Noise in quantum and classical computation & non-locality

Unger, F.P.

Citation for published version (APA):
Chapter 4

Perfect 1-qubit operations and noisy k-qubit unitaries

This chapter is based on the paper

Julia Kempe, Oded Regev, Falk Unger and Ronald de Wolf, Upper bounds on the noise threshold for fault-tolerant quantum computing, ICALP 2008

4.1 Introduction

In the previous chapter we have seen an upper bound on the tolerable noise for quantum computation. It applied to erasure noise only, which means that the experimenter is notified whenever an error happens. Of course this is a strong assumption. A more common (and often more realistic) assumption is that errors happen and the experimenter does not know when. Intuitively it is clear that these kinds of errors are harder to correct, since the experimenter does not know where and when errors happen. And it is very plausible that the thresholds for these kinds of noise should be smaller, see also Section 3.4 for more on this. In this chapter we give stronger bounds for the noise model of depolarizing noise on wires and a slightly weaker set of available gates.

We consider circuits consisting of unitary $k$-qubit gates each of whose input wires is subject to depolarizing noise of strength at least $\varepsilon_k$, as well as arbitrary one-qubit gates that are essentially noise-free. We assume that the output of the circuit is the result of measuring some designated qubit of the final state. The main result is that if the noise $\varepsilon_k$ is strictly larger than $1 - \sqrt{2^{1/k} - 1} = 1 - \Theta(1/\sqrt{k})$ the output of any such circuit of large enough (but constant) depth is essentially independent of its input, thereby making the circuit useless. For the
important special case of \( k = 2 \), our bound is \( \epsilon_2 > 1 - \sqrt{\sqrt{2} - 1} \approx 35.7\% \). It is interesting to note that our bound on the threshold behaves like \( 1 - \Theta(1/\sqrt{k}) \). This matches what is known for classical circuits [38, 37], and therefore probably represents the correct asymptotic behavior. In comparison, the bound for erasure noise from Chapter 3 behaves like \( 1 - 1/k \).

It is known that fault-tolerant quantum computation is impossible (for any positive noise level) without a source of “fresh” qubits. Our model takes care of this by allowing arbitrary one-qubit gates—in particular, this includes gates that take any input, and output a fixed one-qubit state, for instance \( |0\rangle \).

By allowing essentially noise-free one-qubit gates, our model addresses the fact that gates on more than one qubit are generally much harder to implement than one-qubit gates. It should also be noted that the exact value of the constant \( \epsilon_1 \) is inessential and can be chosen to be an arbitrarily small positive constant, see also comments after Theorem 4.2.1.

Note that since our theorem applies to arbitrary starting states, it in particular applies to the case that the initial state is encoded in some good quantum error-correcting code, or that it is some sort of “magic state” [21, 81]. Further, we could even allow operations which add/replace arbitrary states on multiple qubits at any time during the computation. To extend our proof to accommodate for this is straightforward. In all these cases, our theorem shows that the computation becomes essentially independent of the input after sufficiently many levels.

**Weaknesses of the model** Our assumption that all \( k \)-qubit gates are mixtures of unitaries does slightly restrict generality. Not every completely-positive trace-preserving map can be written as a mixture of unitaries. However, we believe that it is still a reasonable assumption. For instance, to the best of our knowledge, all known fault-tolerant constructions can be implemented using such gates (in addition to arbitrary one-qubit gates). Moreover, all known quantum algorithms obtain their speed-up over classical algorithms by using only unitary gates.

A slightly more severe restriction is the assumption that the output consists of just one qubit. Recall that in Chapter 3 we showed that if the noise is above the threshold of \( 1 - 1/k \) than after logarithmically many steps no measurement on all qubits can distinguish any two input states. However, we believe that in many instances the assumption that there is only one output qubit is still reasonable. For instance, this is the case whenever the circuit is required to solve a decision...
4.2 Model and results

Before we state the results, we describe the exact model, recalling definitions from Section 2.2. We consider parallel circuits, composed of \( n \) wires and \( T \) levels of gates\(^3\) (see Figure 4.2). We assume that the number of qubits \( n \) does not change during the computation. Notice that at each level, all qubits must go through some gate (possibly the identity). For each gate, the number of input qubits is the same as the number of output qubits.

