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 $f\colon \{0,1\}^n \to \{0,1\}^m$ 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 ε-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 $g\colon \{0,1\}^n \to \{0,1\}^{m+O(n)}$ 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. Research supported by RGC GRF grant 2150617.

†Efi Arazi School of Computer Science, IDC Herzliya. E-mail: 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 ε 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 ε and $1-\delta$.

The most basic method for amplifying hardness is due to Yao [18]. 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 [11]. 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 [13, 12]. 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 a 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 parallel complexity (more generally, one-way functions computable in logspace) 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 [10]. 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 these functions (as long as the output length is linear in the input length).

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. [17]). 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 $f\colon \{0,1\}^n \to \{0,1\}^m$ 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 ε-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 $f_n\colon \{0,1\}^n \to \{0,1\}^m$ 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 ε 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 $f\colon \{0,1\}^n \to \{0,1\}^m$ has input locality d, and that f cannot be inverted in time $\exp(\tilde{O}(\sqrt{n}\cdot d))$ on an ε-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 $g\colon \{0,1\}^n \to \{0,1\}^{m+O(n)}$ 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 subexponential time attacks are known is Goldreich's one-way function [10]. Given a bipartite graph G with n vertices on the left, m vertices on the right, and regular right-degree d_{out} and given a predicate $P\colon \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}$, the function $f_{G,P}\colon \{0,1\}^n \to \{0,1\}^m$ is defined by setting the i^{th} bit of $f_{G,P}(x)$ to be equal to $P(x_{\Gamma(i,1)}, \ldots, x_{\Gamma(i,d_{\mathsf{out}})})$, where $\Gamma_{(i,j)}$ is the j^{th} neighbor of right vertex i of G. Goldreich proposed setting $m = n$ and considered d_{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_{out}-*parity-blowup* predicates, and assume that G is chosen at random.

Theorem 6.3 (Amplification for Goldreich's function): *For at least half the graphs G, if $f_{G,P}$ is hard to invert by circuits of size $\exp(\tilde{O}(\sqrt{n}))$ on an ε-fraction of inputs, then $f_{G,P}$ is hard to invert by circuits of size $\exp(\tilde{O}(\sqrt{n}))$ on a $(1-\varepsilon)$-fraction of inputs.*

By observing that parity-blowup predicates can be represented by constant degree polynomials over $\mathbb{F}_2$ 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_{out} is constant and $m \geq Dn$, where $D = D(d_{\mathsf{out}})$ is a sufficiently large constant. Invoking a recent result of Bogdanov and Qiao [7], 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 ε 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 [6].

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 $\tilde{\Omega}(1/\varepsilon)$.

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 ε-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 $f\colon \{0,1\}^n \to \{0,1\}^m$ is not 1-1 but regular (i.e. K-to-1 for some K), then there exists a new function $f'\colon \{0,1\}^n \to \{0,1\}^{m'}$, $m' = m + O(n)$ such that if f is hard on a 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 [15, 14]. 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 Lemma 5.3 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 ε-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 [7], we then recover an inverse for $f(x')$.

Our second application concerns functions $f\colon \{0,1\}^n \to \{0,1\}^m$ 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). We prove that such functions are likely to be at most K-to-1 for some constant

K. We obtain a self-amplification for this function by a simple extension of Theorem 3.1. 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 $f\colon \{0,1\}^n \to \{0,1\}^m$ be a function. We say that the ith output $f(x)_i$ *depends* on the jth input x_j if there exists a setting of the inputs $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ such that $f(x_1, \ldots, x_{j-1}, 0, x_{j+1}, \ldots, x_n)_i \neq f(x_1, \ldots, x_{j-1}, 1, x_{j+1}, \ldots, x_n)_i$. We define the *degree* of the jth input to be the number of outputs that depend on the jth 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 *ε-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 $f\colon \{0,1\}^n \to \{0,1\}^m$ is *invertible by circuit size s on an α-fraction of inputs* if there exists a circuit C of size at most s such that $f(C(f(x))) = f(x)$ for at least $\alpha \cdot 2^n$ inputs $x \in \{0,1\}^n$.

In the uniform setting, a function family $f = \{f_n \colon \{0,1\}^n \to \{0,1\}^{m(n)}\}$ is *invertible in (randomized) time $t(n)$ on an $\alpha(n)$-fraction of inputs* if there exists a (randomized) algorithm A that runs in time $t(n)$ such that $f_n(A(1^n, f_n(x))) = f_n(x)$ 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) for infinitely many n. (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 $f \colon \{0,1\}^n \to \{0,1\}^m$ be any 1-1 function, and let d_j be the degree of the jth input. Set $\Delta = \sum_{j=1}^n d_j^2$.

Theorem 3.1. *Let $f = \{f_n \colon \{0,1\}^n \to \{0,1\}^{m(n)}\}$ be a 1-1 function family. Suppose f can be inverted in time $\exp(O(\sqrt{r\Delta}\log n))$ on a e^{-r}-fraction of inputs ($r = r(n)$). Then f can be inverted in time $\exp(O(\sqrt{r\Delta}\log n))$ on a $(1-e^{-r})$-fraction of inputs.*[1]

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.

