

# Input Locality and Hardness Amplification

Andrej Bogdanov\*      Alon Rosen†

## Abstract

We establish new hardness amplification results for one-way functions in which each input bit influences only a small number of output bits (a.k.a. input-local functions). Our transformations differ from previous ones in that they approximately preserve input locality and at the same time retain the input size of the original function.

Let  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  be a one-way function with input locality  $d$ , and suppose that  $f$  cannot be inverted in time  $\exp(\tilde{O}(\sqrt{n} \cdot d))$  on an  $\varepsilon$ -fraction of inputs. Our main results can be summarized as follows:

- If  $f$  is injective then it is equally hard to invert  $f$  on a  $(1 - \varepsilon)$ -fraction of inputs.
- If  $f$  is regular then there is a function  $g: \{0, 1\}^n \rightarrow \{0, 1\}^{m+O(n)}$  that is  $d + O(\log^3 n)$  input local and is equally hard to invert on a  $(1 - \varepsilon)$ -fraction of inputs.

A natural candidate for a function with small input locality and for which no sub-exponential time attacks are known is Goldreich’s one-way function. To make our results applicable to this function, we prove that when its input locality is set to be  $d = O(\log n)$  certain variants of the function are (almost) regular with high probability.

In some cases, our techniques are applicable even when the input locality is not small. We demonstrate this by extending our first main result to one-way functions of the “parity with noise” type.

**keywords:** one-way function, input locality, hardness amplification, parity with noise

---

\*Department of Computer Science and Engineering and Institute for Theoretical Computer Science and Communications, Chinese University of Hong Kong. E-Mail: [andrejb@cse.cuhk.edu.hk](mailto:andrejb@cse.cuhk.edu.hk). Research supported by RGC GRF grant 2150617.

†Efi Arazi School of Computer Science, IDC Herzliya. E-mail: [alon.rosen@idc.ac.il](mailto:alon.rosen@idc.ac.il). Research supported by BSF grant 2006317 and ISF grant 334/08.

# 1 Introduction

In this paper we are interested in amplifying the hardness of inverting a one-way function. Our goal is to do so without significantly deteriorating the function’s parallel complexity and/or efficiency. To the best of our knowledge, these objectives are not simultaneously achieved by any of the previous methods for amplifying hardness.

Our results assume the function is regular, and sub-exponentially hard to invert. They crucially rely on it being *input-local*, meaning that each input bit affects only a small number of output bits. Under these assumptions we show how to amplify hardness while preserving the function’s input length and input locality. In some cases we achieve this without modifying the function altogether.

## 1.1 Hardness Amplification

The problem of hardness amplification can be described as follows: given a one-way function  $f(x)$ , construct a function,  $g(y)$ , so that if  $f(x)$  is hard to invert on an  $\varepsilon$  fraction of inputs, then  $g(y)$  is hard to invert on some  $1 - \delta > \varepsilon$  fraction of inputs. Amplification of hardness is established by exhibiting a reduction from the task of inverting  $f$  to the task of inverting  $g$ . The overall quality of the amplification is determined by: (1) the complexity of the construction (in particular, the relationship between  $|x|$  and  $|y|$ ), (2) the complexity of the reduction, and (3) the exact asymptotic relationship between  $\varepsilon$  and  $1 - \delta$ .

The most basic method for amplifying hardness is due to Yao [16]. It consists of independently evaluating the function  $f(x)$  many times in parallel. Using this transformation, it is essentially possible to obtain an arbitrary level of amplification. However, this comes at the cost of significantly blowing up the input size. For instance, if we wish to amplify from error  $\varepsilon > 0$  to error  $1 - \delta > \varepsilon$ , evaluating  $g(y)$  will involve applying  $f(x)$  to  $O((1/\varepsilon) \log(1/\delta))$  small pieces of  $y$ , each of size  $|x|$  (resulting in  $|y| = O(|x| \cdot (1/\varepsilon) \log(1/\delta))$ ).

A better tradeoff between security and efficiency is achieved by Goldreich et al (GILVZ), for the special case of regular one-way functions [9]. In their construction, the evaluation of  $g(y)$  consists of repeatedly applying  $f$  in sequence, where every two successive applications are interleaved with a randomly chosen step on an expander graph. The starting point of  $g$ ’s evaluation is an input  $x$  to  $f$ , and intermediate steps on the graph are determined by an auxiliary random string whose total length is  $O((1/\varepsilon) \log(1/\delta))$ . This results in  $|y| = |x| + O((1/\varepsilon) \log(1/\delta))$ , but renders the evaluation of  $g(y)$  inherently sequential.

A related transformation was analyzed by Haitner et al (HHR), also for the case of regular functions [11, 10]. Their transformation sequentially iterates the function with intermediate applications of a hash function, and has the advantage of not requiring knowledge of the regularity of  $f$ . Similarly to the GILVZ transformation, it is sequential in nature.

One last category of amplification results relies on *random self-reducibility*. It applies to functions that allow an efficient mapping from  $f(x)$  to  $f(y)$ , where  $y$  is a random value from which one can efficiently retrieve  $x$ . When satisfied, random self-reducibility enables very simple worst-case to average-case hardness amplification, without having to modify the original function. However, it is generally not known to be satisfied by one-way functions.

## 1.2 Highly Parallelizable One-Way Functions

Applebaum, Ishai and Kushilevitz (AIK) give strong evidence for the existence of one-way functions that can be evaluated in as little as constant parallel time. They first present one-way functions with constant *output locality*, meaning that each output bit depends on at most a constant number of input bits [3]. These functions are constructed using *randomized encodings*, a tool that allows them to transform well known candidate one-way functions that have low (but not necessarily constant) parallel complexity into ones with constant output locality. They then go on and show that, in some specific cases, the functions resulting from their randomized encodings also satisfy constant input locality [4].

An alternative source for candidate one-way functions with small input and output locality is given by Goldreich [8]. These candidates are arguably more natural than the ones resulting from the AIK transformations. They also seem to offer a more attractive tradeoff between input length and security (as in many cases randomized encodings necessitate a significant blow up in the input size of the original function). Goldreich’s constructions are quite general, and allow flexibility in the choice of the function, both in terms of the way in which inputs are connected to outputs, as well as in the choice of the predicates used to compute the function’s output bits. To date, no sub-exponential time inversion algorithm is known for any variant of his functions.

Known hardness amplification methods are not well suited for functions of the above sort. Being inherently sequential, the GILVZ and HHR transformations do not preserve parallelism. Yao’s transformation, on the other hand, does not increase parallel time, but it does incur a significant loss in efficiency (cf. Lin et al. [15]). This presents us with the challenge of coming up with efficient hardness amplification methods that are well suited for parallelizable functions. Our approach to the problem will be to utilize properties implied by the highly parallel structure of the function, and specifically small input-locality.

## 1.3 Main Results

Let  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  be a one-way function with input locality  $d$ , and suppose that  $f$  cannot be inverted in time  $\exp(\tilde{O}(\sqrt{n} \cdot d))$  on an  $\varepsilon$ -fraction of inputs. Our first main result falls into the category of *self-amplification*, meaning that the hardness amplification does not require modifying the underlying function.

**Theorem 3.1 (Self-amplification for injective functions):** *Suppose that  $f$  is injective. Then,  $f$  cannot be inverted in time  $\exp(\tilde{O}(\sqrt{n} \cdot d))$  on a  $(1 - \varepsilon)$ -fraction of inputs.*

Based on the ideas used in the proof of Theorem 3.1, we prove an analogous theorem for functions of the “parity with noise” type. Specifically, consider a family,  $\{M_n\}$ , of  $m(n) \times n$  matrices with entries in  $\{0, 1\}$  and let  $p \in [0, 1]$  be a parameter. Define a function family  $f_n: \{0, 1\}^n \rightarrow \{0, 1\}^m$  as  $f_n(x, e) = M_n x + e \pmod{2}$ , where  $x$  is a vector chosen uniformly at random from  $\{0, 1\}^n$ , and  $e \in \{0, 1\}^m$  is a vector of hamming weight at most  $2pm$  chosen from the following distribution: Each entry of  $e$  is chosen independently from a  $p$ -biased distribution, conditioned on  $e$  having hamming weight at most  $2pm$ .

We assume that  $\{f_n\}$  is one-way against randomized time  $\exp(\tilde{O}(\sqrt{m}))$  on some  $\varepsilon$  fraction of inputs. We also require that the functions  $f_n$  are 1-1. This happens when  $M_n$  is a generator matrix of a code of minimum distance  $4pm$ . In such a case, the input locality of  $f_n$  will be as large as  $\Omega(n)$ . Nevertheless, we can prove the following analogue of Theorem 3.1.

**Theorem 4.1 (Self-amplification for parity with noise):** *Suppose that  $\{f_n\}$  is injective. Then, (under appropriate constraints on parameters)  $\{f_n\}$  cannot be inverted in randomized time  $\exp(\tilde{O}(\sqrt{m}))$  on a  $(1 - \varepsilon)$ -fraction of inputs.*

To make our results applicable to a wider class of functions, we also consider a generalization of Theorem 3.1 to the case where the function we wish to amplify is regular (every output has the same number of preimages). As before, we assume that the function  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  has input locality  $d$ , and that  $f$  cannot be inverted in time  $\exp(\tilde{O}(\sqrt{n} \cdot d))$  on an  $\varepsilon$ -fraction of inputs. This time, however, we are not able to prove self-amplification and settle for some increase in output length and input locality, while still preserving input length.

**Theorem 5.1 (Amplification for regular functions):** *Suppose that  $f$  is regular. Then, there is a function  $g: \{0, 1\}^n \rightarrow \{0, 1\}^{m+O(n)}$  that is  $d + O(\log^3 n)$  input local and that cannot be inverted in time  $\exp(\tilde{O}(\sqrt{n} \cdot d))$  on a  $(1 - \varepsilon)$ -fraction of inputs.*

A natural candidate for a function with small input locality and for which no sub-exponential time attacks are known is Goldreich’s one-way function [8]. Given a bipartite graph  $G$  with  $n$  vertices on the left,  $m$  vertices on the right, and regular right-degree  $d_{\text{out}}$  and given a predicate  $P: \{0, 1\}^{d_{\text{out}}} \rightarrow \{0, 1\}$ , the function  $f_{G,P}: \{0, 1\}^n \rightarrow \{0, 1\}^m$  is defined by setting the  $i^{\text{th}}$  bit of  $f_{G,P}(x)$  to be equal to  $P(x_{\Gamma(i,1)}, \dots, x_{\Gamma(i,d_{\text{out}})})$ , where  $\Gamma(i,j)$  is the  $j^{\text{th}}$  neighbor of right vertex  $i$  of  $G$ . Goldreich proposed setting  $m = n$  and considered  $d_{\text{out}}$  ranging from a constant to  $O(\log n)$ . He conjectured that when  $G$  is a good expander graph and  $P$  is randomly chosen, with high probability  $f_{G,P}$  is one-way when  $n$  is sufficiently large.

We consider instantiations of Goldreich’s functions with a certain class of balanced predicates, which we call  $d_{\text{out}}$ -parity-blowup predicates, and assume that  $G$  is chosen at random. Relying on Theorem 5.1 we can prove the following.

**Theorem 6.3 (Amplification for Goldreich’s function):** *Suppose that for at least half the graphs  $G$ , the function  $f_{G,P}$  is hard to invert on an  $\varepsilon$ -fraction of inputs for circuits of size  $\exp(\tilde{O}(\sqrt{n}))$ . Then there exists a function  $g: \{0, 1\}^n \rightarrow \{0, 1\}^{n+O(\log(1/\varepsilon))}$  of circuit size  $O(n \log n)$  that is hard to invert on a  $(1 - \varepsilon)$ -fraction of inputs by circuits of size  $\exp(\tilde{O}(\sqrt{n}))$ .*

By observing that parity-blowup predicates can be represented by constant degree polynomials over  $GF(2)$  we can apply the randomized encodings of AIK [3], and obtain a function with constant output locality and slightly longer input and output length.

Finally, we state a result that applies in the setting where  $d_{\text{out}}$  is constant and  $m \geq Dn$ , where  $D = D(d_{\text{out}})$  is a sufficiently large constant. Invoking a recent result of Bogdanov and Qiao [6], we prove that for any  $P$  and with high probability over the choice of  $G$  if  $f_{G,P}$  is hard to invert on an  $\varepsilon$  fraction of inputs in time  $\exp(\tilde{O}(\sqrt{n}))$ , then  $f_{G,P}$  is hard to invert on a  $1 - \varepsilon$  fraction of inputs in time  $\exp(\tilde{O}(\sqrt{n}))$ .