We assume the circuit is composed of \( k \)-qubit gates that are probabilistic mixtures of unitary operations, as well as arbitrary (i.e., all completely-positive trace-preserving) one-qubit gates. In particular, it is possible to do intermediate

\[^3\text{So, we call the parts of the skeleton graph from Definition 3.2.1 “levels.”}\]
one-qubit measurements. We assume the output of the circuit is the outcome of a measurement of a designated output qubit in the computational basis. Finally, we assume that the circuit is subject to noise as follows. Recall that $p$-depolarizing noise on a certain qubit replaces that qubit by the completely mixed state with probability $p$, and does not alter the qubit otherwise. Formally, this is described by the superoperator $E$ acting on a qubit $\rho$ as $E(\rho) = (1 - p)\rho + pI/2$. We assume that each one-qubit gate is followed by at least $\varepsilon_1$-depolarizing noise on its output qubit, where $\varepsilon_1 > 0$ is an arbitrarily small constant. Thus one-qubit gates can be essentially noise-free. We also assume that each $k$-qubit gate is preceded by at least $\varepsilon_k$-depolarizing noise on each of its input qubits, where $\varepsilon_k > 1 - \sqrt{2^{1/k} - 1}$.

**Main results** We prove the following main result

**4.2.1. Theorem.** Fix any $T$-level quantum circuit as above. Then for any two states $\rho$ and $\tau$, the probabilities of obtaining measurement outcome 1 at the output qubit starting from $\rho$ and starting from $\tau$, respectively, differ by at most $2^{-\Omega(T)}$.

In other words, for any $\eta > 0$, the probability of measuring 1 at the output qubit of a circuit running for $T = O(\log(1/\eta))$ levels is independent of the input (up to $\pm \eta$). This makes the output essentially independent of the starting state, and renders long computations “essentially useless”.

As pointed out in the introduction, $\varepsilon_1$ can be chosen to be an arbitrarily small but positive constant. The value of $\varepsilon_1$ only affects the constant in the $\Omega(\cdot)$ of Theorem 4.2.1. The reason we require $\varepsilon_1 > 0$ is a technicality which simplifies the statement of our result. However, for $\varepsilon_1 = 0$ the statement of Theorem 4.2.1 is just wrong: One could choose input states $\rho := |0\rangle\langle 0| \otimes \rho'$ and $\tau := |1\rangle\langle 1| \otimes \tau'$, do nothing for $T$ levels (i.e., apply noise-free one-qubit identity gates on all wires) and then measure the first qubit in the computational basis. Clearly, from this measurement outcome one can exactly tell which of the two states $\rho, \tau$ was input. Nevertheless, it is possible to let $\varepsilon_1 = 0$, if we slightly change the model and additionally require that every path from the input to the output qubit goes through enough $k$-qubit gates. Our proof can easily be adapted to this case.

Of special interest from an experimental point of view is the case $k = 2$, for which our bound becomes about $35.7\%$. Furthermore, for the case in which the only allowed two-qubit gate is the controlled-NOT (CNOT) gate, we can improve our bound further to about $29.3\%$, as we show in Section 4.5. This case is interesting both theoretically and experimentally. Note also that the CNOT gate together with all one-qubit gates forms a universal set [10]. The same noise-bound applies if we additionally allow controlled-Y and controlled-Z gates.

### 4.3 Preliminaries

We first recall some definitions.
The Pauli matrices are
\[
\begin{align*}
\mathbb{I} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
\end{align*}
\]
and we define \( \mathcal{P} = \{ \mathbb{I}, X, Y, Z \} \) and \( \mathcal{P}_* = \{ X, Y, Z \} \). We use \( \mathcal{P}^n \) to denote the set of all tensor products of \( n \) one-qubit Pauli matrices. For a Pauli matrix \( S \in \mathcal{P}^n \) we define its support, denoted \( \text{supp}(S) \), to be the qubits on which \( S \) is not identity.

We sometimes use superscripts to indicate the qubits on which certain operators act. Thus \( \mathbb{I}^A \) denotes the identity operator applied to the qubits in set \( A \).

The set of all \( 2^n \times 2^n \) Hermitian matrices forms a \( 4^n \)-dimensional real vector space. On this space we consider the Hilbert-Schmidt inner product, given by \( \langle A, B \rangle := \text{Tr}(A^\dagger B) = \text{Tr}(AB) \). Note that for any \( S, S' \in \mathcal{P}^n \), \( \text{Tr}(SS') = 2^n \) if \( S = S' \) and \( \text{Tr}(SS') = 0 \) otherwise. Hence, \( \mathcal{P}^n \) is an orthogonal basis of this space. It follows that we can uniquely express any Hermitian matrix \( \delta \) in this basis as
\[
\delta = \frac{1}{2^n} \sum_{S \in \mathcal{P}^n} \hat{\delta}(S)S
\]
where \( \hat{\delta}(S) := \text{Tr}(\delta S) \) are the (real) coefficients.

We now state some easy observations which will be used in the proof of our main result. First, by the orthogonality of \( \mathcal{P}^n \), it follows that for any \( \delta \),
\[
\text{Tr}(\delta^2) = \frac{1}{2^n} \sum_{S \in \mathcal{P}^n} \hat{\delta}(S)^2.
\]
This easily leads to the following observation.