We prove the theorem for deterministic time inverters; the extension to randomized time inverters is straightforward.

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 $\sqrt{2rn}d$ coordinates. Therefore we can invert f at $y = f(x)$ by flipping the given set of $\sqrt{2rn}d$ 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.

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 $d_1, \ldots, d_n$ be positive numbers. Let*

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

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

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

To prove Theorem 3.1, the special case where all d_j equal 1 and $p = 1/2$ suffices. We will need the more general form for later applications.

Proof. For z in $\{0,1\}^n$ equipped with the p-biased distribution, let

$$d(z) = \min_{x \in S} \sum_{j \in [n]\colon x_j \neq z_j} d_j.$$

We can think of $d(z)$ as the ℓ_1 distance between z and the set X, where the distance in coordinate j is scaled by d_j. As we expose the coordinates of $z_1, z_2, \ldots, z_n$ one by one, the sequence $\mathrm{E}[d(z)], \mathrm{E}[d(z) \mid z_1], \ldots, \mathrm{E}[d(z) \mid z_1, \ldots, z_n]$ defines a martingale with

$$\left|\mathrm{E}[d(z) \mid z_1, \ldots, z_j] - \mathrm{E}[d(z) \mid z_1, \ldots, z_{j-1}]\right| \leq d_j$$

for every j and $z_1, \ldots, z_j$, since changing z_j from 0 to 1 can change $d(z)$ by at most d_j. By Azuma's inequality (see for instance [9, Theorem 5.1]), for every $t > 0$

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

Setting $t = \mathrm{E}[d(z)]$ in the first inequality we get $e^{-r} = \Pr[d(z) = 0] < e^{-2t^2/\Delta}$, and therefore $\mathrm{E}[d(z)] < \sqrt{r\Delta/2}$. Setting $t = \sqrt{r\Delta/2}$ in the second inequality we conclude that

$$\Pr[z \notin Z] = \Pr[d(z) \geq \sqrt{2r\Delta}] \leq \Pr[d(z) \geq \mathrm{E}[d(z)] + \sqrt{r\Delta/2}] < e^{-r}. \qquad \square$$

Alternatively, Lemma 3.2 can be easily derived from Talagrand's inequality.

Proof of Theorem 3.1. Let $\varepsilon = e^{-r}$. We prove the contrapositive. Assume A inverts f_n on an ε-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 any one of these y' satisfies $f_n(x') = y'$, take this y' and 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

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

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 jth input. Now let

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

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 extends to non-injective families, as long as the preimage size is not too large: If f is at most e^r-to-one and can be inverted in time $\exp(O(\sqrt{r\Delta}\log n))$ on an e^{-r} fraction of inputs, then it can be inverted in time $\exp(O(\sqrt{r\Delta}\log n))$ on an $1-e^{-r}$ fraction of inputs.

Remark 2. Theorem 3.1 and Remark 1 also generalize 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 $f_n\colon \{0,1\}^n \to \{0,1\}^{m(n)}$ 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 e 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 $\{f_n\colon f_n(x,e) = M_n x + e\}$ is 1-1 and $r < pm/10$. If $\{f_n\}$ can be inverted in randomized time $\exp(O(\sqrt{rm}\log m))$ on a e^{-r} fraction of inputs, then it can be inverted in randomized time $\exp(O(\sqrt{rm}\log m))$ on a $1-e^{-r}$ fraction of inputs.*

The idea of the proof is to use the random self-reducibility of $f_n(x,e)$ in terms of the first parameter x. The success of the inverter on an input (x,e) is essentially independent of x: If for a fixed e the inverter succeeds on a non-negligible fraction of x, by re-randomizing it can be made to succeed on almost all x. Once x is taken out of consideration, $f_n(x,e)$ becomes an input-local function of e and we can follow the lines of the proof of Theorem 3.1.

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:

On input $y \in \{0,1\}^{m(n)}$,
    For all $y' \in \{0,1\}^{m(n)}$ that are within hamming distance $\sqrt{2(r+3)m}$ from y:
        Repeat the following $8/\varepsilon$ times:
            Choose a random $z \in \{0,1\}^n$.
            If $A(y' + M_n z) = (u, e') \neq \perp$ and $e' + (y' + y)$ has hamming weight at most $2pm$, output $(u + z, e' + (y' + y))$.
    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 $A(M_n x + e') = (x, e')$ with probability at least $1/2$ over the randomness of A. 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,

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

By conditioning, $\Pr_{e' \sim \mathcal{R}}[E] \geq \Pr_{e' \sim \mathcal{R}_c}[E] \cdot \Pr_{e' \sim \mathcal{R}}[\mathrm{wt}(e') \leq 2pm]$ for any event E and so

$$\Pr_{e' \sim \mathcal{R}}\left[\Pr_{x \sim \mathcal{U}}[(x, e') \in S] \geq \frac{\varepsilon}{2}\right] \geq \frac{\varepsilon}{2} \cdot \Pr_{e' \sim \mathcal{R}}[\mathrm{wt}(e') \leq 2pm] \geq \frac{\varepsilon}{2} \cdot (1 - e^{-\Omega(pm)}) \geq \frac{\varepsilon}{4},$$