## 1.4 Applicability

Generally speaking, our results are not applicable to functions that are obtained via the randomized encodings of AIK. This is because these encodings typically incur at least a quadratic blow up in the input size. Thus, even if the original function is exponentially hard to invert, we cannot hope to prove that the resulting function is more than  $\exp(O(\sqrt{n}))$  hard to invert (at least not based on the hardness of the original function).

It is conceivable that in some specific cases the randomized encodings can be performed in a way that does not significantly increase the input length of the original function. However, even if such cases exist, we are currently not aware of any natural candidate one-way function that would potentially satisfy Theorem 3.1's hypothesis. While AIK give several injective functions with constant output locality, none of these seems to have small input locality, and moreover they are all known to be invertible in time less than  $\exp(\tilde{O}(\sqrt{n}))$  (e.g., ones that are based on the hardness of factoring and of finding discrete-logarithms). Other, presumably harder to invert, candidates are not known to be injective (though they may be regular, making Theorem 5.1 applicable).

Nevertheless, we feel that Theorem 3.1 is worth stating and proving. First of all, the fact that we could not think of any appropriate example does not mean that such does not exist. Secondly, the proof of the theorem contains the core ideas behind our reductions, and gives us the opportunity to present them without any irrelevant complications. Finally, and most importantly, using the main ideas of the theorem, we are able to prove an analogous result for functions of the "parity with noise" type, which are generally not known to be invertible in less than  $\exp(O(n/\log n))$  time [5].

As we mentioned above, there is no known sub-exponential time algorithm that succeeds in inverting Goldreich's function on a non-negligible fraction of inputs. Applebaum, Barak, and Wigderson [2] prove that, when based on  $d$ -parity blowup predicates, the output of Goldreich's function is pseudorandom against linear functions, low-degree polynomials, and constant-depth circuits. In light of this, it currently seems reasonable to conjecture that no algorithm can invert such variants of the function on a small  $\varepsilon = \varepsilon(n)$  fraction of inputs in time  $\exp(\tilde{O}(\sqrt{n} \cdot d))$ . Under this assumption, we obtain a function with poly-logarithmic input locality and constant output locality that cannot be inverted by algorithms with comparable running time on a significantly larger,  $(1 - \varepsilon)$ , fraction of inputs.

Even though not stated explicitly in Section 1.3, our reductions offer a concrete tradeoff between the running time of the reduction and the error  $\varepsilon = \varepsilon(n)$  we are able to amplify from. The actual overhead incurred by the reduction is  $\exp(O(\sqrt{n} \cdot \log(1/\varepsilon) \cdot d \cdot \log n))$ . Thus, assuming that the original function is hard to invert in roughly this time, we can amplify starting from errors as small as say  $\varepsilon(n) = 2^{-n^{O(1)}}$ . Note that previous amplification methods are not applicable for such ranges of parameters, even if we assume sub-exponential hardness. This is because the input lengths of the functions resulting from their transformations grows proportionally to  $\tilde{O}(1/\varepsilon)$ .

## 1.5 Ideas and Techniques

Our self-amplification result is based on the following simple idea. Suppose  $f$  is a 1-1 function with input locality  $d$  and  $x$  and  $x'$  are two inputs that differ in one coordinate. Suppose we can invert  $f(x)$ . Then with a little bit more work we can invert  $f(x')$ : By input locality,  $f(x)$  and  $f(x')$  can differ in at most  $d$  coordinates. We change  $d$  coordinates of  $f(x')$  until we find  $f(x)$ , recover  $x$ , and change  $x$  in one coordinate to recover  $x'$ .

By repeating this argument  $r$  times, we can invert  $f(x')$  where  $x$  and  $x'$  are within distance  $r$  using  $O(n^{dr})$  invocations to the original inverter. So if we can invert  $f$  at  $x$ , we can also invert  $f$  at any  $x'$  within distance  $r$  of  $x$ . Therefore, assuming  $f$  is easy to invert on some set that covers an  $\varepsilon$ -fraction of  $\{0, 1\}^n$ , we can also invert  $f$  at any input within distance  $r$  of this set. By setting  $r = O(\sqrt{n})$ , we obtain Theorem 3.1, the self-amplification result for 1-1 functions.

**Amplifying regular functions.** The assumption that  $f$  is 1-1 is important in this argument. If  $f$  was not 1-1, the inverter could return some other preimage which is very far from  $x$  and therefore also far from  $x'$ . In Theorem 5.1 we show that if the function  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  is not 1-1 but regular (i.e.  $K$ -to-1 for some  $K$ ), then there exists a new function  $f': \{0, 1\}^n \rightarrow \{0, 1\}^{m'}$ ,  $m' = m + O(n)$  such that if  $f$  is hard on an small fraction of inputs, then  $f'$  is hard on almost all of its inputs.

The transformation from  $f$  to  $f'$  effectively isolates inputs by applying an appropriate hash function. Hashing is a standard way to reduce a regular function to a 1-1 function [13, 12]. However, applying a pairwise-independent hash increases input locality by  $\Omega(\log K)$  (see Section 5.1) and makes Theorem 3.1 inapplicable when  $K$  is large. In Claim 1 we describe a new construction of a hash function which increases input locality only by  $O((\log n)^3)$  and maps most preimages of  $f$  to unique values. Combining this hash with Theorem 3.1, we obtain Theorem 5.1, our amplification result for regular input-local functions.

**Parity with noise.** In Section 4 we apply our ideas to show self-amplification for functions of the parity with noise type. Although these functions do not have low-input locality, we are able to apply our techniques. The reason is that these functions consists of two parts: A linear component, which is randomly self reducible, and the noise component, which is input-local. By combining an application of Theorem 3.1 to the noise component with a random self-reduction on the linear component, we prove Theorem 4.1.

**Goldreich's function.** As we explain in Section 6, Goldreich's function is unlikely to be 1-1 (except in special cases which are easy to invert), so Theorem 3.1 does not apply directly. However, we show that when  $m/n$  is a sufficiently large constant, if  $f(x_1) = f(x_2)$ , then  $x_1$  and  $x_2$  must be substantially correlated. Assuming  $f$  can be inverted on an  $\varepsilon$ -fraction of inputs, using our self-reduction from Theorem 3.1, for most  $x'$  we can invert  $f(x)$  at some  $x$  that is close to  $x'$ . The inverse we obtain may not be equal to  $x$ , but it will be substantially correlated to  $x'$ . Using a result of Bogdanov and Qiao [6], we then recover an inverse for  $f(x')$ .

Our second application concerns functions  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  where  $m = n$ , but the degree is  $O(\log n)$  and the predicate that  $f$  is based on is a “parity blowup” predicate (see Section 6). First we prove that such functions are likely to be at most  $K$ -to-1 for some

constant  $K$ . Using the hash from Claim 1, we obtain a new function  $f'$  that is almost 1-1 and is almost as hard to invert. Finally, using the randomized encodings of Applebaum et al. [3], we can transform  $f'$  into a function with constant *output* locality at a polylogarithmic cost in the input and output length.

## 1.6 Open Questions

We believe it is interesting to investigate if our methods apply to a wider class of candidate one-way functions. In Section 6 we show that our amplification methods apply to variants of Goldreich's function where either (1) the degree is constant but the output to input length ratio is sufficiently large, or (2) the function is length-preserving, but the degree is logarithmic (so the function is not output-local) and the predicate is of a special form.

It would be interesting to investigate the range of parameters where the function is length-preserving and the degree is constant. We conjecture that when the predicate is balanced, such functions are “almost  $2^{cn}$ -to-1” for some constant  $c$ , in the sense that for most  $x$ ,  $f(x)$  has  $2^{cn \pm o(n)}$  preimages. If this was the case, we could apply Theorem 5.1 (and Corollary 5.2) to obtain very hard to invert functions with better locality parameters.

## 2 Definitions

Let  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  be a function. We say that the  $i$ th output  $f(x)_i$  *depends* on the  $j$ th input  $x_j$  if there exists a setting of the inputs  $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$  such that  $f(x_1, \dots, x_{j-1}, 0, x_{j+1}, \dots, x_n)_i \neq f(x_1, \dots, x_{j-1}, 1, x_{j+1}, \dots, x_n)_i$ . We define the *degree* of the  $j$ th input to be the number of outputs that depend on the  $j$ th input. We say  $f$  has *input locality*  $d$  if the degree of every input is at most  $d$ . We define the *degree* of an output as the number of inputs it depends on and the *output locality* as the maximum degree of an output.

We say that  $f$  is  $K$ -*to-1* if for every  $x \in \{0, 1\}^n$ , there exist exactly  $K$  inputs  $x' \in \{0, 1\}^n$  such that  $f(x') = f(x)$ . We say  $f$  is *regular* if it is  $K$ -to-1 for some  $K$ . We say  $f$  is *at most*  $K$ -*to-1* (resp., *at least*  $K$ -*to-1*) if for every  $x$  there are at most  $K$  (resp., at least  $K$ )  $x'$  such that  $f(x') = f(x)$ . We say  $f$  is  $\varepsilon$ -*close to*  $K$ -*to-1* if for at least a  $(1 - \varepsilon)$  fraction of the inputs  $x \in \{0, 1\}^n$ , there exist exactly  $K$  inputs  $x' \in \{0, 1\}^n$  such that  $f(x') = f(x)$ .

In this work we consider both uniform and non-uniform constructions of one-way functions. The security of such functions can be defined against deterministic, randomized, and non-uniform inverters. We do not attempt to state our results in the most general setting. Instead, we use the definition that is most natural for the proof, in order to avoid distracting technical issues. For our purposes, it will be sufficient to define non-uniform one-way functions against non-uniform adversaries and uniform one-way functions against uniform (possibly randomized) adversaries.

In the non-uniform setting, we say  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  is *hard against circuits of size*  $s$  *on an*  $\alpha$ -*fraction of inputs* if for every circuit  $C$  of size at most  $s$ ,  $f(C(f(x))) = f(x)$  for at most  $\alpha \cdot 2^n$  inputs  $x \in \{0, 1\}^n$ .

In the uniform setting, a function family  $f = \{f_n: \{0, 1\}^n \rightarrow \{0, 1\}^{m(n)}\}$  is *one-way against (randomized) time  $t(n)$  on an  $\alpha(n)$ -fraction of inputs* if (1)  $f$  is computable in deterministic polynomial time and (2) for every (randomized) algorithm  $A$  that runs in time  $t(n)$  and every sufficiently large  $n$ ,  $f_n(A(1^n, f_n(x))) = f_n(x)$  for at most an  $\alpha(n) \cdot 2^n$  fraction of inputs  $x \in \{0, 1\}^n$  (and with probability at most  $1/2$  over the coin tosses of  $A$ ). (To simplify notation, we will omit the length parameter  $1^n$  as an input to the inverter in our proofs.)

### 3 Self-amplification for 1-1 functions

Let  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  be any 1-1 function, and let  $d_j$  be the degree of the  $j$ th input. Set  $\Delta = \sum_{j=1}^n d_j^2$ .

**Theorem 3.1.** *Let  $f = \{f_n: \{0, 1\}^n \rightarrow \{0, 1\}^{m(n)}\}$  be a 1-1 function family. Suppose  $f$  is one-way against time  $\exp(O(\sqrt{r\Delta} \log n))$  on a  $e^{-r}$ -fraction of inputs ( $r = r(n)$ ). Then  $f$  is one-way against time  $\exp(O(\sqrt{r\Delta} \log n))$  on a  $(1 - e^{-r})$ -fraction of inputs.<sup>1</sup>*

When  $f$  is a 1-1 function with input locality  $d$ , we get that if  $f$  is one-way against time  $\exp(O(\sqrt{rn} \cdot d \log n))$  for a  $e^{-r}$ -fraction of inputs, then the same function is also one-way for a  $(1 - e^{-r})$ -fraction of inputs.

The proof is based on the following idea. For simplicity let us consider the case where the degree of every input is at most  $d$ . Assume that  $f$  can be inverted in time  $\exp(O(\sqrt{r\Delta} \log n))$  on an  $e^{-r}$ -fraction of inputs and let  $S'$  be the set of inputs on which this inversion algorithm succeeds. Let us consider all inputs  $x$  that are within hamming distance  $\sqrt{2rn}$  from  $S'$ . By a standard probabilistic argument (Lemma 3.2, based on Theorem 7.5.3 in [1]) it follows that at least  $1 - e^{-r}$  fraction of inputs  $x$  have this property. Now if  $x$  and  $x' \in S'$  differ in at most  $\sqrt{2rn}$  coordinates, then  $y = f(x)$  and  $y' = f(x')$  will differ in at most  $\sqrt{2rnd}$  coordinates. Therefore we can invert  $f$  at  $y = f(x)$  by flipping the given set of  $\sqrt{2rnd}$  coordinates on which  $y$  and  $y'$  differ, inverting  $f$  at  $y'$  to obtain  $x'$ , and then moving back from  $x'$  to  $x$  by changing at most  $\sqrt{2rn}$  coordinates.