1. **Observation (Unitary preserves sum of squares).** For any unitary \( U \) and any Hermitian matrix \( \delta \), if we denote \( \delta' = U\delta U^\dagger \), then
\[
\sum_{S \in \mathcal{P}^n} \hat{\delta'}(S)^2 = 2^n\text{Tr}(\delta'^2) = 2^n\text{Tr}(U\delta U^\dagger U\delta U^\dagger) = 2^n\text{Tr}(\delta^2) = \sum_{S \in \mathcal{P}^n} \hat{\delta}(S)^2.
\]
This also shows that the operation of conjugating by a unitary matrix, when viewed as a linear operation on the vector of Pauli coefficients, is an orthogonal transformation.

2. **Observation (Tracing out qubits).** Let \( \delta \) be some Hermitian matrix on a set of qubits \( W \). For \( V \subseteq W \), let \( \delta_V = \text{Tr}_{W \setminus V}(\delta) \). Then,
\[
\hat{\delta}(S^W \setminus V) = \text{Tr}(\delta \cdot S^W \setminus V) = \text{Tr}(\delta_V \cdot S) = \hat{\delta}_V(S).
\]

3. **Observation (Noise in the Pauli basis).** Applying \( p \)-depolarizing noise \( \mathcal{E} \) to the \( j \)-th qubit of Hermitian matrix \( \delta \) changes the coefficients as follows:
\[
\mathcal{E}(\hat{\delta})(S) = \begin{cases} 
\hat{\delta}(S) & \text{if } S_j = \mathbb{I} \\
(1 - p)\hat{\delta}(S) & \text{if } S_j \neq \mathbb{I}
\end{cases}
\]
In other words, $\mathcal{E}$ “shrinks” by a factor $1 - p$ all coefficients that have support on the $j$-th coordinate.

4. **Observation.** Let $\rho$ and $\tau$ be two one-qubit states and let $\delta = \rho - \tau$. Consider the two probability distributions obtained by performing a measurement in the computational basis on $\rho$ and $\tau$, respectively. Then the difference in the probabilities of obtaining the outcome 1 given $\rho$ respectively $\tau$ is

$$\frac{1}{2} |\hat{\delta}(Z)|.$$

**Proof:** The difference in the probabilities of obtaining the outcome 1 is given by

$$|\text{Tr}(\rho - \tau) \cdot |1\rangle \langle 1|) = \left| \text{Tr} \left( \delta \cdot \frac{\mathbb{I} - \frac{1}{2}Z}{2} \right) \right| = \frac{1}{2} |\text{Tr}(\delta \cdot Z)| = \frac{1}{2} |\hat{\delta}(Z)|,$$

where we have used $\text{Tr}(\delta) = 0$.

The last observation follows immediately from the convexity of the function $x^2$.

5. **Observation (Convexity).** Let $p_i$ be any probability distribution, and $\delta_i$ a set of Hermitian matrices. Let $\delta = \sum_i p_i \delta_i$. Then

$$\sum_{S \in \mathcal{P}^n} \hat{\delta}(S)^2 \leq \sum_i p_i \sum_{S \in \mathcal{P}^n} \hat{\delta}_i(S)^2.$$

**Proof:** Follows immediately from the convexity of the function $x^2$.

### 4.4 Proof of Theorem 4.2.1

In this section we prove Theorem 4.2.1. The rough idea is the following. Fix two arbitrary initial states $\rho$ and $\tau$. Our goal is to show that after applying the noisy circuit, the state of the output qubit is nearly the same with both starting states. Equivalently, we can define $\delta = \rho - \tau$ and show that after applying the noisy circuit to $\delta$, the “state” of the output qubit is essentially 0 (the noisy circuit is a linear operation, and hence there is no problem in applying it to $\delta$, which is the difference of two density matrices). In order to show this, we will examine how the coefficients of $\delta$ in the Pauli basis evolve through the circuit. Initially we might have many large coefficients. Our goal is to show that the coefficients of the output qubit are essentially 0. This is established by analyzing the balance between two opposing forces: noise, which shrinks coefficients by a constant factor (as in Observation 3), and gates, which can increase coefficients. As we saw in Observation 1, unitary gates preserve the sum of squares of coefficients. They can, however, “concentrate” several small coefficients into one large coefficient. One-qubit operations need not preserve the sum of squares (a good example is the
gate that resets a qubit to the $|0\rangle$ state), but we can still deal with them by using a known characterization of one-qubit gates. This characterization allows us to bound the amount by which one-qubit gates can increase the Pauli coefficients, and very roughly speaking shows that the gate that resets a qubit to $|0\rangle$ is “as bad as it gets”.

