Electronic Colloquium on Computational Complexity, Report No. 164 (2010)

Better gates can make fault-tolerant computation impossible

Falk Unger *

November 4, 2010

Abstract

We consider fault-tolerant computation with formulas composed of noisy Boolean gates with two input wires. In our model all gates fail independently of each other and of the input. When a gate fails, it outputs the opposite of the correct output. It is known that if all gates fail with probability at least $\beta_2 = (3-\sqrt{7})/4 \approx 8.856\%$, fault-tolerant computation is not possible. On the other hand, if all gates fail with probability $\epsilon < \beta_2$ and ϵ is the same for all gates, then fault-tolerant computation is possible. The assumption that all gates fail with *exactly* the same probability is pretty strong and unrealistic in real-world scenarios. Furthermore, one might be tempted to think that it can be removed easily, since making gates "better" should not hurt. Surprisingly, this is not the case, as we show in this work: there is a constant $\alpha_2 < \beta_2$ such that almost all functions cannot be computed by formulas, if the noise rate of each individual gate is selected adversarially in the range $[0, \alpha_2]$. Hence, while a hardware manufacturer who consistently produces bad gates with noise rate α_2 can always achieve reliable computation, a manufacturer who can only ensure that all gates have noise rates at most α_2 cannot.

1 Introduction

Essentially all physical devices for information processing can fail. It was realized early on by von Neumann [19] that it is still possible to combine unreliable devices in such a way that arbitrary computation is possible. Since then, a huge amount of work in many different areas has been devoted to the design of systems that still perform reliably even if some of their parts fail, but also to finding the limitations of such fault-tolerant designs.

In this work we consider computation by formulas, which are circuits in which each gate has exactly one output wire. Each gate has two Boolean input wires and computes a Boolean output. Every gate G fails independently of the other gates (and independently of the input) with some error probability ϵ_G, i.e. with probability $1-\epsilon_G$

*UC Berkeley, falk.unger@gmail.com

ISSN 1433-8092

it outputs the correct result, and with probability ϵ_G it outputs the opposite. For this model there is a noise bound of $\beta_2 = (3 - \sqrt{7})/4 \approx 8.856\%$ in the following way: if all gates fail with probability at least β_2, then fault-tolerant computation is not possible [32]. If all gates fail with the *same* probability $\epsilon < \beta_2$, then fault-tolerant computation is possible, see Evans and Pippenger [6]. The assumption that all gates fail with the same probability is somewhat unsatisfactory, since this means that a hardware manufacturer who can only produce unreliable gates must do so in a reliable way, i.e. all gates must have the same error rate. At first, one might be tempted to think that the requirement that all gates fail with the same probability is easy to remove, since if a computation works correctly if all gates have noise rate $\epsilon < \beta_2$, then lowering the noise rates of some gates should not hurt. We show that this intuition is wrong.

Theorem 1. *There is a $0 \leq \alpha_2 < \beta_2$ such that for any $\Delta > 0$ and almost all Boolean functions*[1] *there is no formula F that computes f with bias Δ, if the noise rate of each individual gate in F is chosen adversarially in $\{0, \alpha_2\}$.*

We mentioned earlier that this implies that a hardware manufacturer who can only ensure an upper bound of α_2 on the noise rates cannot ensure that the hardware always works, while a hardware manufacturer who produces consistently bad gates with error exactly α_2 can always achieve reliable computation.

Note that a hardware manufacturer, who can first produce gates, then test them (i.e. determine their noise rate) and use only gates with similar noise rates, is not affected by our result.[2] Our result would however apply to a manufacturer who cannot test their gates before deciding whether to use them. In particular, this applies to the way modern computer chips are produced, where gates are "printed" to a sample, and once there, they cannot easily be removed.

It is possible to consider more adversarial settings. For example, the noise rate of a gate might depend on the input to the gate (and previous gates) and also on the input to the function being computed. Also in these models with an adaptive adversary (see Pippenger [23] for exact definitions) it is possible to do fault-tolerant computation with constant noise rates. In particular, this also means that the smallest α_2 which satisfies Theorem 1 is greater than zero, and hence our model does allow fault-tolerant computation. We do not know what the smallest possible value for α_2 is.

More adversarial models will almost certainly lead to much lower acceptable noise rates. In the example model mentioned above, the adversary might decide not to apply noise, if the inputs to the gate are already "wrong", and hence prevent that the output of the gate flips back to its "correct" value. The surprising fact about Theorem 1 is that even a non-adaptive adversary who can essentially only make gates "better", can prevent fault-tolerant computation. Incidentally, we also do not know about any previous results which establish a better upper bound on the acceptable noise rate than β_2 for any other (more adversarial) model.

[1] "Almost all" means all Boolean functions that depend on more than a constant number $C(\Delta)$ of input bits.

[2] It is not shown explicitly in [6] that this will work, but inspecting their proof one can see that their scheme also works if all gates have noise rate less than β_2 and the noise rates are similar enough.

We prove our result only for formulas but not for circuits. Generally, the known noise bounds for fault-tolerant computation are weaker for circuits, see Section 1.1. However, we believe that the only reason for this is that circuits are harder to analyze and that in particular Theorem 1 is also true for circuits.

More discussion of our result and model is deferred to Section 3.6.

1.1 Context and related work

Noise and fault-tolerance has been studied in many contexts, for example fault-tolerant computation by circuits [19, 5, 22, 28, 11] with threshold analyses for classical computation in [19, 14, 9, 8, 6, 32, 15, 12] and for quantum computation in [1, 17, 18, 2, 27, 25, 16, 3, 26], communication protocols [24, 29, 20], decision trees [28, 20, 7, 10], learning theory [30, 31, 13, 4] but also others (some referenced in the cited work).

Most of these are positive results, showing that also under noise the desired tasks can be accomplished. We do not know about any previous results which prove that fault-tolerant computation becomes impossible if the noise rates of some gates become too small. However, there are results which suggest that in other settings a similar phenomenon might occur, some of which we review now.

In the same model as ours, but for formulas with gates of fan-in k and k odd, Evans and Schulman [9] proved the tight noise bound $\beta_k = \frac{1}{2} - \frac{2^{k-2}}{k\binom{k-1}{k/2-1/2}}$: if all gates fail independently with the same fixed probability $\epsilon < \beta_k$, then any function can be computed, and if each gate fails with some probability at least β_k (which does not need to be the same for all gates), universal computation is not possible. For $k = 3$ the result was first established by Hajek and Weller [14]. To establish the lower bound for β_k they also strongly rely on the assumption that all gates have the same error probability and it is likely that also for gates with fan-in $k > 2$ this assumption is necessary. For even $k > 2$ tight noise bounds in the above sense are not known. However, we believe that a similar approach to ours (previously used in [32]), using a suitable potential function, can also lead to tight (non-adversarial) noise bounds.

It is also worth noticing that all results mentioned—including the present—only work for formulas. The best upper bounds on the acceptable noise rates for circuits with at most k input wires are $\frac{1}{2} - \frac{1}{2\sqrt{k}}$ by Evans and Schulman [8]. We believe that the bounds for formulas also hold for circuits, and that the only reason why upper bounds for circuits are worse is because they are harder to analyze.