We first state and prove the probabilistic inequality which is the technical heart of our argument. We prove the inequality in slightly more general form for later applications.

**Lemma 3.2.** *Consider the space  $\{0, 1\}^n$  with the  $p$ -biased distribution (i.e., each coordinate takes value 1 independently at random with probability  $p$ ) for some  $p \in [0, 1]$ . Let  $X \subseteq \{0, 1\}^n$  be any set of measure  $e^{-r}$  and let  $d_1, \dots, d_n$  be positive numbers. Let*

$$Z = \{z: \sum_{j \in [n]: x_j \neq z_j} d_j \leq \sqrt{2r\Delta} \text{ for some } x \text{ in } X\}.$$

where  $\Delta = \sum_{i=1}^n d_i^2$ . Then  $Z$  has measure at least  $1 - e^{-r}$ .

---

<sup>1</sup>Usually, hardness amplification results are stated in terms of two parameters, the initial hardness  $\varepsilon$  and the “derived hardness”  $(1 - \delta)$ . Since the complexity of our inverter is dictated by the minimum of  $\varepsilon$  and  $\delta$ , without loss of generality we state our results for the special case  $\varepsilon = \delta = e^{-r}$ .

*Proof.* Define  $d(z) = \min_{x \in S} \sum_{j \in [n]: x_j \neq z_j} d_j$ . Then any change in  $z_j$  changes  $d(z)$  by at most  $d_j$ . By Azuma's inequality, we have

$$\Pr[d(z) \leq \mathbb{E}[d(z)] - t] < e^{-2t^2/\Delta} \quad \text{and} \quad \Pr[d(z) \geq \mathbb{E}[d(z)] + t] < e^{-2t^2/\Delta}$$

Setting  $t = \mathbb{E}[d(z)]$ , from the first inequality we get  $e^{-2t^2/\Delta} > e^{-r}$ , and therefore  $t < \sqrt{r\Delta/2}$ . From the second one,  $\Pr[z \notin Z] = \Pr[d(z) \geq \sqrt{2r\Delta}] < e^{-r}$ .  $\square$

Alternatively, Lemma 3.2 follows from a simple application of Talagrand's inequality.

*Proof of Theorem 3.1.* Let  $\varepsilon = e^{-r}$ . We prove the contrapositive. Assume  $A$  inverts  $f_n$  on an  $\varepsilon$ -fraction of inputs in time  $\exp(O(\sqrt{r\Delta} \log m))$ . We construct an algorithm  $B$  that inverts  $f_n$  on a  $(1 - \varepsilon)$ -fraction of inputs as follows: On input  $y$ , perform the following procedure: For every set of at most  $\sqrt{2r\Delta}$  coordinates of  $[m]$ , flip the bits of  $y$  in these coordinates to obtain  $y'$  and compute  $x' = A(y')$ . If  $f_n(x') = y'$ , flip all possible sets of  $\sqrt{2r\Delta}$  bits of  $x'$  to obtain  $x$ . If  $f_n(x) = y$ , output  $x$ . The running time of  $B$  is

$$\binom{m}{\sqrt{2r\Delta}} \cdot (\text{running time of } A) + \binom{n}{\sqrt{2r\Delta}} \cdot (\text{eval. time of } f_n) = \exp(O(\sqrt{r\Delta} \log n)).$$

We now argue that  $B$  inverts  $f$  on a  $(1 - \varepsilon)$ -fraction of inputs. Let  $S'$  be the set of those  $x'$  such that  $A(f(x')) = x'$ . For each  $j \in [n]$ , let  $d_j$  denote the degree of the  $j$ th input. Now let

$$S = \{x: \sum_{j \in [n]: x_j \neq x'_j} d_j \leq \sqrt{2r\Delta} \text{ for some } x' \text{ in } S'\}.$$

If  $x'$  is in  $S'$  and  $x$  is its closest element in  $S$ , then  $f(x)$  and  $f(x')$  differ in at most  $\sqrt{2r\Delta}$  coordinates. Moreover,  $x$  and  $x'$  can also differ in at most this many coordinates. It follows that if  $x$  is in  $S$ , then  $B$  successfully inverts  $f(x')$ . By Lemma 3.2,  $S$  contains at least a  $1 - \varepsilon$  fraction of inputs.  $\square$

**Remark 1.** The proof of Theorem 3.1 easily generalizes to function families that are  $e^{-r}/2$ -close to 1-1. A non-uniform version, where “running time” is replaced by “circuit size”, is also straightforward. We will use these extensions in our applications in Sections 5 and 6.

Theorem 3.1 gives a non-trivial result only when the sum of the squares of the input degrees  $D$  is at most  $o(n^2/\log n)$ . This assumption could be violated even if there is a single input of  $f$  whose degree is  $\Omega(n)$ . It is natural to ask if the self-amplification argument could be modified so as to allow for a small number of inputs that have unusually large degree.

We argue that this is unlikely to be the case: In Appendix A, we give an example showing that if non-trivial self-amplification can be achieved for functions where all but one of their inputs have degree at most  $d + 1$ , then every function of input locality  $d$  has a non-trivial inversion algorithm.

## 4 Linear functions with noise

We now state a self-amplification result for functions of the “parity with noise” type. We consider the following type of function. Let  $\{M_n\}$  be a family of  $m(n)$  by  $n$  matrices with entries in  $\{0, 1\}$  and  $p \in [0, 1]$  be a parameter. We define the function family  $f_n: \{0, 1\}^n \rightarrow \{0, 1\}^{m(n)}$  as follows:

$$f_n(x, e) = M_n x + e$$

where  $x$  is a vector chosen uniformly at random from  $\{0, 1\}^n$ , and  $e \in \{0, 1\}^m$  is a vector of hamming weight at most  $2pm$  chosen from the following distribution: Each entry of  $e$  is chosen independently from a  $p$ -biased distribution, conditioned on  $r$  having hamming weight at most  $2pm$ . The matrix multiplication and vector addition are performed modulo two.

We will consider functions  $f_n$  that are 1-1. This happens when  $M_n$  is a generator matrix of a code of minimum distance  $4pm$ . In such a case, the input locality of  $f_n$  will be as large as  $\Omega(n)$ . Nevertheless, we can prove an analogue of Theorem 3.1 in this setting. One difference is that our self-amplification argument here is randomized, so we require that the function family is hard to invert even for randomized adversaries.

**Theorem 4.1.** *Suppose the function family  $\{f_n: f_n(x, e) = M_n x + e\}$  is 1-1 and one-way against randomized time  $\exp(O(\sqrt{rm} \log m))$  on a  $e^{-r}$  fraction of inputs. Assume  $r < pm/10$ . Then  $\{f_n\}$  is one-way against randomized time  $\exp(O(\sqrt{rm} \log m))$  on a  $1 - e^{-r}$  fraction of inputs.*

*Proof.* Let  $\varepsilon = e^{-r}$ . Let  $A$  be an algorithm that inverts  $\{f_n\}$  in time  $\exp(O(\sqrt{rm} \log m))$  on a  $1 - e^{-r}$  fraction of inputs with probability  $1/2$ . Without loss of generality, we will assume that when  $A(y)$  fails to output an inverse for  $y$ , it outputs the special symbol  $\perp$ . Consider the following algorithm  $B$  for inverting  $f_n$ :

- 1 On input  $y \in \{0, 1\}^{m(n)}$ ,
- 2     For all  $y' \in \{0, 1\}^{m(n)}$  that are within hamming distance  $\sqrt{2(r+3)m}$  from  $y$ :
- 3         Repeat the following  $32/\varepsilon$  times:
- 4             Choose a random  $z \in \{0, 1\}^n$ .
- 5             If  $A(y' + M_n z) = (u, e') \neq \perp$  and  $e' + (y' + y)$  has hamming weight at most  $2pm$ , output  $(u + z, e' + (y' + y))$ .
- 6     Otherwise, output  $\perp$ .

Algorithm  $B$  makes  $O(1/\varepsilon) \cdot \binom{m}{O(\sqrt{rm})}$  calls to  $A$ , from where the running time of  $B$  follows. We now argue that  $B$  inverts  $f_n$  on a  $1 - \varepsilon$  fraction of inputs with probability  $1/2$ . For notational convenience we define the following distributions on  $\{0, 1\}^m$ . Let  $\mathcal{R}$  denote the distribution where each entry is chosen independently at random with probability  $p$ , and let  $\mathcal{R}_c$  denote the distribution  $\mathcal{R}$  conditioned on having hamming weight at most  $2pm$ . We use  $\mathcal{U}$  to denote the uniform distribution on  $\{0, 1\}^n$ .

Let  $S$  be the set of pairs  $(x, e')$ , where  $x \in \{0, 1\}^n$  and  $e' \in \{0, 1\}^m$  is of hamming weight at most  $2pm$ , such that  $A(M_n x + e') = (x, e')$ . By assumption, we have that

$$\Pr_{x \sim \mathcal{U}, e' \sim \mathcal{R}_c} [(x, e') \in S] \geq \varepsilon.$$

By Markov's inequality,

$$\Pr_{e' \sim \mathcal{R}_c}[\Pr_{x \sim \mathcal{U}}[(x, e') \in S]] \geq \frac{\varepsilon}{2}.$$

By definition of the distribution  $\mathcal{R}_c$ , we then have that

$$\Pr_{e' \sim \mathcal{R}}[\Pr_{x \sim \mathcal{U}}[(x, e') \in S]] \geq \frac{\varepsilon}{2} \cdot \Pr_{e' \sim \mathcal{R}}[\text{wt}(e') \leq 2pm] \geq \frac{\varepsilon}{2} \cdot e^{-\Omega(pm)} \geq \frac{\varepsilon}{4},$$

where the second to last inequality follows by a Chernoff bound for sufficiently large  $n$ . We will say  $e'$  is *good* if  $\Pr_{x \sim \mathcal{U}}[(x, e') \in S] \geq \varepsilon/2$  and the weight of  $e'$  is at most  $1.5pm$ . Using the Chernoff bound and the assumption  $r < pm/10$ , we have that

$$\Pr_{e' \sim \mathcal{R}}[e' \text{ is good}] \geq \frac{\varepsilon}{8}.$$

We will now argue the following two claims:

1. If  $e'$  is good and  $y' = M_n x + e'$ , then  $\Pr_{A, z}[A(y' + M_n z) = (x + z, e')] \geq \varepsilon/16$ , and
2. With probability  $1 - \varepsilon$ , over a choice of  $e \sim \mathcal{R}_c$ ,  $e$  is within hamming distance  $O(\sqrt{rm})$  of some good  $e'$ .

To put the two parts together, by the second claim a random output  $y$  is of the form  $x + e$ , where  $e$  is within hamming distance  $O(\sqrt{rm})$  of some good  $e'$ . In this case, step 2 of the algorithm will find  $y' = x + e'$ . By the first claim, a run of the loop 4-5 will succeed in producing an inverse for  $A(y' + M_n z)$  with probability at least  $\varepsilon/16$ . Repeating the loop  $32/\varepsilon$  times increases the success probability to  $1/2$ . Since  $f_n$  is 1-1, if  $A$  produces an inverse for  $y' + M_n z$ , this inverse must be  $(u, e') = (x + z, e')$ . Since  $y' + M_n z = M_n u + e'$ , it follows that  $y = M_n(u + z) + e' + (y' + y)$ , so step 5 will produce a correct inverse for  $y$ , as long as  $e' + (y' + y)$  has hamming weight at most  $2pm$ . This follows because  $e$  is good (so it has hamming weight at most  $1.5pm$ ) and  $y'$  is within distance  $O(\sqrt{rm})$  from  $y$ .

The first claim follows by a simple self-reduction argument: If  $z$  is chosen at random from  $\{0, 1\}^n$ , then for any fixed  $x \in \{0, 1\}^n$ ,  $x + z$  is uniformly distributed. Since  $e'$  is good, it follows that  $A(M_n(x + z) + e') = (x + z, e')$  with probability at least  $\varepsilon/8$  over the choice of  $z$  and  $1/2$  over the choice of  $A$ . Multiplying the two probabilities together, we obtain that  $\Pr_{A, z}[A(y' + M_n z) = (x + z, e')] \geq \varepsilon/16$ .

