangle = i^{a_j} X^a_j S_j Z^a_j Z^b_j |\psi\rangle = i^{a_j} X^a_j Z^b_j S_j |\psi\rangle,$$

where we note that $i^{a_j}$ is a global phase.

To simulate the application of a CNOT gate on two logical qubits, with classical shares $a_1 b_1$ and $a_2 b_2$, we update $a_2 := a_1 \oplus a_2$, $b_1 := b_1 \oplus b_2$ and CNOT 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 (see also Figure 4.3 on page 54)

$$\text{CNOT}_2^1 (X \otimes I) = (X \otimes X) \text{CNOT}_2^1$$
$$\text{CNOT}_2^1 (I \otimes X) = (I \otimes X) \text{CNOT}_2^1$$
$$\text{CNOT}_2^1 (Z \otimes I) = (Z \otimes I) \text{CNOT}_2^1$$
$$\text{CNOT}_2^1 (I \otimes Z) = (Z \otimes Z) \text{CNOT}_2^1.$$

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 construct, see also Figure 5.1: for each of Alice’s input bits $x_j$, Alice sets her classical share to $a_j := x_j, b_j := 0$ and Bob sets his quantum share to $|0\rangle_j$; for each of Bob’s input bits $y_j$, Alice sets her classical share to $a_j := b_j := 0$ and Bob sets his quantum share to $|y_j\rangle_j$. If the $j$-th input bit to circuit $C$ is an ancilla qubit (initialized to $|0\rangle$) then Alice sets $a_j := x_j, b_j := 0$ and Bob sets the $j$-th qubit to $|0\rangle_j$. Note that the logical state encoded in this way is $|x\rangle |y\rangle |0 ... 0\rangle$, where $|0 ... 0\rangle$ denotes the ancilla qubits.

With this representation, Alice and Bob can simulate the execution of circuit $C$ on input $|x\rangle |y\rangle |0 ... 0\rangle$ without any communication as explained above. In
5.3. The power of Clifford circuits

In particular, they can obtain the shares of the output qubit of $C$, which without loss of generality we assume to be the first qubit of $C$. For Bob to obtain the result of measuring the (logical) 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 in the computational basis. 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 with high enough precision on a classical computer. Then, upon receipt of the classical bit from Alice, he can exactly determine the output probabilities of his measurement to determine which outcome is more likely.

The next Corollary characterizes exactly all functions computable by Clifford circuits. From Lemma 5.3.1 we get that this set is very limited and far from being universal.

5.3.2. COROLLARY. All functions $f : \{0, 1\}^n \to \{0, 1\}$ 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 not depending on the input bits and $c \in \{0, 1\}$.

Proof: It is clear that all functions $f$ of this form can be computed by a Clifford circuit. We now also prove the reverse.

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 5.3.1, where we give Alice the whole input, i.e., with the notation before Lemma 5.3.1 this means $S = \{1, \ldots, n\}$.

Inspecting the proof of Lemma 5.3.1 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 $a_1$ she sends over 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 $|\psi^0\rangle = |0\ldots0\rangle$. 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 circuit is either 0, 1/2 or 1 (see [68] page 463). It cannot be 1/2 in our case, because that would mean that the circuit does not compute $f$. 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 [2] that there is a log-space machine which transforms a Clifford circuit $C$ into a classical circuit $C'$.
consisting only of CNOT and NOT gates, with the property that \( C \) accepts the all zero state \( |0\rangle^\otimes n \) iff \( C' \) accepts the (classical) all zero input. 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 [2] we see that we can compute the bit \( c \) in the proof of Corollary 5.3.2 in log-space and it is also clear that the circuit Alice uses to compute \( a_i \) can be computed in log-space.

5.3.3. Remark. It is trivial to extend Lemma 5.3.1 to functions with \( m \) output bits, if the communication complexity of the function is also higher than \( m \), resulting in a scheme that uses \( m \) bits of communication.

5.4 Simulating 1-qubit unitaries by Clifford gates

We want to extend Lemma 5.3.1, by replacing CLIFFORD with CLIFFORD*. We show in Lemma 5.4.1 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 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 the 1-qubit Clifford gates in Bloch sphere representation.