Before continuing with the proof, we introduce some terminology. From now on we use the term qubit to mean a wire at a specific time, so there are $(T + 1)n$ qubits (although during the proof we will also consider qubits that are located between a gate and its associated noise). We say that a set of qubits $V$ is consistent if we can meaningfully talk about a “state of the qubits of $V$” (see Figure 4.2). More formally, we define a consistent set as follows. The set of all qubits at time 0 and all its subsets are consistent. If $V$ is some consistent set of qubits, which contains all input qubits $IN$ of some gate (possibly a one-qubit identity gate), then also $(V \setminus IN) \cup OUT$ and all its subsets are consistent, where $OUT$ denotes the gate’s output qubits. Note that here we think of the noise as being part of the gate. For a consistent set $V$ and a state (or more generally, a Hermitian matrix) $\rho$, we denote the state of $V$ when the circuit is applied with the initial state $\rho$, by $\rho_V$. In other words, $\rho_V$ is the state one obtains by applying some initial part of the circuit to $\rho$, and then tracing out from the resulting state all qubits that are not in $V$.

If $v$ is a qubit, we use $\text{dist}(v)$ to denote its distance from the input, i.e., the level of the gate just preceding it. The qubits of the starting state have $\text{dist}(v) = 0$. For a nonempty set $V$ of qubits we define $\text{dist}(V) = \min\{\text{dist}(v) \mid v \in V\}$, and extend it to the empty set by $\text{dist}(\emptyset) = \infty$. Note that $\text{dist}(V)$ does not increase if we add qubits to $V$.

In the rest of this section we prove the following lemma, showing that a certain invariant holds for all consistent sets $V$.

**4.4.1. Lemma.** For all $\varepsilon_1 > 0$ and $\varepsilon_k > 1 - \sqrt{2^{1/k} - 1}$ there exists a $\theta < 1$ such that the following holds. Fix any $T$-level circuit in our model, let $\rho$ and $\tau$ be some arbitrary initial states, and let $\delta = \rho - \tau$. Then for every consistent $V$,

$$\sum_{S \in P^V} \widehat{\delta}_V(S)^2 \leq 2 \cdot 2^{|V|} \cdot \theta^{\text{dist}(V)}, \quad (4.1)$$

or equivalently,

$$\text{Tr}(\delta_V^2) \leq 2 \cdot \theta^{\text{dist}(V)}.$$

In particular, if we consider the consistent set $V$ containing the designated output qubit at time $T$, then we get that $\widehat{\delta}_V(Z)^2 \leq 4\theta^T$. By Observation 4, this implies Theorem 4.2.1.
4.4.1 Proof of Lemma 4.4.1

The proof of the invariant is by induction on the sets $V$. At the base of the induction are all sets $V$ which only contain qubits at time 0. All other sets are handled in the induction step. In order to justify the inductive proof, we need to provide an ordering on the consistent sets $V$ such that for each $V$, the proof for $V$ uses the inductive hypothesis only on sets $V'$ which appear before $V$ in the ordering. As will become apparent from the proof, if we denote by latest($V$) the maximum time at which $V$ contains a qubit, then each $V'$ for which we use the induction hypothesis has strictly less qubits than $V$ at time latest($V$). Therefore, we can order the sets $V$ first in increasing order of latest($V$) and then in increasing order of the number of qubits at time latest($V$).

**Base case**

Here we consider the case that $V$ is fully contained within time 0. If $V = \emptyset$ then both sides of the invariant are zero, so from now on assume $V$ is nonempty. In this case $\text{dist}(V) = 0$. The matrix $\delta_V$ is the difference of two density matrices, say $\delta_V = \rho_V - \tau_V$, and hence $\text{Tr}(\delta_V^2) = \text{Tr}(\rho_V^2) + \text{Tr}(\tau_V^2) - 2\text{Tr}(\rho_V\tau_V) \leq 2$, and the invariant is satisfied.

**Induction step**

Let $V''$ be any consistent set containing at least one qubit at time greater than zero. Our goal in this section is to prove the invariant for $V''$. Consider any of the qubits of $V''$ located at time latest($V''$) and let $G$ be the gate that has this qubit as one of its output qubits. We now consider two cases, depending on whether $G$ is a $k$-qubit gate or a one-qubit gate.

**Case 1: $G$ is a $k$-qubit gate.** Here we consider the case that $G$ is a probabilistic mixture of $k$-qubit unitaries. First note that by Observation 5 it suffices to prove the invariant for $k$-qubit unitaries. So assume $G$ is a $k$-qubit unitary acting on the qubits $\mathcal{A} = \{A_1, \ldots, A_k\}$. Let $\mathcal{A}' = \{A'_1, \ldots, A'_k\}$ be the qubits after the $\epsilon_k$-noise but before the gate $G$ and $\mathcal{A}'' = \{A''_1, \ldots, A''_k\}$ the qubits after $G$ (see Figure 4.2). By our choice of $G$, $\mathcal{A}'' \cap V'' \neq \emptyset$. Define $V' = (V'' \setminus \mathcal{A}') \cup \mathcal{A}'$ and $V = (V'' \setminus \mathcal{A}'') \cup \mathcal{A}$. Note that $V$ and its subsets are consistent sets with strictly fewer qubits than $V''$ at time latest($V''$), and hence we can apply the induction hypothesis to them.