An area in which fault-tolerance is particularly important is *quantum computation*, since here the physical components in which the information is stored have to be very small, and with current technology it is very hard to operate on them accurately. It is likely that results similar to ours also hold for quantum computers. Apart from the fact that this would be in analogy with our result, another small indicator is that among the currently best rigorous lower bounds on the fault-tolerance threshold [2, 27] the higher bound by Reichardt [27] is also proven under the assumption that gates fail independently of each other and the noise rates are the same for all gates.

It is also worth mentioning that a certain "guaranteed" amount of noise might be

helpful to *speed up* computation. For example, if P≠BPP,[3] then circuits in which all gates are guaranteed to have some (small enough) amount of noise could use this noise to distill random bits and hence these circuits would be able solve problems in BPP\P more efficiently than noise-free (deterministic) circuits. Thus, also in this setting lower noise rates might be harmful.

Also in learning theory a similar phenomenon is known: without giving any precise definitions, we simply state that Goldman and Sloan [13] show that certain concept classes are not PAC-learnable, if the samples provided to the learning algorithm undergo some noise process, in which the exact noise rates are not known, only an upper bound on them. However, Decatur and Gennaro [4] show that if the noise rates are known to the learning algorithm, then these concept classes become PAC-learnable.

Our proof extends an approach in [32], in which it is shown that fault-tolerant computation is not possible if all gates fail with probability at least β_2. This result can easily be reproved from one of our Lemmas. However, the potential function we are using in this paper (see later) is nicer and leads to simpler calculations.

1.2 Organization of the paper

Section 2 contains some standard definitions. Section 3 contains the proof Theorem 1. In Section 3.1 we explain the proof on a high level and also provide some important notation. Section 3.6 contains some extensions and final remarks.

2 Definitions

A *circuit* is composed of gates. Each *gate* has a certain number of input wires, which is called the *fan-in* of the gate. The wires can take Boolean values 0 or 1. A gate computes an output bit as a Boolean function of its input bits. A *formula* is a particular type of circuit in which the gates are connected in a tree, with the output gate at the root and the input bits at the leaves. In particular, this means that each gate has exactly one output wire.

A (perfect) PARITY gate with input bits x_1 and x_2 outputs 0 if $x_1 = x_2$ and 1 otherwise. A (perfect) NOR gate outputs 1 if $x_1 = x_2 = 0$ and 0 otherwise. We say that a gate G with fan-in 2 is of *NOR-type* if it can be obtained from a NOR gate by applying NOT gates to the input / output wires, i.e. there is an odd number of inputs $x_1, x_2 \in \{0, 1\}$ which are mapped to 1.

We call a gate ϵ-*noisy* if it outputs the correct result with probability $1 - \epsilon$ and with probability ϵ it outputs the opposite. For any $\epsilon \leq 1/2$ we define the function

$$\eta_\epsilon(x) = (1 - 2\epsilon)x + \epsilon.$$

If x is the probability that some Boolean variable is 0, then $\eta_\epsilon(x)$ is the probability that it is 0 after it has gone through an ϵ-noisy bit-flip channel. Most of the time we use η_ϵ with $\epsilon = \beta_2$, in which case we will omit the subscript, i.e. $\eta(x) = (1-2\beta_2)x+\beta_2$.

[3] P is the class of all functions which can be efficiently computed by deterministic algorithms. BPP is the class of functions which can be efficiently computed by randomized algorithms. It is widely believed [21] that P=BPP, but the final answer is not yet known.

We say that a formula F with noisy gates computes the function f with bias $\Delta > 0$ if for all $x \in f^{-1}(0)$, $y \in f^{-1}(1)$: $\Pr[F(x) = 0] \geq \Delta + \Pr[F(y) = 0]$.

A Boolean function $f : \{0,1\}^n \to \{0,1\}$ *depends* on the i-th input bit x_i if there is some setting of the other bits, such that flipping x_i flips the function value.

3 Proof

3.1 Outline and notation

In our proof we will show that under the conditions of Theorem 1 all Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that depend on sufficiently many input bits cannot be computed with bounded bias $\Delta > 0$. For such functions, we will fix a particular input bit x_i which f depends on, and fix all other bits such that flipping x_i flips the output, see the proof of Theorem 1 later. Assume that f is computed by a formula F with noisy gates that fail independently. Then, for each gate in F with input wires A and B and output wire C we can define

$$a = \frac{1}{2}\Pr[A = 0 \mid x_i = 0] + \frac{1}{2}\Pr[A = 0 \mid x_i = 1] \quad (1)$$

$$\delta_a = \Pr[A = 0 \mid x_i = 0] - \Pr[A = 0 \mid x_i = 1] \quad (2)$$

and analogously for B and C. The variable a can be seen as the average probability of A being 0. We call δ_a the *bias* of A. It is clear that for every wire A we always have $|\delta_a| \leq \min\{2a, 2(1-a)\}$, an observation we will use repeatedly.

The lemmas in the next section will use this notation, and implicitly assume a function f and a particular choice of an input bit x_i. Our lemmas hold for all f, F and x_i. In the proof of Theorem 1 we will instantiate f, F and x_i.

We define a *potential function* q as

$$q(x) = \frac{1}{(1-x)x + (11 - 4\sqrt{7})/18}, \quad (3)$$

where $(11 - 4\sqrt{7})/18 \approx 0.023$. See Figure 1 for a plot.

It is convex, symmetric around $1/2$ and bounded between positive constants for $x \in [0,1]$. In particular, $q_{max} := \max_{x\in[0,1]} q(x) = q(0) = 18/(11 - 4\sqrt{7}) < 50$ and $q_{min} := \min_{x\in[0,1]} q(x) = q(1/2) > 2$. Furthermore, $1/q(\cdot)$ is concave and also bounded between positive constants for $x \in [0,1]$.

The main argument in our proof (Lemma 1) is that for any ϵ-noisy gate G with input wires A, B and output wire C it holds

$$|\delta_c| q(c) \leq \theta \max\{|\delta_a| q(a), |\delta_b| q(b)\}. \quad (4)$$

for $\theta = 1$ if $\epsilon \geq \beta_2$. For wire A we also say that $|\delta_a| q(a)$ is the *extractable information*, i.e. the information about x_i that can be extracted by noisy formulas. We can interpret (4) as saying that a gate with noise at least β_2 cannot increase the "extractable information". Furthermore, we will see that this inequality is only tight if $\delta_a, \delta_b \to 0$,