Combining Lemmas 5.3.1 and 5.4.1, we get that for all circuits with CLIFFORD*-gates and any distribution of its input bits among Alice and Bob, the output of the circuit can be obtained with a single bit of communication (Lemma 5.4.2). Using the fact that there are functions which require communication more than one bit, we get our main result (Theorem 5.4.3): The set of gates in CLIFFORD* cannot be universal. We also generalize our result to the case that the inputs are quantum states.

We first show how one can simulate arbitrary 1-qubit gates with depolarizing noise \( \hat{\theta} = (6 - 2\sqrt{2})/7 \) with a probabilistic mixture of Clifford operations.

5.4.1. Lemma. 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
\[
E_U(\rho) = \sum_{C \in \mathcal{C}} p_C U_C \rho U_C^*
\]
and \( U_C \) is a Clifford operation corresponding to the Clifford rotation matrix \( C \).
**Proof:** Using Section 2.5.1 and Section 5.2 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

$$ (1 - \hat{\theta})S = \sum_{C \in \mathcal{C}} p_C C. \quad (5.7) $$

We will prove this latter statement. Define the *Clifford polytope* $P := \text{conv}(\mathcal{C}) = \left\{ S \mid S = \sum_{C \in \mathcal{C}} p_C C, p_C \geq 0, \sum_{C \in \mathcal{C}} p_C = 1 \right\}$ (5.8)
as the convex hull of the 24 Clifford rotation matrices in $\mathbb{R}^{3\times3}$. We have to prove (see also Figure 5.2)

$$ (1 - \hat{\theta})S \in P \text{ for any } S \in SO(3). \quad (5.9) $$

Figure 5.2: The polytope $P$ (schematically in two dimensions)

Rotation matrices (corresponding to 1-qubit unitaries) are depicted by patches of a circle. The polytope $P$ spanned by the Clifford operators $C_i$ is depicted by the rectangle, with facets $F_i \in \mathcal{F}$. For every rotation matrix $S$ on the circle there is a smallest value $p$ such that shrinking $S$ by a factor $(1 - p)$ gives a point inside $P$. Then $\hat{\theta}$ is the maximum of such $p$ over all rotation matrices $S$. 
For this we use the fact that the Clifford polytope can be alternatively described by its facet description:

\[ P = \{ S \in \mathbb{R}^{3\times3} \mid \langle F, S \rangle \leq 1 \text{ for all } F \in \mathcal{F} \}, \quad (5.10) \]

where

\[ \mathcal{F} := \{ C_1 BC_2 | C_1, C_2 \in \mathcal{C}, B \in \{ B_1, B_1^T, B_2 \} \}, \quad (5.11) \]

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

One can use the software from [46] for computing the facet description (5.10); we will give a direct proof in Appendix B. In view of (5.10), our claim (5.9) is equivalent to

\[ (1 - \hat{\theta}) \langle F, S \rangle \leq 1 \text{ for all } S \in SO(3), F \in \mathcal{F}. \quad (5.12) \]

Let \( F \in \mathcal{F} \) be of the form \( F = C_1 BC_2 \) where \( C_1, C_2 \in \mathcal{C} \). As \( \langle F, S \rangle = \langle C_1^T SC_2^T, B \rangle \) and \( C_1^T SC_2^T \in SO(3) \), (5.12) is equivalent to

\[ \langle S, B \rangle \leq \frac{1}{1 - \hat{\theta}} = 2\sqrt{2} - 1 \text{ for all } B \in \{ B_1, B_2 \}, S \in SO(3). \quad (5.13) \]

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 (5.13) for \( B = B_2 \). Write \( S \in \mathbb{R}^{3\times3} \) as

\[ S = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}. \quad (5.14) \]