Recall that our goal is to prove the invariant Eq. (4.1) for $V''$. To begin, using Observation 2,

$$\sum_{S \in \mathcal{P}V''} \delta_{V''}(S)^2 \leq \sum_{S \in \mathcal{P}V'' \cup \mathcal{A}''} \delta_{V'' \cup \mathcal{A}''}(S)^2.$$ (4.2)
4.4. Proof of Theorem 4.2.1

Figure 4.2: An example showing the sets \( V, V', \) and \( V'' \) for a two-qubit gate \( G \).

Because \( G \) (which maps \( \delta_V \) to \( \delta_{V \cup A'} \)) is unitary, it preserves the sum of squares of \( \hat{\delta} \)-coefficients (see Observation 1), so the right hand side of (4.2) is equal to

\[
\sum_{S \in P^{V'}} \left( \sum_{R \in P^{A'}} \delta_{V'}(RS)^2 \right).
\]

Since the only difference between \( \delta_V \) and \( \delta_{V'} \) is noise on the qubits \( A_1, \ldots, A_k \), using Observation 3 and denoting \( \mu = 1 - \epsilon_k \), we get that the above is at most

\[
\sum_{S \in P^{V \setminus A}} \mu^2 \left( \sum_{R \in P^{A \setminus a}} \delta_{V \cup a}(RS)^2 \right).
\]

where the equality follows by noting that for any fixed \( S \) and any \( R \in P^A \), the term \( \delta_{V}(RS)^2 \), which appears with coefficient \( \mu^2 \) on the left hand side, appears with the same coefficient \( \sum_{a \in \text{supp}(R)} \mu^2 (1 - \mu^2)^{|a|} \) on the right hand side. By rearranging and using Observation 2 we get that the above is equal to

\[
\sum_{a \subseteq A} \mu^2 (1 - \mu^2)^{|a|} \left( \sum_{S \in P^{V \setminus A} \cup a} \delta_{(V \setminus A) \cup a}(S)^2 \right)
\]

\[
\leq \sum_{a \subseteq A} \mu^2 (1 - \mu^2)^{|a|} \cdot 2 |(V \setminus A) \cup a| \cdot \theta_{\text{dist}}((V \setminus A) \cup a)
\]

where we used the inductive hypothesis. Note that \( \text{dist}((V \setminus A) \cup a) \geq \text{dist}(V) \), so the above is

\[
\leq 2 \cdot 2^{V \setminus A} \cdot \theta_{\text{dist}}(V) \sum_{a \subseteq A} \mu^2 (1 - \mu^2)^{|a|}
\]

\[
= 2 \cdot 2^{V \setminus A} \cdot \theta_{\text{dist}}(V) \left( (1 - \mu^2) + 2 \mu^2 \right)^k
\]

\[
= 2 \cdot 2^{V \setminus A} \cdot \theta_{\text{dist}}(V) \left( 1 + \mu^2 \right)^k. \quad (4.3)
\]
Note that $|V \setminus A| \leq |V''| - 1$ and $\text{dist}(V'') - 1 \leq \text{dist}(V)$, so the right hand side is bounded by

$$\leq 2 \cdot 2^{|V''| - 1} \cdot \theta^{\text{dist}(V'')} - 1 (1 + \mu^2)^k.$$ 

Since $\epsilon_k > 1 - \sqrt{2^k/k} - 1$, we have that $(1 + \mu^2)^k \leq 2\theta$ if $\theta$ is close enough to 1, so we can finally bound the last expression by

$$\leq 2 \cdot 2^{|V''|} \cdot \theta^{\text{dist}(V'')}$$

which proves the invariant for $V''$.

**Case 2:** $G$ is a one-qubit gate. Before proving the invariant, we need to prove the following property of completely-positive trace-preserving (CPTP) maps on one qubit.