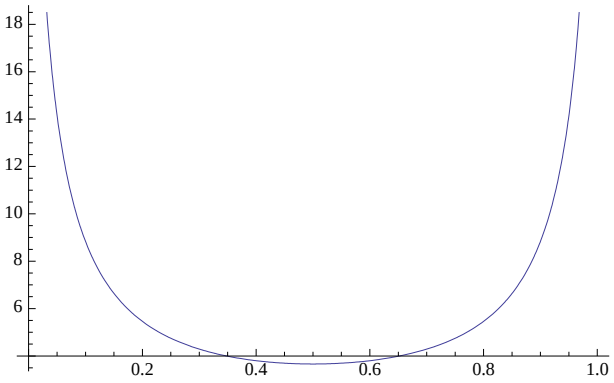

Figure 1: Graph of $q(x)$

G is a NOR gate (or of NOR-type) and a, b satisfy certain properties, e.g. if G is a NOR gate then $a = b = \hat{x}$ with

$$\hat{x} = \frac{1+\sqrt{7}}{6} \approx 0.61.$$

We will chop the formula F into subformulas S, of depth $2D$, where D is given in Lemma 2. We then show that if an S satisfies all the above conditions approximately (i.e. it is composed of NOR gates and for the input wires A_i of the subformula it holds $\delta_{a_i} \approx 0$ and $a_i \approx \hat{x}$), then the output wire O of the subformula will have much smaller extractable information then any of its input wires if all gates are *noise-free*, see Lemma 2. If S does not satisfy these conditions, then we use (in the proof of Lemma 3) that for some gate in S inequality (4) will hold for some fixed $0 \leq \theta < 1$, if all gates have noise at least β_2. Then some small additional technical arguments (Claim 5) will also imply that the extractable information $|\delta_o|q(o)$ of the output wire O must have decreased sufficiently. A continuity argument (Lemma 4) will then imply that this also holds if all gates fail with probability α_2, for some $0 \leq \alpha_2 < \beta_2$. This implies that the extractable information of any wire decreases "on average" over blocks of depth $2D$.

3.2 Noisy gates

The following lemma is the main technical lemma. A similar lemma was used in [32], in which it was shown that fault-tolerant computation is not possible if all gates fail with probability at least β_2. The current version is slightly stronger, as this will be needed for our later result, and its proof is simpler, which is mainly due to the fact that the potential function $q(\cdot)$ we are using here is nicer.

Lemma 1. *Assume an ϵ-noisy gate G, with input wires A and B and output wire C. Define a, b, c and $\delta_a, \delta_b, \delta_c$ as in Section 3.1. The following inequality*

$$\forall_{\beta_2 \leq \epsilon \leq 1/2} : |\delta_c|q(c) \leq \theta \max\{|\delta_a|q(a), |\delta_b|q(b)\}. \tag{5}$$

holds in all of the following cases:

1. *Inequality (5) holds with* $\theta = 1$.
2. *If* G *is not of NOR-type then* $\exists_{0\le\theta<1}$ *such that inequality (5) holds.*
3. $\forall_{0\le r<1}\exists_{0\le\theta<1}$ *ineq. (5) holds if* $\min\{|\delta_a|q(a), |\delta_b|q(b)\} \le r\max\{|\delta_a|q(a), |\delta_b|q(b)\}$.
4. $\forall_{\delta>0}\exists_{0\le\theta<1}$ *inequality (5) holds if*
   (a) $|\delta_a| \ge \delta$ *or* $|\delta_b| \ge \delta$ *or*
   (b) G *is a NOR gate and* $|a - \hat{x}| \ge \delta$ *or* $|b - \hat{x}| \ge \delta$

The proof of this lemma is a bit tedious and technical, we suggest to skip it during a first read. Essentially, we are optimizing real polynomials.

Before we prove Lemma 1 we show some simple statements. Our first observation says that instead of analyzing a gate G directly, we may analyze G with additional NOT gates on its input / output wires, if we also change the input / output distributions appropriately.

Observation 1. *Let* G *be a gate as in Lemma 1. Let* G' *be as* G*, but with an additional NOT on wire* A*. Then any of the statements of Lemma 1 (apart from (4b)) hold for* G*,* a, b, c *and* $\delta_a, \delta_b, \delta_c$ *if and only if they also hold for* G' *and* a, δ_a *"negated", i.e.* $a' := 1 - a, b' := b, c' := c$ *and* $\delta'_a := -\delta_a, \delta'_b := \delta_b, \delta'_c := \delta_c$*. The same holds analogously for wires* B *and* C*.*

Proof. Note that $q(a) = q(1 - a)$ and hence $|\delta'_a|q(a') = |\delta_a|q(a)$. Therefore, the extractable information on all wires is the same for G and G'. For wires B and C an analogous argument works. □

The next observation is a monotonicity statement about the extractable information $|\delta_c|q(c)$ of a wire C. It basically says that moving the quantities $\Pr[C = 0|x_i = 0]$ and $\Pr[C = 0|x_i = 1]$ apart cannot decrease $|\delta_c|q(c)$.

Observation 2. *Let* C *be a wire whose information content about* x_i *is parameterized by* y, z *as follows:*

$$\begin{aligned} y &= \min\{\Pr[C = 0|x_i = 0], \Pr[C = 0|x_i = 1]\} \\ z &= \max\{\Pr[C = 0|x_i = 0], \Pr[C = 0|x_i = 1]\}. \end{aligned}$$

Let C' *be another wire and let* $y' \le z'$ *be defined analogously. Then it holds that if* $y' \le y$ *and* $z \le z'$ *then* $|\delta_c|q(c) \le |\delta_{c'}|q(c')$.

Proof. We prove that the first derivative of $|\delta_c|q(c)$ with respect to z is positive for $0 \le y \le z \le 1 - 0$. This implies that moving z up to z' cannot decrease $|\delta_c|q(c)$. Analogously, one can show that moving y to y' cannot decrease $|\delta_c|q(c)$ either, which then implies the statement of the observation.

One can calculate that

$$|\delta_c|q(c) = (z - y)q(\frac{y + z}{2}) = \frac{36(y - z)}{9y^2 + 18(z - 1)y + 9z^2 - 18z + 8\sqrt{7} - 22}.$$

The first derivative with respect to z is $-\frac{36\left(27y^2+18zy-36y-9z^2+8\sqrt{7}-22\right)}{\left(-9y^2-18zy+18y-9z^2+18z-8\sqrt{7}+22\right)^2}$. Hence, we need to show

$$27y^2 + 18zy - 36y - 9z^2 + 8\sqrt{7} - 22 < 0 \tag{6}$$

From $0 \leq y \leq z \leq 1$ it is easy to see that the following inequalities hold $27y^2 - 27y \leq 0$, $9zy - 9z^2 \leq 0$, $9zy - 9y \leq 0$ and $-0.83 \approx \sqrt{7} - 22 < 0$. Adding these together gives (6). □

Proof. (Lemma 1) First we note that it is enough to prove the lemma for $\epsilon = \beta_2$, because increasing ϵ decreases $|\delta_c|$ and moves c closer to 1/2, i.e. it also decreases $q(c)$. We will assume throughout that $|\delta_b|q(b) \leq |\delta_a|q(a)$.

G is not of NOR-type It is enough to prove item 2. We may assume that G is a (noisy) PARITY gate: the cases that G is constant (outputting noisy 0 or 1) or G outputs one of its input wires or a NOT of it, can be reduced to a PARITY gate with $\delta_a = 0$ and $a \in \{0, 1\}$ or $\delta_b = 0$ and $b \in \{0, 1\}$. This leaves us with the case that G is a (noisy) PARITY gate, but with additional (noise-free) NOT gates on its wires: note that a NOT gate on an input wire of a PARITY gate is equivalent to a NOT gate on the output wire. Hence, using Observation 1 we may assume that G *is* a PARITY gate, and also further that $a \geq 1/2, b \geq 1/2$.

