Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 103 (2022)

Almost Chor–Goldreich Sources and Adversarial Random Walks

Dean Doron
Department of Computer Science
Ben Gurion University
deand@bgu.ac.il

Dana Moshkovitz[*]
Department of Computer Science
University of Texas at Austin
danama@cs.utexas.edu

Justin Oh[†]
Department of Computer Science
University of Texas at Austin
sjo@cs.utexas.edu

David Zuckerman[‡]
Department of Computer Science
University of Texas at Austin
diz@cs.utexas.edu

Abstract

A Chor–Goldreich (CG) source [CG88] is a sequence of random variables $X = X_1 \circ \ldots \circ X_t$, where each $X_i \sim \{0,1\}^d$ and X_i has δd min-entropy conditioned on any fixing of $X_1 \circ \ldots \circ X_{i-1}$. The parameter $0 < \delta \leq 1$ is the entropy rate of the source. We typically think of d as constant and t as growing. We extend this notion in several ways, defining *almost* CG sources. Most notably, we allow each X_i to only have conditional *Shannon entropy* δd.

We achieve pseudorandomness results for almost CG sources which were not known to hold even for standard CG sources, and even for the weaker model of Santha–Vazirani sources [SV86]: We construct a *deterministic condenser* that on input X, outputs a distribution which is close to having *constant entropy gap*, namely a distribution $Z \sim \{0,1\}^m$ for $m \approx \delta dt$ with min-entropy $m - O(1)$. Therefore, we can simulate any randomized algorithm with small failure probability using almost CG sources with *no* multiplicative slowdown. This result extends to randomized *protocols* as well, and any setting in which we cannot simply cycle over all seeds, and a "one-shot" simulation is needed. Moreover, our construction works in an online manner, since it is based on random walks on expanders.

Our main technical contribution is a novel analysis of random walks, which should be of independent interest. We analyze walks with adversarially correlated steps, each step being entropy-deficient, on good enough lossless expanders. We prove that such walks (or certain interleaved walks on two expanders), starting from a fixed vertex and walking according to $X_1 \circ \ldots \circ X_t$, accumulate most of the entropy in X.

[*] Supported in part by NSF Grant CCF-1705028 and CCF-2200956.
[†] Supported in part by NSF Grant CCF-1705028 and the David L. Miller and Mary H. Miller Graduate Fellowship.
[‡] Supported in part by NSF Grant CCF-1705028 and CCF-2008076 and a Simons Investigator Award (#409864).

ISSN 1433-8092

Contents

1 Introduction

Randomness is an incredibly useful resource. The use of randomness is sometimes provably essential (e.g., in cryptography or property testing), and sometimes we conjecture it is not, prominently in time-bounded randomized algorithms. Yet, it is often the case that randomized algorithms outperform deterministic ones. However, true randomness is scarce, and often we may only be able to access a weak, defective source of randomness. This motivates the problem of simulating randomized algorithms that expect to receive true randomness, using only weak sources of randomness.

The most natural way to use a weak random source is to convert it into a high quality random source. An extractor does exactly this. Specifically, a (deterministic) extractor for a class of sources $\mathcal{X}$ over n bits is a function $\mathsf{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ such that for any $X \in \mathcal{X}$ it holds that $\mathsf{Ext}(X)$ is close, in total variation distance, to U_m, the uniform distribution on m bits. Deterministic extractors are only possible for some restricted classes of sources.

For general sources $\mathcal{X}$, randomness extraction is possible with the addition of a short random seed $Y \sim \{0,1\}^\ell$, independent of X. It is not hard to see that simulation of randomized algorithms given a weak randomness source can be done by cycling over all seeds; see the well known Lemma 2.10. For a running time T, that simulation takes $2^\ell(T + t_{\mathsf{Ext}})$ time, where t_{Ext} is the time it takes to compute the extractor. Since typically $t_{\mathsf{Ext}} \le T$, we denote by 2^ℓ the simulation's *slowdown*, and naturally we want to minimize it. Generally, the distributions that we could hope to extract from are modeled as an arbitrary probability distribution with some amount of min-entropy [CG88, Zuc90], also known as k-sources.[1] Unfortunately, we have a lower bound of $\ell \ge \log n + O(1)$ on the seed length of extractors for arbitrary k-sources over n bits, so simulating $\mathbf{BPP}$ with weak sources using extractors must incur at least $\Omega(n)$ slowdown.[2]

Previous research focused on two extremes: sources where deterministic extraction is possible, and hence there's a negligible slowdown, and simulations giving an $\Omega(n)$ slowdown. A basic natural question is to ask whether anything can be done in between these extremes.

1. Are there natural weak sources where deterministic extraction is impossible, but where an $o(n)$ or even constant slowdown is possible?

It turns out that an affirmative answer to this question can be inferred from previous results, as we will discuss later. However, for some applications, such as in one-shot scenarios like cryptography and interactive proofs, one cannot cycle over all seeds. In other applications, even a constant slowdown is undesirable. In such settings, a deterministic transformation is essential. We therefore ask what is feasible deterministically.

2. Are there natural weak sources where deterministic extraction is impossible, but nevertheless it is possible to deterministically transform the source into a random variable that is essentially as useful as uniform randomness in many settings?

[1] We say that X is a k-source if its min-entropy is at least k, i.e., if every sequence x occurs in X with probability at most 2^{-k}.

[2] Note that the slowdown is (at least) linear in n, and the number of random coins is $m < n$. The difference between n and m naturally depends on the entropy k that the source has. For the precise lower bounds on the parameters of extractors for arbitrary k-sources, see [RT00]. In terms of explicit results, for $k = \Omega(n)$, a simulation with linear slowdown follows from [Zuc07], and for arbitrary k-s we can get a polynomial slowdown (e.g., from [GUV09, LRVW03]).

We answer this question in the affirmative for Santha-Vazirani (SV) and Chor-Goldreich (CG) sources, and generalizations of such sources, which we call Shannon CG sources and almost CG sources, by giving constructions of deterministic condensers with constant entropy gap.

Additionally, in some situations one may not know the ultimate length of a weak random source, or one may wish to extend the length of a given transformed random variable while preserving its useful properties. This leads us to ask:

3. Can the deterministic transformations from Question 2 be computed in an online manner?

This online extraction question is of interest in cryptography [DGSX21a, DGSX21b]. We also answer this question in the affirmative for our generalized notions of CG sources.

Our algorithms take a very natural approach: perform a random walk using the source as a sequence of instructions. For arbitrary sources with entropy rate $1/2$, a random walk may not mix at all: each random step may be followed by an adversarial step that reverses the random step. This raises the question:

4. Do random walks mix well in some sense for any natural weak sources with entropy rate below $1/2$?

We show that indeed it is possible to get *good mixing properties for random walks using SV sources and their generalizations*. That is, for an adversarial random walk on a sufficiently high quality expander, it suffices that each step has a small amount of fresh entropy for the walk to mix quite well. We give an overview of our analysis, which is readily applicable even beyond the scope of pseudorandomness, in Section 1.5.

1.1 Santha–Vazirani Sources and Chor–Goldreich Sources

Santha–Vazirani (SV) sources [SV86] are sequences of random bits in which the conditional distribution of each bit given the previous ones can be partially controlled by an adversary. Namely, $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}$, is a δ-SV source if for any i and any prefix $a \in \{0,1\}^{i-1}$ and $b \in \{0,1\}$, it holds that $\Pr[X_i = b | X_{[1,i-1]} = a] \leq 1 - \delta/2$.[3] Chor and Goldreich [CG88] generalized the SV model by considering each $X_i \sim \{0,1\}^d$ and assuming that no sequence of d bits has too high a probability of being output. Formally, X is a δ-CG source if for any i and any prefix $a \in \{0,1\}^{d(i-1)}$, it holds that $H_\infty(X_i | X_{[1,i-1]} = a) \geq \delta d$, where H_∞ denotes the min-entropy. We typically think of d being constant and t growing.[4]

Santha and Vazirani showed that there is no deterministic extractor for SV sources that's better than outputting the first bit[5] [SV86] (see also [RVW04]). Chor and Goldreich showed an even stronger result for CG sources.

Theorem 1.1 ([CG88]). *The class of δ-CG sources does not admit deterministic extraction.*

We first observe that a constant-length seed suffices to extract from CG sources (and thus SV sources). The proof is actually given in [NZ96, Lemma 10], although there is no theorem statement to this effect (because the focus in [NZ96] was on general min-entropy sources).

[3] We denote $X_{[1,i-1]} = X_1 \circ \ldots \circ X_{i-1}$. Note that the X_i-s are not assumed to be independent.

[4] This is in contrast with "block-sources", which is the term often used when t is very small and d is large.

[5] We note that some variations of SV sources do admit better deterministic extraction. See [BEG17].

Theorem 1.2 (informal; follows from [NZ96])**.** *For any constants $0 < \varepsilon, \delta \leq 1$, there exists an ε-error extractor for δ-CG sources, with seed length $\ell = O(1)$.*

This was improved to CG sources with subconstant δ in [SZ99, Lemma 5.3], but again there is no theorem statement. Since we believe many are not aware of this result, for completeness, we include a proof in Appendix A.2 that puts it in a more general framework.

By the previously mentioned connection, Theorem 1.2 gives a simulation using CG sources with constant slowdown.[6] However, there are scenarios where even constant seed is undesirable. This work shows that there is a way to *deterministically transform* such generalized CG sources, in an online manner, into a random variable that is *essentially as useful* as a nearly uniform random variable in many scenarios. In a bit more detail, surprisingly, we show that one can simulate low-error randomized algorithms, and in general *biased distinguishers*, in a "one-shot" manner. In particular, we have the following theorem.

Theorem 1 (informal; follows from Theorem 2 and Claim 1.6)**.** *There exists a deterministic, efficient, function* Cond *such that the following holds. Given a δ-CG source $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^{d=O(1)}$, for any randomized algorithm A and any input w such that $A(w,y)$ errs with probability $O_{\delta,d}(\varepsilon^2)$ (over a uniform $y \sim U$), it holds that $A(w, \mathsf{Cond}(x))$ errs with probability ε (over $x \sim X$).*

The one-shot simulation via CG sources (and later we will see that such a simulation is possible with a much richer class of sources) is possible in light of our deterministic condensers, overviewed in Section 1.2 (see also the discussion in Section 1.3). We continue with the very natural generalization of CG sources that we study.

***Shannon* CG Sources.** Instead of requiring that each X_i, conditioned on every prefix, has at least δd min-entropy, we only require the conditional X_i have δd Shannon entropy.[7]

While Shannon CG sources seem more general than the *almost* CG sources we define next, it turns out that strong enough results for almost CG sources imply results for Shannon CG sources. Thus, much of the technical focus of this work is on almost CG sources, with the case of Shannon CG sources following as a corollary.

***Almost* CG Sources.** Instead of requiring that each X_i, conditioned on every prefix, has at least δd min-entropy, we only require the conditional X_i to be γ-*close* to some source with entropy rate δ.

Definition 1.3 (almost CG source, I)**.** *We say that $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, is a γ-almost δ-CG source if for any i and any prefix $a \sim X_{[1,i-1]}$, it holds that $X_i | \left\{X_{[1,i-1]} = a\right\}$ is γ-close, in total variation distance, to a source with δd min-entropy.*

The definition of almost CG sources is also quite natural. In particular, considering γ-s which can be much larger than 2^{-d} is very natural and has several advantages. In particular, it is often the

[6] We give a brief overview of the construction of Theorem 1.2. Given $X_1 \circ \ldots \circ X_t$, we use a constant-sized seed Y to extract, in a "strong" sense (say, using universal hashing) a uniform Z_1 from $X_{[1,a]}$ where $a = O(1)$. Then, we use Z_1 as a seed to extract from $X_{[a+1,b]}$ to get Z_2, where $[a+1,b]$ is roughly twice as long as $[1,a]$. Continuing this way for $s = O(\log t)$ times, we use Z_s as a seed to extract from a suffix of X of length $\Omega(dt)$. The output of the final extraction is the output of the extractor.

[7] Recall that one always have that $H(X) \geq H_\infty(X)$, for $H(\cdot)$ being the Shannon entropy. In fact, one can easily find X-s with nearly maximal Shannon entropy, but extremely low min-entropy, or even smooth min-entropy.

case that the X_i-s are a result of some prior transformations, which almost always incur some error. In fact, we already demonstrate such an example in this work. In Section 1.5.2, we will see that in order to condense from an (almost) δ-CG source, we will first "condense" the original source into a γ-almost δ' CG source with $\delta' > \delta$, and some $\gamma > 0$. In Definition 1.10 we will further extend our definition of almost CG sources.

The techniques of [NZ96] also work to give a constant-seeded extractor for almost CG sources as defined in Definition 1.3.

Theorem 1.4 (informal; see Appendix A.3)**.** *For any constants $0 < \varepsilon, \delta, \gamma \leq 1$, and $\gamma \geq 0$, there exists an ε-error extractor for γ-almost δ-CG sources, with seed length $\ell = O(1)$.*

For the formal statement, see Corollary A.8. Although this generalization is not hard, we stress that it was not known, and in particular requires some observations about almost CG sources provided in this work (see Lemma 3.3). Later on, we'll discuss even further extensions of CG-sources, for which the techniques of [NZ96] completely fail, while ours do not.

1.2 Deterministic Condensing from Almost CG Sources

Recall that we have the following parameters:

1. d is the length of each block, and t is the number of blocks (so X is distributed over $n = dt$ bits.);
2. Each block X_i is γ-close to having δ entropy rate; and,
3. m denotes the output length of our extractor (and later condenser).

Later, we will study two additional extensions for CG sources: Those with some λ-fraction of *damaged* blocks, for which we have no guarantee, and those in which for every good block, it is only guaranteed that all but some ρ-fraction of prefixes give rise to a (close to) high-entropic block.

While an extractor aims to purify a weak source X into a nearly-uniform source, a *condenser* aims to improve the source's quality, namely by increasing the entropy rate [RR99]. Formally, $\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ is a (k', ε) condenser for a class of sources $\mathcal{X}$ distributed over $\{0,1\}^n$ if for any $X \in \mathcal{X}$ and an independent and uniform $Y \sim \{0,1\}^\ell$, it holds that $\mathsf{Cond}(X,Y)$ is ε-close to a source with k' min-entropy. When $\ell = 0$, we say the condenser is *deterministic* (or seedless), and that $\mathcal{X}$ admits deterministic condensing.

The entropy *rate* of a condenser is $\frac{k'}{m}$, and we want it to be larger than $\frac{k}{n}$, where k is the min-entropy in each $X \in \mathcal{X}$. When the rate is very close to 1, i.e., when k' is very close to m, it makes sense to measure the additive difference $m - k'$.

Definition 1.5 (entropy gap)**.** *The* entropy gap *of a random variable $Z \sim \{0,1\}^m$ is $\Delta = m - H_\infty(Z)$. We say that a (k', ε) condenser* Cond *has entropy gap Δ if its output is ε-close to a source with entropy gap Δ. (Note that an extractor has entropy gap 0.)*

Condensers were proven incredibly useful as building blocks for extractors (e.g., in [RSW06, TUZ07, GUV09, Zuc07, BKS+10]). Regardless, they are also of great independent interest, because:

1. They can achieve parameters that are *unattainable* for extractors, and in particular,

2. There are classes of sources that admit deterministic condensing and (provably) do not admit deterministic extraction.

For Item 1, we give as an example the fact that for arbitrary weak sources, condensers can achieve smaller entropy loss[8] and a smaller seed length. The latter fact was used for the construction of full-fledged extractors and pseudorandom generators (see [BDT19, DMOZ20]).

Our focus in this work is on the intriguing phenomenon described in Item 2. Recall that the class of CG sources do not admit deterministic extraction. Our main result is that not only do CG sources, and even almost CG sources, admit deterministic condensing, but we are able to construct explicit condensers for such sources with *constant entropy gap*!

Theorem 2 (see also Theorem 6.1)**.** *For any constants $\delta, \varepsilon, \gamma > 0$, any constant integer $d \geq 1$, the following holds. For any positive integer t, there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^{m=\Omega(\delta dt)}$$

such that given an almost δ-CG source X with smoothness parameter γ, $\mathsf{Cond}(X)$ is ε-close to an $(m - O(\log \frac{1}{\varepsilon}))$-source.

We view Theorem 2 as quite striking. It states that even a stream of constant-length random strings where each element locally appears *essentially deterministic* (for example, consider $d = 1000$ and $\delta d = 0.01$), can be readily transformed, without any additional resources, into a random variable that is almost as useful as nearly uniform randomness in many applications.

Deterministic extraction (and thus condensing) is known for several classes of sources. Some have more algebraic structure, such as uniform distributions on affine subspaces or varieties (see [CGL22, Dvi12] and references therein), where others are arguably better models of random sources obtained from natural physical phenomena, such as bit-fixing sources, samplable sources, small-space sources or local sources ([TV00, KRVZ06, DW12, Vio14, CG22] are just few examples). Our study of CG sources and almost CG sources adds to the very short list of natural classes of sources which admit deterministic condensing (even explicitly) but do not admit deterministic extraction. In [BGM22], Ball, Goldreich, and Malkin considered the problem of condensing and extracting from two *somewhat dependent* sources. They showed that if X and Y are weak sources such that each source has bounded influence on the outcome of the other source[9], or that the mutual information $I(X, Y)$ is bounded, then condensing from X and Y is possible, whereas extraction is not. A more contrived example is a certain type of block sources which appear in [BCDT19].

1.3 Simulating True Randomness with Almost CG Sources

The deterministic condenser guaranteed by Theorem 2 implies a constant-seed extractor as in Theorem 1.4. This is because there are explicit extractors for sources with constant entropy gap Δ that have seed length $O(\Delta)$ [GW97] (see Theorem 2.12; there are even explicit extractors with seed length $O(\log(\Delta/\varepsilon)$ [RVW02], but they don't further improve our seed length asymptotically). We now state our more general constant-seed extractor that works even for almost CG sources.

[8] The entropy loss of a condenser or an extractor is the difference between the input entropy and the output entropy. When $\mathcal{X}$ is the set of all k-sources, the entropy loss of a seeded extractor $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ is $k + d - m$, and the entropy loss of a (k', ε) seeded condenser $\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ is $k + d - k'$. In seeded condensers, the entropy loss can be zero, which is impossible for extractors (see [RT00, AT19]).

[9] For a discussion about the notion of bounded influence, see [BGM22, Section 2.2], or Definition 4.1 in the ECCC version of [BGM22].

Theorem 3 (see also Theorem 7.1)**.** *For any constants $\delta, \varepsilon, \gamma > 0$, any constant integer $d \geq 1$, the following holds. For any positive integer t there exists an explicit function*

$$\mathsf{Ext}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=O(1)} \to \{0,1\}^{m=\Omega(\delta dt)}$$

such that given an almost δ-CG source X with smoothness parameter γ, and an independent uniform $Y \sim \{0,1\}^{\ell}$, it holds that $\mathsf{Ext}(X,Y) \approx_{\varepsilon} U_m$.[10]

We now focus on ways in which our deterministic condenser is better than the constant-seed extractor (even for exact CG sources). We give a *one-shot simulation* of randomized protocols with almost CG sources for biased distinguishers, and particularly, a no-overhead simulation of **BPP** algorithms that err with small probability. This wasn't known even for CG sources, or even for SV sources. We discuss this next.

The Usefulness of Constant Entropy Gap. While constant seed is needed to simulate a **BPP** algorithm with error $\frac{1}{3}$ using CG sources, what if we start with an algorithm that has a very small constant error? What if we wish to simulate a *protocol* rather than an algorithm, and we cannot simply cycle over all seeds? Our next discussion is devoted to what can be done with nonzero, yet very small, entropy gap.

Consider the following simple observation.

Claim 1.6 (see, e.g., [DPW14])**.** *Let $Z \sim \{0,1\}^m$ be $\frac{\varepsilon}{2}$-close to some random variable with $m - \Delta$ min-entropy. Then, for any $\mathrm{BAD} \subseteq \{0,1\}^m$ with density at most $\rho(\mathrm{BAD}) \leq 2^{-\Delta-1}\varepsilon$, it holds that $\Pr[Z \in \mathrm{BAD}] \leq \varepsilon$.*

Thus, Theorem 2 implies that we can sample roughly $\frac{m}{\delta}$ bits from an almost CG source, apply our condenser, and simulate a randomized algorithm that uses m bits of randomness. As long as the algorithm's error is small enough compared to our condenser's entropy gap, we can simulate it to within a (larger) error ε, and the *only overhead we have is computing the condenser*. This is the essence of Theorem 1. We note that sources with small entropy gap were recently used to simulate algorithms that err rarely in the computational setting, where computational entropy is used rather than the min-entropy of Claim 1.6 (see [DMOZ20]).

Additionally, we observe that Claim 1.6 and Theorem 1 suggest an *alternative* method for simulating **BPP** algorithms with constant overhead. Given a randomized algorithm A that errs with probability at most $\frac{1}{3}$, simply amplify the algorithm to error probability $2^{-\Delta-1}\varepsilon$ by considering A' that repeats A on fresh randomness a constant number of times and takes the majority vote. Then, one can simply run A' using Z as the randomness. Note this method is different than the standard one as it does not require computing an extractor at all. In other words, modulo different constant error probabilities, a source with constant entropy gap is essentially as useful as a nearly uniform source for **BPP** algorithms.

Sources with small Δ have found applications in cryptography (see, e.g., [BDK+11, DRV12, DY13, DPW14]), and our one-shot generation of constant-gap sources from almost CG sources make the latter useful for those applications. In [DPW14], Dodis, Pietrzak, and Wichs considered the notion of *biased distinguishers*, which is well-motivated in cryptography, and studied extractors

[10] We remark that the output length $m = \Omega(\delta dt)$ can in fact be stated as $m = (1-\theta)\delta dt$ where θ is an arbitrary small constant, by slightly strengthening the constraints on the constructions' parameters. For simplicity and readability, we do not give the constraints' dependence on θ.

that are only guaranteed to fool biased distinguishers rather than arbitrary ones. (This is also related to "slice extractors.")

Definition 1.7 (unpredictability extractor, [DPW14])**.** *A function $D\colon \{0,1\}^m \times \{0,1\}^\ell \to \{0,1\}$ is a μ-distinguisher if $\mathbb{E}[D(U_m, Y)] \le \mu$, where (U_m, Y) is uniform over $\{0,1\}^m \times \{0,1\}^\ell$. A function $\mathsf{UExt}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ is a (k,μ,ε)-unpredictability extractor if for any k-source $X \sim \{0,1\}^n$ and any μ-distinguisher D, we have that $\mathbb{E}[D(\mathsf{UExt}(X,Y),Y)] \le \varepsilon$, where Y is uniform over $\{0,1\}^\ell$ and independent of X.*

Dodis et al. showed that condensers with small entropy gap are equivalent to unpredictability extractors [DPW14].[11] This follows from the connection between sources with small entropy gap and biased distinguishers, essentially rephrasing Claim 1.6: For any $Z \sim \{0,1\}^m$ which is ε-close to having $m - \Delta$ min-entropy, and a μ-distinguisher $D\colon \{0,1\}^m \to \{0,1\}$, it holds that $\mathbb{E}[D(Z)] \le \varepsilon + 2^\Delta \mu$. While Dodis et al. discussed seeded primitives and arbitrary weak source, the connection between constant entropy gap and biased distinguishers readily follows to our setting as well. Concretely, Theorem 2 gives deterministic unpredictability extractors for almost CG sources.[12] We believe the notion of a deterministic unpredictability extractor is a very natural one and may find applications beyond the ones that stem from [DPW14].

To conclude this section, we mention a work by Gavinsky and Pudlák on deterministic condensers for SV sources [GP20]. There, they studied the less-standard notion of errorless condensers, and showed that no such determinstic condenser exists for (standard) SV sources. We do allow error, which evidently does enable deterministic condensing. (Allowing error also enables seeded extraction from general weak sources, and is the standard model in pseudorandomness.) They also gave a seedless condenser for a more restrictive model than SV sources, although it doesn't have constant entropy gap.

