Noise in quantum and classical computation & non-locality

Unger, F.P.

Citation for published version (APA):
Chapter 8

Limits on non-locality from communication complexity

This chapter is based on the paper

Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Methot, Alain Tapp, and Falk Unger, Limit on nonlocality in any world in which communication complexity is not trivial, Physical Review Letters 96(25), 2006.

8.1 Introduction

Quantum mechanics is a physical theory which is hugely successful in describing very small physical systems, on the scale of atoms. Its foundations were laid out in the 1920s and 1930s. At first quantum mechanics was not accepted immediately by all physicists, because of its counter-intuitive properties and predictions. Most notably, Albert Einstein rejected quantum mechanics with the famous words “Gott würfelt nicht”, which translates to “God does not play dice”. Nowadays, quantum mechanics is widely accepted, mainly due to the fact that quantum mechanics predicts outcomes of many experiments where classical physics fails. For example, later in this chapter we will describe (an abstract setting of) an experiment, so called Bell inequality violations, which can be explained by quantum mechanics but not by classical mechanics. Because of experiments of this kind, quantum mechanics is nowadays widely believed to accurately describe the world in the “small”.

Nevertheless, despite its successes in experiments, quantum mechanics is often considered mysterious, due to its counter-intuitive predictions and intriguing properties. One of these properties is entanglement. In our example, based on a
violation of a Bell inequality, we will see that two separated, non-communicating parties possessing entanglement can create correlations which are not attainable without entanglement. However, it follows from the axioms of quantum mechanics that also entanglement does not allow for all (causal\footnote{explained later}) correlations which are in principle conceivable in the physical world. The question we want to address in this chapter is whether this limitation of achievable correlations is more than merely a consequence of quantum mechanics. We will explain why limitations of possible correlations can be seen as a “natural axiom”, which every reasonable physical theory (including quantum mechanics) should obey. In our argument we show that under a reasonable assumption about the physical world—namely that communication complexity (explained in Section 2.4 or later in this chapter) is non-trivial—restrictions on achievable correlations are indeed necessary.

This chapter is about the axioms of physics. Since they cannot be stated as rigorously as mathematical theories, the first part (in particular the introduction) of this chapter will be less rigorous than the other chapters. After we have set the stage and defined a suitable mathematical model, the proofs presented will be rigorous.

**CHSH inequality**

We will start by explaining the CHSH inequality, a particular type of *Bell inequality*. Assume two parties, Alice and Bob, share a quantum state $|\psi\rangle$ and they can both perform two different measurements on their respective parts of $|\psi\rangle$, with binary outcomes 0 or 1. Let Alice’s choice of the measurement be denoted by $x \in \{0, 1\}$ and Bob’s by $y \in \{0, 1\}$. Let Alice’s and Bob’s outcomes be $a, b \in \{0, 1\}$, respectively. Tsirelson [93] proved a bound on the correlation between Alice’s and Bob’s outcomes, see also Figure 1.3:

$$\frac{1}{4} \sum_{x,y} \Pr[x \cdot y = a \oplus b] \leq \varphi = \frac{1}{2} + \frac{\sqrt{2}}{4} \approx 85\%.$$ \hspace{1cm} (8.1)

Here $x \cdot y$ is the logical AND of $x$ and $y$ and $a \oplus b$ is the logical XOR of $a$ and $b$. This means that $x \cdot y = a \oplus b$ is satisfied if and only if (1) $x = y = 1$ and $a \neq b$ or if (2) $x$ and $y$ are not both 1 and $a = b$.

In quantum mechanics, this bound is actually tight and can be attained, see [96, 95] for a derivation. Furthermore, it is even possible to achieve that

$$\forall_{x,y} \Pr[x \cdot y = a \oplus b] = \varphi = \frac{1}{2} + \frac{\sqrt{2}}{4},$$ \hspace{1cm} (8.2)

which means that it is possible to guarantee *optimal worst-case behaviour*.\footnotetext[1]{explained later}
However, classical physics, using the so-called local hidden-variable model (LHV), only allows a correlation of up to $3/4$. Bell [12] and Clauser, Horne, Shimony and Holt [25] proved for classical LHV-theories

$$\frac{1}{4} \sum_{x,y} \Pr[x \cdot y = a \oplus b] \leq \frac{3}{4},$$

(8.3)

which is known as the CHSH inequality. It is a special kind of Bell inequality. It can be used to test local hidden-variable theories because it follows also from CHSH that in a local realistic theory (i.e. under a local hidden-variable model) Alice and Bob cannot succeed with probability greater than $3/4$ if they are space-like separated. If correlations greater than $3/4$ can be detected in an experiment, then we conclude that the world cannot be local realistic and in particular not classical.