It remains to prove the second claim. Let  $S'$  be the set of those  $e'$  that are good. Then  $\Pr_{e' \sim \mathcal{R}}[e' \in S'] \geq \varepsilon/8$ . By Lemma 3.2, it follows that

$$\Pr_{e \sim \mathcal{R}}[e \text{ is within hamming distance } \sqrt{2(r+3)m} \text{ from } S'] \geq 1 - \varepsilon/8.$$

Passing to the distribution  $\mathcal{R}_c$ , we have that

$$\begin{aligned} \Pr_{e \sim \mathcal{R}_c}[e \text{ is within hamming distance } \sqrt{2(r+3)m} \text{ from } S'] \\ \geq 1 - \varepsilon/8 - \Pr_{e \sim \mathcal{R}}[\text{wt}(e) > 2pm] \geq 1 - \varepsilon \end{aligned}$$

by a Chernoff bound, when  $n$  is sufficiently large.  $\square$

## 5 Hardness amplification for regular functions

Theorem 3.1 shows how to achieve self-amplification for functions with small input locality that are 1-1. The assumption that the function is 1-1 was crucial in the argument for the following reason. Suppose  $f$  is a 1-1 function with input locality  $d$  and  $x$  and  $x'$  are two inputs that differ in exactly one coordinate. Suppose we can invert  $f(x)$ . Then with a little bit more work we can invert  $f(x')$ : Since  $f(x)$  and  $f(x')$  can differ in at most  $d$  coordinates, we change  $d$  coordinates of  $f(x')$  until we find  $f(x)$ , recover  $x$ , and move back from  $x$  to  $x'$ .

An important point in this argument is that because  $f$  is 1-1, the inversion algorithm is guaranteed to return  $x$  and not some other preimage for  $f(x)$ . If  $f$  were not 1-1, the inverter could return some other preimage which is very far from  $x$  and therefore also far from  $x'$ . So in general we do not know how to achieve self-amplification for input-local functions that are not 1-1.

We now argue that if  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  is not 1-1 but regular, then there exists a new function  $f': \{0, 1\}^n \rightarrow \{0, 1\}^{m'}$ ,  $m' = m + O(n)$  such that if  $f$  is hard on an small fraction of inputs, then  $f'$  is hard on almost all of its inputs. Moreover, the input locality of  $f'$  is not much larger than the input locality of  $f$ .

To simplify notation, let  $\alpha(d, r, \log n) = (d + r + (\log n)^3) \cdot (\log n)$ .

**Theorem 5.1.** *Suppose there exists a  $K$ -to-1 function  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  with input locality  $d$  which is hard against circuits of size  $\exp(O(\sqrt{rn} \cdot \alpha(d, r, \log n)))$  on a  $e^{-r}$ -fraction of inputs. Then there exists  $f': \{0, 1\}^n \rightarrow \{0, 1\}^{m+\log_2 K+O(r)}$ , with input locality  $d + O(r + (\log n)^3)$  which is hard against circuits of size  $\exp(O(\sqrt{rn} \cdot \alpha(d, r, \log n)))$  on a  $(1 - e^{-r})$ -fraction of inputs. Moreover, if  $f$  is computable by a circuit of size  $s$ , then  $f'$  is computable by a circuit of size  $s + O(n(\log n)^3)$ .*

The construction of  $f'$  from  $f$  is non-uniform. In fact, our proof provides a randomized construction but for simplicity we present the argument in the non-uniform setting. We follow the standard approach of turning a general function into an almost 1-1 function via hashing [13, 12]. The function  $f'$  will have the form  $f'(x) = (f(x), h(x))$ , where  $h$  is a suitably chosen hash function that does not increase input locality by much. If  $f$  is regular, then  $f'$  will be almost 1-1 in the sense that for most  $x$ ,  $f(x)$  has a unique preimage. Moreover, if  $f$  has input locality  $d$ , then  $f'$  will have input locality  $d + O(r + (\log n)^3)$ . We then amplify the hardness of  $f$  using Theorem 3.1 (and Remark 1).

Theorem 5.1 can be combined with the randomized encodings of Applebaum et al. [3, 4] to obtain a hardness amplification result that preserves *output locality*, at the expense of increasing the input length by logarithmic factors.

**Corollary 5.2.** *Suppose there exists a regular function  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  with input locality  $d_{\text{in}}$  and output locality  $d_{\text{out}} \geq 3$  that is hard against circuits of size  $\exp(O(\sqrt{rn} \cdot \alpha(d_{\text{in}}, r, \log n)))$  on a  $e^{-r}$ -fraction of inputs. Then there is a function  $f': \{0, 1\}^{n'} \rightarrow \{0, 1\}^{m'}$ , where  $n' = O(n(\log n)^3)$  and  $m' = m + O(n(\log n)^3)$  with output locality  $d_{\text{out}}$  that is hard against circuits of size  $\exp(O(\sqrt{rn} \cdot \alpha(d_{\text{in}}, r, \log n)))$  on a  $(1 - e^{-r})$ -fraction of inputs. If  $f$  is computable by a circuit of size  $s$ , then  $f'$  is computable by a circuit of size  $s + O(n(\log n)^3)$ .*

## 5.1 A hash with small input locality

A standard way to reduce a  $K$ -to-1 one-way function to a 1-1 one-way function is by hashing. Namely, we would like to define  $f'(x) = (f(x), h(x))$ , where  $h: \{0, 1\}^n \rightarrow \{0, 1\}^{\log_2 K + O(1)}$  is a pairwise independent hash function. However, known constructions of pairwise independent hash functions have input locality as large as  $\Omega(\log_2 K)$ . This is in fact necessary: Mansour et al [14] showed that pairwise independent hash functions have *average sensitivity*  $\Omega(n)$ . By averaging, it follows that the input locality of such functions must be  $\Omega(\log_2 K)$ .

We need to construct a function  $f'$  from  $f$  which preserves not only the hardness of  $f$  but also its small input locality. Our function  $f'$  will also have the form  $f'(x) = (f(x), h(x))$ , where  $h$  is a suitably chosen hash function. However, our hash function  $h$  will only be approximately pairwise independent, chosen in a manner to have small input locality.

We note that Applebaum et al. [4] (Appendix C in the journal version) give a different construction of an “almost pairwise-independent” hash function. However, the almost pairwise independence property they establish for their construction, while sufficient for their application, appears too weak to derive Claim 1.

**Claim 1.** *Suppose  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  is at most  $K$ -to-1, where  $2^{k-1} \leq K < 2^k$ . Then there exists a function  $h: \{0, 1\}^n \rightarrow \{0, 1\}^{k+3r+3}$  such that the function  $f'(x) = (f(x), h(x))$  is  $e^{-r}/2$ -close to 1-1. Moreover,  $h$  is a linear function over  $\{0, 1\}^n$  with input locality  $O(r) + \min\{k, O((\log n)^3)\}$ .*

We now prove Claim 1. The construction of  $h$  will be probabilistic.

**Construction of  $h$ .** Assume that  $f$  is at most  $K$ -to-1, where  $2^{k-1} \leq K < 2^k$ . The function  $h$  has the form  $h(x) = (h_a(x), h_b(x))$  where

$$\begin{aligned} h_a(x) &= (a_k \cdot x + a'_k, a_{k-1} \cdot x + a'_{k-1}, \dots, a_{k_0+1} \cdot x + a'_{k_0+1}) \\ h_b(x) &= (b_1 \cdot x + b'_1, b_2 \cdot x + b'_2, \dots, b_{3r+k_0+3} \cdot x + b'_{3r+k_0+3}). \end{aligned}$$

and  $k_0 = \min\{8(\log n)^2, k\}$ . (In particular, if  $k < 8(\log n)^2$ ,  $h$  only consists of the  $h_b$  part.)

To generate a random  $h$ , we choose the vectors  $a_i, b_i \in \{0, 1\}^n$  from the following distributions: Each  $a_i$  is chosen independently at random from the  $p_i$ -biased distribution over  $\{0, 1\}^n$ , where  $p_i = 4(\log n)^2/i < 1/2$ . Each  $b_i$  is chosen independently at random from the uniform distribution over  $\{0, 1\}^n$ , and  $a'_i, b'_i$  are uniformly random bits.

We now argue that if  $f$  is regular, then with probability at least  $1/2$  over the choice of  $h$ ,  $f'$  is regular over all but an  $e^{-r}/2$  fraction of its inputs.

The proof will have two stages. In the first stage, we argue that for all but an  $e^{-r}/8$  fraction of inputs  $x$ , there are at most  $2^{r+k_0}$  inputs  $x'$  such that  $(f(x), h_a(x)) = (f(x'), h_a(x'))$ . In the second stage, we finish the proof by showing that  $h_b$  hashes all but an  $e^{-r}/8$  fraction of those  $x$ s uniquely.

**The first stage.** If  $k_0 = k$ , the conclusion is trivial, so let us assume that  $k_0 < k$ . Let us fix an input  $x$  and let  $S = \{x' : f(x) = f(x')\}$ . Without loss of generality, we may assume

that  $2^{k-1} \leq |S| < 2^k$ . (If  $|S|$  is smaller, we disregard the effect of the first few hashes in  $h_a$ .) We consider the following sequence of random sets defined recursively:  $S_k = S$ ,  $T_i = \{x' \in S_i : a_i \cdot x' = a_i \cdot x\}$  and

$$S_{i-1} = \begin{cases} T_i, & \text{if } (1 - 1/n)|S_i|/2 \leq |T_{i-1}| \leq (1 + 1/n)|S_i|/2 \\ S_i, & \text{otherwise.} \end{cases}$$

Here is the intuition for this definition: We want to think of the  $i$ th hash  $a_i$  as “successful” if it decreases the size of siblings of  $x$  by roughly a factor of two (not much more and not much less). If all but  $r$  of the hashes are successful, then the size of  $S_0$  can not be much more than  $2^r$ , and so  $x$  will not have more than  $2^r$  siblings that map to  $(f(x), h_a(x))$ . It is sufficient to show that the probability that more than  $r$  of the hashes fail to be successful is quite small.

Notice that by definition of the sets  $S_i$ , it must be that  $S_i \geq \prod_{j=i}^k (1 - 1/n) \cdot 2^{i-1} \geq 2^{i-2}$ . So we are left with the following question: Given a set  $S$  of size at least  $2^{i-2}$ , how likely is it to be split successfully at stage  $i$ ?

**Lemma 5.3.** *Assume  $|R| \geq 2^{i-2}$ . Let  $a \sim \{0, 1\}_p^n$ ,  $b \sim \{0, 1\}_{1/2}$ , where  $p = 4(\log n)^2/i < 1/2$ . Then for  $n$  sufficiently large and any  $\varepsilon > 0$ ,*

$$\Pr[\#\{y \in R : a \cdot y + b = 0\} \notin (1 \pm \varepsilon)|R|/2] \leq \frac{1}{n^4 \varepsilon^2}.$$

Applying this lemma with  $R = S_i$  and  $\varepsilon = 1/n$ , we have that each hash is successful with probability at least  $1 - 1/n^2$ , and the events are independent of one another. By a union bound, the probability of having more than  $r$  unsuccessful splits is at most  $\binom{n}{r} \cdot (1/n^2)^r \leq n^{-r} < e^{-r}/8$ . So for any  $x \in \{0, 1\}^n$ ,

$$\Pr[|S_{k_0}| \geq 2^{k_0+r}] \leq e^{-r}/8.$$

*Proof of Lemma 5.3.* Let  $X = \sum_{y \in R} (-1)^{a \cdot y + b}$ . Then  $E[X] = 0$  and

$$\begin{aligned} E[X^2] &= \sum_{y, z \in R} E[(-1)^{a \cdot (y+z)}] \\ &\leq |R| \max_z \sum_{y \in R} E[(-1)^{a \cdot (y+z)}] \\ &= |R| \max_z \sum_{y \in R_z} E[(-1)^{a \cdot y}] \\ &= |R| \max_z \sum_{y \in R_z} (1 - 2p)^{|y|} \end{aligned}$$

where  $R_z = \{y+z : y \in R\}$ , and  $|y|$  denotes the hamming weight of  $y$ . By Hölder’s inequality, we have that

$$\sum_{y \in R_z} (1 - 2p)^{|y|} \leq |R|^{1-1/K} \left( \sum_{y \in \{0,1\}^n} (1 - 2p)^{K|y|} \right)^{1/K} = |R| \cdot \left( \frac{(1 + (1 - 2p)^K)^n}{|R|} \right)^{1/K}$$

for every  $K \geq 1$ , and so

$$\mathbb{E}[X^2] \leq |R|^2 \cdot \left( \frac{(1 + (1 - 2p)^K)^n}{|R|} \right)^{1/K}$$

Let  $K = (\ln n)/2p$ . Then

$$\mathbb{E}[X^2] \leq |R|^2 \cdot \left( \frac{(1 + 1/n)^n}{|R|} \right)^{1/K} \leq |R|^2 \cdot \left( \frac{e}{|R|} \right)^{1/K} \leq \frac{|R|^2}{n^4},$$

because  $p = 4(\log n)^2/i$  and  $|R| \geq 2^{-i+2}$ . By Chebyshev's inequality, it follows that

$$\Pr[\#\{y \in R: a \cdot y + b = 0\} \notin (1 \pm \varepsilon)|R|/2] = \Pr[|X| > \varepsilon|R|] \leq 1/n^4\varepsilon^2. \quad \square$$

**The second stage and conclusion.** Now fix an  $x$  such that  $|S_{k_0}| < 2^{k_0+r}$ . We now argue that by the end of the second stage,  $x$  is very likely to have a unique hash:

$$\Pr[\exists x' \in S_{k_0} - \{x\}: h(x') = h(x)] \leq \sum_{x' \in S_{k_0} - \{x\}} \Pr[h(x') = h(x)] < e^{-r}/8.$$

Putting the analysis of both stages together, it follows that by the end of stage 2, for any specific  $x$ ,

$$\Pr_h[\exists x': f'(x') = f'(x)] \leq e^{-r}/4.$$

Averaging over  $x$  and applying Markov's inequality, we get that for at least half the functions  $h$ ,

$$\Pr_x[\exists x': f'(x') = f'(x)] \leq e^{-r}/2.$$

Now let us calculate the locality of a typical function  $h$ . For any fixed input bit, say  $x_1$ , let  $Y_a$  and  $Y_b$  be the number of occurrences of  $x_1$  in  $h_a$  and  $h_b$  respectively. Then  $\mathbb{E}[Y_a] = \sum_{i=k_0}^k 4(\log n)^2/i \leq 4(\log n)^3$  and  $\mathbb{E}[Y_b] = (3r + k_0 + 3)/2$ , so  $\mathbb{E}[Y_a + Y_b] = O((\log n)^3 + r)$ . By Chernoff bounds and a union bound, we get that with probability at least  $3/4$ , no input bit has more than  $O((\log n)^3 + r)$  occurrences in  $h$ .