1.4 On Almost CG Sources and the Smoothness Parameter

Before presenting our technique, let us further discuss the smoothness parameter γ. Towards this end, let us introduce the notion of smooth min-entropy, which we implicitly used above. For a smoothness parameter $\alpha > 0$, we let $H_\infty^\alpha(X) = \max_{X' : |X - X'| \le \alpha} H_\infty(X')$.[13] Using this terminology, the i-th block in our almost CG source satisfies $H_\infty^\gamma(X_i | X_{[1,i-1]} = a) \ge \delta d$ for any prefix $a \sim X_{[1,i-1]}$, and the output of the condenser satisfies $H_\infty^\varepsilon(\mathsf{Cond}(X)) \ge m - O(1)$.

One could imagine the the setting of $\gamma > 0$ to be a technical extension, but successfully handling this regime draws highly nontrivial consequences. First, note that we *cannot* reduce the $\gamma > 0$ setting to the $\gamma = 0$ case via a union-bound type argument, since $\gamma t \gg 1$. It turns out that this is not simply a matter of proof technique.

Claim 1.8 (informal; see Claim 3.14)**.** *There exists an almost δ-CG source with smoothness parameter γ which is far from any $(1-2\gamma)\delta$-CG source.*

Despite this, our technique does handle constant γ-s. Moreover, we emphasize that an almost CG source with $\gamma > 0$ over $dt = n$ bits may not even have $\Omega(\delta n)$ bits of entropy. To see this, consider

[11] The use of biased distinguishers is also explicit in the recents works of [CT21, SV22].

[12] We note that [DPW14] cared about the entropy *loss*. Our condensers lose roughly a small constant fraction of the entropy, which is much more that what is attainable for seeded condensers with small entropy gap.

[13] The distance here is the total variation distance. See Section 2.1.

the source $X = X_1 \circ \ldots \circ X_t$ such that for each $i \in [t]$, X_i is zero with probability γ, and an arbitrary δd-source over $\{0,1\}^d \setminus \{0\}$. Thus, $\Pr[X = 0] = \gamma^t$ and so $H_\infty(X) \leq t \log \frac{1}{\gamma}$. Still, our condenser outputs a source which is close to having roughly δn bits of entropy! This implies that such an X must have ample *smooth* min-entropy. Indeed, this is the case.

Claim 1.9 (informal; see Claim 3.13). *Every almost δ-CG source over n bits with smoothness parameter γ has smooth min-entropy $(1 - 2\gamma)\delta n$.*

Such a claim follows from a technique similar to "entropy flattening" (see, e.g., [GV99]), where the min-entropy of a distribution X is improved by taking multiple independent copies of X.

Handling Shannon Entropy. Handling $\gamma > 0$ enables us to extend our results to Shannon CG sources. Given a Shannon δ-CG source, we show that by grouping every $O(1)$ consecutive blocks, we get an almost $\Omega(\delta^2)$-CG sources with smoothness parameter γ that is exponentially-small in the number of grouped blocks (see Corollary 3.11). Then, we can easily apply our results for almost CG sources. See Theorems 6.4 and 7.3 for the precise condensing and extraction results. Note that the transition from Shannon entropy to min-entropy necessarily induces error, so $\gamma > 0$ is crucial here.

Handling Damaged Blocks. Our random-walks based condensing method is flexible enough to handle damaged blocks too. Namely, we allow some λ-fraction of the i-s to have *completely arbitrary* conditional distributions.

Definition 1.10 (almost CG source, II). *A (γ, λ)-almost δ-CG source is a sequence of random variables $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, such that for at least $(1-\lambda)t$ of the i-s, it holds that $H_\infty^\gamma(X_i | X_{[1,i-1]} = a) \geq \delta d$ for any prefix $a \sim X_{[1,i-1]}$.*

When the damage pattern is arbitrary, we can condense to within $O(\lambda dt)$ entropy gap (i.e., we lose d bits of entropy for each damaged block). Corollary 4.14 handles the $\lambda > 0$ setting as well. We remark that the [NZ96, SZ99] technique would fail for even one damaged block. Moreover, when the damaged locations are "nicely distributed", our technique regains the $O(1)$ entropy gap. We elaborate it on this more in Section 1.5.4, and give the technical details in Theorems 5.4, 6.4, 7.3 and 7.4.

1.5 Our Technique: A New Analysis of Adversarial Random Walks

Our main technical contribution is a new analysis of adversarial random walks. Let's begin our discussion with exact Chor-Goldreich sources. Spectral analysis has been the main tool to analyze random walks on expanders. However, it doesn't seem to work for CG sources with rate below $1/2$. This is because there is no specialized method for CG sources; existing spectral methods that work for CG sources also work for general min-entropy sources, and general sources with rate below $1/2$ do not mix at all (recall that each random step may be followed by an adversarial step that reverses the random step). Moreover, even for general sources with rate above $1/2$ a random stopping time is required, which amounts to a linear number of seeds. We hope to condense without a seed or extract with a constant number of seeds.

Furthermore, spectral methods generally exploit the Markovian nature of random walks. However, an adversarial random walk is not Markovian. That is, the distribution of the next step depends not only on the walk's current node, but also on the path it took to get there. Indeed, although it is true that the distribution of the next step from a given node v is a convex combination of instruction distributions over all the paths that end at v, the memory in the walk still presents a challenge.

Our approach uses expansion directly. We therefore use the highest quality expanders: bipartite lossless expanders.

Definition 1.11 (balanced lossless expander)**.** *We say a D-left-regular bipartite graph $G = ([M], [M], E)$ is a $(K_{\mathsf{max}}, \varepsilon)$ lossless expander if for all subsets $S \subseteq [M]$ of size at most K_{max}, the neighborhood set $\Gamma_G(S)$ has size at least $(1 - \varepsilon)D|S|$.*

For technical purposes, we will actually require that the right degree of the lossless expander be small as well. For a high-level understanding of our work, it suffices to assume that the expander is biregular.

For numerous applications a modest vertex expansion is not enough, and lossless expansion is essential.[14] An explicit construction of balanced (and somewhat imbalanced) constant-degree lossless expanders was given by Capalbo, Reingold, Vadhan, and Wigderson [CRVW02].[15] As a pseudorandomness primitive, it is instructive to think of $\Gamma_G \colon \{0,1\}^m \times \{0,1\}^d \to \{0,1\}^m$, the neighborhood function of G, as a *lossless conductor* (where we use $\{0,1\}^m \equiv [M]$).

Definition 1.12 (balanced lossless conductor)**.** *A function $\mathsf{LC} \colon \{0,1\}^m \times \{0,1\}^d \to \{0,1\}^m$ is a $(k_{\mathsf{max}}, \varepsilon)$ lossless conductor if for any $k \le k_{\mathsf{max}}$, a k-source X, and an independent and uniform $Y \sim \{0,1\}^d$, it holds that $H_\infty^\varepsilon(\mathsf{LC}(X,Y)) \ge k + d$.*[16]

That is, the output distribution "absorbs" the d bits of entropy from the seed, up to an ε error. Intuitively, the larger the vertex expansion, the less freedom the adversary has to skew the distribution over the next step. We soon make this intuition more concrete.

Our first construction, which works for large δ-s, goes as follows. Given an almost CG source $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim [D]$, we walk, from a fixed node, along a $(t+1)$-partite graph with a copy of G between each two layers (the graph's size M is chosen as a function of the source's parameters). Namely, we start at some fixed $Z_0 \in [M]$, and for each $i \in [t]$, let

$$Z_i = \Gamma_G(Z_{i-1}, X_i),$$

and output $\mathsf{Cond}(X) = Z_t$.

For an exact δ-CG source, this amounts to a random walk where an adversary, after seeing previous steps, chooses D^δ nodes among the D neighbors, and the random walker steps to a random node among these D^δ nodes. We are able to show:

Theorem 4 (informal; see Theorem 4.8)**.** *Let $X_1 \circ \ldots \circ X_t$ be a δ-CG source, with each $X_i \sim \{0,1\}^d$. Let G be a sufficiently good $D = 2^d$-regular expander. Then, for any $\eta > 0$, the last step Z_t of a random walk on G, performed as above, is η-close to a $k - O(d + \log \frac{1}{\eta})$-source.*

The proof is nontrivial, and we discuss it next.

[14] Examples can be found in coding theory, data structures, algorithms, storage models, and proof complexity (see the references in [CRVW02], and [BGI$^+$08, DK08, CCLO22, LH22] for more recent works).

[15] For very small sets, Alon showed that lossless expansion follows from high girth. See also [AC02]. In the regime where $M \ll N$, the degree needs to be super-constant, and explicit constructions for this regime are known (e.g., [TUZ07, GUV09]).

[16] The correct equivalence would be to lossless *condensers* if we allow the construction itself to depend on k (see [TUZ07]). For the sake of our discussion, this difference won't matter, and in the technical sections we will not use the lossless condensers/conductors terminology.

Evading the Union Bound. The naive approach to analyze the output distribution after t steps is to follow the definition of conductors. However, conductors only guarantee that the output distribution is ε-close to a distribution with appropriate entropy. Thus, even disregarding the correlation between source and seed, such an argument naturally forces us to union bound over the error of each step. Indeed, one can even show that if each instruction comes from a δd-source, and one wishes to add exactly δd entropy, then such a union bound is necessary. Our ultimate solution avoids this union bound issue, and in doing so, only argues that the entropy gain at each step is $0.9\delta d$ instead.[17]

Expansion of Weight Functions. As usual in analyzing random walks, we need to handle real nonnegative probabilities. It is standard to do this using eigenvalues, but there is a loss in going from expansion to eigenvalues, or other analytic tools such as hypercontractivity. These analytic methods don't seem to capture lossless expansion.

We give a simple way to capture lossless expansion by directly generalizing the combinatorial definition of expansion to nonnegative real numbers, which doesn't seem to have been considered before. Specifically, let 1_S denote the indicator function of a set S. Then $1_{\Gamma(S)}(v) = \vee_{w \in \Gamma(v)} 1_S(w)$. To generalize this to weight functions (nonnegative real valued functions), we replace the OR with a max. We then show that the expansion of weight functions with support size at most K exactly equals the expansion of sets with size at most K. This enables us to capture the effect of lossless expansion. We can even generalize this weighted notion to unique neighbor expansion, although it is not necessary for the proof.

1.5.1 The ℓ_q Norm as a Progress Measure

Recall that spectral analysis typically uses the ℓ_2 norm as a measure of progress. While the ℓ_2 norm doesn't appear to work in our setting, we manage to use the ℓ_q norm as a progress measure, for some suitable $q = 1 + \alpha$. That is, we show that the ℓ_q norm of the vertex distribution decreases by a suitable multiplicative factor at each step.

Theorem 5 (informal; see Lemma 4.6)**.** *Let $G = (U = [M], V = [M], E)$ be a bipartite D-regular (K, ε) lossless expander with error $\varepsilon = \frac{1}{D^\beta}$. For any $0 < \alpha < \beta$, set $q = 1 + \alpha$ and let $\delta \geq 1 - \beta + \alpha$.*

Let p_U be a probability distribution over U and let r_u, for each $u \in U$, be a distribution over $\{0,1\}^d \equiv [D]$, each being a δd source. For any $u \in U$ and $v \in V$ let $r_u(u, v)$ denote the probability that the edge leading from u to v is chosen under r_u. Namely, for G's labelling function $\ell \colon E \to [D]$ we denote $r_u(u, v) \equiv r_u(\ell(u, v))$. Define p_V as the induced probability distribution on V. That is, $p_V(v) = \sum_{u \in \Gamma(v)} r_u(u, v) p_U(u)$. Then,

$$\|p_V\|_q^q \leq \frac{8}{D^{\delta\alpha}} \cdot \|p_U\|_q^q,$$

as long as $\|p_U\|_q^q$ is not already smaller than $1/K^\alpha$.

The ℓ_q-norm is a *proxy measure for min-entropy*, since any distribution p such that $\|p\|_q^q \leq 2^{-\alpha k}$ is ε-close to a distribution with entropy $k - \frac{1}{\alpha} \log \frac{1}{\varepsilon}$ (see Corollary 2.3). Thus, Theorem 5 implies that every step on a lossless expander, according to a δd source, adds roughly δd bits of entropy to the vertex distribution, up to a "saturation" point of roughly $k = \log K$ bits of entropy. Since we have

[17] Or $(1 - \theta)\delta d$ for an arbitrary constant θ close to 0, at the expense of modifying some constraints in the construction.

explicit constructions wherein $k = m - O(1)$, a saturated vertex distribution already has constant entropy gap.

One advantage of using the q-norm is that it allows us to better control the error term corresponding to the small lossy part of the lossless expander. For example, certain nodes on the right may have high degree, causing their probability after a step of a random walk to be large. This problem is exacerbated by the adversarial nature of a random walk via an almost-CG-source, which can assign up to γ probability to edges leading to high degree right nodes. By considering the q-norm for a sufficiently small α, we have a measure of entropy that is less sensitive to such error, all while still ensuring that the entropy gained at each step is roughly the same as the entropy in each instruction.

To prove Theorem 5, since the distribution of the random walk's vertex may not be uniform, we generalize set expansion and unique neighbor expansion to apply to "weight functions" and probability distributions. We then apply Jensen's inequality with a nonstandard choice of coefficients that heavily weights the term where we gain. This gives a simple analysis of adversarial random walks that uses expansion directly.

Overall, our analysis gives a "spectral-like" analysis of random walks even when such techniques cannot be directly applied. In addition to its application in deterministic condensing, we believe that this analysis of entropy gain via random walks from correlated and nonuniform steps is interesting on its own.

Handling Smoothness. Up until now, we did not address the smoothness parameter γ thoroughly. Quite surprisingly, it turns out that our technique based on the ℓ_q-norm analysis is flexible enough to support constant γ-s without substantial changes. Indeed, when dealing with such instructions, we extend Theorem 5 and show that the ℓ_q- norm decrease factor is now roughly $\frac{1}{D^{\delta\alpha}} + D^{\alpha}\gamma$. In fact, there are cases where this factor is tight. This seems unfortunate, because we are now seemingly only gaining less than $\log \frac{1}{\gamma}$ min-entropy at each step, or in other words, lose the vast majority of the desired δd bits.

The trick to overcome this is to simply pick α sufficiently small in the ℓ_q-norm analysis (recall that we set $q = 1 + \alpha$). Indeed, by choosing $\alpha \approx \frac{1}{d} \log \frac{1}{\gamma}$, we see that γ is then comparable to $\frac{1}{D^{\delta\alpha}}$. Under the assumption that $\gamma \leq 2^{-O(1/\delta)}$, the decrease factor can be made to be $D^{-0.9\delta\alpha}$. Thus, we once again gain roughly 90% of the entropy at each step. Setting α this small only results in a loss of roughly $O(d)$ bits of entropy over the entire walk.[18] For the precise norm evolution with an arbitrary γ, in Corollaries 4.11 and 4.12, we set α accordingly.

Additionally, we observe that the assumption $\gamma \leq 2^{-O(1/\delta)}$ is quite mild, as γ only depends on δ and not d. Thus, for sufficiently large d-s, $\gamma \gg D^{-O(1)}$. We note that setting α to be a small constant, say $\alpha = 1/6$, *would* require $\gamma \leq D^{-O(1)}$ in order to argue that $0.9\delta d$ bits of entropy is gained at each step. We view our setting of parameter α as a way that allows us to avoid treating each instruction source as pessimistically as a $\log \frac{1}{\gamma}$-source.

The Limit of Our First Construction. We now explain why our first construction only works for large enough δ. For concreteness, assume that we are at some $Z_{i-1} \sim \{0,1\}^m$ with $H_\infty(Z_{i-1}) = k$, and walk according to $X_i \sim [D]$ having entropy δd (assume for now that $\gamma = 0$). For simplicity,

[18] A key point here is that the closer α is to 1, the larger we can allow our ℓ_q-norm bound to be in order to get high entropy. See Corollary 2.3.

assume that Z_i is flat over some set $S \subseteq [M]$ of size $K = 2^k \leq K_{\mathsf{max}}$, recalling that we walk over a sequence of $(K_{\mathsf{max}}, \varepsilon)$ bipartite lossless expanders with M vertices, arranged in series. It may be informative to simply think of the walk as over a single $(K_{\mathsf{max}}, \varepsilon)$ undirected lossless expander.

While any large enough subset of S or of the edges leaving S has nice properties (for example, at least $1 - 2\varepsilon$ fraction of the vertices in S have a *unique neighbor*), there can still be ε-fraction of the KD edges leaving S that behaves badly. In particular, εKD of the edges may lead to vertices that have *many* incoming edges from S. Assume for simplicity that each node in S has the same number of bad edges, namely εD edges from each node in S lead to heavy vertices. When $D^{\delta} \leq \varepsilon D$, an adversarial X_i can potentially, for each node, be supported *only* on instructions that lead to bad edges. In this case, Z_{i+1} may have accumulate neither min-entropy, nor smooth min-entropy. Thus, we must consider the case where $D^{\delta} \gg \varepsilon D$.

This raises the question of how small can we take ε to be as a function of D, or alternatively, how large can we take δ to be given an existing lossless expander. Non-explicitly, we have Γ_G-s with a great seed length, namely $d = 1 \cdot \log \frac{1}{\varepsilon} + O(1)$, in which case we can take $\varepsilon \ll D^{-(1-\delta)}$ even when $\delta > 0$ is arbitrarily small. In [CRVW02], however, the required seed length is $d = \frac{1}{\beta} \log \frac{1}{\varepsilon}$ for some constant $\beta < \frac{1}{2}$.[19] Denoting $\delta_{\mathsf{thr}} = 1 - \beta$, we see that we can only hope to handle almost δ-CG sources with $\delta > \delta_{\mathsf{thr}}$, and we do indeed achieve this. We note that both in [CRVW02] and in an optimal lossless expander, $K_{\mathsf{max}} = \Omega_D(M)$, which is good enough to lead to constant entropy gap.

1.5.2 Our Two-Level Construction

We handle general $\delta > 0$ via a two-level process: We first walk over a small, *optimal* lossless expander in order to simulate an instruction with sufficiently large δ, and then "flush" it as a step in the big CRVW graph over M vertices.

We are given $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d \equiv [D]$. We let $H = ([D_{\mathsf{crvw}}], [D_{\mathsf{crvw}}], E)$ be an optimal lossless expanders with degree D, and we can choose its error ε to be very close to $1/D$. The number of vertices in H corresponds to the degree of our standard CRVW graph G over M vertices, and we choose D_{crvw} to be quasi-polynomial in D.[20] For the exact choice of parameters for G and H, see Section 5. Now:

- For some parameter $b = \mathrm{poly}(d)$, we group consecutive blocks of X into "super-blocks" $X'_1 \circ \ldots X'_{t/b}$, each X'_i containing b blocks of length d each.

- For each $i \in [t/b]$, we use X'_i as instructions to a separate random walk on H, starting from some fixed node each time. Denote by Z_i the final node reached after the b steps.

- We show that $Z = Z_1 \circ \ldots \circ Z_{t/b}$ is itself an almost CG source, but this time with $\delta > \delta_{\mathsf{thr}}$. Thus, we can use Z as instructions for G!

Fortunately, H is constant-sized, so we can find it in constant time. Using optimal constant-sized ingredients is also a key idea in the [CRVW02] construction itself.

[19] The actual β is around $\frac{1}{6}$, and $\beta < \frac{1}{2}$ is an inherent barrier for their construction.

[20] One can also think of H as an ε-error optimal lossless conductor $H \colon \{0,1\}^{\mathrm{poly}(d)} \times \{0,1\}^d \to \{0,1\}^{\mathrm{poly}(d)}$ with seed length $d = \log \frac{1}{\varepsilon} + O(1)$.

1.5.3 Removing the Constraints on d and γ

So far, we discussed how to condense from a γ-almost δ-CG source when $\gamma < 2^{O(1/\delta)}$ and $d > \mathrm{poly}(1/\delta)$.[21] To obtain Theorem 2, which has no such constraints, we observe that grouping the instructions of the CG source into blocks of length $\mathrm{poly}(1/\delta)$ yields a new CG source with roughly the same entropy rate, but with sufficiently large instruction length, and smoothness error exponentially small in $1/\delta$. The fact that grouping instructions into blocks improves the smoothness error follows quite easily from the observation that sampling a heavy instruction (one whose probability is at most γ) at step i is independent of sampling heavy instructions in previous steps. Thus, the number of heavy instructions sampled over many i follows Chernoff-like tail bounds. See Lemma 3.3 for details.

1.5.4 Suffix-Friendliness

While our technique is flexible enough to recover from damaged blocks and suffer only the expected decrease in entropy per damaged block, it cannot achieve constant entropy gap, if, say, all the damaged blocks are at the end. However, if at any step we can guarantee that we won't encounter too many damaged blocks from now on, we *can* regain constant entropy gap. Roughly speaking, the favorable case is that the λ-fraction of bad blocks is nicely distributed in the sense that each suffix contains at most λ-fraction of bad blocks (up to an additive term). We call this property *suffix friendliness* (see the precise definition in Definition 3.4), and show that we can deterministically condense from such sources to within constant entropy gap in Section 4.3.3. Moreover, we observe that given an almost CG source with $\lambda = 0$, a *random* pattern that damages each block with probability roughly λ, leads to a suffix friendly almost CG source with "bad blocks" parameter λ (see Lemma 3.5).

1.5.5 The Construction's Runtime

Recall that the simulation slowdown is also affected by the time it takes to compute the extractor, or condenser (in the "one-shot" simulation setting). Our online manner of condensing, together with the fact that the primitives we use (namely, the CRVW expander and the GW extractor) are efficient, makes our construction efficient as well. In particular, in Appendix C we achieve a near-quadratic runtime in the TM model. In the RAM model, in which each machine word can store integers up to $N = 2^n$ and perform arithmetic in $\mathbb{F}_q$ for a prime $q \le N$ at unit cost, our construction takes *linear* time.

1.6 On Supporting Bad Prefixes

We extended δ-CG sources to handle smoothness γ and λ fraction of bad blocks. One can also try and further relax the notion of CG sources in the following way: Instead of requiring that for each non-damaged block X_i, for *any* prefix $a \sim X_{[1,i-1]}$ it holds that $X_i | \{X_{[1,i-1]} = a\}$ is γ-close to having entropy rate δ, we require it only for *most* prefixes. Concretely, what if we allow some ρ-fraction of the prefixes to lead to instructions having low entropy? (See Definitions 8.3 and 8.6, also for the Shannon-entropy variant.)

[21] We did not mention the constraint on d explicitly, however the intuition is clear: the raw amount of entropy in an instruction, δd, should be at least 1.

That extension seems *too* permissive, at least in some regime of parameters. We show that any random variable $X \sim \{0,1\}^n$ with $H(X) \geq (1-\zeta)n$ is already an almost $\Omega(1)$-CG source with error parameters γ, λ, and ρ, all roughly equal to $\zeta^{\Omega(1)}$. Moreover, with a constant seed, we show that we can even increase the (smooth) entropy rate from an arbitrary $\Omega(1)$ to $1-\zeta$, at the cost of increasing λ and ρ. Thus, since we provably cannot condense or extract from high Shannon entropy with constant seed, we have an inherent barrier to handling $\rho > 0$ alongside a comparable, nonzero λ. We discuss it further, and give the precise details, in Section 8.

1.7 Extension: Online Condensing and Maintaining Constant Entropy Gap

Unlike other condensers, our construction is an "online" one. That is, the construction makes one pass over the randomness stream $X_1 \circ \ldots \circ X_t$ in order to form the required instructions, and never needs to store more than a constant number of bits in memory before updating the location in the big graph. Moreover, we don't even need to know the number of blocks ahead of time![22]

As given above, it is easy to see that the construction does not ensure constant entropy deficiency in the output distribution *throughout* the random walk, but only at the end, even if there are no corrupted instructions at all ($\lambda = 0$). However, one can *easily adapt* our approach to also work in such a "completely online" fashion. The idea is to walk on graphs of gradually increasing size. Namely, after every constant number of steps (for some fixed constant), we map the current vertex to a vertex in a graph that is a constant times larger (but with the same degree) and continue the walk from there. Although we do not give such a result in full formality, in Appendix B we present an informal theorem and a brief discussion sketching its proof.