The CHSH inequality shows up in many different contexts in quantum mechanics and appears under different names. In this chapter we will discuss it in the framework of non-local boxes (whose exact definition is not important at the moment and will be given in the next section) as we focus on the non-local properties of quantum mechanics. The CHSH inequality can also be cast in terms of XOR games, which are the focus of Chapter 7.

### Violations of CHSH inequality

Within the framework of non-local boxes it is possible to define correlations for which the left-hand side of equation (8.1) becomes 1, but which do not violate causality (see Section 8.2). This was first observed in 1995 by Popescu and Rohrlich and elaborated on in a series of papers [74, 75, 76]. They asked: Why does Quantum Mechanics not allow a higher correlation in equation (8.5) than $\wp = 1/2 + \sqrt{2}/4$? In fact, they constructed a toy theory with correlations of up to 1 in equation (8.1), which did not exhibit any apparent inconsistency.

This question was later answered by Cleve [26] and van Dam [96, 95] who showed that a maximal correlation of 1 in (8.1) would imply some improbable consequences in the real world, namely that all functions have trivial communication complexity of just one bit. Recall from Section 2.4 that for certain functions $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ Alice and Bob have to communicate essentially $n$ bits to compute the value of $f$. An example was the inner product function, but in fact this holds for almost all functions. Indeed, trivial communication complexity seems too good to be true.

When we extend the notion of “trivial” communication complexity to bounded error protocols (either with public random coins or shared entanglement) the inner product function remains nontrivial according to quantum mechanics: In Appendix C we use an argument from [103] to show that Alice and Bob cannot succeed at computing the inner product function on $n$ bits with probability $1 - \epsilon$
if they transmit fewer than \( n - 2 \log_2 \frac{1}{1 - \epsilon} \) bits, even if they share prior entanglement. Hence, if we assume that Nature does not offer “free lunch” by allowing to compute functions on distributed inputs with trivial communication, then the above results by Cleve and van Dam imply that Nature also cannot allow all possible CHSH-correlations.

However, their argument only excludes CHSH-correlations which achieve the maximum value of 1. In this chapter we will improve this and answer a stronger question: Considering that values larger than \( \varphi \) on the right hand side of (8.1) would not violate causality, why do the laws of quantum mechanics only allow correlations of up to \( \varphi \), but not something better? In fact, could it even be that this “magic” value of \( \varphi \) is not only a consequence of quantum mechanics but rather a defining property and should be seen as an axiom [75]? In this chapter we attempt to give a partial answer to this question by generalizing van Dam’s and Cleve’s result: If correlations of more than \( \frac{3 + \sqrt{6}}{6} \approx 90.8\% \) were possible, still all functions with shared inputs could be computed with bounded error with trivial communication complexity (i.e., with just just one bit of communication).

Ideally, we would want to show that for all correlations higher than \( \varphi \approx 85\% \), communication complexity becomes trivial. This indeed would imply that the value of \( \varphi \) may be taken as an axiom of any (reasonable) physical theory. It is an interesting open problem to determine whether our result can be extended up to the value \( \varphi \).

For the precise statement of our result and our proof we use the more modern framework of non-local boxes, introduced by Popescu and Rohrlich [74, Eq. (7)]. This will be explained in the next section. We then prove our main result, Theorem 8.3.1.

### 8.2 Non-local boxes

A non-local box (NLB) is an imaginary device shared between Alice and Bob, who can be arbitrarily far apart. It has an input-output port at Alice’s and another one at Bob’s. Whenever Alice feeds a bit \( x \) into her input port, she instantaneously gets a uniformly distributed random output bit \( a \), locally uncorrelated with anything else, including her own input bit. The same applies to Bob, whose input and output bits we call \( y \) and \( b \), respectively. The “magic” appears in the form of a correlation between the pair of outputs and the pair of inputs: the exclusive-or of the outputs is always equal to the logical AND of the inputs: \( a \oplus b = x \cdot y \). Much like the correlations that can be established by use of quantum entanglement, this device is atemporal: Alice gets her output as soon as she feeds in her input, regardless of whether (and when) Bob feeds in his input, and vice versa. These devices are also known under the name “Popescu-Rohrlich box” or PR-box. The name non-local box derives from the fact that an operation (i.e., choosing an input bit and receiving an output bit) on Alice’s side has an
8.2. Non-local boxes