It is well-known that it is necessary and sufficient for \( S \in SO(3) \) that the column vectors \( a = (a_1, a_2, a_3)^T, b = (b_1, b_2, b_3)^T \) and \( c = (c_1, c_2, c_3)^T \) satisfy

\[ a^T b = 0, \quad c = a \times b, \quad a^T a = 1, \quad b^T b = 1, \quad (5.15) \]

where \( \times \) denotes the vector product, defined as

\[ a \times b := (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)^T. \]

Recall that, for \( a, b, c \) as in (5.15), \( a = b \times c \) and \( b = c \times a \). Using \( c_3 = a_1 b_2 - a_2 b_1 \), we obtain \( \langle B, S \rangle = a_1 - a_2 + b_1 + b_2 - a_1 b_2 + a_2 b_1 \). Therefore our task is now to prove that the optimum value of the program

\[
\begin{align*}
\max & \quad f := a_1 - a_2 + b_1 + b_2 - a_1 b_2 + a_2 b_1 \\
\text{s.t.} & \quad g_1 := a_1^2 + a_2^2 + a_3^2 = 1 \\
& \quad g_2 := b_1^2 + b_2^2 + b_3^2 = 1 \\
& \quad g_3 := a_1 b_1 + a_2 b_2 + a_3 b_3 = 0
\end{align*}
\quad (5.16)
\]
5.4. Simulating 1-qubit unitaries by Clifford gates

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 (5.16). 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 [70]. (Here the gradient vector $\nabla g_i(a, b) = (\frac{\partial}{\partial a_1}, \frac{\partial}{\partial a_2}, \frac{\partial}{\partial a_3}, \frac{\partial}{\partial b_1}, \frac{\partial}{\partial b_2}, \frac{\partial}{\partial b_3})^T g_i(a, b)$ consists of the partial derivatives with respect to the six variables $a_1, \ldots, b_3$.) 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 $(b_1, b_2, b_3)$

$$\begin{pmatrix} 1 - b_2 \\ -1 + b_1 \\ 0 \\ 1 + a_2 \\ 1 - a_1 \\ 0 \end{pmatrix} + 2\lambda_1 a + \lambda_3 b = 0$$

$$\begin{pmatrix} 1 - b_2 \\ -1 + b_1 \\ 0 \\ 1 + a_2 \\ 1 - a_1 \\ 0 \end{pmatrix} + 2\lambda_3 b + \lambda_3 a = 0.$$

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

$$0 = c_1(1 - b_2) + c_2(-1 + b_1) = c_1 - c_2 + a_3$$

$$0 = c_1(1 + a_2) + c_2(1 - a_1) = c_1 + c_2 + b_3.$$

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 $a_1^2 + b_1^2 = 1 = a_2^2 + a_3^2$ 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 (5.16) is equal to $2\sqrt{2} - 1$, which concludes the proof. ■
5.4.2. **Lemma.** Let \( f : \{0, 1\}^n \rightarrow \{0, 1\} \) be a function and \( K \) a quantum circuit for \( f \) with error probability at most \( \epsilon > 0 \) which uses only gates from \( \text{Clifford}^* \) and one final single qubit measurement in the computational basis. Then the randomized worst-case partition communication complexity (defined in Section 2.4 on page 27) of \( f \) is at most one bit, i.e., \( R_{\text{worst}}^\epsilon(f) \leq 1 \).

**Proof:** From Lemma 5.3.1 we know how two parties, Alice and Bob, can simulate perfect Clifford gates. From Lemma 5.4.1 we know how they can jointly simulate the other noisy 1-qubit gates in \( \text{Clifford}^* \), 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.

5.4.3. **Theorem.** The set of gates from \( \text{Clifford} \) together with 1-qubit gates with depolarizing noise more than \( \tilde{\theta} \approx 45\% \) and one single-qubit measurement is not sufficient for arbitrary classical computation.

**Proof:** The result follows by Lemma 5.4.2 and the fact that there are functions with communication complexity greater than 1, for any bounded error.

In fact we have that none of the functions \( f \) with \( R_\epsilon(f) > 1 \) can be computed by \( \text{Clifford}^* \) circuits with error at most \( \epsilon \). From Corollary 5.3.2 we also get that the functions computable by \( \text{Clifford}^* \) are always probabilistic mixtures of parity functions.