1.8 Organization

In Section 2 we give some preliminary definitions and results from previous work, and the connection between small ℓ_q norm and smooth min-entropy. In Section 3 we discuss almost CG sources, both for min-entropy and for Shannon entropy. In Section 4 we establish deterministic condensing for $\delta > \delta_{\mathsf{thr}}$. In particular, after some necessary preparations in Section 4.1, in Section 4.2 we give the analysis of the case where $\gamma = \lambda = 0$, and cover the general setting (including for suffix-friendly sources) in Section 4.3. In Section 5 we give our two-level construction that condenses from any constant rate $\delta > 0$. Section 6 complements this result for Shannon entropy. In Section 7 we give our extraction results that follows easily from previous sections. In Section 8 we discuss the notion of bad prefixes described in Section 1.6. We conclude with a few open problems in Section 9.

2 Preliminaries

2.1 Random Variables and Entropy

The *support* of a random variable X distributed over some domain Ω is the set $x \in \Omega$ for which $\Pr[X = x] \neq 0$, which we denote by $\mathrm{Supp}(X)$.

The *total variation distance* (or, statistical distance) between two random variables X and Y over the same domain Ω is defined as $|X - Y| = \max_{A \subseteq \Omega}(\Pr[X \in A] - \Pr[Y \in A])$. Whenever

[22] In the two-level construction of Section 1.5.2, we first computed all Z_i-s just for the simplicity of exposition. Clearly we can compute Z_i, implement it on the big graph, and continue to compute Z_{i+1} without the need to keep storing Z_i.

$|X - Y| \leq \varepsilon$ we say that X is ε-close to Y and denote it by $X \approx_\varepsilon Y$. We denote by U_n the random variable distributed uniformly over $\{0,1\}^n$. We say a random variable is *flat* if it is uniform over its support. Whenever we write $x \sim A$ for A being a set, we mean x is sampled uniformly at random from the flat distribution over A.

For a function $f \colon \Omega_1 \to \Omega_2$ (even a random one) and a random variable X distributed over Ω_1, $f(X)$ is the random variable distributed over Ω_2 obtained by choosing x according to X and computing $f(x)$. For a set $A \subseteq \Omega_1$, $f(A) = \{f(x) : x \in A\}$. For every $f \colon \Omega_1 \to \Omega_2$ and two random variables X and Y distributed over Ω_1 it holds that $|f(X) - f(Y)| \leq |X - Y|$, and is often referred to as a data-processing inequality.

The (Shannon) entropy of a random variable X is $H(X) = \sum_{x \in \mathrm{Supp}(X)} \Pr[X = x] \log \frac{1}{\Pr[X=x]}$. The min-entropy of X is defined by

$$H_\infty(X) = \min_{x \in \mathrm{Supp}(X)} \log \frac{1}{\Pr[X = x]},$$

and it always holds that $H_\infty(X) \leq H(X)$. For some $\varepsilon > 0$, we define the *smooth min-entropy* of X by

$$H_\infty^\varepsilon(X) = \max_{X' : X' \approx_\varepsilon X} H_\infty(X).$$

We record the following easy claim.

Claim 2.1. *Let $X \sim \{0,1\}^n$ be a random variable such that $X \approx_\varepsilon U_n$. Then, $H_\infty(X) \geq \log \frac{1}{\varepsilon}$.*

A random variable X is an (n, k) source if X is distributed over $\{0,1\}^n$ and has min-entropy at least k. We refer to $\frac{k}{n}$ as the random variable's *entropy rate*. When n is clear from context we sometimes omit it and simply say that X is a k-source.

2.1.1 Distributions as Vectors

We naturally identify a random variable X over some finite domain Ω with the corresponding distribution mass vector p_X in $\mathbb{R}^\Omega$, and often argue that X has large smooth min-entropy when p_X has small ℓ_q-norm.[23] The following lemma gives the exact relation that we use.

Lemma 2.2. *For any $0 < \alpha < 1$, let $q = 1 + \alpha$. Let n be a positive integer, $1 < k \leq n - 1$, and let $\varepsilon > 0$ be such that $\varepsilon^\alpha \leq \frac{1}{2}$. Let p be a distribution over $\{0,1\}^n$ with $\|p\|_q^q \leq 2^{-\alpha k}$. Then, p is ε^α-close to a $k - \log \frac{1}{\varepsilon}$ source.*

Proof: Let B_1 be the set of $x \in \{0,1\}^n$ such that $p(x) > \frac{1}{\varepsilon} 2^{-k}$. We have:

$$2^{-\alpha k} \geq \sum_{x \in \{0,1\}^n} p(x)^{1+\alpha} \geq \sum_{x \in B_1} p(x)^{1+\alpha} \geq \left(\frac{2^{-k}}{\varepsilon} \right)^\alpha \sum_{x \in B_1} p(x).$$

Thus, $\sum_{x \in B_1} p(x) \leq \varepsilon^\alpha$. Let $B_2 \subseteq \{0,1\}^n \setminus B_1$ be the set of x-s for which $\frac{1}{2\varepsilon} 2^{-k} < p(x) \leq \frac{1}{\varepsilon} 2^{-k}$. Note that $|\,\{0,1\}^n \setminus (B_1 \cup B_2)| \geq 2^n - 2\varepsilon 2^k \geq 2^{k+1} \varepsilon^{1+\alpha}$.

[23] We usually identify a random variable with its probability distribution.

Consider the following probability distribution r.

$$r(x) = \begin{cases} 0 & \text{if } x \in B_1, \\ p(x) & \text{if } x \in B_2, \\ p(x) + \frac{\sum_{y \in B_1} p(y)}{|\{0,1\}^n \setminus (B_1 \cup B_2)|} & \text{otherwise.} \end{cases}$$

By construction, r and p are ε^{α}-close. Now, from our bound on the number of elements outside $B_1 \cup B_2$, we have that, for every $x \notin B_1 \cup B_2$,

$$p(x) + \frac{\sum_{y \in B_1} p(x)}{|\,\{0,1\}^n \setminus (B_1 \cup B_2)|} \le \frac{2^{-k}}{2\varepsilon} + \frac{\varepsilon^{\alpha}}{2^{k+1}\varepsilon^{1+\alpha}} \le \frac{2^{-k}}{\varepsilon}.$$

Thus, r is a $(k - \log \frac{1}{\varepsilon})$-source.

■

Invoking Lemma 2.2 with $\varepsilon = (\varepsilon')^{\frac{1}{\alpha}}$, we get the following corollary.

Corollary 2.3. *For any $0 < \alpha < 1$, let $q = 1 + \alpha$. Let n be a positive integer, $1 < k \le n - 1$, and let $0 < \varepsilon \le \frac{1}{2}$. Let p be a distribution over $\{0,1\}^n$ with $\|p\|_q^q \le 2^{-\alpha k}$. Then, p is ε-close to a $k - \frac{1}{\alpha} \log \frac{1}{\varepsilon}$ source.*

2.2 Bipartite Graphs and Lossless Expanders

We say a bipartite graph $G = (V_1, V_2, E)$ is D-regular if it's D *left*-regular. We denote by $\Gamma_G(v)$ the set of neighbors of v in G (whenever $v \in V_1$, $\Gamma_G(v) \subseteq V_2$, and likewise whenever $v \in V_2$). When G is clear from context, we will simply write Γ. When we refer to a step over G, we mean taking a step from V_1 to V_2. Our constructions utilize long walks over G, and specifically we will walk on a layered graph from left to right, with copies of G between consecutive layers. For a D-regular bipartite $G = ([N], [N], E)$, a length-t walk over G starting from $v \in [N]$ according to the instructions $(i_1, \ldots, i_t) \in [D]^t$ is the sequence $(v_0, v_1, \ldots, v_t)$, where v_j is the i_j-th neighbor of v_{j-1}.

Definition 2.4 (bipartite expander)**.** *We say a bipartite graph $G = ([N], [M], E)$ is a (K, A)-*expander *if for all subsets $S \subseteq [N]$ of size at most K, the neighborhood set $\Gamma_G(S)$ has size at least $A \cdot |S|$.*

When G is D-regular we can hope for A to be very close to D up to $K \approx M/D$. When indeed $A = (1 - \varepsilon)D$ we say G is a (K, ε) *lossless expander*.[24] An upper bound on the *right* degree of lossless expanders will also be necessary for our analysis. We will require that each right node has degree at most D^e for some constant exponent $e \ge 1$. For concreteness, we will use the notation (K, ε, e) *lossless expander* to denote a D-regular (K, ε) expander with right degree at most D^e. We will use the lossless expander by Capalbo, Reingold, and Vadhan in its balanced setting of parameters.

Theorem 2.5 ([CRVW02])**.** *There exists a constant $\beta \in (0, 1)$ with $\beta \ge 1/6$ such that the following holds. For every positive integers N and D, there exists an explicit D-regular bipartite graph $G = ([N], [N], E)$ that is a $(K, \varepsilon, e = 100)$ expander for $\varepsilon = \frac{1}{D^{\beta}}$ and $K = \Omega\left(\frac{1}{D^{1+\beta}} N\right)$.*

Remark 2.6. *Although [CRVW02] did not explicitly state an upper bound on the right degree, examining their construction readily shows that the above upper bound holds. Indeed, the construction is essentially a zig zag product between a regular eigenvalue expander, and two special constant sized (polynomial in D) conductors. The largest right degree of the entire construction can be at most the product of the right degrees of the constant sized conductors.*

[24] For brevity, we use K rather than the more standard K_{max}. It is useful to keep in mind that $K = \Omega_D(M)$.

On first reading, we recommend, for simplicity, to think of $e = 1$ (in fact, considering the biregular balanced bipartite expander that is the double cover of D-regular undirected lossless expander will suffice). We will also make use of the fact that optimal lossless expanders have error $\varepsilon = O(1/D)$, and can be biregular.

Theorem 2.7 (nonexplicit lossless expanders)**.** *For every positive integers N and D, there exists a D-regular bipartite graph $G = ([N],[N],E)$ that is a $(K,\varepsilon,1)$ expander for $\varepsilon = O\left(\frac{1}{D}\right)$ and $K = \Omega\left(\frac{N}{D^2}\right)$. By brute-force, such an expander can be found deterministically in time $N^{O(ND)}$.*

It will be convenient to formulate the above theorem as follows, suffering a slight increase in ε.

Corollary 2.8. *There exists a universal constant $c^\star > 1$ such that for any constant $0 < \beta < 1$ and any positive integers $D \geq 2^{\frac{c^\star}{1-\beta}}$ and N, there exists a D-regular bipartite graph $G = ([N],[N],E)$ that is a $(K,\varepsilon,1)$ expander for $\varepsilon = \frac{1}{D^\beta}$ and $K = \frac{N}{c^\star D^2}$. By brute-force, such an expander can be found deterministically in time $N^{O(ND)}$.*

2.2.1 The Expander's Activation Constant δ_{thr}

From here onward, for a given lossless expander, we'll often refer to β as in the statements of Theorem 2.5 and Corollary 2.8 as the "error parameter" of the expander. Also for a given expander, we denote by δ_{thr} the "activation threshold" beyond which a single step via an instruction with δ_{thr} entropy rate adds a decent amount of entropy to the vertex distribution. This activation threshold will depend on the error parameter β. Indeed, as discussed in the overview of our techniques, we will want δ_{thr} to be larger than $1-\beta$. For concreteness, we often think of $\delta_{\mathsf{thr}} = 1 - \beta + \Delta$, where again, Δ is an arbitrary small constant. We now discuss specific settings of these parameters for the lossless expanders we use in our construction.

For optimal lossless expanders, from Corollary 2.8, we can consider β arbitrarily close to 1. Since β is close to 1, and Δ is small, we can also think of $\delta_{\mathsf{thr}} = 1 - \beta + \Delta$ as some arbitrarily small constant. We will show that optimal lossless expanders facilitate entropy gain at each step even when the instructions have arbitrarily small constant entropy rate. Inspecting the [CRVW02] construction, we can see that their lossless expanders can have error parameter $\beta = \frac{1}{6}$ (and even slightly larger). In this case, we can take $\delta_{\mathsf{thr}} = 5/6 + \Delta$. Similarly, we will show that the lossless expanders from Theorem 2.5 facilitate entropy gain at each step when the instructions have entropy rate slightly larger than $5/6$.

As final remarks, note that we can always assume that $K = \Omega\left(\frac{1}{D^2}N\right)$ for both types of expanders. Additionally, we note that the [CRVW02] gives an object stronger than a just vertex expander, namely a lossless *conductor*, but the conductor's vertex expansion properties will suffice for us.

2.3 Seeded Extractors and Condensers

Definition 2.9 (extractor)**.** *A function*

$$\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

is a (k,ε) (seeded) extractor if the following holds. For every (n,k) source X it holds that $\mathsf{Ext}(X,Y) \approx_\varepsilon U_m$, where Y is uniformly distributed over $\{0,1\}^d$ and is independent of X. We say Ext is strong if $(\mathsf{Ext}(X,Y),Y) \approx U_m \times Y$.

As mentioned in the introduction, seeded extractors can be used to simulate randomized algorithms using weak sources.

Lemma 2.10 (see, e.g., [Vad12], Proposition 6.15)**.** *Let $A(w, y)$ be a randomized algorithm deciding some language $L(w)$ such that $A(w, U_m)$ has error δ, and let $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ be an ε-error extractor for $\mathcal{X}$ over n bits. Define $A'(w, x) = \mathrm{maj}_{y \in \{0,1\}^\ell} \{A(w, \mathsf{Ext}(x, y))\}$. Then, for every $X \in \mathcal{X}$, $A'(w, X)$ has error $2(\delta + \varepsilon)$.*[25]

We will use two known constructions of seeded extractors. Recall that a universal family of hash functions is a collection of functions $\mathcal{H} \subseteq \{0,1\}^n \to \{0,1\}^m$ satisfying $\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq 2^{-m}$ for any $x \neq y$. There exist universal family of explicit hash functions of size 2^n.

Theorem 2.11 (Leftover Hash Lemma [ILL89])**.** *Let $X \sim \{0,1\}^n$ be such that $H_\infty(X) \geq k$, let $\varepsilon > 0$, and let $\mathcal{H} = \{h_1, \dots, h_N\} \subseteq \{0,1\}^n \to \{0,1\}^m$ be a universal family of hash functions for $m = k - 2\log\frac{1}{\varepsilon}$ of size 2^n. Define $\mathsf{Ext}_{\mathsf{ILL}}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ by*

$$\mathsf{Ext}_{\mathsf{ILL}}(x, y) = h_y(x).$$

Then, Ext is a strong (k, ε) extractor.

Theorem 2.12 ([GW97])**.** *For every positive integer n, and any $\Delta < n$ and $\varepsilon > 0$, there exists an explicit $(k = n - \Delta, \varepsilon)$ extractor $\mathsf{Ext}_{\mathsf{GW}}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, where $d = O(\Delta + \log\frac{1}{\varepsilon})$ and $m = n - O(\Delta + \log\frac{1}{\varepsilon})$.*

In seeded *condensers*, the goal is to improve the quality of a random source X using few additional random bits, albeit not necessarily into the uniform distribution.

Definition 2.13. *A function*

$$\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

is a (k, k', ε) (seeded) condenser for a class of sources $\mathcal{X}$ over n bits if the following holds. For every source $X \in \mathcal{X}$ it holds that $H_\infty^\varepsilon(\mathsf{Cond}(X, Y)) \geq k'$, where Y is uniformly distributed over $\{0,1\}^d$ and is independent of X. When $d = 0$, we say that $\mathcal{X}$ admits deterministic condensing.

2.4 Martingales

We use a few basic results about martingales. Recall that a martingale with respect to a sequence of random variables $X_1, \dots, X_t$ is a sequence of real random variables $Z_0, \dots, Z_t$ such that for all i, Z_i is a function of $X_1, \dots, X_t$, $\mathbb{E}[|Z_i|] < \infty$, and $\mathbb{E}[Z_{i+1} | X_1, \dots, X_i] = Z_i$. For a sequence of random variables $X_0, \dots, X_t$, we will specifically utilize the Doob martingale, which for a given Y that is a function of $X_0, \dots, X_t$, is the sequence $Z_i = \mathbb{E}[Y | X_0, \dots, X_i]$. It is well known (and easy to verify) that such a sequence $Z_0, \dots, Z_t$ is a martingale (with respect to itself).

We will use the Azuma-Hoeffding Inequality as a tail bound on a martingale.

Theorem 2.14 (Azuma–Hoeffding)**.** *Suppose $Z_0, \dots, Z_t$ is a martingale and that $|Z_i - Z_{i-1}| \leq c_i$. Then, for any $\varepsilon > 0$,*

$$\Pr[X_n - X_0 \geq \varepsilon] \leq e^{\frac{-\varepsilon^2}{2\sum_{i=1}^t c_i^2}}$$

[25] In fact, if $\mathcal{X}$ is the set of k-sources, the error probability can be made much smaller by letting Ext handle slightly smaller entropies. See [Zuc96].

3 Almost Chor–Goldreich Sources

We now give the formal definitions of the generalized CG sources that we work with. The first and main generalization is what we call an *almost CG source*. Such sources are similar to standard CG sources but allow two types of "errors".

1. Instead of each X_i being a δd-source for every prefix, each X_i is only γ-close to being a δd source;
2. Instead of having a good min-entropy guarantee for *every* $i \in [t]$, we have that for at most λt of the i-s, there is no guarantee on the quality of the distribution of X_i regardless of the prefix.

Before formally defining sources of the above form, we first define what it means for a prefix to be good, and for a block $i \in [t]$ to be good.

Definition 3.1 (good step). *Let $0 \leq \gamma, \delta \leq 1$. Let $X = X_1 \circ \ldots \circ X_t$ be a source with each $X_i \sim \{0,1\}^d$. We say that $i \in [t]$ is (γ, δ)-good for X if for all prefixes $(a_1, \ldots, a_{i-1}) \in \{0,1\}^{d(i-1)}$ we have that:*

$$H_\infty^\gamma(X_i | X_1, \ldots, X_{i-1} = a_1, \ldots, a_{i-1}) \geq \delta d.$$

(Note that for $i = 1$ we simply require $H_\infty^\gamma(X_1) \geq \delta d$.)

When γ, δ, and X are clear from context, we will simply call a coordinate i "good" or a "good step" without the quantifiers. We also call i "bad" or a "bad step" if it is not good. Additionally, we use $\mathcal{G}(X)$ as the set of all good i-s.

With this definition, we can define the notion of an almost CG source.

Definition 3.2 (almost CG source). *A (γ, λ)-almost δ-CG source is a sequence of random variables $X_1 \circ \ldots \circ X_t$ with $X_i \in \{0,1\}^d$, such that at least $(1-\lambda)t$ i-s are (γ, δ)-good for X.*

We will eventually show that given an almost CG source, we can deterministically condense it into a distribution on $m = \Omega(\delta d t)$ bits that is close to a $m - O(\lambda)dt - O(1)$-source. In other words, we can condense it into a source with $O(\lambda)dt + O(1)$ additive entropy gap. We remark that to the best of our knowledge, up until know there were no known constructions that deterministically condenses from such sources, even with $\gamma = 0$ and $\lambda = 0$.

The next lemma shows that grouping an almost CG source into blocks of length b results in an almost CG source with a much smaller smoothness error γ.

Lemma 3.3. *Let $X = X_1 \circ \ldots \circ X_t$ be a (γ, λ)-almost δ-CG source, with $X_i \sim \{0,1\}^d$. For any positive integer b, consider the distribution $X' = X'_1 \circ \ldots \circ X'_{\lfloor t/b \rfloor}$, where $X'_i = X_{[b(i-1)+1, b(i-1)+b]}$. Then:*

- *X' is a $\left(\gamma' = e^{-(1-\gamma)^2 b/8}, \sqrt{\lambda}\right)$-almost $\left(\frac{1-\gamma-2\sqrt{\lambda}}{2} \cdot \delta\right)$-CG source.*
- *X' is a $\left(\gamma' = e^{-(\gamma(1-\sqrt{\lambda}))^2 b/2}, \sqrt{\lambda}\right)$-almost $\left((1-2\gamma)(1-\sqrt{\lambda})\delta\right)$-CG source.*

Proof: Call the i-th super-block $X'_i = X_{[b(i-1)+1, b(i-1)+b]}$ "good" if less than $\sqrt{\lambda}$ fraction of the steps in the block are bad ones. By an averaging argument, there are at least $1 - \sqrt{\lambda}$ fraction of good blocks overall.

Fix any good $1 \leq i \leq \lfloor t/b \rfloor$ and fix any prefix $a = (a_1, \ldots, a_{b(i-1)}) \sim (X_1, \ldots, X_{b(i-1)})$. We show that the distribution X_i' conditioned on the prefix a is sufficiently close to an appropriate high entropy source. For the rest of this proof, for convenience and brevity, we use X_j to refer to the distribution of $X_{(i-1)b+j}$ conditioned on the fixed prefix a.

Since i is a good block, there are at least $(1-\sqrt{\lambda})b$ good steps within the block. Let $b' = (1-\sqrt{\lambda})b$. For each $j \in [b']$, define $Y_j(x_1, \ldots, x_b)$ as the indicator random variable that is 1 if and only if for the j-th good step in the block (call it $s(j)$),

$$\Pr[X_{s(j)} = x_{s(j)} | X_{[1,s(j)-1]} = x_{[1,s(j)-1]}] \geq D^{-\delta}.$$

Let $Y = \sum Y_j$. We define the Doob martingale

$$Z_j = \mathbb{E}[Y | X_1, \ldots, X_{s(j)}]$$

with the convention that $Z_0 = \mathbb{E}[Y]$. Note further that $Z_b = Y$. We can conclude that

$$Z_0 = \mathbb{E}[Y] = \sum_j \mathbb{E}[Y_j] \leq \gamma b' = \gamma(1-\sqrt{\lambda})b.$$

Further, we know that $|Z_j - Z_{j-1}| \leq 1$ for all j. For the first bullet point, by the Azuma–Hoeffding inequality (Theorem 2.14), we get

$$\Pr\left[Y - \gamma(1-\sqrt{\lambda})b > \frac{1-\gamma(1-\sqrt{\lambda})}{2}b\right] \leq \Pr\left[Z_b - Z_0 > \frac{1-\gamma(1-\sqrt{\lambda})}{2}b\right] \leq e^{-(1-\gamma)^2 b/8}.$$

Finally, we observe that any $x_1, \ldots, x_b$ for which $Y(x_1, \ldots, x_b) \leq \frac{1+\gamma(1-\sqrt{\lambda})}{2}b$ has expectation at most $\left(D^{-\delta}\right)^{(1-\sqrt{\lambda})b - \left(\frac{1+\gamma(1-\sqrt{\lambda})}{2}\right)b} \leq D^{-\left(\frac{1-\gamma-2\sqrt{\lambda}}{2}\right)\delta b}$.

For the second conclusion, the proof is identical, but we use the Azuma–Hoeffding Inequality with different parameters:

$$\Pr\left[Y > 2\gamma(1-\sqrt{\lambda})b\right] \leq \Pr\left[Z_b - Z_0 > \gamma(1-\sqrt{\lambda})b\right] \leq e^{-(\gamma(1-\sqrt{\lambda}))^2 b/2}.$$

■

As an interpretation of the above lemma, we note that the first bullet essentially says that the smoothness error is exponentially small in the length of the block. However, the rate of the CG-source suffers (in particular it is at least halved). On the other hand, the second bullet point states that such a loss in the entropy rate need not be necessary, in the right regime of parameters, such as when $\gamma < 1/2$.