Figure 8.1: Non-local box

*instantaneous* effect on Bob’s side as well. In particular, depending on the values $x$ and $a$ which Alice observes, there is a deterministic function $f$ computing Bob’s output $b := f(y)$. The function $f$ is either the identity function, logical negation, constant zero or constant one.

Also inspired by entanglement, this is a *one-shot* device: the correlation appears only as a result of the first pair of inputs fed in by Alice and Bob.

NLBs cannot be used by Alice and Bob to signal instantaneously to one another, i.e., they are *non-signalling*. This is because the outputs that can be observed are purely random from a local perspective.\(^2\) Hence, NLBs are *causal*: they cannot make an effect precede its cause in the context of special relativity.

We are interested in the question of how well the correlation of NLBs can be approximated. A *approximation of an NLB* with success probability $p$ is an NLB with the property that

$$\forall_{x,y} \sum_{x,y} \Pr[x \cdot y = a \oplus b] \geq p.$$  \hspace{1cm} (8.4)

Alternatively, we can define approximate NLBs operationally, by saying that an approximate NLB is “obtained” from a (perfect) NLB by flipping Bob’s output bit with probability $1 - p$.

\(^2\)Remember that Bob’s output is “created” even if Alice has not yet input a bit into her input port and vice versa. It is even possible to demand that locally the output bit $a$ of an NLB is always uniformly random, if $b$ is not yet determined; and vice versa.
Although originally presented differently, the CHSH inequality can be recast in terms of imperfect NLBs. The availability of shared entanglement allows Alice and Bob to approximate NLBs with success probability
\[ \varphi = \cos^2 \frac{\pi}{8} = \frac{2 + \sqrt{2}}{4} \approx 85.4\% . \]

Tsirelson proved the optimality of the CHSH inequality, which translates into saying that quantum mechanics does not allow for a success probability greater than \( \varphi \) at the game of simulating NLBs [93]. See also [24] for an information-theoretic proof of the same result.

In the next section we attempt to show why the axioms of quantum mechanics are such that they do not allow to approximate NLBs with success probability greater than \( \varphi \).

### 8.3 Main result

Our main theorem is stated below and proved in the rest of this chapter. It shows that even the availability of imperfect NLBs would dramatically change the picture of communication complexity discussed in Section 2.4: It would make the randomized communication complexity of all functions trivial. Indeed, most computer scientists would consider a world in which randomized communication complexity is trivial to be as surprising as a modern physicist would find the violation of causality.

**8.3.1. Theorem.** In any world in which it is possible, without communication, to implement an approximate NLB that works correctly with probability greater than \( \frac{3 + \sqrt{6}}{6} \approx 90.8\% \), i.e.
\[ \forall x, y \Pr[x \cdot y = a \oplus b] > \frac{3 + \sqrt{6}}{6} \approx 90.8\% , \]

every Boolean function has trivial probabilistic communication complexity of just one bit.

### 8.4 Proof

To prove this theorem, we introduce the notion of distributed computation and the notion of bias for such computations. Then, we explain how compute any Boolean function with small bias and show how to amplify this “natural” bias by having Alice and Bob calculate it many times and taking the majority. We determine how imperfect a majority gate can be and still increase the bias. Finally, we construct a majority gate with the use of NLBs, and we determine to what extent we can allow them to be faulty.
8.4.1 Distributed computation

8.4.1. Definition. A bit $c$ is distributed if Alice has bit $a$ and Bob bit $b$ such that $c = a \oplus b$.

8.4.2. Definition. A Boolean function $f$ is distributively computed by Alice and Bob if, given inputs $x$ and $y$, they can produce a distributed bit equal to $f(x, y)$.

8.4.3. Definition. A Boolean function is biased if it can be distributively computed without any communication and with probability strictly greater than $\frac{1}{2}$.