**4.4.2. Lemma.** For any CPTP map $G$ on one qubit there exists a $\beta \in [0, 1]$ such that the following holds. For any Hermitian matrix $\delta$, if we let $\delta'$ denote the result of applying $G$ to $\delta$, then we have

$$\hat{\delta}(X)^2 + \hat{\delta}(Y)^2 + \hat{\delta}(Z)^2 \leq (1 - \beta) \cdot \hat{\delta}(I)^2 + \beta \cdot (\hat{\delta}(X)^2 + \hat{\delta}(Y)^2 + \hat{\delta}(Z)^2).$$

**Proof:** The proof is based on the characterization of trace-preserving completely-positive maps on one qubit given in Section 2.5.2 on page 28, which we recall now. Any one-qubit gate $G$ can be written as a convex combination of gates of the form $U_1 \circ J \circ U_2$. Here $U_1$ and $U_2$ are one-qubit unitaries (acting on the density matrix by conjugation), and $J$ is a one-qubit map that in the Pauli basis has the form

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\
0 & \lambda_1 & 0 & 0 \\
0 & 0 & \lambda_2 & 0 \\
t & 0 & 0 & \lambda_1 \lambda_2 \end{pmatrix}$$

for some $\lambda_1, \lambda_2 \in [-1, 1]$ and $t = \pm \sqrt{(1 - \lambda_1^2)(1 - \lambda_2^2)}$.

First observe that by the convexity of the square function, it suffices to prove the lemma for $G$ of the form $U_1 \circ J \circ U_2$ (with the resulting $\beta$ being the appropriate average of the individual $\beta'$s). Next note that since $U_1$ and $U_2$ are unitary, they act on the vector of coefficients $(\hat{\delta}(X), \hat{\delta}(Y), \hat{\delta}(Z))$ as an orthogonal transformation, and hence leave the sum of squares invariant. This shows that it suffices to prove the lemma for a map $J$ as above. For this map,

$$\hat{\delta}(X)^2 + \hat{\delta}(Y)^2 + \hat{\delta}(Z)^2 = \lambda_1^2 \hat{\delta}(X)^2 + \lambda_2^2 \hat{\delta}(Y)^2 + (t \hat{\delta}(I) + \lambda_1 \lambda_2 \hat{\delta}(Z))^2.$$ 

Assume without loss of generality that $\lambda_1^2 \geq \lambda_2^2$. Applying Cauchy-Schwarz to the two 2-dimensional vectors $(\pm \sqrt{1 - \lambda_1^2} a, \lambda_1 b)$ and $(\sqrt{1 - \lambda_2^2}, \lambda_2)$, we get that
for any \( a, b \in \mathbb{R} \), \((ta + \lambda_1 \lambda_2 b)^2 \leq (1 - \lambda_1^2)a^2 + \lambda_1^2 b^2\). Hence the above expression is upper bounded by

\[
\lambda_1^2 \hat{\delta}(X)^2 + \lambda_2^2 \hat{\delta}(Y)^2 + (1 - \lambda_1^2)\hat{\delta}(I)^2 + \lambda_1^2 \hat{\delta}(Z)^2
\]

and we complete the proof by choosing \( \beta = \lambda_1^2 \).

Let \( A \) be the qubit \( G \) is acting on, and recall that our goal is to prove the invariant for the set \( V'' \). Denote by \( A' \) the qubit of \( G \) after the gate but before the \( \epsilon_1 \) noise, and by \( A'' \) the qubit after the noise. As before, by our choice of \( G \), we have \( A'' \in V'' \). Let \( A = \{A\}, A' = \{A'\}, A'' = \{A''\} \). Define \( V' = (V'' \setminus A'') \cup A' \) and \( V = (V'' \setminus A'') \cup A \) and notice that \( |V| = |V'| = |V''| \). By using Lemma 4.4.2, we obtain a \( \beta \in [0, 1] \) such that

\[
\sum_{S \in P^{V''}} \hat{\delta}_{V''}(S)^2 \\
\leq \sum_{S \in P^{V'' \setminus A'}} \left( \hat{\delta}_{V'}(I S)^2 + (1 - \epsilon_1)^2 \sum_{R \in P^{A'}} \hat{\delta}_{V'}(R S)^2 \right) \\
\leq \sum_{S \in P^{V'' \setminus A}} \left( (1 + (1 - \epsilon_1)^2(1 - 2\beta))\hat{\delta}_{V}(I S)^2 + (1 - \epsilon_1)^2 \beta \sum_{R \in P^{A}} \hat{\delta}_{V}(R S)^2 \right).
\]

By applying the induction hypothesis to both \( V \setminus A \) and \( V \), we can upper bound the above by

\[
(1 + (1 - \epsilon_1)^2(1 - 2\beta)) \cdot 2 \cdot 2^{|V''| - 1} \cdot \theta^{\text{dist}(V \setminus A)} + (1 - \epsilon_1)^2 \beta \cdot 2 \cdot 2^{|V|} \cdot \theta^{\text{dist}(V)} \\
\leq \frac{1 + (1 - \epsilon_1)^2}{2\theta} \cdot 2 \cdot 2^{|V''|} \cdot \theta^{\text{dist}(V''},
\]

where we used that \( |V| = |V''| \), and \( \text{dist}(V'') - 1 \leq \text{dist}(V) \leq \text{dist}(V \setminus A) \). Hence the invariant remains valid if we choose \( \theta < 1 \) such that \( 1 + (1 - \epsilon_1)^2 \leq 2\theta \).