Therefore, there exists a hash function  $h$  that has input locality  $O((\log n)^3 + r)$  and such that  $f'$  is 1-1 on all but  $e^{-r}/2$  fraction of its inputs.

**Remark 2.** The conclusion of Claim 1 also holds under the weaker assumption that the function  $h$  is  $e^{-r}/4$ -close to at most  $K$ -to-1. We will use this generalization in Section 6.

## 5.2 Proof of Theorem 5.1

We now prove Theorem 5.1. To do so, first we show that the transformation from  $f$  to  $f'$  is hardness-preserving in the following sense: If  $f$  is hard to invert on an  $e^{-r}$ -fraction of inputs, then  $f'$  is hard to invert on an  $\Omega(e^{-r})$ -fraction of inputs. Since  $f'$  is almost 1-1, we can apply self-amplification to conclude that  $f'$  is in fact hard on a  $1 - e^{-r}$  fraction of inputs.

**Claim 2.** Assume  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  is  $K$ -to-1 where  $2^{k-1} \leq K < 2^k$ . Let  $f'$  and  $h$  be as in Claim 1. Assume that  $f'$  can be inverted on an  $(1 - e^{-r}/400)$ -fraction of inputs by a circuit of size  $s$ . Then  $f$  can be inverted on a  $(1 - e^{-r})$ -fraction of inputs by a circuit of size  $O(s \cdot r \cdot 2^{3r})$ .

Before proving Claim 2, let us show how it implies Theorem 5.1 and Corollary 5.2.

*Proof of Theorem 5.1.* Suppose  $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$  is a regular function with input locality  $d$  which is hard against circuits of size  $\exp(O(\sqrt{rn} \cdot \alpha(d, r, \log n)))$  on a  $e^{-r}$ -fraction of inputs. Let  $f'(x) = (f(x), h(x))$ , where  $h$  is chosen as in Claim 1. It is easy to check that  $f'$  has the desired input locality and circuit complexity.

Now suppose  $f'$  can be inverted by a circuit of size  $\exp(O(\sqrt{rn} \cdot \alpha(d, r, \log n)))$  on a  $e^{-r}$  fraction of its inputs. By Claim 1,  $f'$  is  $e^{-r}/2$ -close to 1-1. By Theorem 3.1 and Remark 1,  $f'$  can be inverted on a  $(1 - e^{-r}/400)$ -fraction of inputs by a circuit of size  $s = \exp(O(\sqrt{rn} \cdot \alpha(d, r, \log n)))$ . By Claim 2,  $f$  can then be inverted on a  $(1 - e^{-r})$  fraction of inputs by circuits of size  $\exp(O(\sqrt{rn} \cdot \alpha(d, r, \log n)))$ .  $\square$   $\square$

*Proof of Corollary 5.2.* Since  $h(x)$  is a linear function, we can apply the randomized encoding of Applebaum et al. to reduce its output locality at the cost of increasing the input and output length of  $f'$ . Specifically, we perform the following transformation on  $f'$  to obtain a new function  $f''$ . Suppose the  $i$ th output  $h(x)_i$  has the form

$$h(x)_i = x_{i1} + x_{i2} + \dots + x_{ik_i}.$$

We introduce new inputs  $r_{i1}, r_{i2}, \dots, r_{i(k_i-1)}$  and replace the output  $h(x)_i$  by the sequence of outputs:

$$(x_{i1} + r_{i1}, r_{i1} + x_{i2} + r_{i2}, \dots, r_{i(k_i-1)} + x_{ik_i}).$$

It is easy to check that  $f''$  has the desired input and output length, and its output locality is  $\max\{d_{\text{out}}, 3\}$ .

Applebaum et al. [3, 4] show that if  $f''$  can be inverted on an  $\varepsilon$ -fraction of inputs by a circuit of size  $s$ , then  $f'$  can be inverted on a  $\Omega(\varepsilon)$ -fraction of inputs by a circuit of size  $O(s/\varepsilon)$ . Plugging in  $\varepsilon = e^{-r}$  and  $s = \exp(O(\sqrt{rn} \cdot \alpha(d_{\text{in}}, r, \log n)))$ , the corollary follows.  $\square$

*Proof of Claim 2.* Let  $\varepsilon = e^{-r}$  and let  $A'$  be circuit that inverts  $f'$  on a  $(1 - \varepsilon/4)$ -fraction of inputs. We will argue that the following randomized circuit  $A$  inverts  $f$  on a  $(1 - \varepsilon)$ -fraction of inputs:

- 1 On input  $y \in \{0, 1\}^{m(n)}$ ,
- 2 Repeat the following  $O(r \cdot 2^{3r})$  times:
- 3 Choose a random  $z \in \{0, 1\}^{k+3r+3}$ .
- 4 If  $A'(y, z)$  outputs  $x$  such that  $f'(x) = (y, z)$ , output  $x$ .
- 5 Otherwise, output  $\perp$ .

We will argue that for at least a  $1 - 3\varepsilon/4$  fraction of images  $y$  of  $f$ ,  $A(y)$  returns an inverse for  $y$  with probability at least  $1 - \varepsilon/4$ . By an averaging argument, we can fix the randomness of  $A$  so that it finds an inverse for a  $1 - \varepsilon$  fraction of  $ys$ .

Say an output  $y \in \{0, 1\}^m$  (where  $y = f(x)$  for some  $x$ ) is *bad* if

$$\#\{h(x): f(x) = y \text{ and } A'(f'(x)) \in f'^{-1}(f'(x))\} < 2^{k-6}.$$

We will first argue that if  $y$  is not bad, then  $A(y)$  returns an inverse for  $y$  with probability  $1 - \varepsilon/4$ . We will then argue that at most a  $3\varepsilon/4$  fraction of the  $ys$  are bad.

If  $y$  is not bad, then there exist at least  $2^{k-2}$  choices of  $z \in \{0, 1\}^{k+3r+3}$  such that  $f(A'(y, z)) = y$ . In such a case,  $A$  succeeds in finding an inverse for  $y$ . The probability of choosing such a  $z$  in one of the repetitions in round 2 is at least  $1 - (1 - 2^{k-6}/2^{k+3r+3})^{O(r \cdot 2^{3r})} > 1 - \varepsilon/4$ , as desired.

It remains to argue that at most a  $3\varepsilon/4$  fraction of the  $ys$  are bad. To argue this, we cover the bad  $ys$  by two sets  $Y_1$  and  $Y_2$ , and argue that  $|Y_1| + |Y_2| \leq \frac{3}{4}\varepsilon \cdot (2^n/K)$ . We define:

$$\begin{aligned} Y_1 &= \{y: \#\{h(x): f(x) = y\} < 2^{k-5}\} \\ Y_2 &= \{y: \#\{h(x): A'(f'(x)) \notin f'^{-1}(f'(x))\} \geq 2^{k-6}\}. \end{aligned}$$

If  $y$  is bad, then  $y \in Y_1 \cup Y_2$ . We now bound the fraction of  $ys$  in each one of these sets.

Suppose  $y \in Y_1$  and consider the set of  $x$  such that  $f(x) = y$ . There are at least  $K$  such  $x$ , where  $K \geq 2^{k-1}$ . On the other hand, these  $xs$  can take at most  $2^{k-5}$  values of  $h(x)$ . It follows that for at least  $K - 2^{k-5}$  such  $x$ , there exists an  $x' \neq x$  such that  $f(x) = f(x')$  and  $h(x) = h(x')$ . Therefore each  $y \in Y_1$  gives rise to  $K - 2^{k-5} \geq (15/16)K$  preimages  $x$  such that  $f(x) = y$  and  $f'(x) = f'(x')$  for some  $x = x'$ .

It follows that there are at least  $(15/16)K \cdot |Y_1|$  values of  $x$  on which  $f'$  is not 1-1. By Claim 1,  $f'$  is  $\varepsilon/2$ -close to 1-1. So it must be that  $|Y_1| \leq \frac{8}{15}\varepsilon \cdot (2^n/K)$ .

Now suppose  $y \in Y_2$ . For every such  $y$ , there are at least  $2^{k-6}$  values of  $h(x)$  such that  $A'$  fails to invert  $f(x)$ . So there must be at least  $2^{k-6}$  values of  $x$  where  $f(x) = y$  for which the inversion fails. It follows that there are at least  $2^{k-6}|Y_2|$  values of  $x$  for which the inverter  $A'$  fails. By assumption, the number of such values is at most  $(\varepsilon/400) \cdot 2^n$ . It follows that  $|Y_2| \leq \frac{1}{5} \cdot (2^n/K)$ .

We have that  $|Y_1| + |Y_2| \leq \frac{8}{15}\varepsilon \cdot (2^n/K) + \frac{1}{5} \cdot (2^n/K) \leq \frac{3}{4}\varepsilon \cdot (2^n/K)$ , concluding the proof.  $\square$

$\square$

## 6 Goldreich's function on a random graph

We now consider two applications of our techniques to the candidate one-way function proposed by Goldreich [8]. Given a bipartite graph  $G$  with  $n$  vertices on the left,  $m$  vertices on the right, and regular right-degree  $d_{\text{out}}$  and a predicate  $P: \{0, 1\}^{d_{\text{out}}} \rightarrow \{0, 1\}$ , the function  $f_{G,P}$  from  $\{0, 1\}^n$  to  $\{0, 1\}^m$  is defined by

$$f_{G,P}(x)_i = \text{the } i\text{th bit of } f(x) = P(x_{\Gamma(i,1)}, \dots, x_{\Gamma(i,d_{\text{out}})})$$

where  $\Gamma_{(i,j)}$  is the  $j$ th neighbor of right vertex  $i$  of  $G$ .

Goldreich [8] considered such constructions for the setting of parameters  $m = n$  and  $d_{\text{out}}$  ranges from a constant to  $O(\log n)$ . He conjectured that when  $G$  is a good expander graph and  $P$  is a randomly chosen predicate, with high probability  $f_{G,P}$  is one-way.

Cook et al. [7] showed that when  $G$  is random and  $P$  is suitably chosen,  $f_{G,P}$  is secure against adversaries that implement *myopic algorithms*. Bogdanov and Qiao [6] studied a variant of Goldreich's function in the setting where  $G$  is random,  $d$  is constant, and  $m = Dn$ , where  $D = D(d_{\text{out}})$  is a sufficiently large constant. They showed that for a large class of predicates  $P$  (those that correlate with one or a pair of their inputs) and for most  $G$ ,  $f_{G,P}$  can be inverted on most inputs. It is conceivable that  $f_{G,P}$  could be one-way for all predicates  $P$  that are not linear and do not belong to the class ruled out by Bogdanov and Qiao.