If we additionally allow perfect stabilizer operations \( \text{Stab} \) and perfect classical control, we can state the following theorem.

5.4.4. **Theorem.** Any computation using

1. gates \( \text{Clifford}^* \)
2. perfect stabilizer operations \( \text{Stab} \)
3. perfect classical control and classical side computation

in which the input state is a computational basis state, can be simulated by a classical computer with at most polynomial overhead.

**Proof:** Follows immediately by Lemma 5.4.1 and the Gottesman-Knill Theorem.
5.5 Lower bound on $\hat{\theta}$

We now discuss whether it is possible to improve Theorems 5.4.3 and 5.4.4.

Theorem 5.4.3 states that fault-tolerant quantum computing is not possible if we have depolarizing noise at least $\hat{\theta} \approx 45\%$ on 1-qubit gates even if we can use perfect gates from CLIFFORD in our fault-tolerant circuit design. Is this optimal? Could it be that with less than $\hat{\theta}$ noise on the single-qubit gates and perfect gates from CLIFFORD still no fault-tolerant circuit design is possible? This is still an open question since we do not know if Theorem 5.4.3 is tight.

In contrast to this, the second result (Theorem 5.4.4) is tight, which was pointed out by Ben Reichardt [80]. The argument builds upon magic-state distillation, introduced in [21], and goes as follows. Assume we have at our disposal noisy $\pi/8$-gates $T'$, with depolarizing noise strictly less than $\hat{\theta}$, i.e. $T'(\rho) = (1-p)T\rho T^\dagger + pI/2$ with $p < \hat{\theta}$, where $T$ is the perfect $\pi/8$ gate, see equation (5.19). Then apply $T'$ to the second half of an EPR-pair and measure the observable $Z \otimes Z$, which can be implemented as a measurement in the computational basis with additional gates from CLIFFORD. If the outcome is $-1$ throw away the state and do the experiment again. If the outcome is $+1$, apply a CNOT from the first to the second qubit, which gives

$$\frac{1}{2} \left( I + \frac{1-p}{1-p/2} \frac{1}{\sqrt{2}} X + \frac{1-p}{1-p/2} \frac{1}{\sqrt{2}} Y \right) \otimes |0\rangle\langle 0|.$$  
(5.17)

Using the result from [83, 82] an arbitrary supply of qubits in the state of the first qubit of (5.17) can be used to distill magic states in the $H$-direction, which together with stabilizer operations is sufficient for quantum computation. We do not know if this also holds for gates other than the $\pi/8$-gate.

5.6 Discussion and extensions

In this section we will discuss certain extensions and generalizations of our results.

Best gates

From the proof of Lemma 5.4.1 we see that the rotation matrix $S$ which achieves the optimal value, is

$$\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$  
(5.18)
Multiplying from the right by the Clifford-matrix $\text{diag}(1, -1, -1)$ we get a rotation around the $z$-axis by $\pi/4$. The $\pi/8$-gate

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

(5.19)

performs a rotation of $\pi/4$ around the $z$-axis. So, the $\pi/8$-gate and its symmetric versions are the ones which need the most depolarizing noise to be simulated by gates from Clifford.

### Worst case noise

In Lemma 5.4.1 we asked with how much depolarizing noise all 1-qubit unitary gates are equivalent to probabilistic mixtures of Clifford gates. Similarly to [98] one can also ask how much arbitrary noise is needed to make every gate a mixture of Cliffords. More precisely what is the value $\tilde{\theta} = \sup_{U \in SU(2)} p_U$, where $p_U$ is the infimum of all $p$ such that there is a completely positive trace-preserving 1-qubit quantum operation $E_U$ with the property that the noisy implementation of $U$

$$U' : \rho \mapsto (1 - p)U\rho U^\dagger + pE_U(\rho)$$

becomes a probabilistic mixture of Clifford operation.