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 $\Pr_{A,z}[A(y' + M_n z) = (x + z, e')] \geq \varepsilon/4$, 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 $\varepsilon/4$. Repeating the loop $8/\varepsilon$ 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 self-reduction argument. Suppose e' is good. Since for any uniformly random z and fixed $x \in \{0,1\}^n$, $x + z$ is also uniformly random, $(x + z, e')$ is in S with probability at least $\varepsilon/2$ over a uniformly random choice of z. It follows that $A(M_n(x + z) + e') = (x + z, e')$ with probability at least $\varepsilon/2$ over the choice of z and $1/2$ over the randomness of A, so with probability at least $\varepsilon/4$ altogether.

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

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

by a Chernoff bound since $r < pm/10$, when n is sufficiently large. □

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 $f\colon \{0,1\}^n \to \{0,1\}^m$ is not 1-1 but regular, then there exists a new function $f'\colon \{0,1\}^n \to \{0,1\}^{m'}$, $m' = m + O(n)$ such that if f is hard on a 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. *For every K-to-1 function $f\colon \{0,1\}^n \to \{0,1\}^m$ with input locality d there exists a function $f'\colon \{0,1\}^n \to \{0,1\}^{m+\log_2 K+O(r)}$, with input locality $d + O(r + (\log n)^3)$ so that if 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 inputs, then f can be inverted by a circuit 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 [15, 14]. 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 2).

The construction of f' 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. *For every regular function $f\colon \{0,1\}^n \to \{0,1\}^m$ with input locality d_{in} and output locality $d_{\mathsf{out}} \geq 3$ there exists a function $f'\colon \{0,1\}^{n'} \to \{0,1\}^{m'}$, where $n' = O(n(\log n)^3)$ and $m' = m + O(n(\log n)^3)$ with output locality d_{out} so that if f' can be inverted by circuits of size $\exp(O(\sqrt{rn} \cdot \alpha(d_{\mathsf{in}}, r, \log n)))$ on a e^{-r} fraction of inputs, then f can be inverted by circuits of size $\exp(O(\sqrt{rn} \cdot \alpha(d_{\mathsf{in}}, r, \log n)))$ 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 $s + O(n(\log n)^3)$.*

We construct and analyze the hash function family with small input locality in Section 5.1 and prove Theorem 5.1 and Corollary 5.2 in Section 5.2.

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 $h\colon \{0,1\}^n \to \{0,1\}^{\log_2 K+O(1)}$ 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. [16] 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 pair-

wise independence property they establish for their construction, while sufficient for their application, appears too weak to derive Lemma 5.3.

Lemma 5.3. *Suppose $f\colon \{0,1\}^n \to \{0,1\}^m$ is at most K-to-1, where $2^{k-1} \le K < 2^k$. Then there exists a function $h\colon \{0,1\}^n \to \{0,1\}^{k+3r+3}$ 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 $\mathbb{F}_2^n$ with input locality $O(r) + \min\{k, O((\log n)^3)\}$.*

To explain our construction, fix some $x \in \{0,1\}^n$ and consider the set S of its siblings x' such that $f(x) = f(x')$. Now look at the one-bit linear hash $h(x) = a \cdot x$, where $a \in \{0,1\}^n$ and $a \cdot x$ is the inner product of a and x modulo 2. For a random a', in expectation half of the x' in S satisfy $h(x') = 0$ and half satisfy $h(x') = 1$. If a was uniformly random, pairwise independence and Chebyshev's inequality would give that a roughly half-half split occurs with high probability, and so a random one-bit hash is very likely to halve the size of S. So after about k independent applications of the one-bit hash, x is likely to be isolated from all its siblings.

The resulting hash function is not input local because a uniformly random a is likely to be dense. We would like to choose a to be a sparse random vector instead. In such a case the pairs $(h(x_1), h(x_2))$ where $x_1, x_2 \in S$ will not be independent anymore. However, if S is sufficiently large and x_1', x_2' are "typical", then they will differ in many coordinates so even for a random sparse a the values $a \cdot x_1$ and $a \cdot x_2$ will have low correlation, which turns out sufficient for the application of Chebyshev's inequality.

A technical side note: As the number of siblings x' of x shrinks in every application of the one-bit hash, we gradually increase the density of a as we apply more one-bit hashes. The cost of this increase in density is a factor of $O(\log n)$ in the input locality of h.

We now give details of the probabilistic construction of h and prove Lemma 5.3. Assume that f is at most K-to-1, where $2^{k-1} \le K < 2^k$.