8.4.4. Lemma. Provided Alice and Bob are allowed to share random variables, all Boolean functions are biased.

**Proof:** Let $f : \{0, 1\}^n \times \{0, 1\}^n \mapsto \{0, 1\}$ be an arbitrary Boolean function and let Alice and Bob share a uniformly distributed random variable $z \in \{0, 1\}^n$. Upon receiving her input $x$, Alice produces $a = f(x, z)$. Bob's strategy is to test whether $y = z$. If so, he produces $b = 0$; if not, he produces a uniformly distributed random bit $b$. In the lucky event that $y = z$, the bit distributed between Alice and Bob is correct since $a \oplus b = f(x, z) \oplus 0 = f(x, y)$. In all other cases, the distributed bit $a \oplus b$ is uniformly random. Summing up, the distributed bit is correct with probability $\frac{1}{2^n} + \left(1 - \frac{1}{2^n}\right)\frac{1}{2} = \frac{1}{2} + \frac{1}{2^{n+1}} > \frac{1}{2}$. 

8.4.5. Definition. A Boolean function has bounded bias if it can be distributively computed without any communication and with probability bounded away from $\frac{1}{2}$, with probability at least $\frac{1}{2} + \epsilon$ for some $\epsilon > 0$.

The difference between bias and bounded bias is that the probability of being correct in the former case can come arbitrarily close to $\frac{1}{2}$ as the input size increases. In the latter case, there must be some fixed $p > \frac{1}{2}$ such that the probability of being correct is at least $p$ no matter how large the inputs are.

8.4.6. Lemma. Any Boolean function that has bounded bias has trivial probabilistic communication complexity of one bit.

**Proof:** Assume Boolean function $f$ has bounded bias. For all inputs $x$ and $y$, Alice and Bob can produce bits $a$ and $b$, respectively, without communication, such that $a \oplus b = f(x, y)$ with probability at least $p > \frac{1}{2}$. If Bob transmits his single bit $b$ to Alice, she can compute $a \oplus b$, which is correct with bounded error probability. ■
Chapter 8. Limits on non-locality from communication complexity

8.4.2 Bias Amplification

8.4.7. Definition. The non-local majority problem consists in computing the distributed majority of three distributed bits. More precisely, let Alice have bits $x_1, x_2, x_3$ and Bob have $y_1, y_2, y_3$. The purpose is for Alice and Bob to compute $a$ and $b$, respectively, such that

$$a \oplus b = \text{Maj}(x_1 \oplus y_1, x_2 \oplus y_2, x_3 \oplus y_3),$$

where $\text{Maj}(u, v, w)$ denotes the bit occurring the most among $u, v$ and $w$. The computation must be achieved without any communication between Alice and Bob.

Von Neumann proved a statement rather similar to Lemma 8.4.8 below in 1956, albeit not in the context of distributed computation [67]. A more general result appears also in [38]. We sketch the proof nevertheless for the sake of completeness.

8.4.8. Lemma. For any $q$ such that $\frac{5}{6} < q \leq 1$, if Alice and Bob can compute non-local majority with probability at least $q$, then every Boolean function has bounded bias.

Proof: Let $f$ be an arbitrary Boolean function, fix Bob’s input size, and consider any $p > \frac{1}{2}$ so that Alice and Bob can distributively compute $f$ with probability at least $p$. We know from Lemma 8.4.4 that such a $p$ exists (although it may depend on the input size). Let Alice and Bob apply their distributed computational process three times, with independent random choices and shared random variables each time. This produces three distributed bits such that each of them is correct with probability at least $p$. Now, let Alice and Bob compute the non-local majority of these three outcomes with correctness probability at least $q$, which we assumed they can do. Because the overall result will be correct either if most of the distributed outcomes were correct and the distributed majority calculation was performed correctly, or if most of the distributed outcomes were wrong and the distributed majority calculation was performed incorrectly, the probability that the distributed majority as computed yields the correct value of $f$ is at least

$$h(p) = q(p^3 + 3p^2(1-p)) + (1-q)(3p(1-p)^2 + (1-p)^3).$$

Define

$$s = \frac{1}{2} + \sqrt{\frac{6q - 5}{8q - 4}} > \frac{1}{2}.$$

With this definition $h(s) = s$ and $h\left(\frac{1}{2}\right) = \frac{1}{2}$, see also Figure 8.4.2. Further,

$$\frac{d}{dp} h(p) = 6(1-p)p(2q - 1)$$
8.4. **Proof**

Figure 8.2: \( h(p) \) for \( q=0.95 \)

is positive for \( 0 < p < 1 \). Thus, \( p < h(p) < s \) provided \( \frac{1}{2} < p < s \). Because of this and the fact that \( h(p) \) is continuous over the entire range \( \frac{1}{2} < p < s \), iteration of the above process can boost the probability of distributively computing the correct answer arbitrarily close to \( s \). This proves that \( f \) has bounded bias because, given any fixed value of \( q > \frac{5}{6} \), we can choose an arbitrary constant \( t < s \) such that \( t > \frac{1}{2} \) and distributively compute \( f \) with probability at least \( t \), independently of the input size.