### 4.5 Arbitrary one-qubit gates and CNOT gates

In this section we consider the case where CNOT is the only allowed gate acting on more than one qubit. We still allow arbitrary one-qubit gates. The proof follows along the lines of that of Theorem 4.2.1 with one small modification. As before, we will prove that for all \( \epsilon_1 > 0 \) and \( \epsilon_2 > 1 - 1/\sqrt{2} \approx 0.293 \) the invariant, Eq. (4.1), holds. The proof for the case that \( G \) is a one-qubit gate holds without change. We will give the modified proof for the case that \( G \) is a CNOT gate. The idea for the improved bound is to make use of the fact that the CNOT gate merely permutes the 16 elements of \( P \otimes P \), and does not map elements from \( I \otimes P_\ast \) to \( P_\ast \otimes I \) or vice versa (as illustrated in Figure 4.3). As a result we need
Chapter 4. Perfect 1-qubit operations and noisy k-qubit unitaries

The action of CNOT on $\mathcal{P} \otimes \mathcal{P}$ under conjugation, with the control wire corresponding to the first qubit.

Figure 4.3: Action of CNOT on Pauli group

to apply the induction hypothesis on one less term, which in turn improves the bound.

Assume the CNOT acts on qubits $\mathcal{A} = \{A, B\}$, with $\mathcal{A}' = \{A', B'\}$ and $\mathcal{A}'' = \{A'', B''\}$ as before, where again $\mathcal{A}'' \cap V'' \neq \emptyset$. If both $A''$ and $B''$ are contained in $V''$ then the proof of the general case (cf. Eq. (4.3)) already gives a bound of

$$2 \cdot 2^{V' \setminus \mathcal{A}} \cdot \theta \text{dist}(V) (1 + \mu^2)^2 \leq 2 \cdot 2^{V''} \cdot \theta \text{dist}(V'')^{-1} (1 + \mu^2)^2 \leq 2 \cdot 2^{|V''|} \cdot \theta \text{dist}(V'')$$

where the last inequality holds for all $\mu < 1$. Hence it suffices to consider the case that exactly one of $A''$ and $B''$ is in $V''$. Assume without loss of generality that $A'' \in V''$ and $B'' \notin V''$. As before, our goal is to upper bound

$$\sum_{S \in \mathcal{P}^V} \hat{\delta}_{V''}(S)^2 = \sum_{S \in \mathcal{P}^V} \hat{\delta}_{V'' \cup B''}(S \mathbb{I}^B)^2.$$ 

where the equality follows from Observation (2). Because of the property of CNOT mentioned above, we can now upper bound this by

$$\sum_{S \in \mathcal{P}^V \setminus \mathcal{A}'} \left( \hat{\delta}_{V'}(\mathbb{I}^A \mathbb{I}^B S)^2 + \sum_{R \in \mathcal{P}^A} \hat{\delta}_{V'}(R \mathbb{I}^B S)^2 + \sum_{R \in \mathcal{P}^A \otimes \mathcal{P}^B} \hat{\delta}_{V'}(RS)^2 \right).$$

This is the crucial change compared to the case of general two-qubit gates (the latter case also includes a term of the form $\sum_{R \in \mathcal{P}^B} \hat{\delta}_{V'}(\mathbb{I}^A RS)^2$). The rest of the proof is similar to the earlier proof. Using the induction hypothesis we can upper
bound the above by

\[
\sum_{S \in P \setminus A} (\hat{\delta}_V(I^A I^B S)^2 + \mu^2 \sum_{R \in P^A} \hat{\delta}_V(R I^B S)^2 + \mu^4 \sum_{R \in P^A \otimes P^B} \hat{\delta}_V(R S)^2) \\
\leq (1 - \mu^2) \sum_{S \in P \setminus A} \hat{\delta}_{V \setminus A}(S)^2 + (\mu^2 - \mu^4) \sum_{S \in P \setminus \{B\}} \hat{\delta}_{V \setminus \{B\}}(S)^2 + \mu^4 \sum_{S \in P} \hat{\delta}_V(S)^2 \\
\leq (1 - \mu^2) 2 \cdot 2^{\|V \setminus A\|} \theta_{\text{dist}}(V \setminus A) + (\mu^2 - \mu^4) 2 \cdot 2^{\|V \setminus \{B\}\|} \theta_{\text{dist}}(V \setminus \{B\}) + \mu^4 2 \cdot 2^{\|V\|} \theta_{\text{dist}}(V) \\
\leq 2 \cdot 2^{\|V''\|} \theta_{\text{dist}}(V'') \left(\frac{1 + \mu^2}{2} + \mu^4\right) \frac{1}{\theta'} \\
\leq 2 \cdot 2^{\|V''\|} \theta_{\text{dist}}(V'') \left(\frac{1 + \mu^2}{2} + \mu^4\right) \frac{1}{\theta'}
\]

Hence the invariant remains valid as long as \(1 + \mu^2 + \mu^4 \leq \theta < 1\). This can be satisfied as long as \(\mu < 1/\sqrt{2}\), equivalently \(\varepsilon_2 > 1 - 1/\sqrt{2} \approx 0.293\).