If the two input wires of a noiseless PARITY gate are independently 0 with probability p resp. q, then the output wire will be 0 with probability $pq + (1-p)(1-q)$. Thus, in our case

$$\begin{aligned}
\Pr[C = 0 \mid x_i = 0] &= \eta((a + \delta_a/2)\,(b + \delta_b/2) + (1 - a - \delta_a/2)\,(1 - b - \delta_b/2)) \\
\Pr[C = 0 \mid x_i = 1] &= \eta((a - \delta_a/2)\,(b - \delta_b/2) + (1 - a + \delta_a/2)\,(1 - b + \delta_b/2))
\end{aligned}$$

which implies

$$\begin{aligned}
c &= \eta\,(ab + (1-a)(1-b) + \delta_a\delta_b/2) \\
\delta_c &= ((2a-1)\delta_b + (2b-1)\delta_a)\,(1 - 2\beta_2).
\end{aligned}$$

Hence, we need to show $\exists\theta < 1$:

$$|(2a-1)\delta_b + (2b-1)\delta_a|\,(1-2\beta_2)q(\eta\,(ab + (1-a)(1-b) + \delta_a\delta_b/2)) \leq \theta|\delta_a|q(a). \tag{7}$$

We may assume $\delta_a \geq 0$ because flipping the sign of both δ_a and δ_b does not change the statement.

Next we show that we may replace the $q(...)$-term on the lhs of (7) by the rhs of (8), by showing

$$q(\eta\,(ab + (1-a)(1-b) + \delta_a\delta_b/2)) \leq q(\eta\,(ab + 3(1-a)(1-b))). \tag{8}$$

We first notice that $\eta\,(ab + (1-a)(1-b) + \delta_a\delta_b/2) = \eta\,(ab + (1-a)(1-b)) + (1 - 2\beta_2)\delta_a\delta_b/2$. We also have $ab + (1-a)(1-b) \geq 1/2$ (since $a, b \geq 1/2$) and hence also $\eta\,(ab + (1-a)(1-b)) \geq 1/2$. Then we notice that $\forall x \geq 1/2, y \geq 0 : q(x+y) \geq q(x-y)$. Using this with $x = \eta\,(ab + (1-a)(1-b))$ and $y = (1-2\beta_2)\delta_a\delta_b/2$ we see that the lhs of (8) is maximized for $\delta_a\delta_b \geq 0$. Furthermore, $\forall x \geq 1/2, y \geq 0 : \partial q(x+y)/\partial y \geq 0$,

hence the lhs of (8) is maximized if $\delta_a\delta_b$ is maximal. Using that $\delta_a/2 \leq 1-a$ and $\delta_b/2 \leq 1-b$, we get inequality (8).

Furthermore, we may also replace the first occurrence of δ_b in (7) by its maximum $\delta_a q(a)/q(b)$, since $a, b \geq 1/2$. After doing these two replacements, dividing by $q(a)q(\eta\,(ab+3(1-a)(1-b)))$ and cancelling δ_a, inequality (7) follows from the following claim. Its proof is provided after the Lemma.

Claim 1. $\exists_{0\leq\theta<1}\forall_{1/2\leq a,b\leq 1}$:

$$\left(\frac{2a-1}{q(b)}+\frac{2b-1}{q(a)}\right)(1-2\beta_2) \leq \frac{\theta}{q(\eta\,(ab+3(1-a)(1-b)))}. \tag{9}$$

G is of NOR-type W.l.o.g. we may assume that G *is* a NOR gate, by Observation 1. We have

$$\begin{aligned} \Pr[C=0 \mid x_i=0] &= 1-\eta\,((a+\delta_a/2)\,(b+\delta_b/2)) && (10)\\ \Pr[C=0 \mid x_i=1] &= 1-\eta\,((a-\delta_a/2)\,(b-\delta_b/2))\,, && (11) \end{aligned}$$

which implies

$$\begin{aligned} \delta_c &= -\,(a\delta_b+b\delta_a)\,(1-2\beta_2) && (12)\\ c &= 1-\eta\,(ab+\delta_a\delta_b/4)\,. && (13) \end{aligned}$$

W.l.o.g. we may assume $\delta_a \geq 0$: in case $\delta_a < 0$ we can flip the sign of both δ_a and δ_b, which will change neither $|\delta_c|$ nor c. Furthermore, we may then also assume $\delta_b \geq 0$: using that $1-\eta(\cdot)$ is monotonically decreasing, we see from (10) and (11) that $\min\{\Pr[C=0 \mid x_i=0], \Pr[C=0 \mid x_i=1]\}$ is minimized and $\max\{\Pr[C=0 \mid x_i=0], \Pr[C=0 \mid x_i=1]\}$ is maximized when $\delta_b \geq 0$. Hence, by Observation 2 we may assume $\delta_b \geq 0$.

Using $q(c)=q(1-c)$ and (12) and (13) we see that (5) is equivalent to

$$(1-2\beta_2)(a\delta_b+b\delta_a)q\,(\eta\,(ab+\delta_a\delta_b/4)) \leq \theta\delta_a q(a). \tag{14}$$

In order to prove items 1 and 4 it is enough to show (14) where we replace the first occurrence of δ_b by its maximum $\delta_a q(a)/q(b)$. Canceling δ_a and dividing by $q(a)q\,(\eta\,(ab+\delta_a\delta_b/4))$ gives

$$(1-2\beta_2)\left(\frac{a}{q(b)}+\frac{b}{q(a)}\right) \leq \frac{\theta}{q\,(\eta\,(ab+\delta_a\delta_b/4))}. \tag{15}$$

In order to show (15), we show the following claim. Its proof is provided after this Lemma.

Claim 2. *For all* $0 \leq a, b \leq 1$ *the inequality*

$$(1-2\beta_2)\left(\frac{a}{q(b)}+\frac{b}{q(a)}\right) \leq \frac{\theta_1}{q\,(\eta\,(ab))} \tag{16}$$

always holds for $\theta_1 = 1$ *and under the conditions of item 4b it holds for fixed* $0 \leq \theta_1 < 1$. *Furthermore, there is a* $0 \leq \theta_2 < 1$ *for which*

$$(1-2\beta_2)\left(\frac{a}{q(b)}+\frac{b}{q(a)}\right) \leq \frac{\theta_2}{q\,(\eta\,(ab+(1-a)(1-b)))}. \tag{17}$$

The two inequalities in Claim 2 implies (15) for items 1 and 4b by concavity of $1/q(\cdot)$, $0 \le \delta_a/2 \le 1-a$ and $0 \le \delta_b/2 \le 1-b$.

With this we can also show item 3: if our parameters satisfy the conditions in item 4b we are done. Otherwise, we have $|a - \hat{x}| \le 1/1000$. We already know that (14) always holds with $\theta = 1$ (by item 1 of our Lemma) if the first δ_b is replaced by $\delta_a q(a)/q(b)$ (which yielded inequality (15)). Hence, if $\delta_b \le r\delta_a q(a)/q(b)$ for some $0 \le r < 1$, then we see that if the first occurrence of δ_b is replaced by $r\delta_a q(a)/q(b)$, then $|a - \hat{x}| \le 1/1000$ and the fact that $q(\cdot)$ is bounded between positive constants implies that there is a $\theta < 1$ such that (14) holds after this replacement.