3.1 Suffix-Friendly Almost CG Sources

Ultimately, we hope to condense almost CG sources into sources with *constant*, additive, entropy gap, since one can extract from such sources using only a constant number of auxiliary random bits (see Theorem 2.12). Looking ahead, we won't be able to do so unless we pose some restriction on the bad steps. This is since we condense in an "online" manner. Thus, if for example, all bad steps are at the end, we may lose roughly λdt bits of entropy overall. However, if we further require a good fraction of good steps from all *suffixes*, we can evade this problem. With this motivation in mind, we define the following.

Definition 3.4 (suffix-friendly almost CG source). *A $(\gamma, \lambda, \Lambda)$-suffix-friendly-almost δ-CG source is a sequence of random variables $X_1 \circ \ldots \circ X_t$ with $X_i \sim \{0,1\}^d$, such that for every suffix $X_j, \ldots, X_t$, for all but at most $(t-j+1)\lambda + \Lambda$ i-s between j and t, we have that i is (γ, δ)-good for X.*

When δ and γ are clear from context, we may refer to X as being suffix-friendly with parameters λ and Λ.

Such a definition is indeed natural: If each step is corrupted independently with probability λ, the resulting distribution will be suffix-friendly with parameters $O(\lambda)$ and $\Lambda = O(1/\lambda)$, with high probability:

Lemma 3.5. *Let $X = X_1 \circ \ldots \circ X_t$ be a $(\gamma, 0)$-almost δ-CG source. Let $Y = Y_1 \circ \ldots \circ Y_t$ be a sequence of independent Bernoulli random variables with $\Pr[Y_i = 1] = \lambda < \frac{1}{2}$. Suppose $X' = X'_1 \circ \ldots \circ X'_t$ is formed as follows. Independently, for each $i \in [t]$, we do the following.*

- *If $Y_i = 0$, we set the conditional distributions*

$$X'_i \mid \left\{(X'_1, \ldots, X'_{i-1}) = (a_1, \ldots a_{i-1})\right\} = X_i \mid \{(X_1, \ldots, X_{i-1}) = (a_1, \ldots a_{i-1})\}$$

  for every prefix $(a_1, \ldots a_{i-1}) \in \left(\{0,1\}^d\right)^{i-1}$.

- *If $Y_i = 1$, we set the conditional distribution $X'_i \mid \left\{(X'_1, \ldots, X'_{i-1}) = (a_1, \ldots a_{i-1})\right\}$ arbitrarily for every prefix $a_1, \ldots a_{i-1} \in \left(\{0,1\}^d\right)^{i-1}$.*

Then, with probability at least $\frac{9}{10}$ over the choice of Y, X' is a $(\gamma, 2\lambda, \Lambda)$-suffix-friendly-almost δ-CG source, where $\Lambda = O\left(\frac{\log(1/\lambda)}{\lambda}\right)$.[26]

Proof: For any $\ell \in [t]$ we say that X' corrupts ℓ if $Y_\ell = 1$. Fixing $j \in [t]$, we first bound the probability that X' corrupts more than $(t-j+1)\lambda + \Lambda$ coordinates in the suffix $X_j, \ldots, X_t$. For convenience, denote $i = t - j + 1$. Clearly, we can assume that $i > \Lambda$, so write $i = \Lambda + i'$ for $i' \in [t - j - \Lambda + 1]$. Let Z_i denote the number of corruptions in the suffix of length i. By Chernoff,

$$\Pr[Z_i \geq \Lambda + 2\lambda(\Lambda + i')] \leq \Pr[Z_i \geq 2\lambda i] \leq e^{-\lambda i/3} = e^{-\lambda\Lambda/3 - \lambda i'/3}.$$

To consider all suffixes, we union-bound over the above probability for all i'-s ranging from 1 to $t - j - \Lambda + 1 \leq t - \Lambda$. However, we will treat different i'-s differently. First, note that, for a sufficiently large t, there exists $i^\star$ for which $e^{-\frac{\lambda}{3} i^*} \leq \frac{1}{20} \frac{1 - e^{-\lambda/3}}{e^{-\lambda/3}}$. In fact, one can verify that $i^\star = O\left(\frac{\log 1/\lambda}{\lambda}\right)$.

For every $i' > i^\star$, we write $i' = i^\star + i''$, and can then union-bound over the events that $Z_{\Lambda + i'}$ is too large as follows:

$$\sum_{i''=1}^{t-\Lambda-i^*} e^{-\frac{\lambda}{3}\Lambda - \frac{\lambda}{3}(i^* + i'')} \leq \sum_{i''=1}^{\infty} e^{-\frac{\lambda}{3} i^* - \frac{\lambda}{3} i''} \leq \frac{1}{20} \frac{1 - e^{-\lambda/3}}{e^{-\lambda/3}} \sum_{i''=1}^{\infty} e^{-\frac{\lambda}{3} i''} \leq \frac{1}{20}.$$

For the case of $i' \leq i^\star$, since $i^\star = O\left(\frac{\log 1/\lambda}{\lambda}\right)$, we observe that for some $\Lambda = O(\log(1/\lambda)/\lambda)$ it holds that that $e^{-\frac{\lambda}{3}\Lambda} \cdot i^\star \leq \frac{1}{20}$. Overall, the probability that any suffix $X_j, \ldots, X_t$ has more than $2\lambda(t - j + 1) + \Lambda$ bad steps is at most $\frac{1}{10}$. ■

[26] The choice of $\frac{9}{10}$ is arbitrary, and the analysis can be easily extended to any success probability close to 1.

We can also prove a similar result to Lemma 3.3 about reducing the smoothness error γ for suffix friendly almost CG sources.

Lemma 3.6. *Let $X = X_1 \circ \ldots \circ X_t$ be a $(\gamma, \lambda, \Lambda)$-suffix-friendly-almost δ-CG source, with $X_i \sim \{0,1\}^d$. Let $0 < \theta < 1$ be any constant. For any positive integer b, consider the distribution $X' = X'_1 \circ \ldots \circ X'_{\lfloor t/b \rfloor}$, where $X'_i = (X_{b(i-1)+1}, \ldots, X_{b(i-1)+b})$. Then:*

- *X' is a $(\gamma = e^{-(1-\gamma)^2 b/8}, \sqrt{\lambda}, \frac{\Lambda}{\theta\delta b})$-suffix-friendly-almost $\left(\frac{1-\gamma-2\sqrt{\lambda}-2\theta\delta}{2}\right)$ δ-CG source.*
- *X' is a $(\gamma = e^{\frac{-(\gamma(1-\sqrt{\lambda}-\theta\delta))^2 b}{2}}, \sqrt{\lambda}, \frac{\Lambda}{\theta\delta b})$-suffix-friendly-almost $\left((1-2\gamma)(1-\sqrt{\lambda}-\theta\delta)\right)$ δ-CG source.*

Proof: Similarly to the proof of Lemma 3.3, call the j-th super-block, $X'_j = X_{(j-1)b+1} \circ \cdots \circ X_{(j-1)b+b}$ "good" if less than $\sqrt{\lambda} + \theta\delta$ fraction of the X_i-s in the block are bad steps.

We will show that for any suffix of the X'_j-s, there are at most $\frac{\Lambda}{\theta\delta b} + \sqrt{\lambda}(t-j)$ bad blocks. Consider any suffix of length s of the X'_j-s. There are at most $\Lambda + \lambda b s$ bad steps in X in this suffix. By an averaging argument, for any a, there are at most

$$\frac{\left(\frac{\Lambda}{s} + \lambda b\right) s}{a}$$

blocks with more than a bad steps in them. Setting $a = \sqrt{\lambda} b + \theta\delta b$ tells us that the number of bad blocks in the suffix is at most

$$\frac{\Lambda}{\sqrt{\lambda} b + \theta\delta b} + \frac{\lambda b}{\sqrt{\lambda} b + \theta\delta b} s \leq \frac{\Lambda}{\theta\delta b} + \sqrt{\lambda} s.$$

The proof then proceeds as in Lemma 3.3. For each good block, there are at least $b - a = (1 - \sqrt{\lambda} - \theta\delta) b$ good steps. The probability that these steps sample heavy instructions (those with probability more than $D^{-\delta}$) is at most γ. Thus, the expected fraction of heavy instructions from good steps is at most $\gamma(1 - \sqrt{\lambda} - \theta\delta)$. Applying Theorem 2.14 with either $\varepsilon = \frac{1-\gamma(1-\sqrt{\lambda}-\theta\delta)}{2}$ or $2\gamma(1 - \sqrt{\lambda} - \theta\delta)$ yields the two bullet points. ■

3.2 From Shannon Entropy to Min-Entropy

A weaker definition than almost CG sources are sources where each step has high *Shannon* entropy.

Definition 3.7 (Shannon CG source). *A λ-almost δ-Shannon-CG source is a sequence of random variables $X_1 \circ \ldots \circ X_t$ with $X_i \sim \{0,1\}^d$, such that for all but at most λt i-s, for all $a \in (\{0,1\}^d)^{i-1}$,*

$$H(X_i | X_1, \ldots, X_{i-1} = a) \geq \delta d.$$

One reason extractors are able to deal with min-entropy rather than Shannon entropy is because sources with high Shannon entropy could still output a constant outcome 99% of the time. When you only have one shot to extract a truly random output, such a source is useless. However, in the setting of a source that is in fact a sequence of many constant-length sources, each having high Shannon entropy, there are intuitively many chances for the source to output "good randomness". Considering many of the constant-length sources at once by grouping them into blocks, there is a very small probability of getting a high-probability outcome for the entire block. In other words, the block is close to a high min-entropy source.

We prove this formally here, giving a reduction from Shannon CG sources to almost CG sources. We begin with a simple claim about Shannon entropy that states that sources with high Shannon entropy are essentially smoothed min-entropy sources with error parameter close to 1.

Claim 3.8. *Let $X \sim \{0,1\}^d$. For any $\eta, \xi > 0$, define $A \subseteq \{0,1\}^d$ as $A = \{x : \Pr[X = x] \geq \eta\}$ and suppose that $\Pr[X \in A] \geq 1 - \xi$. Then,*

$$H(X) \leq \xi d + \log \frac{1}{\eta} + 1.$$

Proof: Write

$$H(X) = \sum_{x \in A} \Pr[X = x] \log \frac{1}{\Pr[X = x]} + \sum_{x \in \bar{A}} \Pr[X = x] \log \frac{1}{\Pr[X = x]}. \tag{1}$$

For the first term, note that

$$\begin{aligned}\sum_{x \in A} \Pr[X = x] \log \frac{1}{\Pr[X = x]} &= \sum_{x \in A} \Pr[X = x | X \in A] \Pr[X \in A] \log \frac{1}{\Pr[X = x | X \in A] \Pr[X \in A]} \\ &= \sum_{x \in A} \Pr[X = x | X \in A] \Pr[X \in A] \log \frac{1}{\Pr[X = x | X \in A]} \\ &\quad + \sum_{x \in A} \Pr[X = x | X \in A] \Pr[X \in A] \log \frac{1}{\Pr[X \in A]} \\ &\leq H(X | \{X \in A\}) + \Pr[X \in A] \log \frac{1}{\Pr[X \in A]},\end{aligned}$$

and observe that $H(X | \{X \in A\}) \leq \log \frac{1}{\eta}$ since $|A| \leq \frac{1}{\eta}$. Similarly, we can bound the second term of Equation (1) by

$$\sum_{x \in \bar{A}} \Pr[X = x] \log \frac{1}{\Pr[X = x]} \leq \xi \cdot H(X | \{X \in \bar{A}\}) + \Pr[X \notin A] \log \frac{1}{\Pr[X \notin A]},$$

and we can trivially bound $H(X | \{X \in \bar{A}\}) \leq d$. Finally, note that

$$\Pr[X \in A] \log \frac{1}{\Pr[X \in A]} + \Pr[X \notin A] \log \frac{1}{\Pr[X \notin A]} \leq 1,$$

and the claim follows by collecting terms. ■

Corollary 3.9. *Let $\delta > 0$. There exists $d^\star = d^\star(\delta) = O(1/\delta)$ such that for all $d > d^\star$, if X is a distribution on $\{0,1\}^d$, with $H(X) \geq \delta d$, then the total weight of elements x s.t. $\Pr[X = x] \geq \frac{1}{D^{\delta/3}}$ is at most $1 - \frac{\delta}{3}$.*

Proof: Suppose not, then by Claim 3.8, set with $\xi = \delta/3$ and $\eta = 1/D^{\delta/3}$, we get:

$$H(X) \leq \xi d + \log \frac{1}{\eta} + 1 \leq \frac{2}{3}\delta d + 1.$$

If $d^\star = \frac{3}{\delta} = O(1/\delta)$, then for any $d > d^\star$ the above expression is at most δd. ■

From the above corollary we immediately see that we can view Shannon CG sources as almost CG sources with very high smoothness error γ

Corollary 3.10. *Let $\delta > 0$. There exists $d^\star = d^\star(\delta) = O(1/\delta)$ such that for any $d > d^\star$ the following holds. Let $X = X_1 \circ \ldots \circ X_t$ be a λ-almost δ-Shannon-CG source, with $X_i \in \{0,1\}^d$. Then X is a $(\gamma = 1 - \delta/3, \lambda)$-almost $(\delta/3)$-CG source.*

Similarly, if X is a (λ, Λ)-suffix-friendly-almost δ-Shannon-CG source, then, X is a $(\gamma = 1 - \delta/3, \lambda, \Lambda)$-almost $(\delta/3)$-CG-source.

Using the first bullet of Lemma 3.3 or Lemma 3.6, and Corollary 3.10, we see that we can convert Shannon CG sources to almost CG sources with small smoothness parameter γ.[27]

Corollary 3.11. *Let $X = X_1 \circ \ldots \circ X_t$ be a λ-almost δ-Shannon-CG source, with $X_i \sim \{0,1\}^d$. Suppose d is sufficiently large as in Corollary 3.9. Suppose further that $\sqrt{\lambda} \le \delta/12$. For any positive integer b, consider the distribution $X' = X'_1 \circ \ldots \circ X'_{\lfloor t/b \rfloor}$, where $X'_i = (X_{b(i-1)+1}, \ldots, X_{b(i-1)+b})$. Then, X' is a $(\gamma = e^{-\delta^2 b/72}, \sqrt{\lambda})$-almost $(\delta^2/36)$-CG source.*

One can show the following suffix-friendly variant of the corollary by applying the first bullet of Lemma 3.6 with $\theta = 1/12$.

Corollary 3.12. *Let $X = X_1 \circ \ldots \circ X_t$ be a (λ, Λ)-suffix-friendly-almost δ-Shannon-CG source, with $X_i \sim \{0,1\}^d$. Suppose d is sufficiently large as in Corollary 3.9. Suppose further that $\sqrt{\lambda} \le \delta/24$. For any positive integer b, consider the distribution $X' = X'_1 \circ \ldots \circ X'_{\lfloor t/b \rfloor}$, where $X'_i = (X_{b(i-1)+1}, \ldots, X_{b(i-1)+b})$. Then, X' is a $(\gamma = e^{-\delta^2 b/72}, \sqrt{\lambda}, \frac{12\Lambda}{\delta b})$-suffix-friendly-almost $(\delta^2/36)$-CG source.*

In Section 8 we will show how considering Shannon CG sources can help explain why it might be difficult to deterministically condense from *even more generalized* notions of the almost CG sources defined above.

3.3 On Almost δ-CG Sources and δ-CG Sources

As discussed in the introduction, we show that, although any $(\gamma, 0)$-almost δ-CG source is close to *some* (general) $(1 - 2\gamma)\delta$-rate weak source (see Claim 3.13 below), there are $(\gamma, 0)$-almost δ-CG sources that are far from *any* $(1 - 2\gamma)\delta$-CG source (see Claim 3.14). This tells us that although it is plausible that one can extract roughly δdt bits of entropy from such a source (as we do), one cannot do so by simply applying a technique for condensing standard CG sources. The two claims below establish the above discussion formally.

Claim 3.13. *Let $X = X_1 \circ \ldots \circ X_t$ be a $(\gamma, 0)$-almost δ-CG source. Then, X is ε-close to a $(1-2\gamma)\delta dt$-source, where $\varepsilon = e^{-\gamma^2 t/2}$.*

Proof: The proof is nearly identical to that in Lemma 3.3, considering the entire $X_1, \ldots, X_t$ as a single block. ■

Claim 3.14. *For any positive integers t, d, and any $1 > \delta > 0$ and $\gamma \le \frac{1}{4}$ such that $d \ge \frac{4\log(1/\gamma)}{\delta}$, there exists a $(\gamma, 0)$-almost δ-CG source $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, that is $\left(1 - e^{-\gamma t/8} - 2^{-\delta d/4+1}\right)$-far from any $(1 - 2\gamma)\delta$-CG source.*

[27] This can be seen as a quantitative manifestation of the following phenomena: Let $X^{(1)}, \ldots, X^{(t)}$ be independent copies of some random variable X. Then, $\frac{1}{t}H_\infty(X_1 \circ \ldots \circ X_t)$ approaches $H(X)$ as t tends to inifnity.

Proof: Consider the $(\gamma, 0)$-almost δ-CG source $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, where for every i, and every prefix $a = a_1, \ldots, a_{i-1}$, the distribution of X_i conditioned on a is the following convex combination.

- With probability γ, the output is the fixed zero string 0^d.
- With probability $1 - \gamma$, the output is a sample from an arbitrary δd-source whose support does not contain 0^d.

Let Y be *any* $(1-2\gamma)\delta$-CG source. Consider the test $T\colon \{0,1\}^{dt} \to \{0,1\}$ for which $T(x_1, \ldots, x_t) = 1$ iff at least $\gamma t/2$ of the x_i-s are 0^d. The expected number of x_i-s that are 0^d under X is γt. By Chernoff, $\Pr[T(X) = 1] \geq 1 - e^{-\gamma t/8}$.

On the other hand, under Y, we know that the expected number of x_i-s that are 0^d is at most $D^{-(1-2\gamma)\delta}t$. By Markov's inequality (note that now we don't necessarily have independence), the probability of seeing more than $\gamma t/2$ zeros is at most

$$\frac{2^{-(1-2\gamma)\delta d}t}{\gamma t/2} = \frac{2^{-(1-2\gamma)\delta d+1}}{\gamma} \leq 2^{-\delta d/4+1}.$$

Thus, $|\mathbb{E}[T(X)] - \mathbb{E}[T(Y)]| \geq 1 - e^{-\gamma t/8} - 2^{-\delta/4+1}$, implying the same lower bound on $|X - Y|$ as well. ■

4 Deterministic Condensing via Lossless Expanders

In this section we formalize our technique for deterministic condensing from almost CG sources when $\delta \geq \delta_{\mathsf{thr}}$. (Later we will show that we can also *extract* from *suffix-friendly* almost CG sources; even later we'll show how to do all of this for arbitrary $\delta > 0$). At a high level, we simply treat the almost CG source as instructions for a random walk on a lossless expander of degree D starting from an arbitrary fixed node. We analyze how the distribution on the expander's vertices evolves with each step by bounding its ℓ_q-norm for a suitable $1 < q < 2$.

4.1 Additional Framework

There are two main concepts to formalize in order to handle even the case of standard CG sources. First, since we use the ℓ_q norm as a proxy measure for entropy, we need to argue that vertex expansion implies a good multiplicative factor ℓ_q-norm decrease when the next instruction has sufficiently large entropy $\delta \geq \delta_{\mathsf{thr}}$, and the current distribution does not have too small ℓ_q norm already (roughly $\frac{1}{K^\alpha}$). The framework in this section requires only the expansion property of left sets, and requires no upper bound on the right degree.

We begin by generalizing some properties of vertex expansion to expansion of *weight functions* as follows. Consider a bipartite graph $G = (U, V, E)$. A weight function is simply an assignment of nonnegative real numbers to the vertices of U or V. In other words, we consider functions $w\colon U \to [0, \infty)$, or $w\colon V \to [0, \infty)$, and we denote by $|w|$ its ℓ_1 norm: $\sum_{u \in U} w(u)$ or $\sum_{v \in V} w(v)$. Given a weight function on U, we can define the "neighboring" weight function on V.

Definition 4.1. *Let $G = (U, V, E)$. Let $w\colon U \to [0, \infty)$ be a weight function on U. We define $\mathcal{N}(w)\colon V \to [0, \infty)$ as:*

$$\mathcal{N}(w)(v) = \max_{u \in \Gamma(v)} w(u)$$

When the weight function w is clear from context, for any $v \in V$ we may use the notation u_v to denote $\arg\max_{u \in \Gamma(v)} w(u)$ *(with ties broken arbitrarily).*

Observe that the above notion generalizes the notion of neighbor sets. That is, when w corresponds to the indicator function of a set of nodes in $S \subset U$, then $\mathcal{N}(w)$ is the indicator function of $\Gamma(S) \subset V$.

Lemma 4.2. *Let $G = (U, V, E)$ be a (K, ε, e)-expander for any e. For all weight functions $w\colon U \to [0, \infty)$ supported on at most K nodes, $|\mathcal{N}(w)| \geq (1 - \varepsilon)D|w|$.*

Proof: Write $w = w_1 + \ldots + w_t$, where each w_i is a multiple of an indicator function on a set of size S_i at most K, and $S_i \supseteq S_{i+1}$ for all i. Observe that in this case, $|\mathcal{N}(w)| = \sum_i |\mathcal{N}(w_i)|$. Since $|\Gamma(S_i)| \geq (1 - \varepsilon)D|S_i|$, we have $|\mathcal{N}(w_i)| \geq (1 - \varepsilon)D|w_i|$. Thus,

$$|\mathcal{N}(w)| = \sum_i |\mathcal{N}(w_i)| \geq \sum_i (1 - \varepsilon)D|w_i| = (1 - \varepsilon)D|w|.$$

■

In Appendix D, we also generalize the notion of "unique neighbor expansion" in expander graphs using similar arguments (although this is not necessary for our analysis).

We now generalize Lemma 4.2 to work for distributions with large ℓ_q-norm for $q = 1 + \alpha$, rather than just weight functions supported on at most K nodes. We first prove a simple claim about such distributions.

Claim 4.3. *Let $0 < \alpha < 1$ and let $q = 1 + \alpha$. Let $K > 1$. Suppose p is a probability distribution over a finite domain U such that $\sum_{u \in U} p(u)^q \geq \frac{1}{K^\alpha}$. For any $r > 1$, let $T_r \subseteq U$ be the set of heaviest rK elements of U according to p. Then,*

$$\sum_{u \in T_r} p(u)^q \geq \left(1 - \frac{1}{r^\alpha}\right) \sum_{u \in U} p(u)^q.$$

Proof: First, for every $u \in U \setminus T_r$, $p(u) \leq \frac{1}{rK}$. We have:

$$\sum_{u \in U \setminus T_r} p(u)^q \leq \max_{u \in U \setminus T_r} p(u)^\alpha \sum_{u \in U \setminus T_r} p(u) = \left(\frac{1}{rK}\right)^\alpha.$$

Thus,

$$\begin{aligned}\sum_{u \in T_r} p(u)^q &= \sum_{u \in U} p(u)^q - \sum_{u \in U \setminus T_r} p(u)^q \\ &= \left(1 - \frac{\sum_{u \in U \setminus T_r} p(u)^q}{\sum_{u \in U} p(u)^q}\right) \sum_{u \in U} p(u)^q \geq \left(1 - \frac{1}{r^\alpha}\right) \sum_{u \in U} p(u)^q.\end{aligned}$$

■

We can now generalize Lemma 4.2 to hold not only when a weight function w is supported on at most K nodes, but also when the weight function represents the contribution of each node to the ℓ_q-norm of a distribution.