We establish two results regarding local hardness amplification of Goldreich's function. Informally, we show that

1. In the setting where  $d$  is constant and  $m \geq Dn$ , where  $D = D(d_{\text{out}})$  is a sufficiently large constant, for any  $P$  and with high probability over the choice of  $G$  if  $f_{G,P}$  is hard to invert on an  $e^{-r}$  fraction of inputs in time  $\exp(O(\sqrt{rn} \cdot d_{\text{out}} \cdot \log n))$ , then  $f_{G,P}$  is hard to invert on a  $1 - e^{-r}$  fraction of inputs in time  $\exp(O(\sqrt{rn} \cdot d_{\text{out}} \cdot \log n))$ .
2. When  $d_{\text{out}} = O(\log n)$  and  $m = n$ , for a certain class of predicates  $P$  and with high probability over  $G$ , if  $f_{G,P}$  is hard to invert on a  $e^{-r}$  fraction of inputs, then there exists a function  $f' : \{0, 1\}^{n'} \rightarrow \{0, 1\}^{m'}$ , where  $n', m' = n \cdot \text{polylog} n$ , of similar complexity to  $f$  and constant output locality that is hard to invert on a  $1 - e^{-r}$  fraction of inputs.

Our result applies to all  $O(\log n)$ -*parity-blowup* predicates, which we define as follows. Let  $P_c : \{0, 1\}^c \rightarrow \{0, 1\}$  be any balanced predicate, where  $c$  is some constant. The  $d_{\text{out}}$ -parity-blowup of  $P_c$  is the predicate  $P : \{0, 1\}^{d_{\text{out}}} \rightarrow \{0, 1\}$  which is obtained by replacing each of the variables in  $P_c$  by a parity of  $\lfloor d_{\text{out}}/c \rfloor$  inputs, where all the inputs are distinct. Applebaum, Barak, and Wigderson [2] showed that the output of Goldreich's function based on such predicates is pseudorandom against linear functions, low-degree polynomials, and constant-depth circuits.

The random graph  $G$  is chosen from the following distribution: For each of the  $m$  right vertices of  $G$ , choose all of its  $d_{\text{out}}$  neighbors independently at random among the  $n$  left vertices of  $G$ . We will call such graphs  $(n, m, d_{\text{out}})$  *random graphs*.

## 6.1 Self-reducibility for functions with long output

**Theorem 6.1.** *Let  $D \geq 2^{Kd_{\text{out}}}$  where  $K$  is a sufficiently large constant, and  $P : \{0, 1\}^{d_{\text{out}}} \rightarrow \{0, 1\}$  be any predicate. Let  $G$  be an  $(n, m, d_{\text{out}})$  random graph. With probability at least  $1 - o(1)$  over the choice of  $G$ , if  $f_{G,P}$  is hard for circuits of size  $\exp(O(\sqrt{rn} \cdot Dd_{\text{out}} \log n))$  on an  $e^{-r}$ -fraction of inputs, then  $f_{G,P}$  is hard for circuits of size  $\exp(O(\sqrt{rn} \cdot Dd_{\text{out}} \log n))$  on a  $1 - e^{-r}$ -fraction of inputs.*

We prove Theorem 6.1 by an argument similar to the one used in the proof of Theorem 3.1. The principal obstacle to applying Theorem 3.1 here is that with high probability, the function  $f_{G,P}$  is not 1-1. There are several reasons for this. One reason is that  $f_{G,P}$  is likely to contain inputs that do not appear in any output. A more important reason is that unless the predicate  $P$  is linear, for most inputs  $x$ , it is likely that there is a linear number of coordinates  $i$  such that the  $i$ th coordinate does appear in the output, but changing the value of  $x_i$  does not change the value of  $f_{G,P}(x)$ .

We show that although  $f_{G,P}$  is unlikely 1-1, with high probability every pair of inputs that map to the same output is highly correlated (or anticorrelated), that is they agree (or disagree) in value on most of the coordinates. Using the argument from the proof of Theorem 3.1, we show that if  $f_{G,P}$  can be inverted on an  $\varepsilon$ -fraction on inputs by a circuit of suitable size, then for a  $1 - \varepsilon$  fraction of inputs  $x$ , it is possible to find an  $x'$  such that  $x$  and  $x'$  are highly correlated. We then use a result of Bogdanov and Qiao [6] which says that for most inputs  $x$ , given  $x'$  that is correlated with  $x$ , then we can invert  $f_{G,P}(x)$ .

**Claim 3.** *Assume that  $D > 2 \cdot 6^{d_{\text{out}}}$ . Let  $P: \{0, 1\}^{d_{\text{out}}} \rightarrow \{0, 1\}$  be any nonconstant predicate. Consider the function  $f_{G,P}: \{0, 1\}^n \rightarrow \{0, 1\}^m$ , where  $m = Dn$ . With probability  $1 - 2^{-n}$  over the choice of  $G$ , for any pair of inputs  $x, x' \in \{0, 1\}^n$ , if  $f_{G,P}(x) = f_{G,P}(x')$ , then  $x$  and  $x'$  either agree on at least a  $2/3$ -fraction of coordinates or they disagree on at least a  $2/3$ -fraction of coordinates.*

**Theorem 6.2** (Bogdanov and Qiao [6]). *Let  $K$  be a sufficiently large constant and  $D > 2^{-Kd_{\text{out}}}$ . Let  $P: \{0, 1\}^{d_{\text{out}}} \rightarrow \{0, 1\}$  be any predicate. Then there is an algorithm  $A$  such that with probability  $1 - o(1)$  over the choice of  $G$ , the following holds. Consider the function  $f_{G,P}: \{0, 1\}^n \rightarrow \{0, 1\}^m$ , where  $m = Dn$ . For a  $1 - 2^{-n \cdot 2^{-\Omega(d_{\text{out}})}}$  fraction of assignments  $x$  and any assignment  $x'$  such that  $x$  and  $x'$  agree on at least a  $4/9$  fraction of coordinates, on input  $G, P, f(x)$  and  $x'$ ,  $A$  outputs an inverse for  $f_{G,P}(x)$ . The running time of  $A$  is polynomial in  $2^d \cdot n$ .*

We will also make use of the following technical claim concerning the degree distribution in a random graph. For convenience of notation, from now on we set  $d_{\text{out}} = d$ .

**Claim 4.** *Let  $G$  be an  $(n, m = Dn, d)$  random graph. Assume  $Dd > 6$ . Let  $d_i$  denote the degree of left vertex  $i$  and  $\Delta = d_1^2 + \dots + d_n^2$ . Then*

$$\Pr[\Delta > 4(Dd)^2 n] < \frac{2}{Ddn}.$$

*Proof.* Let  $X_{ij}$  denote the number of edges between vertices  $i \in [n]$  and  $j \in [m]$ . Then  $\mathbb{E}[X_{ij}] \leq d/n$  and  $\text{Var}[X_{ij}] \leq d/n$ . Then  $d_i = \sum_{j \in [m]} X_{ij}$ , and the variables  $X_{ij}$  and  $X_{i'j'}$  are independent whenever  $j \neq j'$ . Therefore for every  $i$ ,

$$\mathbb{E}[d_i^2] = \mathbb{E}[d_i]^2 + \text{Var}[d_i] = \left( \sum_{j \in [m]} \mathbb{E}[X_{ij}] \right)^2 + \sum_{j \in [m]} \text{Var}[X_{ij}] \leq (m \cdot d/n)^2 + m \cdot d/n = D^2 d^2 + Dd.$$

And so  $E[\Delta] \leq (D^2d^2 + Dd)n \leq 2D^2d^2n$ . To bound the deviation from  $\Delta$ , we apply the second moment method. We have

$$\begin{aligned} \text{Var}[\Delta^2] &= \text{Var}\left[\sum_{i \in [n]} \left(\sum_{j \in [m]} X_{ij}\right)^2\right] \\ &= E\left[\sum_{i,i',j,j',k,k'} X_{ij}X_{ik}X_{i'j'}X_{i'k'}\right] - E\left[\sum_{i,j,k} X_{ij}X_{ik}\right] E\left[\sum_{i',j',k'} X_{i'j'}X_{i'k'}\right] \\ &= \sum_{i,i',j,j',k,k'} (E[X_{ij}X_{ik}X_{i'j'}X_{i'k'}] - E[X_{ij}X_{ik}]E[X_{i'j'}X_{i'k'}]), \end{aligned}$$

where  $i, i'$  range over  $[n]$ , and  $j, j', k, k'$  range over  $[m]$ . When  $i \neq i'$ , the variables  $X_{ij}X_{ik}$  and  $X_{i'j'}X_{i'k'}$  are negatively correlated and so the term is not positive. When  $i = i'$ , by independence, the only terms in the summation that may not vanish are those where  $|\{j, k, j', k'\}| < 4$ . For any fixed  $i$ , There are at most  $\binom{4}{t}m^t$  terms where  $|\{j, k, j', k'\}| \leq t$ , and for each such term we have

$$E[X_{ij}X_{ik}X_{i'j'}X_{i'k'}] \leq E[X_{11}]^t \leq (d/n)^t$$

from where

$$\text{Var}[\Delta^2] \leq n \cdot \sum_{t=0}^3 \binom{4}{t} m^t (d/n)^t \leq 4D^3d^3 + 6D^2d^2 + 4Dd \leq 8nD^3d^3.$$

Applying Chebyshev's inequality to  $\Delta$ , we obtain the theorem.  $\square$

*Proof of Theorem 6.1.* If  $P$  is a constant predicate, the conclusion holds trivially, so we assume that  $P$  is non-constant. Also, assume that  $r < 2^{-Kd}n - 2$  for a sufficiently large constant  $K$ , for otherwise the result is trivial.

Let  $G$  be a graph for which the conclusions of Claim 3, Theorem 6.2 and Claim 4 hold. This comprises a  $1 - o(1)$  fraction of graphs  $G$ .

Let  $A$  be a circuit that inverts  $f = f_{G,P}$  on an  $e^{-r}$ -fraction of inputs and let  $I$  be the inverter from Theorem 6.2.

We will argue that the following algorithm  $B$  inverts  $f_{G,P}$  on a  $1 - e^{-r}$ -fraction of inputs: On input  $y \in \{0, 1\}^m$ , for any set of at most  $\sqrt{2(r+2)\Delta}$  coordinates of  $[m]$ , flip  $y$  in those coordinates to obtain  $y''$ . If  $A(y'')$  produces an inverse  $x''$ , run  $I$  on inputs  $(G, P, y, x'')$  and  $(G, P, y, \overline{x''})$ . If either run produces an inverse  $x$  for  $y$ , output  $x$ .

Let  $S'$  be the set of all  $x'$  such that  $A(f(x'))$  returns an inverse  $x''$  for  $f(x')$ . By Claim 3,  $x'$  and  $x''$  must agree on at least a  $2/3$  fraction of coordinates. Now let

$$S = \{x : \sum_{j \in [n]: x'_j \neq x_j} d_j \leq \sqrt{2(r+2)\Delta} \text{ for some } x' \text{ in } S'\}.$$

By Lemma 3.2, since  $|S'| \geq e^{-r} \cdot 2^n$ , we get that  $|S| \geq (1 - e^{-r}/2) \cdot 2^n$ . Now let  $x \in S$  be such that  $I(G, P, f(x), x'')$  returns an inverse for  $f(x)$  whenever  $x$  and  $x''$  agree on at least a  $4/9$  fraction of coordinates. By Theorem 6.2, all but a  $2^{-2^{Kd}n} < e^{-r}/2$  fraction of  $x$ s have

this property. Consider such an  $x$ . By definition of  $S$ , there exists an  $x' \in S'$  such that  $\sum_{j \in [n]: x'_j \neq x_j} d_j \leq \sqrt{2(r+2)\Delta}$ . Take the  $x'$  that is closest to  $x$ . Then  $x$  and  $x'$  can differ in at most  $\sqrt{2(r+2)\Delta}$  coordinates. By Claim 3, either  $x'$  and  $x''$  or  $x'$  and  $\overline{x''}$  must agree on at least  $2n/3$  of their coordinates. By the triangle inequality, it follows that  $x$  and either  $x''$  or  $\overline{x''}$  must agree on  $2n/3 + \sqrt{2(r+2)\Delta} \leq 4n/9$  of their coordinates. Therefore, at least one of  $I(G, P, f(x), x'')$  and  $I(G, P, f(x), \overline{x''})$  is an inverse for  $f(x)$ .  $\square$

*Proof of Claim 3.* Let  $f = f_{G,P}$ . Fix  $x$  and  $x'$ . Let  $p_{aa'}$  be the fraction of coordinates  $i \in [n]$  such that  $x_i = a$  and  $x'_i = a'$ . The fraction of coordinates on which  $x$  and  $x'$  agree is  $p_{00} + p_{11}$ . We will prove the following two claims:

1. If  $1/3 \leq p_{00} + p_{11} \leq 2/3$ , then there exist  $aa', bb' \in \{0, 1\}^2$  that differ in exactly one coordinate such that  $p_{aa'} \geq 1/6$  and  $p_{bb'} \geq 1/6$ .
2. If  $p_{aa'}, p_{bb'} \geq 1/6$  and  $aa', bb'$  differ in exactly one coordinate, then  $\Pr[f(x) = f(x')] \leq 2^{-3n}$ .