Construction of h. The function h has the form $h(x) = (h_a(x), h_b(x))$ where

$$h_a(x) = (a_k \cdot x, a_{k-1} \cdot x, \ldots, a_{k_0+1} \cdot x)$$
$$h_b(x) = (b_1 \cdot x, b_2 \cdot x, \ldots, b_{3r+k_0+3} \cdot x).$$

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 $\{0,1\}^n$.

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 inputs x uniquely.

The first stage. If $k_0 = k$, the conclusion of the first stage is vacuous, so let us assume that $k_0 < k$. Let us fix an input x and let $S = \{x' \colon f(x) = f(x')\}$. Without loss of generality, we may assume that $2^{k-1} \le |S| < 2^k$. (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: $S_k = S$, $T_i = \{x' \in S_i \colon a_i \cdot x' = a_i \cdot x)\}$ and

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

Here is the intuition for this definition: We want to think of the ith 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 $|S_i| \ge (1-1/n)^{k-i} \cdot 2^{i-1} \ge 2^{i-3}$. So we are left with the following question: Given a set S of size at least 2^{i-3}, how likely is it to be split successfully at stage i?

Lemma 5.4. *Assume $|R| \ge 2^{i-3}$. Let a be chosen from the p-biased distribution on $\{0,1\}^n$, where $p = 4(\log n)^2/i < 1/2$. Then for n sufficiently large and any $\varepsilon > 0$,*

$$\Pr\bigl[\#\{y \in R \colon a \cdot y = 0\} \notin (1 \pm \varepsilon)|R|/2\bigr] \le \frac{1}{n^4 \varepsilon^2}.$$

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 $\binom{n}{r} \cdot (1/n^2)^r \le n^{-r} < e^{-r}/8$. So for any $x \in \{0,1\}^n$,

$$\Pr\bigl[|S_{k_0}| \ge 2^{k_0 + r}\bigr] \le e^{-r}/8.$$

Proof of Lemma 5.4. Let $X = \sum_{y \in R} (-1)^{a \cdot y + a'}$, where a' is a uniformly random bit. Notice that X counts the imbalance between those $y \in R$ that are hashed to 0 and 1, respectively. By linearity of expectation and uniformity of a', $\mathrm{E}[X] = 0$. We will upper bound the second moment of X and use Chebyshev's inequality to argue concentration.

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

Here, $R_z = \{y + z \colon y \in R\}$, and $|y|$ is the hamming weight of y. By Hölder's inequality

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

for every $K \geq 1$, where 1_{R_z} is the indicator for R_z. The last expression simplifies to give

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

and so

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

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

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

because $p = 4(\log n)^2/i$ and $|R| \geq 2^{i-3}$. By Chebyshev's inequality, it follows that

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

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:

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

where the probability is over the choice of b_i, $1 \leq i \leq 3r + k_0 + 3$. Putting the analysis of both stages together, it follows that by the end of stage 2, for any specific x,

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

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

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

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 $\mathrm{E}[Y_a] = \sum_{i=k_0}^{k} 4(\log n)^2/i \leq 4(\log n)^3$ and $\mathrm{E}[Y_b] = (3r + k_0 + 3)/2$, so $\mathrm{E}[Y_a + Y_b] = O((\log n)^3 + r)$. 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.

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 1. *Assume $f\colon \{0,1\}^n \to \{0,1\}^m$ is K-to-1 where $2^{k-1} \le K < 2^k$. Let f' and h be as in Lemma 5.3. 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 1, let us show how it implies Theorem 5.1 and Corollary 5.2.

Proof of Theorem 5.1. Suppose $f\colon \{0,1\}^n \to \{0,1\}^m$ 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 Lemma 5.3. 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 Lemma 5.3, f' is $e^{-r}/2$-close to 1-1. By Theorem 3.1 and Remark 2, 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 1, f can then be inverted on a $(1 - e^{-r})$ fraction of inputs by circuits of size $\exp(O(\sqrt{rn} \cdot \alpha(d, r, \log n)))$. □

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 ith output $h(x)_i$ has the form

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

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

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

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

Applebaum et al. [3, 4] show that if f'' can be inverted on an ε-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 $s = \exp(O(\sqrt{rn} \cdot \alpha(d_{\mathsf{in}}, r, \log n)))$, the corollary follows. □

Proof of Claim 1. Let $\varepsilon = e^{-r}$ and let A' be circuit that inverts f' on a $(1-\varepsilon/400)$-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 $y \in \{0,1\}^m$ is *bad* if the number of hash values of preimages of y that A' inverts successfully is small:

$$\#\{t\colon t = h(x),\ f(x) = y, \text{ and } A'(f'(x)) \in f'^{-1}(f'(x)) \text{ for some } x \in \{0,1\}^n\} < 2^{k-6}.$$

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-6} 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:

$$Y_1 = \{y\colon \#\{t\colon t = h(x) \text{ and } f(x) = y \text{ for some } x\} < 2^{k-5}\}$$
$$Y_2 = \{y\colon \#\{t\colon t = h(x) \text{ and } A'(f'(x)) \notin f'^{-1}(f'(x)) \text{ for some } x\} \geq 2^{k-6}\}.$$

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 inputs x 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 $x \neq x'$.