Lemma 4.4. *Let $G = (U, V, E)$ be a (K, ε, e)-expander for any e. Let $0 < \alpha, \eta < 1$, and set $q = 1 + \alpha$. Let p be any distribution over U such that $\sum_{u \in U} p(u)^q \geq \frac{1}{(\eta K)^\alpha}$. Then,*

$$\sum_{v \in V} \max_{u \in \Gamma(v)} p(u)^q \geq (1 - \varepsilon - \eta^\alpha) D \sum_{u \in U} p(u)^q.$$

Proof: Let $T \subseteq U$ be the heaviest $\frac{1}{\eta} \cdot \eta K = K$ nodes. Note that: $\sum_{v \in V} \max_{u \in \Gamma(v)} p(u)^q \geq \sum_{v \in \Gamma(T)} \max_{u \in \Gamma(v)} p(u)^q$. Let $w(\cdot)$ be a weight function on U such that $w(u) = p(u)^q$ for $u \in T$, and $w(u) = 0$ otherwise. Notice that $|\mathcal{N}(w)| = \sum_{v \in \Gamma(T)} \max_{u \in \Gamma(v)} p(u)^q$. Since w is supported on K nodes, Lemma 4.2 implies

$$\sum_{v \in V} \max_{u \in \Gamma(v)} p(u)^q \geq (1 - \varepsilon) D |w|.$$

Finally, Claim 4.3 tells us:

$$(1 - \varepsilon) D |w| \geq (1 - \varepsilon)(1 - \eta^\alpha) D \sum_{u \in U} p(u)^q \geq (1 - \varepsilon - \eta^\alpha) D \sum_{u \in U} p(u)^q.$$

■

We utilize the framework we have developed so far to prove the following useful corollary.

Corollary 4.5. *Let $G = (U, V, E)$ be a bipartite D-regular (K, ε, e)-expander for any e. Let $\alpha > 0$, and let p be a probability distribution on U with $\|p\|_q^q \geq \frac{1}{\varepsilon K^\alpha}$. Then,*

$$\sum_{v \in V} \sum_{u \in \Gamma(v) \setminus u_v} p(u)^q \leq 2 \varepsilon D \sum_{u \in U} p(u)^q.$$

Proof: Applying Lemma 4.4 with $\eta = \varepsilon^{1/\alpha}$ gives us

$$|\mathcal{N}(w)| \geq (1 - 2\varepsilon) D |w|.$$

Rearranging gives the result. ■

4.2 The Analysis for Standard Sources

To demonstrate some of our key ideas in a familiar setting, we now show how to deterministically condense from a standard CG source (i.e., $\gamma = \lambda = 0$) with our technique. We first show how the ℓ_q-norm decreases at each step.

Lemma 4.6. *Let $G = (U, V, E)$ be a bipartite D-regular $(K, \varepsilon = \frac{1}{D^\beta}, e)$-expander for any e. For any $0 < \alpha < \frac{\beta}{e}$, set $q = 1 + \alpha$ and let $\delta \geq 1 - \beta + e\alpha$.*

Let p_U be a probability distribution over U and let r_u, for each $u \in U$, be a distribution over $\{0, 1\}^d \equiv [D]$, each being a δd source. For any $u \in U$ and $v \in V$ let $r_u(u, v)$ denote the probability that the edge leading from u to v is chosen under r_u. That is, for G's labelling function $\ell \colon E \to [D]$, we denote $r_u(u, v) \equiv r_u(\ell(u, v))$. Define p_V as the induced probability distribution on V. That is,

$$p_V(v) = \sum_{u \in \Gamma(v)} r_u(u, v) p_U(u). \tag{2}$$

Suppose that $\|p_U\|_q^q \geq \frac{1}{\varepsilon K^\alpha}$. Then,

$$\|p_V\|_q^q \leq \frac{8}{D^{\delta\alpha}} \cdot \|p_U\|_q^q.$$

Proof: By definition, for each $v \in V$,

$$p_V(v) = r_{u_v}(u_v, v) p_U(u_v) + \sum_{u \in \Gamma(v) \setminus u_v} r_u(u, v) p_U(u). \tag{3}$$

Now, note that:

$$\sum_{v \in V} r_{u_v}(u_v, v) p_U(u_v)^q \leq \sum_{v \in V} \sum_{u \in \Gamma(v)} r_u(u, v) p_U(u)^q = \sum_{u \in U} \sum_{v \in \Gamma(u)} r_u(u, v) p_U(u)^q \leq \sum_{u \in U} p_U(u)^q. \tag{4}$$

Note that Equation (4) is true for *any* distributions r_u and p_U.

The point is we want to raise the first term of Equation (3) to the q, as this will be our gain. Thus, for each fixed v, noting that v has at most D^e neighbors, we apply Jensen's inequality $(\sum_i \lambda_i x_i)^q \leq \sum_i \lambda_i x_i^q$ for $\sum_{i \in [D^e]} \lambda_i = 1$, with:

- $\lambda_1 = 1/2$,
- $\lambda_i = \frac{1}{2(D^e - 1)}$ for $i \in \{2, \ldots, D^e\}$,
- $x_1 = 2 r_{u_v}(u_v, v) p_U(u_v)$, and,
- $x_i = 2(D^e - 1) r_u(u, v) p_U(u)$ where u is the i-th neighbor of v (and 0 if there is no i'th neighbor) for $i \in \{2, \ldots, D^e\}$.

Thus, noticing that $\sum_i \lambda_i x_i = p_V(v)$, we have:

$$\begin{aligned}
\sum_{v \in V} p_V(v)^q &\leq \sum_{v \in V} \left(\frac{1}{2} \left(2 r_{u_v}(u_v, v) p_U(u_v) \right)^q + \sum_{u \in \Gamma(v) \setminus u_v} \frac{1}{2(D^e - 1)} \left(2(D^e - 1) r_u(u, v) p_U(u) \right)^q \right) \\
&\leq 2^\alpha \sum_{v \in V} r_{u_v}(u_v, v)^\alpha \cdot r_{u_v}(u_v, v) p_U(u_v)^q + 2^\alpha D^{e\alpha} \sum_{v \in V} \sum_{u \in \Gamma(v) \setminus u_v} r_u(u, v)^q p_U(u)^q \\
&\leq 2^\alpha \frac{1}{D^{\delta\alpha}} \sum_{v \in V} r_{u_v}(u_v, v) p_U(u_v)^q + 2^\alpha D^{e\alpha} \frac{1}{D^{\delta q}} \sum_{v \in V} \sum_{u \in \Gamma(v) \setminus u_v} p_U(u)^q \\
&\leq \frac{2^\alpha}{D^{\delta\alpha}} \sum_{u \in U} p_U(u)^q + 2^\alpha D^{e\alpha} \frac{1}{D^{\delta q}} \cdot 2\varepsilon D \sum_{u \in U} p_U(u)^q \\
&\leq \frac{8}{D^{\delta\alpha}} \cdot \|p_U\|_q^q .
\end{aligned}$$

In the second to last inequality, we used Corollary 4.5. In the last inequality, we used the fact that if $\delta \geq 1 - \beta + e\alpha$, then

$$\varepsilon D^{1 + e\alpha - \delta q} = D^{1 - \beta + e\alpha - \delta - \delta\alpha} < D^{-\delta\alpha}. \tag{5}$$

■

We next give a general lemma that states that, for *any* distribution on nodes, and *any* collection of distributions on edges, the ℓ_q-norm cannot increase too much. This will be useful toward analyzing the case when a good step occurs, but the ℓ_q-norm is already small enough, or in the next section, when a bad step occurs at any time.

Lemma 4.7. *Let $G = (U, V, E)$ be any bipartite D-regular graph with maximum right degree at most D^e. Let p_U be any probability distribution on U. Let r_u, for each $u \in U$, be any distributions over $\{0,1\}^d \equiv [D]$. Let p_V be the induced probability distribution on V:*

$$p_V(v) = \sum_{u \in \Gamma(v)} r_u(u,v) p_U(u).$$

Then,

$$\|p_V\|_q^q \leq D^{e\alpha} \cdot \|p_U\|_q^q.$$

Proof: Using Jensen's inequality, we get

$$\begin{aligned}\|p_V\|_q^q &= \sum_{v \in V} \left(\sum_{u \in \Gamma(v)} r_u(u,v) p_U(u) \right)^q = \sum_{v \in V} \left(\sum_{u \in \Gamma(v)} \frac{1}{D^e} D^e \cdot r_u(u,v) p_U(u) \right)^q \\ &\leq D^{e\alpha} \sum_{v \in V} \sum_{u \in \Gamma(v)} r_u(u,v)^q p_U(u)^q \leq D^{e\alpha} \sum_{v \in V} \sum_{u \in \Gamma(v)} r_u(u,v) p_U(u)^q = D^{e\alpha} \sum_{u \in U} p_U(u)^q.\end{aligned}$$

■

We now show how to use a lossless expander with error $\varepsilon = 1/D^\beta$ to condense a δ-CG source for sufficiently large δ (relative to $1-\beta$). On first reading it may be instructive to think of $1 - \beta = \delta/2$ and $\Delta = \delta/2$. The interpretation of the following theorem is that the guarantee from Lemma 4.6 implies that a decent amount of entropy is gained at each step of a random walk (assuming D^δ is large compared to $8^{1/\alpha}$). Thus if the total entropy gained after t steps is comparable to the "capacity" $k = \log K$ of the lossless expander, then the final distribution of the random walk would have entropy close to that capacity.

Theorem 4.8. *Let $1 > \beta > \Delta > 0$ be constants. Let $\delta \geq \delta_{\mathsf{thr}} = 1 - \beta + \Delta$. Let $X_1 \circ \ldots \circ X_t$ be a δ-CG source, with each $X_i \sim \{0,1\}^d$. Let $G = (U = [N], V = [N], E)$ be a D-regular $(K = 2^k, \varepsilon = \frac{1}{D^\beta}, e)$-expander for some constant e, and where $d = \log D \geq \frac{30e}{\Delta\delta}$. Further, suppose that*

$$0.9\delta dt \geq k - \frac{e}{\Delta} \log \frac{1}{\varepsilon} = k - e\frac{\beta}{\Delta} d.$$

Consider the distribution on the vertices of G after a random walk according to $X_1, \ldots, X_t$ starting from an arbitrary node. Namely, let $Z_0 \sim [N]$ be concentrated on an arbitrary fixed node, and for $i \in [t]$ let

$$Z_i = \Gamma_G(Z_{i-1}, X_i).$$

Then, for any $\eta > 0$, Z_t is η-close to a $\left(k - \left(\frac{\beta}{\Delta} + 1\right) ed - \frac{e}{\Delta} \log \frac{1}{\eta}\right)$-source.

Proof: Let $p_i \in \mathbb{R}^V$ denote the distribution of Z_i. For any $i \in [t]$ and $v \in V$,

$$p_i(v) = \sum_{u \in \Gamma(v)} \Pr\left[X_i = \ell_G(u,v) | Z_{i-1} = u\right] \cdot p_{i-1}(u),$$

where $\ell_G \colon E \to [D]$ is the labeling of the edges. In order to apply Lemma 4.6, note that for any $i \in [t]$ and $u \in U$, $X_i | \{Z_{i-1} = u\}$ is a δd-source. This is since Z_{i-1} is a deterministic function of $X_1, \ldots, X_{i-1}$, so $X_i | \{Z_{i-1} = u\}$ is a convex combination of $X_i | \left\{X_{[1,i-1]} = a_{[1,i-1]}\right\}$ for some $(a_1, \ldots, a_{i-1})$-s, each of which is a δd-source, by the definition of a δ-CG source.

Set $\alpha = \Delta/e$ and $q = 1 + \alpha$. We first claim that there must exist a timestep $s \in [t]$ such that $\|p_s\|_q^q \leq \frac{1}{\varepsilon K^\alpha}$. Suppose not. Then, at every timestep i, we can apply Lemma 4.6 (with $\alpha = \Delta/e < \beta/e$). Since $\|p_0\|_q^q = 1$, we have that $\|p_t\|_q^q \leq \left(\frac{8}{D^{\delta\alpha}}\right)^t$. However, by our assumption that $\frac{30e}{\Delta\delta} \leq d$, we have $8^{1/\alpha} \leq D^{0.1\delta}$. Also by our assumption, we know that $D^{.9\delta t} \geq \varepsilon^{1/\alpha} K$. Therefore :

$$\|p_t\|_q^q \leq \left(\frac{8}{D^{\delta\alpha}}\right)^t = \left(\frac{8^{1/\alpha}}{D^\delta}\right)^{\alpha t} \leq \left(\frac{1}{D^{0.9\delta t}}\right)^\alpha \leq \left(\frac{1}{\varepsilon^{1/\alpha} K}\right)^\alpha,$$

a contradiction.

Now, let $\ell \in [t]$ be the *last* timestep in which $\|p_\ell\|_q^q \leq \frac{1}{\varepsilon K^\alpha}$. There are two cases to consider.

1. $\ell = t$, in which case: $\|p_t\|_q^q \leq \left(\frac{1}{\varepsilon^{1/\alpha} K}\right)^\alpha$.

2. $\ell < t$. In this case, we must have that $\|p_{\ell+1}\|_q^q > \left(\frac{1}{\varepsilon^{1/\alpha} K}\right)^\alpha$. By Lemma 4.7, the $\ell + 1$-th step can only increase the norm by a factor of $D^{e\alpha}$. Since ℓ is the last time the norm is too small to apply Lemma 4.6, we can apply Lemma 4.6 to every step after $\ell + 1$, and so the norm must decrease every step after $\ell + 1$. Thus we have:

$$\left(\frac{D^e}{\varepsilon^{1/\alpha} K}\right)^\alpha \geq \|p_{\ell+1}\|_q^q \geq \ldots \geq \|p_t\|_q^q > \left(\frac{1}{\varepsilon^{1/\alpha} K}\right)^\alpha$$

   In particular we have $\|p_t\|_q^q \leq \left(\frac{D^e}{\varepsilon^{1/\alpha} K}\right)^\alpha$.

In either case, by Corollary 2.3, for any $\eta > 0$, Z_t is η-close to a $\left(k - \frac{\beta}{\alpha} d - ed - \frac{1}{\alpha} \log \frac{1}{\eta}\right)$-source. ■

Recall that in both an optimal lossless expander and in the expander from [CRVW02], $k = n - O(d)$, so Z_t above is close to a source with *constant entropy gap*. Moreover, when the size of the expander is chosen properly in comparison to the amount of entropy in the source (i.e. tightness in the constraint $0.9\delta dt = k - \frac{1}{\Delta} \log \frac{1}{\varepsilon}$), then the entropy loss is roughly $0.1\delta dt$.

4.3 The General Case

We now show how to handle the case when each conditional distribution is only γ-close to having $\delta \geq \delta_{\mathsf{thr}}$ entropy rate, and there are at most λ fraction of bad steps in the source. To handle the error parameter γ, we modify the proof of Lemma 4.6 to show that the ℓ_q norm essentially decreases by a factor of $O\left(\frac{1}{D^{\delta\alpha}} + \gamma D^{e\alpha}\right)$. We then choose α small enough so that $\gamma D^{e\alpha}$ is comparable to $\frac{1}{D^{\delta\alpha}}$ and so the decrease in ℓ_q norm is similar to that in Lemma 4.6 (i.e., overall $O\left(\frac{1}{D^{\delta\alpha}}\right)$). To handle the case of a bad step, we apply Lemma 4.7 and show that since there are λt such steps, overall they do not affect the entropy of the random walk too much.

4.3.1 Handling the Case of $\gamma > 0$

The following lemma shows in general how using a distribution that is γ-close to a δd-source affects the ℓ_q norm.

Lemma 4.9. *Let $G = (U, V, E)$ be a bipartite D-regular (K, ε, e)-expander for some e. For any $\alpha > 0$, let $q = 1 + \alpha$, and fix some $\gamma > 0$.*

Let p_U be a probability distribution on U and let r_u for each $u \in U$ be a collection of distributions over $\{0,1\}^d \equiv [D]$, each γ-close to a δd source. Suppose that $\|p_U\|_q^q \geq \frac{1}{\varepsilon K^\alpha}$. Let p_V be the induced probability distribution on V:

$$p_V(v) = \sum_{u \in \Gamma(v)} r_u(u, v) p_U(u).$$

Then,

$$\|p_V\|_q^q \leq \left(\frac{2^\alpha}{D^{\delta\alpha}} + 2^q \varepsilon D^{1-\delta+e\alpha-\delta\alpha} + 2^q q\gamma + 2^q D^{e\alpha} q\gamma \right) \|p_U\|_q^q .$$

Proof: The proof is similar to that of Lemma 4.6. We divide the contribution of each right hand node $v \in V$ into a contribution from the heaviest left hand neighbor, and the contribution from the rest of the neighbors. We can apply Jensen's inequality in the same way to work directly with sums of terms of the form $r_u(u, v)^q p_U(u)^q$. We'll show that certain sums of this form are close to sums of terms of the form $a_u(u, v)^q p_U(u)^q$ for the δd-source a_u that each r_u is close to. Toward this end, we first prove a small claim:

Claim 4.10. *For every $r_u(\cdot)$, let $a_u(\cdot)$ be the corresponding δd-source it is γ-close to. For every $v \in V$, let T_v be an arbitrary subset of $\Gamma(v)$. Then,*

$$\sum_{v \in V} \sum_{u \in T_v} r_u(u, v)^q p_U(u)^q \leq \sum_{v \in V} \sum_{u \in T_v} a_u(u, v)^q p_U(u)^q + 2q\gamma \cdot \sum_{u \in U} p_U(u)^q.$$

Proof: We will show that

$$\left| \sum_{v \in V} \sum_{u \in T_v} r_u(u, v)^q p_U(u)^q - \sum_{v \in V} \sum_{u \in T_v} a_u(u, v)^q p_U(u)^q \right| \leq 2q\gamma \cdot \sum_{u \in U} p_U(u)^q.$$

First, note that the collection of subsets $T_v \subseteq \Gamma(v)$ for $v \in V$ naturally induces a collection of subsets $S_u \subseteq \Gamma(u)$ where $S_u = \{v : u \in T_v\}$. Thus, it suffices to show that

$$\left| \sum_{u \in U} \sum_{v \in S_u} r_u(u, v)^q p_U(u)^q - \sum_{u \in U} \sum_{v \in S_u} a_u(u, v)^q p_U(u)^q \right| \leq 2q\gamma \cdot \sum_{u \in U} p_U(u)^q.$$

Now note that the Lipschitz constant of the function $f(x) = x^q$ on $[0, 1]$ is q. In other words, for every $x, y \in [0, 1]$, $|x^q - y^q| \leq q \cdot |x - y|$. We use this fact, together with the triangle inequality and the fact that r_u and a_u are γ-close, to get:

$$\begin{aligned}\left|\sum_{u\in U}\sum_{v\in S_u} r_u(u,v)^q p_U(u)^q - \sum_{u\in U}\sum_{v\in S_u} a_u(u,v)^q p_U(u)^q\right| &\le \sum_{u\in U}\sum_{v\in S_u} \left|r_u(u,v)^q p_U(u)^q - a_u(u,v)^q p_U(u)^q\right| \\ &\le \sum_{u\in U} p_U(u)^q \sum_{v\in\Gamma(u)} \left|r_u(u,v)^q - a_u(u,v)^q\right| \\ &\le \sum_{u\in U} p_U(u)^q \sum_{v\in\Gamma(u)} q\left|r_u(u,v) - a_u(u,v)\right| \\ &\le 2q\gamma\cdot\sum_{u\in U} p_U(u)^q.\end{aligned}$$

■

Now again, we write p_V as:

$$p_V(v) = r_{u_v}(u_v,v)p_U(u_v) + \sum_{u\in\Gamma(v)\setminus u_v} r_u(u,v)p_U(u).$$

Again, for each v, we apply Jensen's inequality in the same way as in the proof of Lemma 4.6.

$$\begin{aligned}\sum_{v\in V} p_V(v)^q &\le \sum_{v\in V}\left(\frac{1}{2}\left(2r_{u_v}(u_v,v)p_U(u_v)\right)^q + \sum_{u\in\Gamma(v)\setminus u_v}\frac{1}{2(D^e-1)}\left(2(D^e-1)r_u(u,v)p_U(u)\right)^q\right) \\ &\le 2^\alpha\sum_{v\in V} r_{u_v}(u_v,v)^q p_U(u_v)^q + 2^\alpha D^{e\alpha}\sum_{v\in V}\sum_{u\in\Gamma(v)\setminus u_v} r_u(u,v)^q p_U(u)^q.\end{aligned}$$

We show that:

$$\sum_{v\in V} r_{u_v}(u_v,v)^q p_U(u_v)^q \le \left(\frac{1}{D^{\delta\alpha}} + 2q\gamma\right)\|p_U\|_q^q, \tag{6}$$

and that:

$$\sum_{v\in V}\sum_{u\in\Gamma(v)\setminus u_v} r_u(u,v)^q p_U(u)^q \le \left(\frac{2\varepsilon D}{D^{\delta q}} + 2q\gamma\right)\|p_U\|_q^q. \tag{7}$$

For the first summand, by Claim 4.10 we have:

$$\begin{aligned}\sum_{v\in V} r_{u_v}(u_v,v)^q p_U(u_v)^q &\le \sum_{v\in V} a_{u_v}(u_v,v)^q p_U(u_v)^q + 2q\gamma\cdot\sum_{u\in U} p_U(u)^q \\ &= \sum_{v\in V} a_{u_v}(u_v,v)^\alpha\cdot a_{u_v}(u_v,v)p_U(u_v)^q + 2q\gamma\cdot\sum_{u\in U} p_U(u)^q \\ &\le \frac{1}{D^{\delta\alpha}}\sum_{v\in V} a_{u_v}(u_v,v)p_U(u_v)^q + 2q\gamma\cdot\sum_{u\in U} p_U(u)^q \\ &\le \frac{1}{D^{\delta\alpha}}\sum_{u\in U} p_U(u)^q + 2q\gamma\cdot\sum_{u\in U} p_U(u)^q.\end{aligned}$$

In the last inequality, we used Equation (4). Next:

$$\sum_{v\in V}\sum_{u\in\Gamma(v)\setminus u_v} r_u(u,v)^q p_U(u)^q \le \sum_{v\in V}\sum_{u\in\Gamma(v)\setminus u_v} a_u(u,v)^q p_U(u)^q + 2q\gamma\cdot\sum_{u\in U} p_U(u)^q$$
$$\le \frac{2\varepsilon D}{D^{\delta q}}\sum_{u\in U} p_U(u)^q + 2q\gamma\cdot\sum_{u\in U} p_U(u)^q,$$

where in the last inequality, we used Corollary 4.5. Putting Equation (6) and Equation (7) together gives:

$$\sum_{v\in V} p_V(v)^q \le \left(2^\alpha\left(\frac{1}{D^{\delta\alpha}}+2q\gamma\right)+2^\alpha D^{e\alpha}\left(\frac{2\varepsilon D}{D^{\delta q}}+2q\gamma\right)\right)\|p_U\|_q^q$$
$$\le \left(\frac{2^\alpha}{D^{\delta\alpha}}+2^q\varepsilon D^{1-\delta+e\alpha-\delta\alpha}+2^q q\gamma+2^q D^{e\alpha}q\gamma\right)\|p_U\|_q^q.$$

■

We present two corollaries that address the q-norm decrease when d is large or small relative to $1/\gamma$. The first corollary tells us that for a large enough d (relative to $1/\gamma$), we can choose α to be roughly $\frac{1}{d}\log\frac{1}{\gamma}$.