We prove the first claim by a case analysis. Assume  $1/3 \leq p_{00} + p_{11} \leq 2/3$ . Therefore,  $1/3 \leq p_{01} + p_{10} \leq 2/3$  as well. Let  $p_0 = p_{00} + p_{01}$  and  $p_1 = p_{10} + p_{11}$ . Then  $p_0$  (resp.  $p_1$ ) is the fraction of coordinates  $i$  on which  $x_i = 0$  (resp.  $x_i = 1$ ). Without loss of generality, let us assume that  $p_0 \geq 1/2$ . Then: (1) If  $p_0 \geq 5/6$ , the inequalities imply that  $p_{00} \geq 1/6$  and  $p_{01} \geq 1/6$ . (2) If  $1/2 \leq p_0 < 5/6$  and  $p_{01} < 1/6$ , we must have  $p_{00} \geq 1/6$  and  $p_{10} \geq 1/6$ . (3) If  $1/2 \leq p_0 < 5/6$  and  $p_{00} < 1/6$ , we must have  $p_{01} \geq 1/6$  and  $p_{11} \geq 1/6$ . (4) If none of the above assumptions hold, then  $p_{00} \geq 1/6$  and  $p_{01} \geq 1/6$ .

We now prove the second claim. Without loss of generality, assume that  $aa' = 00$  and  $bb' = 01$ . Since  $P$  is not constant, there must exist an input  $z \in \{0, 1\}^d$  such that  $P(0) \neq P(z)$ . Now let us look at the  $i$ th output of  $f$  when  $G$  is chosen at random. The probability that  $(x_{\Gamma(i,1)}, \dots, x_{\Gamma(i,d)}) = 0$  and  $(x'_{\Gamma(i,1)}, \dots, x'_{\Gamma(i,d)}) = z$  is at least  $(1/6)^d$ , since each of the pairs  $(0, z_k)$ ,  $1 \leq k \leq d$ , is sampled with probability at least  $1/6$ . Since every output of  $G$  is chosen independently, we have that

$$\Pr[f(x) = f(x')] \leq (1 - (1/6)^d)^m \leq e^{-(1/6)^d m} \leq 2^{-3n}$$

as long as  $m \geq 2 \cdot 6^d$ .  $\square$

## 6.2 Amplification for certain length-preserving functions

Let  $P: \{0, 1\}^c \rightarrow \{0, 1\}$  be a balanced predicate. The  $d_{\text{out}}$ -parity-blowup of  $P$  is the predicate obtained by replacing each variable in  $P$  by  $\lfloor d_{\text{out}}/c \rfloor$  variables, where all the new variables are disjoint.

**Theorem 6.3.** *Let  $c \geq 3$  and  $d_{\text{out}} = \max\{130c \log n, 4c^2\}$ . Let  $P$  be the  $d_{\text{out}}$ -parity-blowup of some balanced predicate on  $c$  bits and let  $G$  be an  $(n, n, d_{\text{out}})$ -random graph. Suppose that for at least half the graphs  $G$ ,  $f_{G,P}: \{0, 1\}^n \rightarrow \{0, 1\}^n$  is hard to invert on an  $e^{-r}$ -fraction of inputs against circuits of size  $\exp(O(r^{3/2} \sqrt{n} \log n))$ . Then there exists*

1. A function  $f': \{0, 1\}^n \rightarrow \{0, 1\}^{n+O(r)}$  of circuit size  $O(n \log n + r)$  that is hard on a  $1 - e^{-r}$ -fraction of inputs for circuits of size  $\exp(O(r^{3/2}\sqrt{n} \log n))$ .
2. A function  $f'': \{0, 1\}^{n'} \rightarrow \{0, 1\}^{m'}$ , where  $n', m' = O(n \cdot d(n)^c)$ ,  $f''$  is hard on a  $1 - e^{-r}$ -fraction of inputs for circuits of size  $\exp(O(r^{3/2}\sqrt{n} \log n))$  and where every output of  $f''$  depends on at most  $c + 1$  inputs.

To prove part 1, we first show that the function  $f_{G,P}$  is likely to be  $e^{-r}/4$ -close to  $O(e^r)$ -to-1. Using Claim 1, we then transform  $f_{G,P}$  into a function  $f': \{0, 1\}^n \rightarrow \{0, 1\}^{n+O(r)}$  which is  $e^{-r}/2$ -close to 1-1. Using the self-reduction of Theorem 3.1, we then argue that if  $f'$  is easy to invert on an  $e^{-r}$  fraction of inputs, then it is also easy to invert on a  $1 - e^{-r}$ -fraction of inputs, and so  $f_{G,P}$  is also easy to invert on the same fraction of inputs by circuits of similar size. To prove part 2, we observe that parity-blowup predicates can be represented constant degree polynomials over  $GF(2)$ . Applying the randomized encodings of Applebaum et al. [3] to these polynomials, we obtain a function with constant output locality and slightly longer input and output length.

**Bounding the preimage sizes** We now argue that in the setting of parameters considered in this section, Goldreich's function is likely to have small preimage size on most of the inputs. The main technical tool is the following claim which bounds the collision probability of  $f_{G,P}$ . For convenience of notation, from now on we set  $d_{\text{out}} = d$ .

**Claim 5.** *Assume  $P: \{0, 1\}^d \rightarrow \{0, 1\}$  is  $c$ -Fourier-sparse and  $d \geq 130c \log n$ . Let  $G$  be an  $(n, n, d)$ -random graph. Then  $\Pr_{G,x,y}[f_{G,P}(x) = f_{G,P}(y)] \leq K \cdot 2^{-n}$ , where  $K$  is some universal constant.*

The notion of  $c$ -Fourier sparsity is somewhat technical and is defined below. The  $d$ -parity blowup of any balanced predicate  $P: \{0, 1\}^c \rightarrow \{0, 1\}$  is  $c$ -Fourier-sparse, whenever  $d \geq 4c^2$ .

From Claim 5 and Markov's inequality, it follows that for at least  $3/4$  of the graphs  $G$ ,

$$\Pr_x[\#\{y: f_{G,P}(x) = f_{G,P}(y)\} \geq 16K \cdot e^r] \leq e^{-r}/4 \quad (1)$$

and so for at least  $3/4$  of the graphs  $G$ ,  $f_{G,P}$  is  $e^{-r}/4$ -close to at most  $16Ke^r$ -to-1.

*Proof of Theorem 6.3.* Let  $G$  be an  $(n, n, d)$  random graph and  $P$  be the  $d$ -blowup of some predicate on  $c$  bits. By ((1)), with probability at least  $3/4$  over the choice of  $G$ ,  $f_{G,P}$  is  $e^{-r}/4$ -close to at most  $16Ke^r$ -to-1. By a Chernoff bound and union bound, with probability  $3/4$  over the choice of  $G$ ,  $f_{G,P}$  has input locality  $O(c \log n)$ . Putting the two together, we have that with probability at least  $1/2$ ,  $f_{G,P}$  has both properties.

From the assumption, there must exist a graph  $G$  such that  $f = f_{G,P}$  is hard to invert on a  $e^{-r}$ -fraction of inputs by circuits of size  $\exp(O(r^{3/2}\sqrt{n} \log n))$ ,  $f_{G,P}$  is  $e^{-r}/4$ -close to at most  $16Ke^r$ -to-1, and it has input locality  $O(c \log n)$ . By Claim 1 (see Remark 2), there exists a linear function  $h: \{0, 1\}^n \rightarrow \{0, 1\}^{O(r)}$  such that  $f'(x) = (f(x), h(x))$  is  $e^{-r}/2$ -close to 1-1, and  $h$  has input locality  $O(r)$ . It is easy to check that  $f'$  has circuit size  $O(n \log n + r)$ .

To prove part 1, we now argue that if  $f'$  can be inverted on an  $e^{-r}$  fraction of inputs by a circuit of size  $\exp(O(r^{3/2}\sqrt{n}\log n))$ , then  $f$  can be inverted on a  $1 - e^{-r}$  fraction of inputs by a circuit of size  $\exp(O(r^{3/2}\sqrt{n}\log n))$ . By Theorem 3.1 and Remark 1,  $f'$  can be inverted on a  $1 - e^{-r}$  fraction of inputs by circuits of size  $\exp(O(r^{3/2}\sqrt{n}\log n))$ . Let  $A'$  be the inverter for  $f'$ . Consider the following inverter  $A$  for  $f$ : On input  $y \in \{0, 1\}^n$ , run  $A(y, z)$  for all possible  $z \in \{0, 1\}^{O(r)}$ . If any of the runs produces  $x$  such that  $f(x) = y$ , output  $x$ . The circuit size of  $A$  is then  $\exp(O(r^{3/2}\sqrt{n}\log n))$ . By construction of  $A$ , whenever  $A'$  inverts  $f'(x)$ ,  $A$  inverts  $f(x)$ , so  $A$  inverts  $f$  on a  $1 - e^{-r}$  fraction of inputs.

To prove part 2, we apply the randomized encodings of Applebaum et al. [4] to every output of  $f'(x) = (f(x), h(x))$  to construct the function  $f''$ . Since  $P$  is a  $d(n)$ -parity blowup of a predicate on  $c$  bits, each output bit of  $f$  is a  $GF(2)$  polynomial of degree  $c$  that depends on  $d(n)$  variables. We write out each such polynomial  $f_i$  as a sum of monomials  $p_i = m_{i1} + m_{i2} + \dots + m_{it}$ . Then  $t \leq d(n)^c$ . We replace each  $f_i$  by the randomized encoding

$$(m_{i1} + r_{i1}, r_{i1} + m_{i2} + r_{i2}, \dots, r_{i(t-1)} + m_{it})$$

where  $r_{i1}, \dots, r_{i(t+1)}$  are new input variables. Similarly, we replace each output  $h_i$  of  $h$  by its randomized encoding, as in the proof of Corollary 5.2. It is easy to check that  $f''$  has input and output length  $O(n \cdot d(n)^c)$  and locality  $c + 1$ . By the same argument as the proof of Corollary 5.2, if  $f''$  can be inverted on a  $e^{-r}$  fraction of inputs by a circuit of size  $\exp(O(r^{3/2}\sqrt{n}\log n))$ , then  $f'$  can be inverted on a  $\Omega(e^{-r})$  fraction of inputs by circuits of size  $\exp(O(r^{3/2}\sqrt{n}\log n))$ .  $\square$

**Fourier-sparse predicates** We say a predicate  $P: \{0, 1\}^d \rightarrow \{0, 1\}$  is  $c$ -Fourier-sparse ( $c \geq 1$ ) if the following conditions hold:

1.  $\ell_1(\hat{P}) = \sum_{S \subseteq [d]} |\hat{P}_S| \leq 2^{d/8c}$
2. For every  $S \subseteq [d]$  such that  $\hat{P}_S \neq 0$ ,  $|S| \geq d/c$
3. For every  $S, T \subseteq [d]$  such that  $S \neq T$  and  $\hat{P}_S, \hat{P}_T \neq 0$ ,  $|S - T| \geq d/c$ .

It is straightforward to check that the  $d$ -parity-blowup of any balanced predicate  $P: \{0, 1\}^c \rightarrow \{0, 1\}$  is  $c$ -Fourier-sparse, whenever  $d \geq 4c^2$ .

**Proof of Claim 5** We prove this claim in two steps: First, we argue that the desired probability does not increase by much if we replace the good predicate  $P_n$  by a linear predicate over  $d/c$  inputs. We then use known results on ranks of random matrices to bound the collision probability for linear  $P_n$ .

*Proof of Claim 5.* Let  $G = G_n$  and  $P = P_n$ . We rewrite the desired probability as:

$$\Pr_{G,x,y}[f_{G,P}(x) = f_{G,P}(y)] = \mathbb{E}_{x,y}[\Pr_G[f_{G,P}(x) = f_{G,P}(y)]] = \mathbb{E}_{x,y}[\Pr_I[P(x|_I) = P(y|_I)]^n]$$

where  $I$  is a random sequence of  $d$  indices from  $[n]$ . Note that the value of the inner probability only depends on  $x$  and  $y$  through the number of pairs  $x_i y_i$  of types 00, 01, 10,

and 11. Let  $n_{ab}$  be the number of pairs  $x_i y_i$  where  $x_i = a$  and  $y_i = b$  and  $p_{ab} = n_{ab}/n$ . Then  $\mathcal{D} = (p_{00}, p_{01}, p_{10}, p_{11})$  is a probability distribution over  $\{0, 1\}^2$ . Therefore

$$\mathbb{E}_{x,y}[\Pr_I[P(x|_I) = P(y|_I)]^n] = \frac{1}{2^{2n}} \sum_{n_{00}+n_{01}+n_{10}+n_{11}=n} \binom{n}{n\mathcal{D}} \Pr_{uv \sim \mathcal{D}^d}[P(u) = P(v)]^n$$

where  $u, v$  are  $d$ -bit strings,  $\binom{n}{n\mathcal{D}}$  is shorthand for  $\binom{n}{n_{00}, n_{01}, n_{10}, n_{11}}$ , and  $\mathcal{D}^d$  is the distribution on  $uv$  obtained by choosing each pair  $u_i v_i$  independently from the joint distribution  $\mathcal{D}$ .