In this section we will provide some bounds on $\tilde{\theta}$. Let $K \in SU(2)$ be any operation that in Bloch representation maps the $X$-eigenstate $v_X = (1, 0, 0)$ to $u = \frac{1}{\sqrt{3}}(1, 1, 1)$. Note that a probabilistic mixture of 1-qubit Clifford operations $C = \sum_i p_i C_i$ can map $v_X$ only into the octahedron $\mathcal{O}$ spanned by $v_X = (1, 0, 0)$, $v_Y = (0, 1, 0)$ and $v_Z = (0, 0, 1)$ and their negatives $-v_X, -v_Y, -v_Z$ (see also [21]). Note that the state of $\mathcal{O}$ which is closest to $u$ is $\frac{1}{2}(1, 1, 1) = \frac{1}{\sqrt{3}} u$ and their distance is $||u - u/\sqrt{3}||_2 = 1 - \frac{1}{\sqrt{3}}$. The Bloch-state which is furthest away from $u$ is $-u$. All three of these states lie on a line. With this it is clear that the state $v_{\text{noise}}$ which needs the smallest noise $p$, such that $(1-p)u + p v_{\text{noise}}$ is inside the octahedron, is $-u$ and the optimal $p$ is $\frac{1}{2}(1 - \frac{1}{\sqrt{3}})$. This implies $21\% \approx \frac{1}{2}(1 - \frac{1}{\sqrt{3}}) \leq \tilde{\theta}$.

To get an upper bound, recall that by Lemma 5.4.1 for any gate $U \in SU(2)$ the operation

$$U' : \rho \mapsto (1 - p)U\rho U^\dagger + pI/2$$

is a Clifford operation, if $p \geq \tilde{\theta}$. Setting

$$E_U(\rho) = \frac{1}{3} (XU\rho U^\dagger X + YU\rho U^\dagger Y + ZU\rho U^\dagger Z)$$

and noting that for any 1-qubit density matrix $\rho$ it holds

$$\frac{I}{2} = \frac{1}{4} (\rho + X\rho X + Y\rho Y + Z\rho Z)$$
we can rewrite the action of $U'$ also as

$$U' : \rho \mapsto (1 - \frac{3}{4}p)U\rho U^\dagger + \frac{3}{4}p\mathcal{E}_U(\rho).$$

Thus, $\tilde{\theta} \leq \frac{3}{4}\hat{\theta} \approx 34\%$. Note that this is certainly not tight, since all gates, apart from the $\pi/8$-gate (and its symmetric versions), need less than $\tilde{\theta}$ depolarizing noise to make it a probabilistic mix of Clifford operations, which implies they need less than $\frac{3}{4}\tilde{\theta}$ worst case noise. However, as follows from [98], the worst case noise for the $\pi/8$-gate(s) is only $\frac{1}{2} - \frac{1}{2\sqrt{2}} \approx 15\%$.

We leave it as an interesting open question to determine the precise value of $\tilde{\theta}$.

**Different noise models**

The approach we have taken can in principle also be applied to other noise models: For any 1-qubit noise operation $\mathcal{E}$, with Bloch representation $S_\mathcal{E}$ we can compute the minimum value $\theta$ such that for all rotations $R \in \mathbb{R}^{3 \times 3}$ the noisy version $(1 - \theta)R + \theta S_\mathcal{E}$ is inside the Clifford polytope $P$, defined in equation (5.8). However, the actual optimization problems might not be as easy as for depolarizing noise, since depolarizing noise with probability $p$ corresponds to multiplying with $(1 - p)$ in Bloch-representation.

In principle, a similar approach might be possible to calculate how well one can approximate arbitrary (unitary) gates given a gate set $S$ other than Clifford$^*$ under a certain noise model. If $S$ is not universal, this will also give new noise bounds.