Corollary 4.11. *Let $1>\beta>\Delta>0$ be constants. Let $G=(U,V,E)$ be a bipartite D-regular $(K,\varepsilon=\frac{1}{D^\beta},e)$-expander for any constant e. Let $\delta\ge\delta_{\mathsf{thr}}=1-\beta+\Delta$, and $\gamma>0$. Assume that $d>\frac{\log(1/\gamma)}{2\Delta}$.*

Let p_U be a probability distribution on U and let r_u, for each $u\in U$, be a collection of distributions over $\{0,1\}^d\equiv[D]$ each γ-close to a δd source. Let $\alpha\le\frac{\log(1/\gamma)}{2de}$, and set $q=1+\alpha$. Suppose that $\|p_U\|_q^q\ge\frac{1}{\varepsilon K^\alpha}$. If p_V is the induced probability distribution on V, then

$$\|p_V\|_q^q\le\frac{32}{D^{\delta\alpha}}\|p_U\|_q^q.$$

Proof: Recall that by Lemma 4.9 we have

$$\|p_V\|_q^q\le\left(\frac{2^\alpha}{D^{\delta\alpha}}+2^q\varepsilon D^{1-\delta+e\alpha-\delta\alpha}+2^q q\gamma+2^q D^{e\alpha}q\gamma\right)\|p_U\|_q^q.$$

Now, since $d>\frac{\log 1/\gamma}{2\Delta}$ and $\alpha\le\frac{\log 1/\gamma}{2de}$, we know that $\Delta>e\alpha$. Therefore, $\delta>\delta_{\mathsf{thr}}>1-\beta+e\alpha$. Thus, just as in the proof of Lemma 4.6, by Equation (5) we know that $\varepsilon D^{1-\delta+e\alpha-\delta\alpha}\le\frac{1}{D^{\delta\alpha}}$. Moreover, since $\alpha\le\frac{\log(1/\gamma)}{2de}\le\frac{\log(1/\gamma)}{(e+\delta)d}$, we have $D^{e\alpha}\gamma<\frac{1}{D^{\delta\alpha}}$. Thus,

$$\|p_V\|_q^q\le\left(\frac{2^\alpha}{D^{\delta\alpha}}+2^q\frac{1}{D^{\delta\alpha}}+2^{1+q}\frac{1}{D^{\delta\alpha}}+2^{1+q}\frac{1}{D^{\delta\alpha}}\right)\|p_U\|_q^q\le\frac{32}{D^{\delta\alpha}}\|p_U\|_q^q.$$

■

The next corollary tells us that when d is small relative to $1/\gamma$ we can pick α to be anything smaller than Δ/e.

Corollary 4.12. *Let $1 > \beta > \Delta > 0$ be constants. Let $G = (U, V, E)$ be a bipartite D-regular $(K, \varepsilon = \frac{1}{D^\beta}, e)$-expander for any constant e. Let $\delta \geq \delta_{\mathsf{thr}} = 1 - \beta + \Delta$, and let $\gamma > 0$. Assume that $d \leq \frac{\log(1/\gamma)}{2\Delta}$.*

Let p_U be a probability distribution on U and let r_u, for each $u \in U$, be a collection of distributions over $\{0,1\}^d \equiv [D]$ each γ-close to a δd source. Let $\alpha \leq \Delta/e$, and set $q = 1 + \alpha$. Suppose that $\|p_U\|_q^q \geq \frac{1}{\varepsilon K^\alpha}$. If p_V is the induced probability distribution on V, then

$$\|p_V\|_q^q \leq \frac{32}{D^{\delta\alpha}} \|p_U\|_q^q$$

Proof: Again, apply Lemma 4.9, and we use the fact that $\delta > 1 - \beta + e\alpha$ to get that $\varepsilon D^{1-\delta+e\alpha-\delta\alpha} \leq \frac{1}{D^{\delta\alpha}}$. Moreover, $(e + \delta)\alpha \leq 2\Delta$, so together with our assumption $\gamma \leq \frac{1}{D^{2\Delta}}$, this implies $D^{e\alpha}\gamma \leq \frac{1}{D^{\delta\alpha}}$. Thus, we get

$$\|p_V\|_q^q \leq \left(\frac{2^\alpha}{D^{\delta\alpha}} + 2^q \frac{1}{D^{\delta\alpha}} + 2^{1+q} \frac{1}{D^{\delta\alpha}} + 2^{1+q} \frac{1}{D^{\delta\alpha}} \right) \|p_U\|_q^q \leq \frac{32}{D^{\delta\alpha}} \|p_U\|_q^q$$

in this case too. ■

4.3.2 Handling the Case of $\lambda > 0$

We are now ready to handle $\lambda > 0$, in addition to $\gamma > 0$.

Theorem 4.13. *Let $1 > \beta > \Delta > 0$ be constants. Let $\delta > \delta_{\mathsf{thr}} = 1 - \beta + \Delta$, and let $\gamma, \lambda > 0$. Let $X_1 \circ \cdots \circ X_t$ be a (γ, λ)-almost δ-CG source, with each $X_i \sim \{0,1\}^d$. For any positive integers N and K, let $G = (U = [N], V = [N], E)$ be a D-regular $(K = 2^k, \varepsilon = \frac{1}{D^\beta}, e)$-expander for any constant e. Further, suppose that $d \geq \frac{80e}{\Delta\delta}$, $\gamma \leq 2^{-100e/\delta}$, and*

$$(0.9\delta - 2e\lambda)\, dt \geq k - \frac{2e\beta}{\log(1/\gamma)} d^2.$$

Consider the distribution on the vertices of G after a random walk according to $X_1, \ldots, X_t$ starting from an arbitrary node. Namely, let $Z_0 \sim [N]$ be concentrated on a arbitrary fixed node, and for $i \in [t]$ let:

$$Z_i = \Gamma_G(Z_{i-1}, X_i)$$

Then for any $\eta > 0$, Z_t is η-close to a $\left(k - \lambda edt - 2\beta ed^2 - 2ed \log \frac{1}{\eta}\right)$-source.

Proof: As in Theorem 4.8 we let p_i denote the distribution of Z_i, and write

$$p_i(v) = \sum_{u \in \Gamma(v)} \Pr[X_i = (u,v) | Z_{i-1} = u] \cdot p_{i-1}(u).$$

We know that when i is a good step, $\Pr[X_i = (u,v) | Z_{i-1} = u]$ is a convex combination of sources that are γ-close to a δd source and is thus itself γ-close to a δd source. We analyze the ℓ_q norm using different α depending on whether $d > \frac{\log(1/\gamma)}{2\Delta}$ or $d \leq \frac{\log(1/\gamma)}{2\Delta}$.

Case 1. Suppose that $d > \frac{\log(1/\gamma)}{2\Delta}$, and choose $\alpha = \frac{\log(1/\gamma)}{2de}$. We first claim that there must exist some time s when $\|p_s\|_q^q \le \frac{1}{\varepsilon K^\alpha}$. Suppose not, then for every $i \in [t]$ that is a good step, we can apply Corollary 4.11, and for every bad step, we can apply Lemma 4.7. There are at least $(1-\lambda)t$ good steps, and at most λt bad steps. Overall this tells us that

$$\|p_t\|_q^q \le D^{e\lambda\alpha t}\left(\frac{32^{1/\alpha}}{D^\delta}\right)^{(1-\lambda)\alpha t}.$$

Since by hypothesis $\gamma < \frac{1}{2^{100e/\delta}}$, we know that $32^{1/\alpha} = 2^{10de/\log(1/\gamma)} \le D^{0.1\delta}$. Also by hypothesis we have $(0.9\delta - 2\lambda)dt \ge k - \frac{2e\beta}{\log 1/\gamma}d^2 = k - \frac{\beta}{\alpha}d$ and so

$$\|p_t\|_q^q \le \left(\frac{D^{e\lambda}}{D^{0.9\delta(1-\lambda)}}\right)^{\alpha t} \le \left(\frac{D^{2e\lambda}}{D^{0.9\delta}}\right)^{\alpha t} \le \left(\frac{D^{\beta/\alpha}}{K}\right)^{\alpha} = \left(\frac{1}{\varepsilon^{1/\alpha}K}\right)^{\alpha},$$

in contradiction.

Now, let $\ell \in [t]$ be the *last* time that $\|p_\ell\|_q^q \le \frac{1}{\varepsilon K^\alpha}$. There are at most λt bad steps remaining after ℓ, and Lemma 4.7, each such step increases the ℓ_q norm by a factor of at most $D^{e\alpha}$. Thus,

$$\|p_t\|_q^q \le \left(\frac{D^{e\lambda t}}{\varepsilon^{1/\alpha}K}\right)^{\alpha}.$$

By Corollary 2.3, for any $\eta > 0$, Z_t is η-close to a $(k - \lambda edt - \frac{1}{\alpha}\log\frac{1}{\varepsilon} - \frac{1}{\alpha}\log\frac{1}{\eta})$-source.

Case 2. Suppose $d \le \frac{\log 1/\gamma}{2\Delta}$. We then set $\alpha = \Delta/e$. Again, we claim that there must exist some time s when $\|p_s\|_q^q \le \frac{1}{\varepsilon K^\alpha}$. If not, then we can apply Corollary 4.12 for every good step, and Lemma 4.7 for every bad step, giving us

$$D^{e\lambda\alpha t}\left(\frac{32^{1/\alpha}}{D^\delta}\right)^{(1-\lambda)\alpha t} \le \left(\frac{D^{2e\lambda}}{D^{0.9\delta}}\right)^{\alpha t} \le \left(\frac{D^{\frac{2e\beta}{\log\frac{1}{\gamma}}d}}{K}\right)^{\alpha} \le \left(\frac{D^{e\beta/\Delta}}{K}\right)^{\alpha} = \left(\frac{1}{\varepsilon^{1/\alpha}K}\right)^{\alpha},$$

in contradiction. In the second inequality, we used the fact that $d \ge \frac{80e}{\Delta\delta}$, and so $32^{1/\alpha} = 32^{e/\Delta} \le D^{0.1\delta}$. Again there is a *last* time ℓ that $\|p_\ell\|_q^q \le \frac{1}{\varepsilon K^\alpha}$. And again, there are at most λt bad steps remaining after ℓ. Thus,

$$\|p_t\|_q^q \le \left(\frac{D^{e\lambda t}}{\varepsilon^{1/\alpha}K}\right)^{\alpha}.$$

And here too, for any $\eta > 0$, p_t is η-close to a $(k - \lambda edt - \frac{1}{\alpha}\log\frac{1}{\varepsilon} - \frac{1}{\alpha}\log\frac{1}{\eta})$-source.

In Case 1 we chose $\alpha = \frac{\log 1/\gamma}{2de}$, and then $\frac{1}{\alpha} < 2ed$. In Case 2 we chose $\alpha = \Delta/e$. Since by our assumption $d > \frac{80e}{\Delta\delta}$ we have that $\frac{1}{\alpha} < 2ed$ in this case too. Therefore, it is always the case that

$$k - \lambda edt - \frac{1}{\alpha}\log\frac{1}{\varepsilon} - \frac{1}{\alpha}\log\frac{1}{\eta} \ge k - \lambda edt - 2\beta ed^2 - 2ed\log\frac{1}{\eta},$$

as needed. ■

We can now plug-in the explicit lossless expanders of Theorem 2.5, with error parameter $\beta \geq 1/6$, and $e = 100$ to Theorem 4.13 above and get an explicit deterministic condenser for high rate sources.

Corollary 4.14. *Let $d \in \mathbb{N}$, $\delta > 0$ and $\gamma, \lambda \geq 0$ be constants that satisfy the following constraints:*

- $\delta \geq 1 - \beta + \Delta = \frac{11}{12}$,
- $d \geq \frac{80e}{\Delta\delta} \geq 2000e$, *and*
- $\gamma \leq 2^{-100e/\delta}$,

where we chose $\beta = 1/6$ and $\Delta = 1/12$, and we let $e = 100$ as given in Theorem 2.5. Then, for any positive integer t, there exists an explicit function

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^{m=(0.9\delta-2e\lambda)dt+O(1)}$$

such that for any (γ, λ)-almost δ-CG source $X = X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$, and any $\eta > 0$, $\mathsf{Cond}(X)$ is η-close to an $\left(m - \lambda edt - O(d^2) - O\left(d \cdot \log \frac{1}{\eta}\right)\right)$-source.

That is, for any constant η, there exists an η-error deterministic condenser for (γ, λ)-almost δ-CG sources with the above constraints with entropy gap $\lambda edt + O(1)$.

We note that 0.9 can be made arbitrarily close to 1 in Theorem 4.13 and Corollary 4.14. In general, unless stated otherwise, the numbers represented by decimals in this paper can be made arbitrarily close to 1 by sufficiently strengthening the constants in constraints such as $\gamma \leq 2^{O(1/\delta)}$. We keep them as is for convenience and readability.

In Section 6, we will dispense with the lower bound on d and upper bound on γ by grouping together consecutive blocks.

4.3.3 Handling the Case of Suffix Friendliness

Observe that the above condenser cannot hope to achieve constant entropy gap when $\lambda > 0$ is not constant. This is because if all the λ-fraction of bad steps are at the end, each step can reduce the entropy by roughly d bits, and there are no future good steps to regain the lost entropy. In this section we show that by imposing the condition of suffix friendliness, we can still condense to constant entropy gap.

We can give analogues of Theorem 4.13 and Corollary 4.14 in the case of suffix friendly almost CG sources, and show that we can condense such sources to constant entropy gap. We first give a theorem similar to Theorem 4.13 that claims that only a constant amount of entropy is lost in the case of a suffix-friendly CG sources.

Theorem 4.15. *Let $1 > \beta > \Delta > 0$ be constants. Let $\delta > \delta_{\mathsf{thr}} = 1 - \beta + \Delta$, and let $\gamma, \lambda > 0$. Let $X_1, \ldots, X_t$ be a $(\gamma, \lambda, \Lambda)$-suffix-friendly almost δ-CG source, with each $X_i \sim \{0,1\}^d$. For any N and K, suppose $G = (U = [N], V = [N], E)$ is a D-regular $(K = 2^k, \varepsilon = \frac{1}{D^\beta}, e)$-expander for any constant e. Further, suppose that $d \geq \frac{80e}{\Delta\delta}$, $\gamma \leq 2^{-100e/\delta}$, $\lambda \leq \frac{\delta}{6e}$, and*

$$(0.9\delta - 2e\lambda)dt - 2\Lambda ed \geq k - \frac{2e\beta}{\log 1/\gamma} d^2.$$

Consider the distribution on the vertices of G after a random walk according to $X_1, \ldots, X_t$ starting from an arbitrary node. Namely, let $Z_0 \sim [N]$ be concentrated on an arbitrary fixed node, and for $i \in [t]$ let

$$Z_i = \Gamma_G(Z_{i-1}, X_i).$$

Then, for any $\eta > 0$, Z_t is η-close to a $\left(k - 2\beta e d^2 - \left(\frac{6e^2\Lambda}{\delta} + 2e \log \frac{1}{\eta}\right) d\right)$-source.

Proof: As in Theorem 4.13, we can choose $\alpha = \frac{\log(1/\gamma)}{2de}$ or $\alpha = \Delta/e$ depending on whether we are in the case of $d > \frac{\log(1/\gamma)}{2\Delta}$ or $d \leq \frac{\log(1/\gamma)}{2\Delta}$ respectively. Using a similar argument again, we can show that in either case, there must exist a time s such that $\|p_s\|_q^q \leq \frac{1}{\varepsilon K^\alpha}$. Overall, there are at least $(1-\lambda)t - \Lambda$ good steps, and at most $\lambda t + \Lambda$ bad steps. By the constraints on d, δ, γ, and k we can verify again that in either case, each of the good steps decreases $\|p_i\|_q^q$ by a factor of $D^{0.9\delta\alpha}$, and each of the bad steps increases it by $D^{e\alpha}$. Thus, if there was no time s for which $\|p_s\|_q^q \leq \frac{1}{\varepsilon K^\alpha}$, then we can apply Corollary 4.12 or Corollary 4.11 for good steps and Lemma 4.7 for bad steps. Again in either case (and thus either choice of α), by our parameter constraints, we can verify that:

$$\|p_t\|_q^q \leq \left(D^{e\lambda t + e\Lambda}\right)^\alpha \left(\frac{1}{D^{0.9\delta((1-\lambda)t-\Lambda)}}\right)^\alpha \leq \left(\frac{D^{2e\Lambda+2e\lambda t}}{D^{0.9\delta t}}\right)^\alpha \leq \left(\frac{1}{\varepsilon^{1/\alpha}K}\right)^\alpha.$$

In either case, we again let $1 \leq \ell \leq t$ be the *last* time that $\|p_\ell\|_q^q \leq \frac{1}{\varepsilon K^\alpha}$.

Let $\mathfrak{g}$ and $\mathfrak{b}$ be the number of good and bad steps respectively between ℓ and t. Since ℓ is the last time that the ℓ_q norm is sufficiently small, it must be the case that: $e\mathfrak{b} > 0.9\delta\mathfrak{g}$. This is because again, by our setting of parameters, in either case, every good step decreases the $\|p_i\|_q^q$ by a factor of $D^{-0.9\delta\alpha}$ and every bad step increases it by D^α. So if $0.9\delta\mathfrak{g} \geq e\mathfrak{b}$ then there must be a timestep greater then ℓ where the $\|p_i\|_q^q$ is smaller that $\frac{1}{\varepsilon K^\alpha}$. Moreover, by the suffix friendly property, $b \leq \lambda(t - \ell + 1) + \Lambda$. Therefore,

$$t - \ell + 1 = \mathfrak{g} + \mathfrak{b} \leq \frac{e}{0.9\delta}\mathfrak{b} + \mathfrak{b} \leq \frac{3e}{\delta}\mathfrak{b} \leq \frac{3e}{\delta}(\lambda(t - \ell + 1) + \Lambda).$$

Thus, $t - \ell + 1 \leq \frac{\frac{3e}{\delta}\Lambda}{1 - \frac{3e}{\delta}\lambda} \leq \frac{6e\Lambda}{\delta}$, where we used the fact that $\lambda \leq \frac{\delta}{6e}$. Since any step can worsen the $||p_i||_q^q$ by a factor of at most $D^{e\alpha}$,

$$||p_t||_q^q \leq D^{\alpha(t-\ell+1)}||p_\ell||_q^q \leq \left(\frac{D^{6e^2\Lambda/\delta}}{\varepsilon^{1/\alpha}K}\right)^\alpha.$$

Thus, in either case, by our choice of parameters, we have by Corollary 2.3 that Z_t is η-close to a $(k - 2\beta e d^2 - (\frac{6e^2\Lambda}{\delta} + 2e \log 1/\eta)d)$-source. ■

We can again directly use the lossless expander construction of Theorem 2.5 to get a deterministic condenser.

Theorem 4.16. *Let $d, \Lambda > 1$ be constant positive integers and let, $\delta, \gamma, \lambda > 0$ be constants that satisfy the following constraints:*

- $\delta \geq 1 - \beta + \Delta = \frac{11}{12}$,

- $d \geq \frac{80e}{\Delta\delta} \geq 2000$,
- $\gamma \leq 2^{-100/\delta}$, *and*
- $\lambda \leq \frac{\delta}{6e}$,

where we chose $\beta = 1/6$ *and* $\Delta = 1/12$, *and we let* $e = 100$ *as given in Theorem 2.5. Then, for any positive integer* t, *and any positive integer* Λ, *there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^{m=(0.9\delta-2e\lambda)dt+O(1)}$$

such that for any $(\gamma, \lambda, \Lambda)$*-suffix-friendly-almost* δ*-CG source* $X = X_1 \circ \ldots \circ X_t$ *with each* $X_i \sim \{0,1\}^d$, *and any* $\eta > 0$, $\mathsf{Cond}(X)$ *is* η*-close to an* $\left(m - O(d^2) - O(\Lambda \cdot d) - O\left(d \cdot \log \frac{1}{\eta}\right)\right)$*-source.*

That is for constant η, *there exists an* η*-error deterministic condenser for* $(\gamma, \lambda, \Lambda)$*-suffix-friendly-almost* δ*-CG sources with the above constraints with entropy gap* $O(1)$.

5 Deterministic Condensing from Any Rate

In this section, we expand our random-walks based construction to handle an arbitrary min-entropy rate $\delta > 0$. The idea is to first split the source into t/b blocks, each of some constant length b. We then use a constant-sized *optimal* lossless expander H (found via brute force), and run t/b (separate) random walks on H using each of the length-b blocks as a set of random walk instructions. Since optimal lossless expanders allow for deterministic condensing for arbitrarily small δ_{thr}, H will condense each length-b block into a distribution with constant entropy gap. Thus, for sufficiently large b, each of the t/b random walk distributions will be close to a source with entropy rate close to 1 (even conditioned on previous blocks). Then, we can use these distributions as instructions for a random walk on the graph G of Theorem 2.5.

In other words, one can view the condensing procedure as a series of epochs. In each epoch, we walk a constant number of steps on H until the entropy rate of the vertex distribution is sufficiently high. Once the epoch is completed, and the entropy rate is sufficiently high, we can "flush the entropy" from the steps in the epoch into the "big" lossless expander G from [CRVW02] by using the vertex position in H as an instruction for a step in the big graph.

More formally, the construction goes as follows. Let $c^\star$ be the global constant from Corollary 2.8. We are given a (γ, λ)-almost δ-CG source $X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$. Here, $\delta > 0$ is an arbitrary constant, and d, γ, λ satisfy the following. Let $e = 100$.

- $d \geq \max\left\{\frac{10^3}{\delta^2}, \frac{2c^\star}{\delta}\right\}$,
- $\gamma \leq 2^{-100/\delta}$, and,
- $\lambda \leq \frac{1}{10^8 \cdot e^2}\delta^2 = \frac{1}{10^{12}}\delta^2$.

We describe the parameters of the two expander graphs we need.[28]

[28] In the cases where we write 10^a for some constant a, we note that smaller constants in fact suffice. However, we do not present these smaller, yet messier constants, for the sake of readability.

The Small Graph. Set $\beta = 1 - \frac{\delta}{2}$ and $\Delta = \frac{\delta}{3}$. Notice that $\beta \geq \Delta$ and that $\delta_{\mathsf{thr}} = 1 - \beta + \Delta = \frac{2}{3}\delta < \delta$. Set the *epoch length*

$$b = \frac{10^6 \cdot d^3 \cdot e}{\delta},$$

and let

$$d_{\mathsf{CRVW}} = \log D_{\mathsf{CRVW}} = \left(0.9\delta - 2\sqrt{\lambda}\right) db + \frac{2\beta}{\log(1/\gamma)} d^2 + 2d + \log c^\star.$$

Since $d \geq \frac{2c^\star}{\delta}$, by Corollary 2.8 there exists a degree D bipartite graph

$$H = ([D_{\mathsf{CRVW}}], [D_{\mathsf{CRVW}}], E)$$

that is a $\left(K = \frac{D_{\mathsf{CRVW}}}{c^\star D^2}, \varepsilon = \frac{1}{D^\beta}, 1\right)$-lossless expander. We can construct H in constant time since b is constant.