To get item 4a we will show that $\delta_a\delta_b/4 > \delta^2/200$. This will immediately imply (15) for a $0 \le \theta < 1$ by (16), (17) and concavity of $1/q(\cdot)$. If $\delta_b > \delta$, then $\delta_a > \delta q(b)/q(a) \ge \delta \frac{\min_{x\in[0,1]} q(x)}{\max_{x\in[0,1]} q(x)} > \frac{\delta}{25}$. Hence, $\delta_a\delta_b/4 > \delta^2/100$. If $\delta_a > \delta$ we argue: if $\delta_b q(b) \le \delta_a q(a)/2$, then item 4a follows already from item 3 with $r = 1/2$. Otherwise, we can bound $\delta_b > \delta \frac{q(a)}{2q(b)} \ge \frac{\delta}{2} \frac{\min_{x\in[0,1]} q(x)}{\max_{x\in[0,1]} q(x)} > \delta/50$, and hence $\delta_a\delta_b/4 > \delta^2/200$. □

We now provide the proofs of the claims in Lemma 1. They are not difficult, but some of the calculations are a bit tedious to do by hand. Alternatively, one can use a computer algebra system that can manipulate terms symbolically.

Proof. (of Claim 1) Since $\forall_{x\in[0,1]} : 2 < q(x) < 50$, it is enough to show

$$0.04 \le \frac{1}{q(\eta\,(ab + 3(1-a)(1-b)))} - \left(\frac{2a-1}{q(b)} + \frac{2b-1}{q(a)}\right)(1-2\beta_2). \tag{18}$$

Changing variables as $x = ab$, $y = a + b$ and multiplying by 72 the right-hand side becomes the polynomial $576\sqrt{7}x^2 - 2304x^2 - 792\sqrt{7}yx + 3384yx + 648\sqrt{7}x - 2808x + 288\sqrt{7}y^2 - 1260y^2 - 564\sqrt{7}y + 2280y + 269\sqrt{7} - 994$. Since $288\sqrt{7} - 1260 < 0$, we see that the second derivative with respect to y is always negative, so the polynomial is minimized for extremal y. Given x, the maximal y is $y = 1 + x$, and then the polynomial becomes $36\left(-5 + 2\sqrt{7}\right)x^2 + \left(336 - 132\sqrt{7}\right)x - 7\sqrt{7} + 26$, whose minimum is $\frac{1859-703\sqrt{7}}{5-2\sqrt{7}} \approx 3.30$ at $x = \frac{28-11\sqrt{7}}{30-12\sqrt{7}} \approx 0.63$. The minimal y is $y = 2\sqrt{x}$. Substituting this the polynomial becomes $576\left(-4+\sqrt{7}\right)x^2 - 144\left(-47+11\sqrt{7}\right)x^{3/2} + 72\left(-109+25\sqrt{7}\right)x - 24\left(-190+47\sqrt{7}\right)\sqrt{x} + 269\sqrt{7} - 994$, which is a polynomial of degree 4 in $\sqrt{x}$, for which one can show that it is always lower bounded by 3.07 for $x = ab \in [1/4, 1]$. Then (18) follows by $3.07/72 > 0.04$. □

Proof. (of Claim 2) Since $\forall_{x\in[0,1]} : 2 < q(x) < 50$, it is enough to show

$$\kappa \le \frac{1}{q\,(\eta\,(ab))} - (1-2\beta_2)\left(\frac{a}{q(b)} + \frac{b}{q(a)}\right) \tag{19}$$

$$0.0001 \le \frac{1}{q\,(\eta\,(ab + (1-a)(1-b)))} - (1-2\beta_2)\left(\frac{a}{q(b)} + \frac{b}{q(a)}\right) \tag{20}$$

with $\kappa = 0$ and with $\kappa > 0$ under the conditions of item 4b.

We first show (19). Changing variables $x = ab$ and $y = a + b$ and multiplying by 72 the rhs of (19) becomes $36\left(-4+\sqrt{7}\right)x^2 + 36\left(\left(-1+\sqrt{7}\right)y - 3\sqrt{7} + 6\right)x +$

$\left(78-30\sqrt{7}\right)y-7\sqrt{7}+26$, which is linear in y. It is minimized when y is extremal. Given x, the maximal y is $y=1+x$, in which case the polynomial becomes $36\left(-5+2\sqrt{7}\right)x^2-6\left(-43+17\sqrt{7}\right)x-37\sqrt{7}+104$, whose minimum is $\frac{4012-1517\sqrt{7}}{10-4\sqrt{7}}>2$. The minimal y is $y=2\sqrt{x}$. Substituting this the polynomial becomes $36\left(-4+\sqrt{7}\right)x^2+36\left(2\left(-1+\sqrt{7}\right)\sqrt{x}-3\sqrt{7}+6\right)x+2\left(78-30\sqrt{7}\right)\sqrt{x}-7\sqrt{7}+26=36\left(-4+\sqrt{7}\right)(\sqrt{x}-\sqrt{\hat{x}})^2(\sqrt{x}-\sqrt{x_3})(\sqrt{x}-\sqrt{x_4})$ with $\hat{x}=\frac{1}{6}\left(1+\sqrt{7}\right)$ and $x_{3/4}=\frac{1}{6}\left(1+\sqrt{7}\pm 2\sqrt{15-3\sqrt{7}}\right)$, i.e. $x_3\approx 1.49$ and $x_4\approx -0.28$. This implies (19).

Now we show (20). Setting $x=ab$ and $y=a+b$ and multiplying by 72 the rhs of (20) becomes $\left(-576+144\sqrt{7}\right)x^2+\left(540-108\sqrt{7}\right)yx-216x+\left(-144+36\sqrt{7}\right)y^2+\left(222-66\sqrt{7}\right)y-7\sqrt{7}+26$. Since $-144+36\sqrt{7}<0$, we see that its second derivative with respect to y is always negative, so it is minimized when y is extremal. Given x, the maximal y is $y=1+x$. Substituting this the polynomial becomes $36\left(-5+2\sqrt{7}\right)x^2-6\left(-43+17\sqrt{7}\right)x-37\sqrt{7}+104$. We have calculated the minimum of this polynomial before, it is $\frac{4012-1517\sqrt{7}}{10-4\sqrt{7}}>2$. The minimal y is $y=2\sqrt{x}$ and then the polynomial becomes $144\left(-4+\sqrt{7}\right)x^2-216\left(-5+\sqrt{7}\right)x^{3/2}+72\left(-11+2\sqrt{7}\right)x+\left(444-132\sqrt{7}\right)\sqrt{x}-7\sqrt{7}+26$, which is a polynomial of degree 4 in $\sqrt{x}$. One can show that for $x\in[0,1]$ it is always at least 0.014. Hence, the right-hand side of (20) is always at least $0.014/72>0.0001$. □

3.3 Composition of noise-free NOR gates

Lemma 2. *There is a $D>0$ such that the following holds: Consider a full tree NOR^{2D} of depth $2D$ of* noise-free *NOR gates with input wires $A_1,\ldots,A_{2^{2D}}$ and output wire B at the top. If*

$$\forall_{1\le i\le 2^{2D}}:|\hat{x}-a_i|\le 1/1000\wedge\delta_{a_i}\le 1/1000 \tag{21}$$

then $|\delta_b|q(b)\le\frac{1}{2}\max_{1\le i\le 2^{2D}}|\delta_{a_i}|q(a_i)$.