**Allowing some perfect unitaries**

Our threshold theorem says the following. Let $f : \{0, 1\}^n \to \{0, 1\}$ be a function which 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 of perfect Clifford operations and single qubit gates with noise $\tilde{\theta}$ ($\approx 45\%$) that can compute $f$. We can strengthen this result to allow a small number of perfect single-qubit gates as well: Assume that $f$ requires $m$ bits of communication to be computed, i.e., the randomized worst-case partition communication complexity $R^{\text{w}}(f)$ is at least $m$. Then there is no quantum circuit that uses perfect Clifford operations, $s$ perfect single-qubit gates, and single qubit gates with noise $\tilde{\theta}$ that computes $f$, for $2s + 1 < m$ with error at most $\epsilon$. We get this strengthening by changing the simulation of a Clifford circuit in Lemma 5.3.1 in the following way: Whenever Alice and Bob want to perform a perfect single qubit gate on some qubit, Alice sends her classical share $a, b$ of that specific qubit to Bob. Note that Bob now has complete control over this qubit and can perform the perfect gate on that qubit. They then proceed as in Lemma 5.3.1.
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 $m > 2s + 1$.

**Quantum inputs**

Lemma 5.3.1 can actually be extended to the case where Alice and Bob get quantum states as inputs and they are provided with entanglement. The statement is as follows: Suppose they have a quantum circuit $C$ as in Lemma 5.3.1, but they get a quantum state $\rho \in \mathbb{C}^{2^n \times 2^n}$. Let $p_{\rho,i}$, $i = 0, 1$, be the probability that $C$ (which uses one 1-qubit measurement in the computational basis to determine the output) outputs $i$ on input $\rho$. Now, let $\rho$ be arbitrarily partitioned between Alice and Bob, that is, Alice gets the qubits with indices $S \subseteq \{1, \ldots, n\}$ of $\rho$ and Bob the rest and they both know $S$, but they do not know what $\rho$ is. Then it is possible with one classical bit of communication from Alice to Bob that Bob outputs $i$ with probability exactly $p_{\rho,i}$.

To see how this works let without loss of generality Alice’s input qubits be $S = \{1, \ldots, m\}$, $m \leq n$. Alice and Bob then need to share $m$ EPR pairs. We use an “aborted” teleportation scheme to set up a representation as in Lemma 5.3.1, equation (5.3): Call $\rho_i$ the $i$-th qubit of $\rho$. Recall that during the standard protocol (see e.g. [68], page 26) of teleporting $\rho_i$ to Bob, Alice measures at some point two classical bits $a_i, b_i$. In the standard protocol for teleportation she then sends these two bits to Bob, who applies $X^{a_i}Z^{b_i}$ on his share of the $i$-th EPR-pair and this then contains $\rho_i$.

Now, in our aborted version of teleportation Alice does not send the bits $a_i, b_i$ ($1 \leq i \leq m$). She keeps them and for $m < i \leq n$ she additionally initializes bits $a_i, b_i$ to $a_i = b_i = 0$, so that Alice ends up with $2n$ classical bits in total. With this protocol, Alice and Bob obtain the correct representation of the state $\rho$ from Lemma 5.3.1. More precisely, if $\rho' \in \mathbb{C}^{2^n \times 2^n}$ is the state in the $n - m$ qubits given initially to Bob and the $m$ qubits from his shares of the EPR-pairs (after the “aborted” teleportation), then

$$
\rho = \left( \bigotimes_{i=1}^{m} X^{a_i}_i Z^{b_i}_i \right)^\dagger \rho' \left( \bigotimes_{i=1}^{m} X^{a_i}_i Z^{b_i}_i \right),
$$

where the subscript $i$ means that the operator acts on the Hilbert space of Bob’s share of the $i$-th EPR-pair. Note that $\rho'$ is completely in Bob’s hands. This is the same representation as in equation (5.3), just that now the quantum share $\rho'$ is some mixed state, which is not known to Alice and Bob. This is necessary since we also assumed that the logical input state $\rho$, which is encoded in this way, can also be some arbitrary mixed state.

From here, they can then run the same protocol as in the proof Lemma 5.3.1, where in the end Alice sends one classical bit to Bob. Of course, this time Bob has
to do the final measurement and can not just classically simulate the quantum computation since we assumed the state $\rho$ to be arbitrary and not known to Alice and Bob. It is clear that the outcome of his final measurement will have the correct distribution $p_{\rho,i}$. 