### 4.6 Discussion

#### 4.6.1 Comparison with other chapters

In Chapter 3 we have shown an upper bound of \(\varepsilon_k = 1 - 1/k\) on erasure noise. On one hand, this result is stronger than the result from this chapter as it allows arbitrary \(k\)-qubit gates and not just mixtures of unitaries and it holds for erasure noise instead of depolarizing noise. Further, we saw that the result of an arbitrary \(n\)-qubit measurement on the full final state becomes essentially independent of the initial state after \(T = O(\log n)\) levels. On the other hand, the bound in this chapter is better for all values of \(k\). Hence the two results are incomparable.

We will see another bound in Chapter 5, which shows that classical circuits can efficiently simulate any quantum circuit that consists of perfect, noise-free stabilizer operations (meaning Clifford gates (Hadamard, phase gate, CNOT), preparations of states in the computational basis, and measurements in the computational basis), perfect classical control (i.e., the ability to condition future gates on earlier classical measurement outcomes, see page 35) and arbitrary one-qubit unitary gates that are followed by 45.3% depolarizing noise. Hence such circuits are not significantly more powerful than classical circuits. We will also see that this result is tight: If the one-qubits gates have less than 45.3% noise, it is possible to efficiently simulate any (noise-free) quantum circuit. Although this result establishes a tight threshold, it is incomparable to our current result since the result in Chapter 5 applies to a restricted gate set only.
4.6.2 Comments on results and open problems

We believe that a main part of our contribution is introducing a technique for obtaining upper bounds on the fault-tolerance threshold. Namely, we use a Pauli basis decomposition in order to track the state of the computation. We believe this framework will be useful also for further analysis of quantum fault-tolerance. A finer analysis of the Pauli coefficients might improve the bounds we achieve here, and possibly obtain bounds that are tailored to other computational models.

We only analyze depolarizing noise acting independently on each qubit. Depolarizing noise is the “most symmetric” type of one-qubit noise and therefore a natural choice for our analysis. Also, it is a relatively weak type of noise: it is not adversarial and does not have correlations between the errors occurring on different qubits. However, since we are proving an upper bound on the fault-tolerance threshold, this weakness is actually a good thing, making our result stronger. In principle one can extend our results to various other one-qubit noise models, using an analysis similar to the one developed in Lemma 4.4.2. However, not all noise models can actually yield a result like Theorem 4.2.1. For instance, if we have Toffoli gates with only phaseflip errors, then we can do perfect classical computation. Statements like Theorem 4.2.1 are just false in that case.

Open problems In the introduction we mentioned some weaknesses of our model. Of course we would like to prove a result which does not have these restrictions. Further, it would be interesting to extend the result in a couple of other directions. We now summarize some desirable extensions:

- We should make it work for all possible $k$-qubit gates (CPTP maps), rather than just mixtures of unitaries.

- We should allow some classical side-processing, where classical outcomes of intermediate measurements can be used by a classical computer and its results can later be fed back into the circuit. Allowing such “classical control” requires a type of theorem different from the one we have now: if initial states $\rho$ and $\tau$ were bits 0 and 1, respectively, we could just measure this right at the start, store the bit in the classical part without noise, and feed it back into the circuit only at the last step, yielding distinguishable final states. Furthermore, if we allow classical control (and classical side-processing) then it is clearly possible to compute any function just in the classical part of the circuit. Hence, a statement like ours is just not true.

To get a noise bound for the model with classical control, one would need to show that if the noise in the quantum hardware is above a certain threshold, then not all problems in BQP could be solved efficiently, where BQP is the class of problems that can be efficiently solved with a noise-free quantum computer (see also Section 2.3).
4.6. Discussion

- We should relax the assumption that the final output is determined by a measurement on one or a few qubits of the final state. Often in fault-tolerant schemes one encodes each “logical qubit” in a large block of physical qubits, and measures all qubits in that block to obtain the final outcome of the computation. If only some of the qubits in the final measurement are faulty the final result can still be recovered by applying classical post-processing. Our results cannot rule out this approach.

- Last but not least, our upper bounds on the fault-tolerance threshold are still higher than one would expect, and we would like to decrease them further.