**8.4.9. Definition.** The **non-local equality** problem consists in distributively deciding if three distributed bits are equal. More precisely, let Alice have bits \( x_1, x_2, x_3 \) and Bob have \( y_1, y_2, y_3 \). The purpose is for Alice and Bob to compute \( a \) and \( b \), respectively, such that

\[
a \oplus b = \begin{cases} 
1 & \text{if } x_1 \oplus y_1 = x_2 \oplus y_2 = x_3 \oplus y_3 \\
0 & \text{otherwise}.
\end{cases}
\]

The computation of \( a \) and \( b \) must be achieved without any communication between Alice and Bob.

**8.4.10. Lemma.** **Non-local equality can be computed using only two (perfect) non-local boxes.**

**Proof:** The goal is to obtain \( a \) and \( b \) such that:

\[
a \oplus b = (x_1 \oplus y_1 = x_2 \oplus y_2) \land (x_2 \oplus y_2 = x_3 \oplus y_3). \tag{8.5}
\]

First, Alice and Bob compute locally \( x' = x_1 \oplus x_2, y' = y_1 \oplus y_2, x'' = x_2 \oplus x_3 \) and \( y'' = y_2 \oplus y_3 \). Then (8.5) becomes equivalent to \( (x' \oplus y') \land (x'' \oplus y'') = a \oplus b \). Hence, it is sufficient to show how Alice and Bob can compute the **AND** of the distributed bits \( x' \oplus y' \) and \( x'' \oplus y'' \).
By distributivity of the AND over the exclusive-or,
\[(x' \oplus y') \land (x'' \oplus y'') = (x' \land x'') \oplus (x' \land y'') \oplus (x'' \land y') \oplus (y' \land y'').\]

Using two non-local boxes, Alice and Bob can compute distributed bits \(a' \oplus b'\) and \(a'' \oplus b''\) with \(a' \oplus b' = x' \land y''\) and \(a'' \oplus b'' = x'' \land y'\). Setting \(a = (x' \land x'') \oplus a' \oplus a''\) and \(b = (y' \land y'') \oplus b' \oplus b''\) yields (8.5), as desired.

8.4.11. **Lemma.** Non-local majority can be computed using only two (perfect) non-local boxes.

**Proof:** Let \(x_1, x_2, x_3\) be Alice’s input and \(y_1, y_2, y_3\) be Bob’s. For \(i \in \{1, 2, 3\}\), let \(z_i = x_i \oplus y_i\) be the \(i^{th}\) distributed input bit. By virtue of Lemma 8.4.10, Alice and Bob use their two NLBs to compute the non-local equality of their inputs, yielding \(a\) and \(b\) so that \(a \oplus b = 1\) if and only if \(z_1, z_2, z_3\) are equal. Finally, Alice produces \(a' = \overline{a} \oplus x_1 \oplus x_2 \oplus x_3\) and Bob produces \(b' = b \oplus y_1 \oplus y_2 \oplus y_3\). Let
\[z = a' \oplus b' = (\overline{a} \oplus b) \oplus (z_1 \oplus z_2 \oplus z_3)\]
be the distributed bit computed by this protocol. Four cases need to be considered, depending on the number \(\ell\) of 1s among the \(z_i\)’s:

1. if \(\ell = 0\), then \(a \oplus b = 1\) and \(z_1 \oplus z_2 \oplus z_3 = 0\);
2. if \(\ell = 1\), then \(a \oplus b = 0\) and \(z_1 \oplus z_2 \oplus z_3 = 1\);
3. if \(\ell = 2\), then \(a \oplus b = 0\) and \(z_1 \oplus z_2 \oplus z_3 = 0\);
4. if \(\ell = 3\), then \(a \oplus b = 1\) and \(z_1 \oplus z_2 \oplus z_3 = 1\).