It follows that there are at least $(15/16)K \cdot |Y_1|$ values of x on which f' is not 1-1. By Lemma 5.3, 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 $|Y_2| \leq \frac{1}{5}\varepsilon \cdot (2^n/K)$.

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

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 [10]. Given a bipartite graph G with n vertices on the left, m vertices on the right, and regular right-degree d_{out} and a predicate $P\colon \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}$, the function $f_{G,P}$ from $\{0,1\}^n$ to $\{0,1\}^m$ is defined by

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

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

Goldreich [10] considered such constructions for the setting of parameters $m = n$ and d_{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. [8] 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 [7] studied a variant of Goldreich's function in the setting where G is random, d is constant, and $m = Dn$, where $D = D(d_{\mathsf{out}})$ 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 $D = D(d_{\mathsf{out}})$ is a sufficiently large constant, for any P and with high probability over the choice of G if $f_{G,P}$ can be inverted on an e^{-r} fraction of inputs in time $\exp(O(\sqrt{rn} \cdot d_{\mathsf{out}} \cdot \log n))$, then it can be inverted on a $1 - e^{-r}$ fraction of inputs in time $\exp(O(\sqrt{rn} \cdot d_{\mathsf{out}} \cdot \log n))$.

2. When $d_{\mathsf{out}} = O(\log n)$ and $m = n$, for a certain class of predicates P and with high probability over G, if $f_{G,P}$ can be inverted on a e^{-r} fraction of inputs in time $\exp(O(\sqrt{rn}) \log n))$, then it can be inverted on a $1 - e^{-r}$ fraction of inputs in time $\exp(O(\sqrt{rn}) \log n)$.

   Our result applies to all $O(\log n)$-*parity-blowup* predicates, which we define as follows. Let $P_c\colon \{0,1\}^c \to \{0,1\}$ be any balanced predicate, where c is some constant. The d_{out}-parity-blowup of P_c is the predicate $P\colon \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}$ which is obtained by replacing each of the variables in P_c by a parity of $\lfloor d_{\mathsf{out}}/c \rfloor$ 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 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_{out} neighbors independently at random among the n left vertices of G. We will call such graphs (n, m, d_{out}) *random graphs*.

6.1 Self-reducibility for functions with long output

Theorem 6.1. *Let $D \geq 2^{Kd_{\mathsf{out}}}$ where K is a sufficiently large constant, and $P\colon \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}$ be any predicate. Let G be an $(n, m = Dn, d_{\mathsf{out}})$ random graph. With probability at least $1 - o(1)$ over the choice of G, if $f_{G,P}$ can be inverted by circuits of size $\exp(O(\sqrt{rn} \cdot Dd_{\mathsf{out}} \log n))$ on an e^{-r}-fraction of inputs, then $f_{G,P}$ can be inverted by circuits of size $\exp(O(\sqrt{rn} \cdot Dd_{\mathsf{out}} \log n))$ 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 have input bits that do not affect the 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 ith 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 ε-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 [7] which says that for most inputs x, given x' that is correlated with x, we can invert $f_{G,P}(x)$.

Claim 2. *Assume that $D > 3 \cdot 6^{d_{\mathsf{out}}}$. Let $P\colon \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}$ be any nonconstant predicate. Consider the function $f_{G,P}\colon \{0,1\}^n \to \{0,1\}^m$, 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 [7])**.** *Let K be a sufficiently large constant and $D > 2^{-Kd_{\mathsf{out}}}$. Let $P : \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}$ be any predicate. Then there is an algorithm I such that with probability $1 - o(1)$ over the choice of G, the following holds. Consider the function $f_{G,P} : \{0,1\}^n \to \{0,1\}^m$, where $m = Dn$. For a $1-2^{-n \cdot 2^{-\Omega(d_{\mathsf{out}})}}$ fraction of assignments x and any assignment x' such that x and x' agree on at least a $5/9$ fraction of coordinates, on input $G, P, f(x)$ and x', I outputs an inverse for $f_{G,P}(x)$. The running time of I 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 $d_{\mathsf{out}} = d$.

Claim 3. *Let G be an $(n, m = Dn, d)$ random graph. Assume $Dd > 6$. Let d_i denote the degree of left vertex i and $\Delta = d_1^2 + \cdots + d_n^2$. Then*

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

The proofs of Claims 2 and 3 are given at the end of this section. We first show how together with Theorem 6.2 they imply Theorem 6.1.