Proof. We define a function f as $f(y_1,\ldots,y_4)=1-(1-y_1y_2)(1-y_3y_4)=y_1y_2+y_3y_4-y_1y_2y_3y_4$. For a tree of NOR gates of depth 2, whose four input wires are independently 0 with probabilities $y_1,\ldots,y_4$, $f(y_1,\ldots,y_4)$ is the probability that the output wire at the root is 0. We also define its k-fold composition, by setting $f^{(1)}=f$ and $f^{(k+1)}(y_1,\ldots,y_{4^{k+1}})=f(f^{(k)}(y_1,\ldots,y_{4^k}),\ldots,f^{(k)}(y_{3\cdot 4^k+1},\ldots,y_{4^{k+1}}))$. It is easy to see that f is monotonically increasing in each argument if $\forall_i:y_i\in[0,1]$, and so is $f^{(k)}$. Furthermore,

$$\delta_b=f^{(D)}(a_1+\delta_{a_1}/2,\ldots,a_{4^D}+\delta_{a_{4^D}}/2)-f^{(D)}(a_1-\delta_{a_1}/2,\ldots,a_{4^D}-\delta_{a_{4^D}}/2). \tag{22}$$

We want to compute an upper bound on $|\delta_b|$. Define $\hat{\delta}=\max_i|\delta_{a_i}|$. Since f is monotonically increasing in each argument, it is clear that the absolute value of the rhs is maximized if all δ_{a_i} have the same sign, so w.l.o.g. we assume $\forall_i:\delta_{a_i}\ge 0$. It implies $\delta_b\ge 0$. Similarly, we can upper bound δ_b by replacing each δ_{a_i} by $\hat{\delta}$.

Claim 3. $\forall_{0\le y_1,\ldots,y_4\le\hat{x}+3/2000}:f(y_1,\ldots,y_4)\le\max\{y_1,\ldots,y_4\}$.

Proof. We noted before that f is monotonically increasing in each argument, so it is enough to prove the statement for $y := y_1 = y_2 = y_3 = y_4$. We compute $f(y,y,y,y) - y = y(2y - y^3 - 1) = -y(y-1)(y + \frac{1+\sqrt{5}}{2})(y + \frac{1-\sqrt{5}}{2}) \leq 0$ for $y \in [0, \frac{\sqrt{5}-1}{2}]$. Noting that $\hat{x} + 3/2000 < \frac{\sqrt{5}-1}{2}$ proves the claim. □

Claim 4. $\forall_{0 \leq y_1,\ldots,y_4} \forall_{0 \leq \delta}$:

$$f(y_1 + \delta/2, \ldots, y_4 + \delta/2) - f(y_1 - \delta/2, \ldots, y_4 - \delta/2) \leq 4\delta \max\{y_1, \ldots, y_4\}.$$

Proof. Let $y = \max\{y_1, \ldots, y_4\}$. Note that $\prod_i (y_i + \delta/2) \geq \prod_i (y_i - \delta/2)$. Looking at the definition of f, we see that the lhs of the inequality can be upper bounded by $(y_1 + \delta/2)(y_2 + \delta/2) + (y_3 + \delta/2)(y_4 + \delta/2) - (y_1 - \delta/2)(y_2 - \delta/2) + (y_3 - \delta/2)(y_4 - \delta/2)$. The term $(y_1 + \delta/2)(y_2 + \delta/2) - (y_1 - \delta/2)(y_2 - \delta/2)$ is maximized when both y_1 and y_2 are maximally equal to y. The same holds analogously for y_3 and y_4. Plugging this in gives $2(y + \delta/2)^2 - 2(y - \delta/2)^2 = 4y\delta$. □

With a pocket calculator we compute $f^{(11)}(\hat{x} + 3/2000, \ldots, \hat{x} + 3/2000) \approx 0.021 \leq 1/40$. Furthermore, (21) implies that for each input wire A_i it holds $\max\{a_i + \delta_{a_i}, a_i - \delta_{a_i}\} \leq \hat{x} + 3/2000$. Then by Claim 3 we have that for each $s \geq 11$ and for each wire W that is exactly $2s$ levels away from the inputs

$$w \leq 1/40. \tag{23}$$

Claim 4 and (21) implies that for all wires W that are exactly 2×11 levels away from the inputs $\delta_w \leq 4^{11}(\hat{x} + 1/1000)^{11}\hat{\delta}$. Using Claim 4 and (23), we see that for all wires W that are $2(11+6)$ levels ways from the bottom we have $\delta_w \leq 4^{11}(\hat{x} + 1/1000)^{11}(4 \cdot \frac{1}{40})^6 \hat{\delta} \leq \hat{\delta}/50$, which we verify with a pocket calculator. Setting $D = 11 + 6 = 17$ we thus have $\delta_b \leq \max_i\{\delta_{a_i}\}/50$ and then with $\forall_{x \in [0,1]} : 2 \leq q(x) \leq 50$ we get our Lemma.

□

3.4 Blocks of Depth D

We now show that for every formula of depth $2D$ (with D from Lemma 2), one can set the noise rate of all gates to either 0 or β_2, such that the output wire at the root will have smaller extractable information than any of the input wires.

Lemma 3. *There is a constant $0 \leq \gamma < 1$ such that for any formula Q of depth $2D$ (with D as in Lemma 2) which is composed of gates with fan-in 2, has input wires $A_1, \ldots, A_{2^{2D}}$, output wire B and for every input distribution $\{a_i\}_i, \{\delta_{a_i}\}_i$ on the A_i it holds that either (a) if all gates have noise rate 0 or (b) if all gates have noise rate β_2, then*

$$|\delta_b| q(b) \leq \gamma \max_i |\delta_{a_i}| q(a_i). \tag{24}$$

The following fact will be helpful in the proof of the lemma. It defines an interval $[s,t]$ which is stable under the action of a noisy NOR-gate. It implies that if we have a tree of noisy NOR-gates, and all of its input wires are zero with a probability in $[\hat{x} - 2/1000, \hat{x} + 2/1000] \subseteq [s,t]$, then the output wire of the tree is zero with a probability in $[s,t]$.

Fact 1. *Consider any NOR gate with noise β_2, whose input wires are independently 0 with probabilities y_1 resp. y_2 and whose output wire is 0 with probability y_3. Define $s = \frac{250496-499\sqrt{7}}{250000\left(-1+\sqrt{7}\right)}$ and $t = \hat{x} + 2/1000$. It holds $s < \hat{x} - 2/1000$ and if $y_1, y_2 \in [s,t]$ then also $y_3 \in [s,t]$.*

Proof. We calculate $\min_{y_1,y_2\in[s,t]} y_3(y_1,y_2) = \min_{y_1,y_2\in[s,t]} 1-(1-2\beta_2)y_1y_2-\beta_2 = 1-(1-2\beta_2)t^2-\beta_2 = s$. Furthermore, with a pocket calculator we can check that $\max_{y_1,y_2\in[s,t]} y_3(y_1,y_2) = 1-(1-2\beta_2)s^2-\beta_2 < t$. □

The next claim will be useful in the proof of Lemma 3, because it implies that it is enough to find a single "bad" wire W in Q, for which $|\delta_w|q(w) \leq \theta' \max_i |\delta_{a_i}|q(a_i)$.