We see that \(z = 0\) in the first two cases and \(z = 1\) in the last two, so that \(z = \text{Maj}(z_1, z_2, z_3)\) in all cases.

We are now ready to prove our main theorem.

**Proof of Theorem 8.3.1:** Assume NLBs can be approximated with some probability \(p\) of yielding the correct result. Using these approximate NLBs, we can compute non-local majority with probability \(q = p^2 + (1 - p)^2\) since the protocol given in the proof of Lemma 8.4.11 succeeds precisely if none or both of the NLBs behave incorrectly. The result follows from Lemmas 8.4.6 and 8.4.8 because \(q > \frac{5}{6}\) whenever \(p > \frac{3+\sqrt{6}}{6}\).

8.4.12. **Corollary.** In any world in which probabilistic communication complexity is nontrivial, non-local boxes cannot be implemented without communication, even if we are satisfied in obtaining the correct behaviour with probability \(\frac{3+\sqrt{6}}{6} \approx 90.8\%\).
8.5 Discussion

In conclusion, we have shown that in any world in which communication complexity is nontrivial, there is a bound on how much nature can be non-local. This bound, which is an improvement over previous knowledge that non-local boxes could not be implemented exactly \cite{95, 96, 26}, approaches the actual bound \( \varphi \approx 85.4\% \) imposed by quantum mechanics. The obvious open question is to close the gap between these probabilities. A proof that nontrivial communication complexity forbids non-local boxes to be approximated with probability greater than \( \varphi \) would be very interesting, as it would render Tsirelson’s bound \cite{93} inevitable, making it a candidate for a new information-theoretic axiom for quantum mechanics \cite{20}. We will finish with some remarks.

8.5.1 One NLB for majority

Neither non-local majority nor non-local equality can be solved exactly with a single non-local box. Otherwise, entanglement could approximate that NLB well enough to solve the non-local majority problem with probability \( \varphi \approx 0.854 > \frac{5}{6} \) of being correct \cite{25}. It would follow from Lemmas 8.4.6 and 8.4.8 that all Boolean functions have trivial entanglement-assisted communication complexity. But we know this not to be the case for the inner product, as we stated earlier in Section 2.4 and prove in Appendix C.

8.5.2 Fault-tolerance threshold

Quite surprisingly, our results also give bounds on the maximum admissible error for purely classical fault-tolerant computation. This subject was discussed already in more detail in earlier chapters, see in particular Chapter 6. Here we only want to explain this interesting connection.

Suppose that we could transform any classical circuit into a fault-tolerant version that would work with probability bounded away from \( \frac{1}{2} \) even if each gate failed independently with probability \( \frac{1}{4} \). Assume furthermore that the fault-tolerant circuit \( C \) is composed only of unary and binary gates, i.e., gates with at most two input wires. In the proof of Lemma 8.4.10, we showed how to simulate distributed AND gates by use of two NLBs. Similarly, it is easy to see that all other binary gates can be computed distributively with at most two NLBs. (Several gates require no NLBs at all, such as the unary NOT gate and the binary XOR gate, also known as the CNOT gate or PARITY gate.) Now, quantum mechanics provides us with NLBs with correctness probability \( \varphi \), which yields distributed gates that are correct with probability \( (1 - \varphi)^2 + \varphi^2 = \frac{3}{4} \) if two NLBs are needed (e.g. the AND gate), or better if no NLBs are needed.

This allows us to use the assumed fault-tolerant circuit \( C \) in a distributed way and conclude that all Boolean functions have bounded bias, and therefore
trivial quantum probabilistic communication complexity. But this is impossible since most Boolean functions, for example the inner product, require $\Omega(n)$ bits of communication even if Alice and Bob share entanglement and are satisfied with a probability of correct answer bounded away from $\frac{1}{2}$. It follows that circuits cannot in general be fault-tolerant if all gates have at most two input wires and the gates fail with probability $\frac{1}{4}$ or more, even if NOT and XOR gates are perfect.

As an interesting coincidence, the best known upper bound on the error threshold, due to Evans and Schulman [37], states that fault-tolerance is impossible in general for circuits with gates of fan-in at most 2 which fail with probability $1 - \rho = \frac{2 - \sqrt{2}}{4}$ or worse.