The Big Graph. Set $\beta_{\mathsf{CRVW}} = \frac{1}{6}$ and $\Delta_{\mathsf{CRVW}} = \frac{1}{12}$. Let $\gamma_{\mathsf{CRVW}} = \eta = 2^{-200e}$. Finally let:

$$\begin{aligned}
\delta_{\mathsf{CRVW}} &= \frac{k - \sqrt{\lambda} db - 2\beta d^2 - 2d \log \frac{1}{\eta}}{d_{\mathsf{CRVW}}} \\
&= \frac{d_{\mathsf{CRVW}} - 2d - \log c^\star - \sqrt{\lambda} db - 2\beta d^2 - 2d \log \frac{1}{\eta}}{d_{\mathsf{CRVW}}} \\
&= 1 - \frac{\sqrt{\lambda} db + 2d + \log c^\star + 2\beta d^2 + 2d \log \frac{1}{\eta}}{d_{\mathsf{CRVW}}}.
\end{aligned}$$

Our assumption $\lambda \leq \frac{1}{10^8 e}\delta^2$ implies that $\frac{\sqrt{\lambda}}{0.9\delta - 2\sqrt{\lambda}} \leq \frac{1}{72}$. This, combined with our choice of b implies that each of the six terms subtracted from 1 in the above is at most $\frac{1}{72}$. Thus, $\delta_{\mathsf{CRVW}} \geq \frac{11}{12}$. Let

$$G = ([N], [N], E)$$

be the $\left(K_{\mathsf{CRVW}} = \Omega\left(\frac{N}{D^2_{\mathsf{CRVW}}}\right), \varepsilon_{\mathsf{CRVW}} = \frac{1}{D^{1/6}_{\mathsf{CRVW}}}, e\right)$-expander guaranteed to us by Theorem 2.5 with

$$k_{\mathsf{CRVW}} = \left(0.9\delta_{\mathsf{CRVW}} - 2e\sqrt{\lambda}\right) d_{\mathsf{CRVW}} \cdot \frac{t}{b} + \frac{2e\beta_{\mathsf{CRVW}}}{\log(1/\gamma_{\mathsf{CRVW}})} d^2_{\mathsf{CRVW}}.$$

Note this implies that

$$n = \log N = k_{\mathsf{CRVW}} + O(d_{\mathsf{CRVW}}) = \left(0.9\delta_{\mathsf{CRVW}} - 2e\sqrt{\lambda}\right) d_{\mathsf{CRVW}} \cdot \frac{t}{b} + \frac{2e\beta_{\mathsf{CRVW}}}{\log 1/\gamma_{\mathsf{CRVW}}} d^2_{\mathsf{CRVW}} + O(d_{\mathsf{CRVW}}).$$

5.1 The Condenser

Having defined our two expanders, we are ready to describe the construction. First, as notation, for any labeled D-regular graph $G = ([N], [N], E)$, and any sequence of strings $x_1, \ldots, x_t$ with each $x_i \in \{0,1\}^d$, let

$$\mathsf{RW}(G, x_1, \ldots, x_t) \in [N]$$

denote the node reached after walking on G using $x_1, \ldots, x_t$ starting from a fixed arbitrary node, say the first one. That is, $\mathsf{RW}(G, x_1, \ldots, x_t) = v_t$ where the sequence of nodes $v_0, \ldots, v_t$ is defined via $v_0 = 1$ and $v_i = \Gamma_G(v_{i-1}, x_i)$.

Recall that we are given as input (γ, λ)-almost δ-CG source $X_1 \circ \ldots \circ X_t$, with each $X_i \sim \{0,1\}^d$, and with the constraints dictated as above. Our condenser is constructed as follows. Given $x_1, \ldots, x_t \in \{0,1\}^d$,

1. For every $j \in [t/b]$, let $z_j = \mathsf{RW}(H, x_{(j-1)b+1}, \ldots, x_{(j-1)b+b})$.
2. Output $w = \mathsf{RW}(G, z_1, \ldots, z_{t/b})$.

5.2 The Analysis

We first show:

Lemma 5.1. *The sequence $Z_1, \ldots, Z_{t/b}$ is an $(\gamma_{\mathsf{CRVW}}, \sqrt{\lambda})$-almost δ_{CRVW}-CG source.*

Proof: Call the j-th epoch $X_{(j-1)b+1}, \ldots, X_{(j-1)b+b}$ "good" if less than $\sqrt{\lambda}$ fraction of the steps in the epoch are bad ones. By an averaging argument, there are at least $1 - \sqrt{\lambda}$ fraction of good epochs overall. We'll show that for every good epoch j, and for every prefix $z_1, \ldots, z_{j-1} \in \{0,1\}^{(j-1)d_{\mathsf{CRVW}}}$, the conditional distribution $Z_j | \left\{Z_{[1,j-1]} = z_{[1,j-1]}\right\}$ is γ_{CRVW}-close to a $(\delta_{\mathsf{CRVW}} \cdot d_{\mathsf{CRVW}})$-source. First, we note that any prefix $z_1, \ldots, z_{j-1}$ is simply a function of the prefixes $x_1, \ldots, x_{(j-1)b}$. Thus, it suffices to show that when j is a good epoch, for any prefix $x_1, \ldots, x_{(j-1)b}$, the conditional distribution $Z_j | \left\{X_{[1,(j-1)b]} = x_{[1,(j-1)b]}\right\}$ is γ_{CRVW}-close to a $(\delta_{\mathsf{CRVW}} \cdot d_{\mathsf{CRVW}})$-source.

Fix any $x_1, \ldots, x_{(j-1)b}$, and any good epoch j, and consider the (conditional) sequence of random variables

$$X_{(j-1)b+1}, \ldots, X_{(j-1)b+b} | \left\{X_{[1,(j-1)b]} = x_{[1,(j-1)b]}\right\}.$$

Such a sequence is a $(\gamma, \sqrt{\lambda})$-almost δ-CG source because j is a good block and the original $X_1, \ldots, X_t$ is a (γ, λ)-almost δ-CG source. Moreover, recall that

$$Z_j | \left\{X_{[1,(j-1)b]} = x_{[1,(j-1b)]}\right\} = \mathsf{RW}(H, X_{(j-1)b+1}, \ldots, X_{(j-1)b+b}) | \left\{X_{[1,(j-1)b]} = x_{[1,(j-1b)]}\right\}.$$

- $1 > \beta = 1 - \frac{\delta}{2} > \Delta = \frac{\delta}{3} > 0$,
- $\delta > 1 - \beta + \Delta$,
- $d \geq \frac{80}{\delta\Delta} = \frac{10^3}{\delta^2}$,
- $\gamma \leq 2^{-100/\delta}$, and,
- $\left(0.9\delta - 2\sqrt{\lambda}\right) db \geq k - \frac{2\beta}{\log 1/\gamma} d^2$.

Therefore, we can apply Theorem 4.13 with

$$X_{(j-1)b+1}, \ldots, X_{(j-1)b+b} | \left\{X_{[1,(j-1)b]} = x_{[1,(j-1b)]}\right\}$$

as our $(\gamma, \sqrt{\lambda})$-almost δ-CG source. We then get that the conditional distribution on Z_j is η-close to a source with $\left(k - \sqrt{\lambda}db - 2\beta d^2 - 2d \log \frac{1}{\eta}\right) = \delta_{\mathsf{CRVW}} \cdot d_{\mathsf{CRVW}}$ min-entropy. ■

Lemma 5.2. *For any $\eta_{\mathsf{CRVW}} > 0$, the distribution of $W = \mathsf{RW}(G, Z_1, \ldots, Z_{t/b})$ is η_{CRVW}-close to a $\left(k_{\mathsf{CRVW}} - \sqrt{\lambda}edt - \mathrm{poly}(d, 1/\delta) \cdot \log(1/\eta_{\mathsf{CRVW}})\right)$-source.*

Proof: We verify that all conditions of Theorem 4.13 are met in order to apply it with G and the Z_i-s as instructions. Indeed,

- $1 > \beta_{\mathsf{CRVW}} = \frac{1}{6} > \Delta_{\mathsf{CRVW}} = \frac{1}{12}$,
- $\delta_{\mathsf{CRVW}} > \frac{11}{12} = 1 - \beta_{\mathsf{CRVW}} + \Delta_{\mathsf{CRVW}}$,
- $d_{\mathsf{CRVW}} \geq 10^6 e \geq 80 \cdot 12 \cdot \frac{12}{11} \cdot e \geq \frac{80e}{\delta_{\mathsf{CRVW}}\Delta_{\mathsf{CRVW}}}$,
- $\gamma_{\mathsf{CRVW}} = \eta = 2^{-200e} \leq 2^{-100e/\delta_{\mathsf{CRVW}}}$, and,
- $(0.9\delta_{\mathsf{CRVW}} - 2e\sqrt{\lambda})d_{\mathsf{CRVW}}\frac{t}{b} = k_{\mathsf{CRVW}} - \frac{2e\beta_{\mathsf{CRVW}}}{\log(1/\gamma_{\mathsf{CRVW}})}d^2_{\mathsf{CRVW}}$.

Therefore, for any $\eta_{\mathsf{CRVW}} > 0$, $\mathsf{RW}(G, Z_1, \ldots, Z_{t/b})$ is η_{CRVW}-close to a source with min-entropy

$$k_{\mathsf{CRVW}} - \sqrt{\lambda}ed_{\mathsf{CRVW}}\frac{t}{b} - 2\beta_{\mathsf{CRVW}}ed^2_{\mathsf{CRVW}} - 2ed_{\mathsf{CRVW}}\log(1/\eta_{\mathsf{CRVW}}).$$

We note that all but the first two terms are $\mathrm{poly}(d, \frac{1}{\delta})$ (and that the second to last term is $O(\log \frac{1}{\eta_{\mathsf{CRVW}}})$. Additionally, observing that, by our choice of b, $\frac{d_{\mathsf{CRVW}}}{b} \leq d$ yields the result. ■

We can finally state the final theorem about the condenser we construct.

Theorem 5.3. *Let $d > 1$ be a positive integer and let, $\delta, \gamma, \lambda > 0$ be constants that satisfy the following constraints:*

- $d \geq \max\left\{\frac{10^3}{\delta^2}, \frac{2c^\star}{\delta}\right\}$,
- $\gamma \leq 2^{-100/\delta}$, *and,*
- $\lambda \leq \frac{1}{10^8 e^2}\delta^2$.

For any positive integer t, there exists an explicit function $\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^m$ with $m = \Omega(\delta dt)$ such that for any (γ, λ)-almost δ-CG source $X = X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$, and any $\eta > 0$, $\mathsf{Cond}(X)$ is η-close to a $\left(m - \sqrt{\lambda}edt - \mathrm{poly}(d, 1/\delta) \cdot \log(1/\eta)\right)$-source, where $e = 100$ as given in Theorem 2.5.

That is, for constant η, there exists an η-error deterministic condenser for (γ, λ)-almost δ-CG sources with the above constraints with entropy gap $\sqrt{\lambda}edt + O(1)$.

Proof: It's easy to to verify from the construction and Lemma 5.2 that we can take $m = k_{\mathsf{CRVW}} + O(1)$. It only remains to verify that $k = \Omega(\delta dt)$:

$$
\begin{aligned}
k_{\textsf{CRVW}} &= \left(0.9\delta_{\textsf{CRVW}} - 2e\sqrt{\lambda}\right) d_{\textsf{CRVW}} \cdot \frac{t}{b} + \frac{2e\beta_{\textsf{CRVW}}}{\log 1/\gamma_{\textsf{CRVW}}} d_{\textsf{CRVW}}^2 \\
&\geq \left(0.9\delta_{\textsf{CRVW}} - 2e\sqrt{\lambda}\right) d_{\textsf{CRVW}} \cdot \frac{t}{b} \\
&= 0.9\left(k - \sqrt{\lambda}db - 2\beta d^2 - 2d\log\frac{1}{\eta}\right)\frac{t}{b} - 2\sqrt{\lambda}ed_{\textsf{CRVW}}\frac{t}{b} \\
&= 0.9\left(d_{\textsf{CRVW}} - 2d - \log c^\star - \sqrt{\lambda}db - 2\beta d^2 - 2d\log\frac{1}{\eta}\right)\frac{t}{b} - 2e\sqrt{\lambda}d_{\textsf{CRVW}}\frac{t}{b} \\
&\geq 0.9\left(\left(0.9\delta - \sqrt{\lambda}\right)db - \sqrt{\lambda}db - 2\beta d^2 - 2d\log\frac{1}{\eta}\right)\frac{t}{b} - 2e\sqrt{\lambda}d_{\textsf{CRVW}}\frac{t}{b} \\
&\geq (0.81\delta - 2\sqrt{\lambda})dt - 0.9\left(2\beta d^2 + 2d\log\frac{1}{\eta}\right)\frac{t}{b} - 2e\sqrt{\lambda}d_{\textsf{CRVW}}\frac{t}{b} \\
&\geq (0.81\delta - 4e\sqrt{\lambda})dt - 0.9\left(2\beta d^2 + 2d\log\frac{1}{\eta}\right)\frac{t}{b} \\
&\geq 0.8\delta dt - \left(2d^2 + 200ed\right)\frac{t}{b}.
\end{aligned}
$$

In the second to last inequality, we used the fact that for our choice of $d_{\textsf{CRVW}}$ and b, we have $d_{\textsf{CRVW}}/b \leq d$. In the last inequality, we used the fact that $4e\sqrt{\lambda} \leq \frac{4}{10^4}\delta \leq 0.01\delta$. Finally, by our choice of b, we know that $(2d^2/b + 200d/b + 1/b)t \leq 0.01\delta t \leq 0.01\delta dt$. So overall, $k \geq 0.79\delta dt = \Omega(\delta dt)$. ■

We remark that again, the constant 0.79 can be made arbitrarily close to 1 by strengthening the appropriate constraints. Namely by increasing the length of b, increasing the 100 in $\gamma \leq 2^{-100e/\delta}$ and increasing the 10^8 in $\lambda \leq \frac{1}{10^8 e}\delta^2$.

5.3 Condensing to Constant Entropy Gap from Suffix Friendliness and Any Rate

Similarly to the condenser in Section 4.3.3, the above condenser cannot hope to achieve constant entropy gap when $\lambda > 0$. The issue is the same: we cannot win if all the bad steps are at the end. We again show one can resolve such an issue by imposing suffix friendliness, and give a construction of a deterministic condenser for suffix-friendly almost CG sources for arbitrary $\delta > 0$. The full construction and proof is very similar to the non-suffix-friendly case, and we defer the details to Appendix A.1.

Theorem 5.4. *Let $d > 1$, $\delta, \gamma, \lambda > 0$ be constants that satisfy the following constraints:*

- $d \geq \max\left\{\frac{10^3}{\delta^2}, \frac{2c^\star}{\delta}\right\}$,
- $\gamma \leq 2^{-100/\delta}$*, and,*
- $\lambda \leq \frac{1}{10^{12}}\delta^2$.

For any positive integer t, and any positive integer Λ, there exists an explicit function Cond: $\{0,1\}^{n=dt} \to \{0,1\}^m$ *with $m = \Omega(\delta dt)$ such that for any $(\gamma, \lambda, \Lambda)$-suffix-friendly almost δ-CG source $X = X_1, \ldots, X_t$ with each $X_i \sim \{0,1\}^d$, and any $\eta > 0$,* Cond(X) *is η-close to a $m - \mathrm{poly}(d, 1/\delta) \cdot (\Lambda + \log(1/\eta))$.*

That is, for constant η and Λ, there exists an η-error deterministic condenser for $(\gamma, \lambda, \Lambda)$-suffix-friendly-almost δ-CG sources with the above constraints with entropy gap $O(1)$.

6 Condensing from Any d, Any γ, and from Shannon Entropy

In this section we show how to deterministically condense almost CG sources without any constraints on d or γ relative to δ (in Section 5, we required, roughly, $d \geq \frac{1}{\delta^2}$ and $\gamma \leq 2^{-1/\delta}$). The main observation is that simply grouping the instructions of a CG source into blocks both increases d and reduces γ, while (roughly) preserving the entropy rate. For appropriately large blocks, the new CG source meets the constraints of Theorem 5.3 or Theorem 5.4.

As a consequence, we can also show that one can deterministically condense from Shannon CG sources. The result follows immediately from combining Corollary 3.11 or Corollary 3.12 with Theorem 5.3 or Theorem 5.4, respectively.

Theorem 6.1. *Let $\delta > 0$, and $\gamma, \lambda \geq 0$ be constants such that $\lambda \leq \frac{1}{10^{24}}\delta^8$. Let $d \geq 1$ be any positive integer. For any positive integer t, there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^{m=\Omega(\delta dt)}$$

such that for any (γ, λ)-almost δ-CG source $X = X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$, and any $\eta > 0$, $\mathsf{Cond}(X)$ is η-close to an $\big(m - \lambda^{1/4}dt - \mathrm{poly}(d, 1/\delta) \cdot \log(1/\eta)\big)$-source.

Proof (sketch): As usual, let $c^\star$ be the global constant from Corollary 2.8. The first bullet of Lemma 3.3 states that for any b, $X' = X'_1 \circ \cdots \circ X'_{\lfloor t/b \rfloor}$ is a $(\gamma = e^{-(1-\gamma)^2 b/8}, \sqrt{\lambda})$-almost δ'-CG source with $\delta' > \frac{\delta}{8}$. We see that by setting $b = O(c^\star/(\delta^4(1-\gamma)^2))$, all conditions needed to apply Theorem 5.3 with X' are met:

- $db \geq \max\left\{\frac{10^3}{\delta'^2}, \frac{2c^\star}{\delta'}\right\}$,
- $\gamma = e^{-(1-\gamma)^2 b/72} \leq 2^{-100/\delta'}$, and
- $\sqrt{\lambda} \leq \frac{1}{10^{12}}\delta'^2$.

■

Before we continue, let us briefly discuss the hidden constant in the expression for $m = \Omega(\delta dt)$. For large γ-s we cannot hope to get m arbitrarily close to δdt. Indeed, when $\gamma > 1/2$, both bullets of Lemma 3.3 yield a significant entropy loss. In other words, overall, the true amount of (smooth) min-entropy in the input X depends on both γ and δ. However, for $\gamma \ll \frac{1}{2}$ we can use the second bullet of Lemma 3.3 to argue that the new CG source has roughly $(1-2\gamma)\delta$ entropy rate overall. Then, as discussed at the end of Section 5.2, one can adjust the random walk techniques to make m (and thus the output entropy) close to the smooth entropy of X, namely close to δdt.

The following is the analogous result for suffix friendly almost CG sources.

Theorem 6.2. *Let $\delta > 0$, and $\gamma, \lambda \geq 0$ be constants such that $\lambda \leq \frac{1}{10^{24}}\delta^8$. Let $d \geq 1$ be any positive integer. For any positive integer t and any positive integer Λ, there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^{m=\Omega(\delta dt)}$$

such that for any $(\gamma, \lambda, \Lambda)$-suffix-friendly almost δ-CG source $X = X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$, and any $\eta > 0$, $\mathsf{Cond}(X)$ is η-close to an $(m - \mathrm{poly}(d, 1/\delta) \cdot (\Lambda + \log(1/\eta)))$-source.

Finally, we state similar results for condensing from Shannon CG sources.

Theorem 6.3. *Let $\delta > 0$ and $\lambda \geq 0$ be constants such that $\lambda \leq \frac{1}{10^{24}}\delta^8$. Let $d \geq 1$ be any positive integer. For any positive integer t, there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^{m=\Omega(\delta^2 dt)}$$

such that for any λ-almost δ-Shannon-CG source $X = X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$, and any $\eta > 0$, $\mathsf{Cond}(X)$ is η-close to an $\big(m - \lambda^{1/4}dt - \mathrm{poly}(d, 1/\delta) \cdot \log(1/\eta)\big)$-source.

Proof (sketch): If needed, we can first group X into blocks of length b' so that $db' > d^*(\delta)$ where d^* is the constant from Corollary 3.10. The new CG source will have the same rate δ. Thus we can then assume that X is a CG-source with $d > d^*$

As usual, let $c^\star$ be the global constant from Corollary 2.8. Corollary 3.11 states that for any b, $X' = X'_1 \circ \cdots \circ X'_{\lfloor t/b \rfloor}$ is a $(\gamma = e^{-\delta^2 b/72}, \sqrt{\lambda})$-almost δ'-CG source with $\delta' = \frac{\delta^2}{36}$. We see that by setting $b = O(c^\star/\delta^4)$, all conditions needed to apply Theorem 5.3 with X' are once again met. ■

The following is the analogous result for suffix friendly Shannon CG sources, obtained by using Corollary 3.12 and Theorem 5.4.

Theorem 6.4. *Let $\delta > 0$ and $\lambda \geq 0$ be constants such that $\lambda \leq \frac{1}{10^{24}}\delta^8$. Let $d \geq 1$ be any positive integer. For any positive integer t and any positive integer Λ, there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \to \{0,1\}^{\Omega(\delta^2 dt)}$$

such that for any (λ, Λ)-suffix-friendly almost δ-Shannon-CG source $X = X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$, and any $\eta > 0$, $\mathsf{Cond}(X)$ is η-close to an $(m - \mathrm{poly}(d, 1/\delta) \cdot (\Lambda + \log(1/\eta)))$-source.

7 Extracting with Constant Seed Length

In previous sections we have constructed condensers for almost CG sources and Shannon CG sources that output sources with very small entropy gap. Specifically:

1. (γ, λ)-almost δ-CG-sources with $\lambda = 0$,[29] with any constant $\delta > 0$. This follows from Theorem 6.1.

2. $(\gamma, \lambda, \Lambda)$-suffix-friendly almost δ-CG-sources with any constant $\delta > 0$, and any integer constant Λ. This follows from Theorem 6.2.

3. δ-Shannon CG-sources for any constant $\delta > 0$. This follows from Theorem 6.3.

4. (λ, Λ)-suffix-friendly almost δ-Shannon-CG-source for any constant $\delta > 0$ and any integer constant Λ. This follows from Theorem 6.4.

For all of the above sources we can employ the high min-entropy extractor of Theorem 2.12 and get a long close-to-uniform string while investing only a constant-sized uniform seed. We omit the easy proof (which amounts to using the triangle inequality and the fact that applying functions can never increase the statistical distance).

[29] Clearly, when λ is greater than 0 but still a small sub-constant, we can still apply an extractor with a short seed. For brevity, we omit the dependence of the seed length on $\lambda \neq 0$.

Theorem 7.1 (following Item 1 above)**.** *For any constants $\delta, \varepsilon > 0$, any positive integer $d \geq 1$, and any constant $\gamma > 0$, the following holds. For any positive integer t there exists an explicit function*

$$\mathsf{CGExt}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=O_{d,\delta,\varepsilon}(1)} \to \{0,1\}^{m=\Omega(\delta dt)}$$

*such that given a $(\gamma, 0)$-**almost δ-CG-source** X, and an independent uniform $Y \sim \{0,1\}^{\ell}$, it holds that* $\mathsf{CGExt}(X,Y) \approx_{\varepsilon} U_m$.

Theorem 7.2 (following Item 2 above)**.** *For any constants δ, ε, any constant $\Lambda \in \mathbb{N}$, any positive integer $d \geq 1$, any constant $\gamma > 0$, and any $\lambda \leq 10^{-24}\delta^8$, the following holds. For any positive integer t there exists an explicit function*

$$\mathsf{SFCGExt}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=O_{d,\delta,\varepsilon,\Lambda}(1)} \to \{0,1\}^{m=\Omega(\delta dt)}$$

*such that given a $(\gamma, \lambda, \Lambda)$-**suffix-friendly almost δ-CG-source** X, and an independent uniform $Y \sim \{0,1\}^{\ell}$, it holds that* $\mathsf{SFCGExt}(X,Y) \approx_{\varepsilon} U_m$.

Again, we note that in Theorems 7.1 and 7.2, for $\gamma \ll 1/2$ we can make m close to the expected smooth min-entropy of the input source X, i.e., $(1-2\gamma)\delta dt$. Namely, we can output $(1-\theta)(1-2\gamma)\delta dt$ entropy for an arbitrarily small constant $\theta > 0$, at the expense of modifying the other constants. (The GW extractor of Theorem 2.12 has tiny entropy loss.) We do not give details on optimizing the entropy loss in this work.

Theorem 7.3 (following Item 3 above)**.** *For any constants $\delta, \varepsilon > 0$ and any positive integer $d \geq 1$, the following holds. For any positive integer t there exists an explicit function*

$$\mathsf{ShannonExt}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=O_{d,\delta,\varepsilon}(1)} \to \{0,1\}^{m=\Omega(\delta^2 dt)}$$

*such that given a δ-**Shannon-CG-source** X, and an independent uniform $Y \sim \{0,1\}^{\ell}$, it holds that* $\mathsf{ShannonExt}(X,Y) \approx_{\varepsilon} U_m$.