Claim 5. *For every $0 \leq \theta' < 1$, there is a $0 \leq \gamma < 1$ such that for every Q and every input distribution $\{a_i, \delta_{a_i}\}_i$ from the statement of Lemma 3 and when all gates have noise β_2 it holds: if there is one wire W in Q with $|\delta_w|q(w) \leq \theta' \max_i |\delta_{a_i}|q(a_i)$ then $|\delta_b|q(b) \leq \gamma \max_i |\delta_{a_i}|q(a_i)$.*

Proof. Let $depth(W)$ be the number of gates between W and the output wire of Q. We do induction on $depth(W) = 0, \ldots, 2D$, where our induction hypothesis is: The statement of Claim 5 is true if we restrict to wires W with $depth(W) \leq j$. Clearly, this is true for $j = 0$.

Assume that for some wire W at depth $j+1$ it holds that $|\delta_w|q(w) \leq \theta' \max_i |\delta_{a_i}|q(a_i)$. Let G be the gate that has W as an input wire. Let the other input wire of G be V and U its output wire. Note that by item 1 of Lemma 1 we have $|\delta_v|q(v) \leq \max_i |\delta_{a_i}|q(a_i)$. We show that

$$\exists_{0\leq\theta<1} : |\delta_u|q(u) \leq \theta \max_i |\delta_{a_i}|q(a_i). \tag{25}$$

If $|\delta_w|q(w) \leq |\delta_v|q(v)/2$, then item 3 of Lemma 1 implies (25). If $|\delta_w|q(w) > |\delta_v|q(v)/2$, then item 1 of Lemma 1 implies that $|\delta_u|q(u) \leq |\delta_w|q(w)$ and (25) holds with $\theta = \theta'$.

Note that $depth(U) = j$. Hence, using (25) and our induction hypothesis for depth j (and wire U) we see that the induction hypothesis is also true for the wire W at depth $j+1$. □

Proof. (Lemma 3) First we note that by repeatedly applying item 1 of Lemma 1 we have that for all wires X in Q

$$|\delta_x|q(x) \leq \max_i |\delta_{a_i}|q(a_i) \tag{26}$$

Hence, if one gate in Q is not of NOR-type then by setting all noise rates to β_2 there must be one wire W, for which

$$|\delta_w|q(w) \leq \theta \max_i |\delta_{a_i}|q(a_i), \tag{27}$$

where θ is given by item 2 in Lemma 1. Then our lemma follows by Claim 5.

Otherwise, all gates in Q are of NOR-type. We assume that Q is in canonical form: each gate G is a NOR gate with possibly additional NOT gates on its input wires but not on its output wire, because if G has an additional NOT on its output wire,

we can propagate this NOT into the gate whose input is the output of G. (If G is the output gate, we may just apply a NOT gate on its output wire, as this does not change $|\delta_b|q(b)$.) Similarly, we demand that all lowest gates G (who have wires A_i as input) *are* NOR gates (without additional NOT's on its input wire), because if G has an additional NOT on its input wire A_i, we can remove this NOT and set $a_i := 1 - a_i$ and $\delta_{a_i} := -\delta_{a_i}$. This will not change the output of G.

Now, assume that

$$\forall_i : |a_i - \hat{x}| \leq 1/1000 \wedge |\delta_{a_i}| \leq 1/1000 \tag{28}$$

was not true, let us say because of wire A_i. Let G be the gate with input wire A_i. Let its output wire be W. Then by item 4 of Lemma 1 inequality (27) follows for $0 \leq \theta < 1$ given in item 4 of Lemma 1 if all gates have noise rate β_2. Then Claim 5 implies our Lemma. Hence, for the remainder we may assume (28) is true and that Q is in canonical form.

If all gates in Q are NOR gates, then our lemma follows by Lemma 2 when setting all noise rates to 0. Otherwise, there must be at least one gate G which is a NOR gate, but has at least one additional (noise-free) NOT gate on one of its input wires, called it D. Let W be the output wire of G. We break up G into a NOR gate G' and the additional NOT gates on its input wires. Let D' be the output of the NOT gate with input D. Note that below D the formula is a complete binary tree of NOR gates. For any of its input wires A_j it holds by (28) that $\Pr[A_j = 0|x_i = 0] \in [s, t]$, where s, t are as in Fact 1. Hence, applying Fact 1 repeatedly, we see that also $\Pr[D = 0|x_i = 0] \in [s, t]$. Analogously, $\Pr[D = 0|x_i = 1] \in [s, t]$, and hence $d \in [s, t]$ and then $d' = 1 - d \leq 1 - s < 4/10$. Therefore, $|d' - \hat{x}| > 1/1000$, which satisfies the conditions of item 4b in Lemma 1. This also implies (27) for a θ given in item 4b in Lemma 1. Again, Claim 5 then implies our Lemma. □

Lemma 4. *There is a constant $0 \leq \alpha_2 < \beta_2$ such that the statement of Lemma 3 also holds for some $0 \leq \gamma < 1$ if we replace β_2 by α_2.*

Proof. For each formula Q the outputs δ_b and $q(b)$ are polynomials in the noise rate ϵ (and δ_{a_i}, a_i). The derivatives of these polynomials with respect to ϵ are bounded (for valid inputs), as is the derivative of q. Hence, also $\partial \delta_b q(b)/\partial \epsilon$ is bounded, and then the claim follows from Lemma 3. □

3.5 Proof of main theorem

Proof of Theorem 1. Let D be given by Lemma 2. Let $\alpha_2 < \beta_2$ and $0 \leq \gamma < 1$ be as in Lemma 4. Choose $l \in \mathbb{N}$ such that $50\gamma^l \leq \Delta$. Choose a function f which depends on at least 2^{2lD} many input bits. Almost all functions have this property.

Let F be any formula "attempting" to compute f fault-tolerantly. For each wire W in F let $depth(W)$ be the number of gates between W and the output wire of F and let the depth of a gate be the depth of its output wire. Note that there is at least one input variable x_i which f depends on such that all input wires of F carrying x_i have depth at least $2lD$. Fix all other input bits of f such that flipping x_i flips the output.

W.l.o.g. we may assume that up to depth $2lD$ F is a full binary tree, i.e. all gates up to depth l are gates of fan-in exactly 2.[4]

We argue inductively that for all $j = l, \ldots, 0$ it is possible to choose individual noise rates in $\{0, \alpha_2\}$ for all gates with depth between $2jD$ and $2(j-1)D$, such that for wires W with $depth(w) = 2jD$: $|\delta_w| q(w) \leq 50\gamma^{l-j}$. This follows by noting that for $j = l$ we have $|\delta_w| \leq 1$ and then $|\delta_w| q(w) \leq \max_{y \in [0,1]} q(y) \leq 50$. For the inductive step we repeatedly use Lemma 4. Hence, for the output wire O of F we have $|\delta_o| q(o) \leq 50\gamma^l \leq \Delta$. And since $\min_{y \in [0,1]} q(y) = q(1/2) > 1/2$ we have $|\delta_o| < \Delta/2$. □

3.6 Remarks

There is nothing special about the choice $\epsilon_G \in \{0, \alpha_2\}$ for the adversary in the statement of Theorem 1. It is also true that for any $0 \leq \kappa \leq 1/2$ there is an $\alpha < \beta_2$ such that an adversary who can only choose noise rates in $\{\kappa, \alpha\}$ can also prevent fault-tolerant computation, although we do not show this explicitly.