We now divide the summation into two parts: The first part  $E$  will consist of those tuples  $(n_{00}, n_{01}, n_{10}, n_{11})$  such that all  $n_{ab}$  are at most  $5n/6$ , and the second part  $F$  will consist of the rest. We begin with the first part, so let us assume that  $n_{ab} \leq 5n/6$  for all pairs  $ab$ . We then write  $\Pr_{uv \sim \mathcal{D}^d}[P(u) = P(v)] = \frac{1}{2} + \frac{1}{2} \mathbb{E}_{uv \sim \mathcal{D}^d}[(-1)^{P(u) \oplus P(v)}]$  and

$$\begin{aligned} \mathbb{E}_{uv \sim \mathcal{D}^d}[(-1)^{P(u) \oplus P(v)}] &= \sum_{S, T \subseteq [n]} \hat{P}_S \hat{P}_T \mathbb{E}_{uv \sim P^d}[\chi_S(u) \chi_T(v)] \\ &= \sum_{S, T \subseteq [n]} \hat{P}_S \hat{P}_T \beta_1^{|S-T|} \beta_2^{|T-S|} \beta_{12}^{|S \cap T|} \end{aligned}$$

where  $\beta_1 = \mathbb{E}_{ab \sim P}[(-1)^a]$ ,  $\beta_2 = \mathbb{E}_{ab \sim P}[(-1)^b]$ , and  $\beta_{12} = \mathbb{E}_{ab \sim P}[(-1)^{a \oplus b}]$ . We now decompose the summation as

$$\begin{aligned} \mathbb{E}_{uv \sim \mathcal{D}^d}[(-1)^{P(u) \oplus P(v)}] &= \sum_{S \subseteq [n]} \hat{P}_S^2 \beta_{12}^{|S|} + \sum_{S \neq T} \hat{P}_S \hat{P}_T \beta_1^{|S-T|} \beta_2^{|T-S|} \beta_{12}^{|S \cap T|} \\ &\leq \beta_{12}^{d/c} + \ell_1(\hat{P})^2 \cdot \max_{S, T: \hat{P}_S, \hat{P}_T \neq 0} |\beta_1|^{|S-T|} |\beta_2|^{|T-S|}. \end{aligned}$$

From the assumption that  $(n_{00}, n_{01}, n_{10}, n_{11})$  is in  $E$ , it follows that  $\min\{|\beta_1|, |\beta_2|\} \leq 5/6$ , and so

$$\ell_1(\hat{P})^2 \cdot \max_{S, T: \hat{P}_S, \hat{P}_T \neq 0} |\beta_1|^{|S-T|} |\beta_2|^{|T-S|} \leq 2^{d/4c} \cdot (5/6)^{d/c} \leq (0.992)^{d/c}.$$

Now let  $L$  be a linear function on  $d/c$  inputs. We just showed that

$$\mathbb{E}_{uv \sim \mathcal{D}^d}[(-1)^{P(u) \oplus P(v)}] \leq \mathbb{E}_{uv \sim \mathcal{D}^d}[(-1)^{L(u) \oplus L(v)}] + (0.992)^{d/c}$$

whenever  $(n_{00}, n_{01}, n_{10}, n_{11}) \in E$ , and so

$$\begin{aligned} \sum_E \binom{n}{n\mathcal{D}} \Pr_{uv \sim P^d}[P(u) = P(v)]^n &\leq \sum_E \binom{n}{n\mathcal{D}} (\Pr_{uv \sim P^d}[L(u) = L(v)] + (0.992)^{d/c})^n \\ &\leq \sum_E \binom{n}{n\mathcal{D}} e \cdot (\Pr_{uv \sim P^d}[L(u) = L(v)])^n \\ &\leq 2^{2n} e \cdot \Pr_{f_L, x, y}[f_L(x) = f_L(y)] \end{aligned}$$

where we used the fact that  $d \geq 130c \log n$ . Here,  $f_L$  is Goldreich's function based on the linear predicate  $L$ .

We now consider the part of the summation coming from  $F$ :

$$\sum_F \binom{n}{n\mathcal{D}} \Pr_{uv \sim P^d}[P(u) = P(v)]^n \leq \sum_F \binom{n}{n\mathcal{D}} \leq O(n^3) \cdot \max_{n\mathcal{D} \in F} 2^{nH(\mathcal{D})}$$

where  $H(\mathcal{D})$  is the entropy of the distribution  $\mathcal{D}$ . By definition of  $F$ , at least one of the probabilities in  $\mathcal{D}$  is at least  $5/6$ . Under this constraint, the entropy is maximized for the distribution  $\mathcal{D} = (5/6, 1/18, 1/18, 1/18)$ , and  $H(\mathcal{D}) \leq 0.92$ . So the summation inside  $F$  is upper bounded by  $O(n^3 \cdot 2^{0.92n})$ .

Putting everything together, we obtain that

$$\Pr_{G,x,y}[f_{G,P}(x) = f_{G,P}(y)] \leq e \cdot \Pr_{G,x,y}[f_{G,L}(x) = f_{G,L}(y)] + O(n^3 \cdot 2^{-1.08n}).$$

Since  $L$  is a linear predicate, the above collision probability is simply  $2^n \cdot \mathbb{E}[2^{-\text{rank}(M)}]$ , where  $M$  is a random  $\{0, 1\}$  matrix where each row is chosen from the following process: First, sample with repetition  $d/c$  entries from  $[n]$ ; set those entries that appear an odd number of times to 1, and the other entries to 0. Blömer, Karp, and Welzl [?] show that when  $M$  is a matrix whose entries are chosen independently at random with probability  $(\log n)/n$ ,  $\mathbb{E}[2^{-\text{rank}(M)}] = O(1)$ . Using a similar calculation, we can show the statement holds for our distribution as well. Putting everything together, we obtain that

$$\Pr_{f_P,x,y}[f_P(x) = f_P(y)] = O(2^{-n}) + O(n^3 \cdot 2^{-1.08n}) \leq K \cdot 2^{-n}$$

where  $K$  is a sufficiently large constant. □

**Acknowledgements.** We thank the TCC'11 reviewers for useful comments.

## References

- [1] Noga Alon, Joel Spencer. *The Probabilistic Method*. John Wiley, 1992.
- [2] Benny Applebaum, Boaz Barak, Avi Wigderson. Public-key cryptography from different assumptions. In *STOC 2010*. Pages 171-180, 2010.
- [3] Benny Applebaum, Yuval Ishai, Eyal Kushilevitz. Cryptography in NC0. In *FOCS 2004*. Pages 166-175, 2004.
- [4] Benny Applebaum, Yuval Ishai, Eyal Kushilevitz. Cryptography with Constant Input Locality. In *CRYPTO 2007*. Pages 92-110, 2007. Journal version in *J. Cryptology* 22(4). Pages 429-469, 2009.
- [5] Avrim Blum, Adam Kalai, Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. In *STOC 2000*. Pages 435-440, 2000.
- [6] Andrej Bogdanov, Youming Qiao. On the Security of Goldreich's One-Way Function. In *APPROX-RANDOM 2009*. Pages 392-405, 2009.

- [7] James Cook, Omid Etesami, Rachel Miller, and Luca Trevisan. Goldreich's One-Way Function Candidate and Myopic Backtracking Algorithms. In *TCC 2009*. Pages 521-538, 2009.
- [8] Oded Goldreich. Candidate One-Way Functions Based on Expander Graphs. In *Electronic Colloquium on Computational Complexity (ECCC)*, 7(90), 2000.
- [9] Oded Goldreich, Russell Impagliazzo, Leonid A. Levin, Ramarathnam Venkatesan, David Zuckerman. Security Preserving Amplification of Hardness. In *FOCS 1990*. Pages 318-326, 1990.
- [10] Oded Goldreich, Hugo Krawczyk, Michael Luby. On the Existence of Pseudorandom Generators. In *SIAM J. Comput.* 22(6). Pages 1163-1175, 1993.
- [11] Iftach Haitner, Danny Harnik, Omer Reingold. On the Power of the Randomized Iterate. In *CRYPTO 2006*. Pages 22-40, 2006.
- [12] Russell Impagliazzo, Leonid Levin, Michael Luby. Pseudo-random Generation from One-way Functions. In *STOC 1989*. Pages 12-24, 1989.
- [13] Russell Impagliazzo, Michael Luby. One-way Functions are Essential for Complexity Based Cryptography. In *FOCS 1989*. Pages 230-235, 1989.
- [14] Yishay Mansour, Noam Nisan, Prasoorn Tiwari. The Computational Complexity of Universal Hashing. In *Theor. Comput. Sci.* 107(1). Pages 121-133, 1993.
- [15] Henry C. Lin, Luca Trevisan, Hoeteck Wee. On Hardness Amplification of One-Way Functions. In *TCC 2005*. Pages 34-49, 2005.
- [16] Andrew Chi-Chih Yao. Theory and Applications of Trapdoor Functions (Extended Abstract). In *FOCS 1982*. Pages 80-91, 1982.

# Appendix

## A On the input locality requirement

Theorem 3.1 gives a non-trivial result only when the sum of the squares of the input degrees  $D$  is at most  $o(n^2/\log n)$ . This assumption could be violated even if there is a single input of  $f$  whose degree is  $\Omega(n)$ . It is natural to ask if the self-amplification argument could be modified so as to allow for a small number of inputs that have unusually large degree.

We argue that this is unlikely to be the case. We show that if non-trivial self-amplification can be achieved for functions where all but one of their inputs have degree at most  $d + 1$ , then every function of input locality  $d$  has a non-trivial inversion algorithm.

We say a family of functions  $f_n: \{0, 1\}^n \rightarrow \{0, 1\}^{m(n)}$  is  $(t(n), \varepsilon(n))$  *self-amplifying* if the following holds: There exists a constant  $c$  such that for every sufficiently large  $n$ , if  $f_n$  can be inverted on an  $\varepsilon(n)$  fraction of inputs in time  $t(n)$ , then  $f_n$  can be inverted on a  $1 - \varepsilon(n)$  fraction of inputs in time  $t(n)^c$ .

Theorem 3.1 implies that if  $f$  has input locality  $d$ , then  $f$  is  $(\exp(O(\sqrt{rn} \cdot d \log n)), e^{-r})$  self-amplifying for every  $r > 0$ .

The following claim indicates that the input locality requirement is necessary in a strong sense. We say that  $f$  has input locality almost  $d$  if all but one of its inputs have degree at most  $d$ .

**Claim 6.** *Suppose that every function with input locality almost  $d + 1$  is  $(t(n), \varepsilon)$  self-amplifying for some  $\varepsilon < 1/2$ . Then every function family with input locality  $d$  can be inverted on a  $1 - 2\varepsilon$  fraction of inputs in time  $t(n + 1)^c + O(n)$  for some constant  $c$ .*

*Proof.* Let  $f_n: \{0, 1\}^n \rightarrow \{0, 1\}^{m(n)}$  be a function family with input locality  $d$ . We define  $f'_n: \{0, 1\}^{n+1} \rightarrow \{0, 1\}^{m(n)+1}$  by

$$f'_{n+1}(x, b) = \begin{cases} (x, 0), & \text{if } b = 0 \\ (f(x), 1), & \text{if } b = 1. \end{cases}$$

Notice that except for the input  $b$ , the degree of every other input in  $f'_{n+1}$  is larger than the locality of the corresponding input in  $f_n$  by at most one. Therefore,  $f'_{n+1}$  has input locality almost  $d + 1$ .

By assumption,  $f'_{n+1}$  is  $(t(n + 1), \varepsilon)$  self-amplifying. However,  $f'_{n+1}$  is trivially invertible on half its inputs. Since  $\varepsilon < 1/2$ , it follows that  $f'_{n+1}$  is invertible on a  $1 - \varepsilon$  fraction of its inputs in time  $t(n)^c$ . But then  $f_n$  is invertible on a  $1 - 2\varepsilon$  fraction of inputs in time  $t(n + 1)^c + O(n)$  as follows. Given an inverter  $I'$  that inverts  $f'$  on a  $(1 - \varepsilon)$ -fraction of inputs, the following inverter  $I$  inverts  $f$  on a  $(1 - 2\varepsilon)$ -fraction of inputs: On input  $y$ , output the first  $n$  bits of  $I'(y, 1)$ .  $\square$