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 2, Theorem 6.2 and Claim 3 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 $f(x')$ and

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

We will argue that B successfully inverts $f(x)$ when x is 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 $5n/9$ coordinates. By Theorem 6.2, all but a $2^{-2^{Kd}n} < e^{-r}/2$ fraction of inputs x have this property. Consider such an x. By definition of S, there exists an $x' \in S'$ such that $\sum_{j \in [n] \colon x'_j \neq x_j} d_j \leq \sqrt{2(r+2)\Delta}$. 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 2, 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 $\overline{x^*}$ differ on at most $2n/3 + \sqrt{2(r+2)\Delta} \leq 4n/9$ 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 $f(x)$. □

Proof of Claim 2. 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'll show that if $1/3 \leq p_{00} + p_{11} \leq 2/3$, then $\Pr_G[f(x) = f(x')] \leq 2^{-3n}$. By a union bound, it will follow that with probability $1 - 2^{-n}$ over the choice of G, no pair x, x' satisfying $1/3 \leq p_{00} + p_{11} \leq 2/3$ (i.e., agreeing on more than $n/3$ but less than $2n/3$ coordinates) maps to the same output, proving the claim.

Because $p_{00} + p_{11} \geq 1/3$, one of them, call it $p_{aa'}$ is at least $1/6$. On the other hand, $p_{00} + p_{11} \leq 2/3$ gives that $p_{01} + p_{10} \geq 1/3$, so at least one of p_{01} and p_{10}, call it $p_{bb'}$ is at least $1/6$. Observe that aa' and bb' differ in exactly one coordinate.

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 ith output of f when G is chosen at random. The probability that $(x_{\Gamma(i,1)}, \ldots, x_{\Gamma(i,d)}) = 0^d$ and $(x'_{\Gamma(i,1)}, \ldots, x'_{\Gamma(i,d)}) = z$ 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

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

as long as $m \geq 3 \cdot 6^d \cdot n$. □

Proof of Claim 3. For vertex $i \in [n]$ and candidate edge $e \in [Ddn]$, let X_{ie} be an indicator random variable for the event that edge e is incident to vertex i, so that $\mathrm{E}[X_{ie}] = 1/n$. Then $d_i = \sum_{e \in [Ddn]} X_{ie}$, and the variables X_{ie} and $X_{i'e'}$ are independent whenever $e \neq e'$. Therefore for every i,

$$\mathrm{E}[d_i^2] = \mathrm{E}[d_i]^2 + \mathrm{Var}[d_i] \leq \mathrm{E}[d_i]^2 + \mathrm{E}[d_i] = (Dd)^2 + Dd$$

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

$$\begin{aligned}
\mathrm{Var}[\Delta^2] &= \mathrm{Var}\Big[\sum\nolimits_{i \in [n]} \Big(\sum\nolimits_{e \in [Ddn]} X_{ie}\Big)^2\Big] \\
&= \mathrm{E}\Big[\sum_{i,i',e,e',f,f'} X_{ie}X_{if}X_{i'e'}X_{i'f'}\Big] - \mathrm{E}\Big[\sum_{i,e,f} X_{ie}X_{if}\Big]\,\mathrm{E}\Big[\sum_{i',e',f'} X_{i'e'}X_{i'f'}\Big] \\
&= \sum_{i,i',e,e',f,f'} \big(\mathrm{E}[X_{ie}X_{if}X_{i'e'}X_{i'f'}] - \mathrm{E}[X_{ie}X_{if}]\,\mathrm{E}[X_{i'e'}X_{i'f'}]\big),
\end{aligned}$$

where i, i' range over $[n]$, and e, e', f, f' range over $[Ddn]$. When $i \neq i'$, the expression inside the sum is zero or negative. When $i = i'$, by independence, the only terms in the summation that may not vanish are those where at least two of e, f, e', f' are equal, or $|\{e, f, e', f'\}| < 4$. For any fixed i, There are at most $\binom{4}{t}(Ddn)^t$ terms where $|\{j, k, j', k'\}| \leq t$, and for each such term

$$\mathrm{E}[X_{ie}X_{if}X_{i'e'}X_{i'f'}] \leq \mathrm{E}[X_{11}]^t = n^{-t}$$

from where

$$\mathrm{Var}[\Delta^2] \leq n \cdot \sum_{t=0}^{3} \binom{4}{t} (Ddn)^t n^{-t} \leq 4(Dd)^3 + 6(Dd)^2 + 4Dd \leq 8(Dd)^3 n.$$

Applying Chebyshev's inequality to Δ, we obtain the desired statement. □

6.2 Self-reducibility for certain length-preserving functions

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

Theorem 6.3. *Let $c \geq 3$ and $d_{\mathsf{out}} = \max\{130c\lceil \log n \rceil, 4c^2\}$. Let P be the d_{out}-parity-blowup of some balanced predicate on c bits and let G be an (n, n, d_{out})-random graph. Then*

1. *For at least half the graphs G, if $f_{G,P}$ can be inverted on a e^{-r} fraction of inputs by circuits of size $\exp(O(\sqrt{rn}\log n))$, then it can be inverted on a $1-e^{-r}$ fraction of inputs by circuits of size $\exp(O(\sqrt{rn}\log n))$.*

2. *There exist functions* $f'_{G,P}\colon \{0,1\}^{n'} \to \{0,1\}^{m'}$, *where* $n', m' = O(n \cdot d_{\mathsf{out}}^c)$, *every output of* f' *depends on at most* $c+1$ *inputs and for at least half the graphs* G, *if* $f'_{G,P}$ *can be inverted on a* e^{-r} *fraction of inputs by circuits of size* $\exp(O(\sqrt{rn}\log n))$, *then* $f_{G,P}$ *can be inverted on a* $1-e^{-r}$ *fraction of inputs by circuits of size* $\exp(O(\sqrt{rn}\log n))$.