Note that our adversary is actually quite simple, since he only looks at small blocks of gates and then sets all noise rates of the gates in this block to the same value 0 or α_2. Furthermore, the adversary is fault-tolerant himself, i.e. he does not need to get exactly the noise rates he wants. It is enough that whenever the adversary chooses a noise rate ϵ for a gate, that the actual noise rate of that gate is in $[\epsilon, \epsilon + \chi]$, for some small enough $\chi > 0$. This follows by a simple continuity argument similar to the one in the proof of Lemma 4.

Also, it would be interesting to find the smallest α_2 which satisfies Theorem 1. We have proven that $\alpha_2 < \beta_2$, giving a proof of principle. However, it would be practically interesting to see how far the smallest possible α_2 is away from β_2.

So far we have not given any idea of why we chose this particular potential function. In fact, this choice is not unique. In [32] another, but more complicated potential function was used. In the end of [32] there is some explanation of how q must be chosen and there q was basically engineered to satisfy these criteria. Our current potential function also satisfies these criteria, but is much nicer and might suggest a satisfactory interpretation in standard information-theoretic terms. It would be interesting to find such an interpretation.

Acknowledgements

I thank Virginia Vassilevska Williams for suggesting the trick of replacing variables as $x = ab$, $y = a+b$ in the proofs of Claims 1 and 2, which simplifies the analysis. Further, I thank Piyush Srivastava, Thomas Vidick and Ronald de Wolf for useful comments.

[4] If this is not the case for some gate, we can add another input wire to this gate, whose value is ignored when computing the output, and below this new input wire we add a full binary tree of NOR gates of sufficient depth, whose input wires are all constant 0.

References

[1] D. Aharonov and M. Ben-Or. Fault tolerant quantum computation with constant error. In *Proceedings of 29th ACM STOC*, pages 176–188, 1997. quant-ph/9611025.

[2] P. Aliferis, D. Gottesman, and J. Preskill. Quantum accuracy threshold for concatenated distance-3 codes. 2005. quant-ph/0504218.

[3] H. Buhrman, R. Cleve, M. Laurent, N. Linden, A. Schrijver, and F. Unger. New limits on fault-tolerant quantum computation. In *Proceedings of 47th IEEE FOCS*, pages 411–419, 2006. quant-ph/0604141.

[4] S. E. Decatur and R. Gennaro. On learning from noisy and incomplete examples. In *COLT '95: Proceedings of the eighth annual conference on Computational learning theory*, pages 353–360, 1995.

[5] R. L. Dobrushin and S. I. Ortyukov. Upper bound on the redundancy of self-correcting arrangements of unreliable functional elements. *Problems Inform. Transmission*, 13:203–218, 1977.

[6] W. Evans and N. Pippenger. On the maximum tolerable noise for reliable computation by formulas. *IEEE Trans. Inform. Theory*, 44(3):1299–1305, 1998.

[7] W. Evans and N. Pippenger. Average-case lower bounds for noisy boolean decision trees. *SIAM J. Comput.*, 28(2):433–446, 1999.

[8] W. Evans and L. Schulman. Signal propagation and noisy circuits. *IEEE Trans. Inform. Theory*, 45(7):2367–2373, 1999.

[9] W. Evans and L. Schulman. On the maximum tolerable noise of k-input gates for reliable computation by formulas. *IEEE Trans. Inform. Theory*, 49(11):3094–3098, 2003.

[10] U. Feige, P. Raghavan, D. Peleg, and E. Upfal. Computing with noisy information. *SIAM J. Comput.*, 23(5):1001–1018, 1994.

[11] P. Gács and A. Gál. Lower bounds for the complexity of reliable boolean circuits with noisy gates. *IEEE Transactions on Information Theory*, 40:579 – 583, 1994.

[12] J. Gao, Y. Qi, and J. Fortes. Bifurcations and fundamental error bounds for fault-tolerant computations. *Nanotechnology, IEEE Transactions on*, 4(4):395 – 402, 2005.

[13] S. A. Goldman and R. H. Sloan. Can pac learning algorithms tolerate random attribute noise? *Algorithmica*, 14:70–84, 1995.

[14] B. Hajek and T. Weller. On the maximum tolerable noise for reliable computation by formulas. *IEEE Trans. Inform. Theory*, 37(2):388–391, 1991.

[15] J. Han, J. Gao, P. Jonker, Y. Qi, and J. Fortes. Toward hardware-redundant, fault-tolerant logic for nanoelectronics. *Design Test of Computers, IEEE*, 22(4):328 – 339, 2005.

[16] J. Kempe, O. Regev, F. Unger, and R. de Wolf. Upper bounds on the noise threshold for fault-tolerant quantum computing. *Quantum Information and Computation*, 10:0361–0376, 2010.

[17] E. Knill, R. Laflamme, and W. H. Zurek. Resilient Quantum Computation. *Science*, 279(5349):342–345, 1998.

[18] M. Knill. Quantum computing with realistically noisy devices. *Nature*, 434:39–44, 2005.

[19] J. Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, volume 3, pages 43–99. Princeton University Press, Princeton, 1956.

[20] I. Newman. Computing in fault tolerant broadcast networks and noisy decision trees. *Random Struct. Algorithms*, 34(4):478–501, 2009.

[21] N. Nisan and A. Wigderson. Hardness vs. randomness. *Jounal of Computer Systems and Sciences*, 49:149–167, 2 1994.

[22] N. Pippenger. On networks of noisy gates. In *26th Annual Symposium on Foundations of Computer Science*, pages 30–38, 1985.

[23] N. Pippenger. Invariance of complexity measures for networks with unreliable gates. *J. ACM*, 36(3):531–539, 1989.

[24] S. Rajagopalan and L. Schulman. A coding theorem for distributed computation. In *STOC '94: Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 790–799, New York, NY, USA, 1994. ACM.

[25] A. Razborov. An upper bound on the threshold quantum decoherence rate. *Quantum Information and Computation*, 4(3):222–228, 2004. quant-ph/0310136.

[26] B. Reichardt. Quantum universality from Magic States Distillation applied to CSS codes. *Quantum Information Processing*, 4:251–264, 2005.

[27] B. Reichardt. Error-detection-based quantum fault-tolerance threshold. *Algorithmica*, 55:517–556, 2009.

[28] R. Reischuk and B. Schmeltz. Reliable computation with noisy circuits and decision trees-a general n log n lower bound. *Foundations of Computer Science, Annual IEEE Symposium on*, pages 602–611, 1991.

[29] L. J. Schulman. Coding for interactive communication. *IEEE Transactions on Information Theory*, 42(6):1745–1756, 1996.

[30] G. Shackelford and D. Volper. Learning k-dnf with noise in the attributes. In *COLT '88: Proceedings of the first annual workshop on Computational learning theory*, pages 97–103, 1988.

[31] R. Sloan. Types of noise in data for concept learning. In *COLT '88: Proceedings of the first annual workshop on Computational learning theory*, pages 91–96, 1988.

[32] F. Unger. Noise threshold for universality of two-input gates. *IEEE Transactions on Information Theory*, 54(8):3693–3698, 2008.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de