In Claim 4 we show that the function $f_{G,P}$ is likely to be $e^{-r}/4$-close to $O(e^r)$-to-1. Part 1 of Theorem 6.3 then follows from Remark 1 on Theorem 3.1. To prove part 2, we observe that parity-blowup predicates can be represented by constant degree polynomials over $\mathbb{F}_2$. 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 techincal tool is the following claim which bounds the collision probability of $f_{G,P}$. For convenience of notation, from now on we set $d_{\mathsf{out}} = d$.

Claim 4. *Assume* $P\colon \{0,1\}^d \to \{0,1\}$ *is* c*-Fourier-sparse and* $d \geq 130c\log n$ *be an even multiple of* c. *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 $P\colon \{0,1\}^c \to \{0,1\}$ is c-Fourier-sparse, whenever $d \geq 4c^2$.

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

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

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. Part 1 then follows from Theorem 3.1 and Remarks 1 and 2.

To prove part 2, we apply the randomized encodings of Applebaum et al. [4] to every output of f to construct the function f''. Since P is a d_{out}-parity blowup of a predicate on c bits, each output bit of f is a $\mathbb{F}_2$ polynomial of degree c that depends on d_{out} variables. We write out each such polynomial f_i as a sum of monomials $p_i = m_{i1} + m_{i2} + \cdots + m_{it}$. Then $t \leq d(n)^c$. We replace each f_i by the randomized encoding

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

where $r_{i1}, \ldots, r_{i(t+1)}$ are new input variables. It is easy to check that f' has input and output length $O(n \cdot d_{\mathsf{out}}^c)$ 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 $\exp(O(\sqrt{rn}\log n))$, then f can be inverted on a $1-\Omega(e^{-r})$ fraction of inputs by circuits of size $\exp(O(\sqrt{rn}\log n))$, and the conclusion follows from part 1. □

Fourier-sparse predicates We say a predicate $P\colon \{0,1\}^d \to \{0,1\}$ 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 can be verified that the d-parity-blowup of any balanced predicate $P\colon \{0,1\}^c \to \{0,1\}$ is c-Fourier-sparse, whenever $d \geq 4c^2$.

Proof of Claim 4 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. Blömer, Karp, and Welzl calculate this probability in a related random graph model [5]. We give a self-contained proof for our model.

Proof of Claim 4. Let $G = G_n$ and $P = P_n$. We rewrite the desired probability as:

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

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

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

where u, v are d-bit strings, $\binom{n}{n\mathcal{D}}$ is shorthand for $\binom{n}{n_{00},n_{01},n_{10},n_{11}} = n!/(n_{00}!n_{01}!n_{10}!n_{11}!)$, and $\mathcal{D}^d$ is the distribution on pairs of strings (u, v), $u, v \in \{0,1\}^d$ obtained by choosing each coordinate 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 $\Pr_{uv\sim\mathcal{D}^d}[P(u) = P(v)] = \Pr_{uv\sim\mathcal{D}^d}[P(u) \oplus P(v) = 0] = \frac{1}{2} + \frac{1}{2}\mathrm{E}_{uv\sim\mathcal{D}^d}[(-1)^{P(u)\oplus P(v)}]$ and

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

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

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

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

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

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

$$\begin{aligned}\sum_E \binom{n}{n\mathcal{D}} \Pr_{uv\sim\mathcal{D}^d}[P(u) = P(v)]^n &\leq \sum_E \binom{n}{n\mathcal{D}} (\Pr_{uv\sim\mathcal{D}^d}[L(u) = L(v)] + (0.992)^{d/c})^n \\ &\leq \sum_E \binom{n}{n\mathcal{D}} (\Pr_{uv\sim\mathcal{D}^d}[L(u) = L(v)] + 1/2n)^n \\ &\leq \sum_E \binom{n}{n\mathcal{D}} e \cdot (\Pr_{uv\sim\mathcal{D}^d}[L(u) = L(v)])^n \\ &\leq 2^{2n} e \cdot \Pr_{G,x,y}[f_{G,L}(x) = f_{G,L}(y)]\end{aligned}$$

Here, $f_{G,L}$ is Goldreich's function based on the linear predicate L in d/c inputs. The second inequality uses the fact that $d \geq 130c \log n$, and the third inequality follows from the fact that $\Pr_{uv\sim\mathcal{D}^d}[L(u) = L(v)] \geq 1/2$ because d/c is even.

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

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

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 $O(n^3 \cdot 2^{0.92n})$.

Summing the contributions from E and F, we obtain that

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

To finish the proof, we need to show that $\Pr_{G,x,y}[f_{G,L}(x) = f_{G,L}(y)] = O(2^{-n})$. We write

$$\Pr_{G,x,y}[f_{G,L}(x) = f_{G,L}(y)] = \Pr_{G,x}[f_{G,L}(x) = 0] = \mathrm{E}_x\Big[\Big(\frac{1}{2} + \frac{1}{2}(1 - 2\mathrm{wt}(x)/n)^{d/c}\Big)^n\Big]$$

where the first equality follows by the linearity of L, and $\mathrm{wt}(x)$ is the hamming weight of x. Grouping the inputs x according to their hamming weight w and simplifying, we obtain

$$\begin{aligned}\Pr_{G,x,y}[f_{G,L}(x) = f_{G,L}(y)] &= 2^{-2n}\sum_{w=0}^{n}\binom{n}{w}(1+(1-w/n)^{d/c})^n \\ &= 2^{-2n}\sum_{w,k=0}^{n}\binom{n}{w}\binom{n}{k}(1-w/n)^{kd/c} \\ &\leq 2^{-2n}\sum_{w,k=0}^{n}\binom{n}{w}\binom{n}{k}n^{-10wk/n}\end{aligned}$$

where in the last line we used that $1 - w/n \leq e^{-w/n}$ and $d \geq 10c\log n$.

We divide the sum into two parts. The first part consists of those pairs (w,k) such that both w and k are at most $n/10$, and the second part consists of the rest. For the first part we have

$$\sum_{w,k=0}^{n/10}\binom{n}{w}\binom{n}{k} \leq n^2 2^{2H(1/10)n+o(n)} \leq 2^n$$

for n sufficiently large, where H is the binary entropy function. For the second part

$$\begin{aligned}\sum_{w\,\geq\, n/10 \text{ or } k\,\geq\, n/10}\binom{n}{w}\binom{n}{k}n^{-10wk/n} &\leq 2\sum_{w\geq k, w\geq n/10}\binom{n}{w}\binom{n}{k}n^{-10wk/n} \\ &\leq 2\sum_{w\geq k}\binom{n}{w}\binom{n}{k}n^{-k} \\ &\leq 2\cdot 2^n\sum_{k=0}^{n}\binom{n}{k}n^{-k} \\ &= 2\cdot 2^n\cdot(1+1/n)^n = 2e\cdot 2^n.\end{aligned}$$

Adding the contribution of the two parts, we conclude that $\Pr_{G,x,y}[f_{G,L}(x) = f_{G,L}(y)] = O(2^{-n})$. □

Acknowledgements. We thank the TCC'11 and Journal of Cryptology 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] Johannes Blömer, Richard M. Karp, Emo Welzl. The rank of sparse random matrices over finite fields. In *Random Struct. Algorithms* 10(4). Pages 407-419, 1997.

[6] Avrim Blum, Adam Kalai, Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. In *STOC 2000*. Pages 435-440, 2000.

[7] Andrej Bogdanov, Youming Qiao. On the Security of Goldreich's One-Way Function. In *APPROX-RANDOM 2009*. Pages 392-405, 2009.

[8] 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.

[9] Devdatt P. Dubhashi and Alessandro Panconesi, *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[10] Oded Goldreich. Candidate One-Way Functions Based on Expander Graphs. In *Electronic Colloquium on Computational Complexity (ECCC)*, 7(90), 2000.

[11] Oded Goldreich, Russell Impagliazzo, Leonid A. Levin, Ramarathnam Venkatesan, David Zuckerman. Security Preserving Amplification of Hardness. In *FOCS 1990*. Pages 318-326, 1990.

[12] Oded Goldreich, Hugo Krawczyk, Michael Luby. On the Existence of Pseudorandom Generators. In *SIAM J. Comput. 22(6)*. Pages 1163-1175, 1993.

[13] Iftach Haitner, Danny Harnik, Omer Reingold. On the Power of the Randomized Iterate. In *CRYPTO 2006*. Pages 22-40, 2006.

[14] Russell Impagliazzo, Leonid Levin, Michael Luby. Pseudo-random Generation from One-way Functions. In *STOC 1989*. Pages 12-24, 1989.

[15] Russell Impagliazzo, Michael Luby. One-way Functions are Essential for Complexity Based Cryptography. In *FOCS 1989*. Pages 230-235, 1989.

[16] Yishay Mansour, Noam Nisan, Prasoon Tiwari. The Computational Complexity of Universal Hashing. In *Theor. Comput. Sci. 107(1)*. Pages 121-133, 1993.

[17] Henry C. Lin, Luca Trevisan, Hoeteck Wee. On Hardness Amplification of One-Way Functions. In *TCC 2005*. Pages 34-49, 2005.

[18] 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 $f_n\colon \{0,1\}^n \to \{0,1\}^{m(n)}$ 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 5. *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 $f_n : \{0,1\}^n \to \{0,1\}^{m(n)}$ be a function family with input locality d. We define $f'_n\colon \{0,1\}^{n+1} \to \{0,1\}^{m(n)+1}$ 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 $I'(y,1)$. □