Theorem 7.4 (following Item 4 above)**.** *For any constants $\delta, \varepsilon > 0$, any constant $\Lambda \in \mathbb{N}$, any positive integer $d \geq 1$, and any $\lambda \leq 10^{-24}\delta^8$, the following holds. For any positive integer t there exists an explicit function*

$$\mathsf{SFShannonExt}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=O_{d,\delta,\varepsilon,\Lambda}(1)} \to \{0,1\}^{m=\Omega(\delta^2 dt)}$$

*such that given a (λ, Λ)-**suffix-friendly almost δ-Shannon-CG-source** X, and an independent uniform $Y \sim \{0,1\}^{\ell}$, it holds that* $\mathsf{SFShannonExt}(X,Y) \approx_{\varepsilon} U_m$.

Finally, we note that instead of using the GW extractors of Theorem 2.12, one can instead use the high min-entropy extractors of Reingold, Vadhan, and Wigderson [RVW02]. For entropy gap Δ, these extractors attain a seed length of $O(\log(\Delta/\varepsilon))$ instead of the $O(\Delta + \log(1/\varepsilon))$ seed length of Theorem 2.12. When ε is large, this may lead to an improved constant ℓ in the above theorems.

8 On Chor–Goldreich Sources with Bad Prefixes

A very natural, and seemingly useful, way to extend our notion of almost CG sources is to allow *bad prefixes*. Namely, for each $i \in [t]$ (or for most of them), $X_i | \left\{X_{[1,i-1]} = a\right\}$ is close to having high min-entropy only for most a-s in the support of $X_{[1,i-1]}$. We define this notion formally.

Definition 8.1 (good prefix). *Let $\gamma, \delta > 0$. Let $X = X_1 \circ \ldots \circ X_t$ be a source with each $X_i \sim \{0,1\}^d$. For $i \in [t]$, we say that a prefix $(a_1, \ldots, a_{i-1})$ is (γ, δ)-good for X if*

$$H_\infty^\gamma(X_i | X_{[1,i-1]} = a_1, \ldots, a_{i-1}) \geq \delta d.$$

Previously, a good step required high min-entropy conditioned on all prefixes. Now, we only require $1 - \rho$ fraction of good prefixes.

Definition 8.2 (good step). *Let $\gamma, \delta, \rho > 0$. Let $X = X_1 \circ \ldots \circ X_t$ be a source with each $X_i \sim \{0,1\}^d$. We say that $i \in [t]$ is (γ, δ, ρ)-good for X if with probability at least $1 - \rho$ over prefixes $(a_1, \ldots, a_{i-1}) \sim X_{[1,i-1]}$ we have that the prefix is (γ, δ)-good for X. (Note that for $i = 1$ we simply require $H_\infty^\gamma(X_1) \geq \delta d$.)*

Our extended definition then goes as follows.

Definition 8.3 (almost CG source, III). *A (γ, λ, ρ)-almost δ-CG source is a sequence of random variables $X_1 \circ \ldots \circ X_t$ with each $X_i \in \{0,1\}^d$, such that at least $(1 - \lambda)t$ i-s are (γ, δ, ρ)-good for X.*

Naturally, we can also define an almost Shannon CG source, where again, for each i, there is a small probability over prefixes $(X_1, \ldots, X_{i-1}) = (a_1, \ldots, X_{i-1})$ for some small fraction of i, and also for a small fraction of i, there is no guarantee on the quality of the distribution (for any prefix). To begin this definition, we can again, naturally define the notion of a good prefix and a good step.

Definition 8.4 (good Shannon prefix). *Let $\delta > 0$. Let $X = X_1 \circ \ldots \circ X_t$ be a source with each $X_i \in \{0,1\}^d$. For $i \in [t]$, we say that a prefix $(a_1, \ldots, a_{i-1})$ is δ-Shannon-good for X if*

$$H(X_i | X_1, \ldots, X_{i-1} = a_1, \ldots, a_{i-1}) \geq \delta d.$$

When δ and X are clear from context, and it is also clear we are discussing Shannon entropy, we will simply call a prefix "good" without the quantifiers.

Definition 8.5 (good Shannon step). *Let $\delta, \rho > 0$. Let $X = X_1 \circ \ldots \circ X_t$ be a source with each $X_i \in \{0,1\}^d$. We say that $i \in [t]$ is (δ, ρ)-Shannon-good for X if with probability at least $1 - \rho$ over prefixes $(a_1, \ldots, a_{i-1}) \sim (X_1, \ldots, X_{i-1})$ we have that the prefix is δ-Shannon-good for X. (Note that for $i = 1$ we simply require $H_\infty^\gamma(X_1) \geq \delta d$.)*

When δ, ρ and X are clear from context, and it is also clear we are discussing Shannon entropy, we will simply call a coordinate i "good" or a "good step" without the quantifiers. We also call i "bad" or a "bad step" if it is not good. Additionally, we use $\mathcal{G}(X)$ as the set of all good i-s.

Definition 8.6 (almost Shannon CG source). *A (λ, ρ)-almost δ-Shannon-CG source is a sequence of random variables $X_1 \circ \ldots \circ X_t$ with each $X_i \sim \{0,1\}^d$, such that at least $(1 - \lambda)t$ i-s are (δ, ρ)-good for X.*

While we do not know how to handle $\rho > 0$ in a way that extends our result, we argue here that there may be an inherent reason for that lack of success. At least in a certain parameter regime (particularly when $\lambda > 0$), we provably *cannot* extract from such sources with constant seed.

Theorem 8.7. *For any small enough constants $\zeta, \beta > 0$, there exists no constant-seed extractor for (γ, λ, ρ)-almost $(1 - \zeta)$-CG sources where $\gamma = \lambda = \rho = \zeta^\beta$. That is, for any such source $X = X_1 \circ \ldots \circ X_t \sim (\{0,1\}^d)^t$ and any function $g \colon \{0,1\}^{dt} \times \{0,1\}^\ell \to \{0,1\}^m$ where $\ell = O(1)$ and $m = \omega(1)$, it holds that $|g(X, U_\ell) - U_m| \geq \frac{1}{2}$.*

Toward establishing Theorem 8.7, we need the following extension of Corollary 3.11.

Lemma 8.8. *Let $X = X_1 \circ \ldots \circ X_t$ be a (λ, ρ)-almost δ-Shannon-CG source, with $X_i \in \{0,1\}^d$. Let $\delta, \rho > 0$ be constant. For any positive integer b, consider the distribution $X' = X'_1 \circ \ldots \circ X'_{\lfloor t/b \rfloor}$, where $X'_i = X_{[(i-1)b+1, ib]}$. Then, X' is a*

$$\left(\gamma = e^{-\delta^2 b/36} + \rho^{1/4}, \sqrt{\lambda}, \rho^{1/4}\right)$$

almost δ'-CG source for $\delta' = \frac{\delta^2}{18} - \frac{\delta}{6}\left(\sqrt{\rho} + \sqrt{\lambda}\right)$.

We defer the proof to Appendix A.4. Note that the result gives no meaningful lower bound on the entropy rate unless $\frac{\delta}{3} > \sqrt{\rho} + \sqrt{\lambda}$.

Unlike almost CG sources with $\delta = 0$, under the more general definition it turns out that *any* high-entropy weak source is an almost CG source with the appropriate error parameters. This is true even for sources with high Shannon entropy.

Lemma 8.9. *Let $X \sim \{0,1\}^n$ be a random variable with $H(X) \geq (1-\zeta)n$ for $\zeta \leq 2^{-40}$, let d and t be any positive integers such that $d \cdot t = n$, and let $b \geq 6 \ln \frac{1}{\zeta}$ be a positive integer that divides t. Then:*

1. *Writing $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, we have that X is a $(\zeta^{1/4}, \zeta^{1/4})$-almost $(1-\sqrt{\zeta})$-Shannon CG source with block size d.*
2. *Writing $X = X'_1, \circ \ldots \circ X'_{t/b}$, each $X'_i \sim \{0,1\}^{bd}$, we have that X is a $(2\zeta^{1/16}, \zeta^{1/8}, \zeta^{1/6})$ almost $\frac{1}{64}$-CG source.*

Recall that $H(X) \geq H_\infty(X)$, so the above also holds for X-s with $H_\infty(X) \geq (1-\zeta)n$ as well.

Proof: The chain rule for Shannon entropy tell us that

$$\frac{1}{t} H(X) = \frac{1}{t} \sum_{i \in [t]} H(X_i | X_{[1,i-1]}) \geq (1-\zeta)d,$$

where $H(X_i | X_{[1,i-1]}) = \sum_a \Pr[X_{[1,i-1]} = a] \cdot h(i,a)$, and for brevity, we write $h(i,a) = H(X_i | X_{[1,i-1]} = a)$. By an averaging argument,

$$\Pr_{i \sim [t], a \sim X_{[1,i-1]}} \left[h(i,a) \leq (1-\sqrt{\zeta})d\right] \leq \sqrt{\zeta}.$$

By another averaging argument, we can conclude that

$$\Pr_{i \in [t]} \left[\Pr_{a \sim X_{[1,i-1]}} \left[h(i,a) \leq (1-\sqrt{\zeta})d\right] \leq \zeta^{1/4}\right] \geq 1 - \zeta^{1/4},$$

which gives us Item 1. By Lemma 8.8, noting that

$$\frac{(1-\sqrt{\zeta})^2}{18} - \frac{1-\sqrt{\zeta}}{6} \cdot \left(\sqrt{\zeta^{1/4}} + \sqrt{\zeta^{1/4}}\right) \geq \frac{1-2\sqrt{\zeta}}{16} - \frac{\zeta^{1/8}}{3} \geq \frac{1}{64},$$

we have that X' that is formed by grouping b consecutive blocks together, is a $\left(e^{-b/72} + \zeta^{1/16}, \zeta^{1/8}, \zeta^{1/16}\right)$-almost $\frac{1}{64}$-CG source. Having chosen b large enough, we get Item 2. ■

Next, we show that we can invest a constant number of bits to regain the original (smooth) entropy rate. Clearly this should come at additional cost, and indeed we get worse ρ and λ.

Lemma 8.10. *Let $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, be a (γ, λ, ρ)-almost δ-CG source and let $\zeta > 0$ be any constant. Let $\mathsf{Ext}\colon \{0,1\}^d \times \{0,1\}^d \to \{0,1\}^m$ be the extractor from Theorem 2.11 set with error $\varepsilon_{\mathsf{E}} = 2^{-\frac{\delta}{3}d}$. For any $y \in \{0,1\}^d$, denote*

$$Z(y) = \mathsf{Ext}(X_1, y) \circ \ldots \mathsf{Ext}(X_t, y).$$

Let $\tau = \sqrt{\lambda + \rho + 2^{-\frac{\delta\zeta}{3}d}}$. Then, with probability at least $1-\tau$ over $y \sim U_d$, $Z(y)$ is a $(\gamma, \lambda' = \sqrt{\tau}, \rho' = \sqrt{\tau})$-almost $(1-\zeta)$-CG source.

Proof: First, note that the output length of Ext is $m = \delta d - 2\log(1/\varepsilon_{\mathsf{E}}) = \frac{\delta}{3}d$. Fix some $i \in \mathcal{G}(X)$, and denote $\mathcal{H}_i = (X_1, \ldots, X_{i-1})$. Further, for $h \sim \mathcal{H}$ denote $X_{i,h} = X_i | \{\mathcal{H}_i = h\}$. When h is good, we know that $X_{i,h}$ is γ-close to some $X'_{i,h}$ which is a δd-source. Thus,

$$\left(Y, \mathsf{Ext}(X'_{i,h}, Y)\right) \approx_{\varepsilon_{\mathsf{E}}} (Y, U_m)$$

By an averaging argument, there exists a set $B_{i,h} \subseteq \{0,1\}^d$ of density at most $\varepsilon_{\mathsf{E}}^{\zeta}$ such that for every $y \notin B_{i,h}$,

$$\mathsf{Ext}(X'_{i,h}, y) \approx_{\varepsilon_{\mathsf{E}}^{1-\zeta}} U_m.$$

By Claim 2.1, and our aforementioned choice of parameters, $D'_{i,h} = \mathsf{Ext}(X'_{i,h}, y)$ has entropy rate $\frac{1}{m}\log(1/\varepsilon_{\mathsf{E}}^{1-\zeta}) = 1-\zeta$. Denoting $D_{i,h} = \mathsf{Ext}(X_{i,h}, y)$, we know that $D_{i,h} \approx_\gamma D'_{i,h}$.

Denoting the set of good prefixes by H_i, we have established that

$$\forall i \in \mathcal{G}(X)\ \forall h \in H_i\ \forall y \notin B_{i,h},\ H_\infty^\gamma(\mathsf{Ext}(X_{i,h}, y)) \geq (1-\zeta)m.$$

Let $I(i,h,y)$ be the bad event $H_\infty^\gamma(\mathsf{Ext}(X_{i,h}, y)) < (1-\zeta)m$. Collecting error terms, we have that

$$\Pr_{i \sim [t], h \sim \mathcal{H}_i, y \sim U_d}[I(i,h,y)] \leq \lambda + (1-\lambda)\left(\rho + (1-\rho)\varepsilon_{\mathsf{E}}^{\zeta}\right) \leq \tau^2.$$

By an averaging argument, we have a set $B_Y \subseteq \{0,1\}^d$ of bad seeds of density at most τ such that for every $y \notin B_Y$ we get that $\Pr_{i \sim [t], h \sim \mathcal{H}_i}[I(i,h,y)] \leq \tau$. By yet another averaging argument, we get that for every $y \notin B_Y$ there exists a set of bad indices $B_I(y)$ of density at most $\sqrt{\tau}$ such that for every $y \notin B_Y$ and $i \notin B_I(y)$ it holds that $H_\infty^\gamma(\mathsf{Ext}(X_{i,h}), y) \geq (1-\zeta)m$ with probability at least $1-\sqrt{\tau}$ over $h \sim \mathcal{H}_i$. ■

We can now combine Lemmas 8.8 and 8.10 to get our the following corollary.

Corollary 8.11. *Let $X \sim \{0,1\}^n$ be a random variable with $H(X) \geq (1-\zeta)n$ for a constant $\zeta < 2^{-64}$. Then, there exist constant positive integers d and ℓ, and an explicit function*

$$f\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^{m=\Omega(n)},$$

such that with probability at least $1-2\zeta^{1/32}$ over $y \in \{0,1\}^\ell$, $f(X,y) = Z_1 \circ \ldots \circ Z_t$ is a (γ, λ, ρ)-almost $(1-\zeta)$-CG source, where each $Z_i \sim \{0,1\}^d$, $\gamma = 2\zeta^{1/6}$, and $\rho = \lambda = O(\zeta^{1/64})$.

Proof of Theorem 8.7 (sketch): The above corollary shows we can convert, using a constant-length seed, a high entropy source (even a high *Shannon* entropy one) into a high *min-entropy* almost CG source, albeit with large ρ and λ. On the other hand, there exist no constant-seed condensers that condense from $(1-\zeta)n$ min-entropy (out of n bits) to $m - O(1)$ min-entropy (out of m bits) where $m = \Omega(n)$, let alone from Shannon entropy. ■

Thus, it seems plausible that (at least) one of the following holds.

1. The barrier to condensing from $f(X, y)$, for a good y, is that almost CG sources with $\rho > 0$ (or at least, a relatively large ρ) do not admit deterministic condensing, or even condensing with constant seed. That is, we cannot hope for an analogue of Theorem 5.3 when a small fraction of the prefixes are bad. Or,

2. The barrier lies in the fraction of bad steps. That is, without suffix-friendliness, no constant seed condensing exists, even using techniques which do not work in an "online" fashion like ours.

We leave this as a line of inquiry for future research.

9 Open Problems

We present several open problems that arise from our work. Recall that in our notation, a (γ, λ, ρ)-almost δ-CG source is a CG source in which every good conditional distribution is γ-close to a δ-source, there are at most λ bad steps, and there is a weight of at most ρ on bad prefixes at each step. A (γ, λ)-almost δ-CG source is a $(\gamma, \lambda, 0)$-almost δ-CG source.

The first two problems, which concern ρ and λ type errors, follow naturally from the discussion in the previous section.

Open Problem 1: Is there an analogue of Theorem 5.3 to $(\gamma, 0, \rho)$-almost δ-CG sources? That is, given a random walk via such a source, $X_1 \circ \ldots \circ X_t$, is the final vertex distribution $O(\rho)$-close to a distribution with constant entropy gap? (An error of $O(\rho)$ is expected, as one can consider the almost CG source X that outputs a fixed string with probability ρ, and otherwise follows the distribution of a $(\gamma, 0, 0)$-almost δ-CG source.)

Open Problem 2: Is there an analogue of Theorem 5.3 to (γ, λ)-almost δ-CG sources without suffix friendliness? Namely, is there an explicit deterministic condenser that outputs a distribution with constant entropy gap without the suffix-friendliness requirement? (This problem does not ask whether *our* random walk construction achieves this, as we know it cannot.)

Open Problem 3: Relaxing Open Problem 1, is there an analogue of Theorem 5.3 to $(\gamma, 0, \rho)$-almost δ-CG sources, where the output distribution only has high Shannon entropy?

Open Problem 4: Can we improve Corollary 4.11 and Corollary 4.12 so that the q-norm decrease at each step is

$$\|p_V\|_q^q \leq \frac{1}{C^\alpha} \|p_U\|_q^q$$

for some C very close to D and all sufficiently small α? We note that a positive answer will essentially solve Open Problem 3.

10 Acknowledgments

We wish to thank Oded Goldreich for valuable comments and suggestions on a preliminary version of this paper, and Salil Vadhan and Yevgeniy Dodis for very helpful discussions. We also thank an anonymous reviewer for useful comments and for pointing out a minor flaw in a previous version of this manuscript.

References

[AC02] Noga Alon and Michael Capalbo. Explicit unique-neighbor expanders. In *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–79. IEEE, 2002.

[AT19] Nir Aviv and Amnon Ta-Shma. On the entropy loss and gap of condensers. *ACM Transactions on Computation Theory (TOCT)*, 11(3):1–14, 2019.

[BCDT19] Avraham Ben-Aroya, Gil Cohen, Dean Doron, and Amnon Ta-Shma. Two-source condensers with low error and small entropy gap via entropy-resilient functions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2019.

[BDK$^+$11] Boaz Barak, Yevgeniy Dodis, Hugo Krawczyk, Olivier Pereira, Krzysztof Pietrzak, François-Xavier Standaert, and Yu Yu. Leftover hash lemma, revisited. In *Annual Cryptology Conference*, pages 1–20. Springer, 2011.

[BDT19] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. An efficient reduction from two-source to nonmalleable extractors: achieving near-logarithmic min-entropy. *SIAM Journal on Computing*, (0):STOC17–31, 2019.

[BEG17] Salman Beigi, Omid Etesami, and Amin Gohari. Deterministic randomness extraction from generalized and distributed Santha–Vazirani sources. *SIAM Journal on Computing*, 46(1):1–36, 2017.

[BGI$^+$08] Radu Berinde, Anna C. Gilbert, Piotr Indyk, Howard Karloff, and Martin J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 798–805. IEEE, 2008.

[BGM22] Marshall Ball, Oded Goldreich, and Tal Malkin. Randomness extraction from somewhat dependent sources. In *Proceedings of the 13th Innovations in Theoretical Computer Science Conference (ITCS)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[BKS$^+$10] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, ramsey graphs, dispersers, and extractors. *Journal of the ACM (JACM)*, 57(4):20, 2010.

[CCLO22] Xue Chen, Kuan Cheng, Xin Li, and Minghui Ouyang. Improved decoding of expander codes. In *Proceedings of the 13th Innovations in Theoretical Computer Science Conference (ITCS)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[CG22] Eshan Chattopadhyay and Jesse Goodman. Improved extractors for small-space sources. In *Proceedings of the 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 610–621. IEEE, 2022.

[CGL22] Eshan Chattopadhyay, Jesse Goodman, and Jyun-Jie Liao. Affine extractors for almost logarithmic entropy. In *Proceedings of the 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 622–633. IEEE, 2022.

[CRVW02] Michael Capalbo, Omer Reingold, Salil Vadhan, and Avi Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the 34th Annual Symposium on Theory of Computing (STOC)*, pages 659–668. ACM, 2002.

[CT21] Lijie Chen and Roei Tell. Simple and fast derandomization from very hard functions: eliminating randomness at almost no cost. In *Proceedings of the 53rd Annual Symposium on Theory of Computing (STOC)*, pages 283–291. ACM, 2021.

[DGSX21a] Yevgeniy Dodis, Siyao Guo, Noah Stephens-Davidowitz, and Zhiye Xie. No time to hash: On super-efficient entropy accumulation. In *CRYPTO*, volume 12828 of *Lecture Notes in Computer Science*, pages 548–576. Springer, 2021.

[DGSX21b] Yevgeniy Dodis, Siyao Guo, Noah Stephens-Davidowitz, and Zhiye Xie. Online linear extractors for independent sources. In *Proceedings of the 2nd Conference on Information-Theoretic Cryptography (ITC)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[DK08] Domingos Dellamonica Jr. and Yoshiharu Kohayakawa. An algorithmic Friedman–Pippenger theorem on tree embeddings and applications. *The Electronic Journal of Combinatorics*, pages R127–R127, 2008.

[DMOZ20] Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. Nearly optimal pseudorandomness from hardness. In *Proceedings of the 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1057–1068. IEEE, 2020.

[DPW14] Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs. Key derivation without entropy waste. In *Advances in Cryptology–EUROCRYPT 2014*, pages 93–110. Springer, 2014.

[DRV12] Yevgeniy Dodis, Thomas Ristenpart, and Salil Vadhan. Randomness condensers for efficiently samplable, seed-dependent sources. In *Theory of Cryptography Conference*, pages 618–635. Springer, 2012.

[Dvi12] Zeev Dvir. Extractors for varieties. *computational complexity*, 21(4):515–572, 2012.

[DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory (TOCT)*, 4(1):1–21, 2012.

[DY13] Yevgeniy Dodis and Yu Yu. Overcoming weak expectations. In *Theory of Cryptography Conference*, pages 1–22. Springer, 2013.

[GP20] Dmitry Gavinsky and Pavel Pudlák. Santha-Vazirani sources, deterministic condensers and very strong extractors. *Theory of Computing Systems*, 64(6):1140–1154, 2020.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *Journal of the ACM (JACM)*, 56(4):20, 2009.

[GV99] Oded Goldreich and Salil Vadhan. Comparing entropies in statistical zero knowledge with applications to the structure of szk. In *Proceedings. Fourteenth Annual IEEE Conference on Computational Complexity (Formerly: Structure in Complexity Theory Conference)(Cat. No. 99CB36317)*, pages 54–73. IEEE, 1999.

[GW97] Oded Goldreich and Avi Wigderson. Tiny families of functions with random properties: A quality-size trade-off for hashing. *Random Structures & Algorithms*, 11(4):315–343, 1997.

[ILL89] Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Pseudo-random generation from one-way functions. In *Proceedings of the 21st Annual Symposium on Theory of computing (STOC)*, pages 12–24, 1989.

[KRVZ06] Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. In *Proceedings of the 38th Annual Symposium on Theory of Computing (STOC)*, pages 691–700. ACM, 2006.

[LH22] Ting-Chun Lin and Min-Hsiu Hsieh. Good quantum LDPC codes with linear time decoder from lossless expanders. *arXiv preprint arXiv:2203.03581*, 2022.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil Vadhan, and Avi Wigderson. Extractors: Optimal up to constant factors. In *Proceedings of the 35th Annual Symposium on Theory of computing (STOC)*, pages 602–611, 2003.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[RR99] Ran Raz and Omer Reingold. On recycling the randomness of states in space bounded computation. In *Proceedings of the 61st Annual Symposium on Theory of Computing (STOC)*, pages 159–168. ACM, 1999.

[RSW06] Omer Reingold, Ronen Shaltiel, and Avi Wigderson. Extracting randomness via repeated condensing. *SIAM Journal on Computing*, 35(5):1185–1209, 2006.

[RT00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24, 2000.

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, pages 157–187, 2002.

[RVW04] Omer Reingold, Salil Vadhan, and Avi Wigderson. A note on extracting randomness from Santha-Vazirani sources. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2004.

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33(1):75–87, 1986.

[SV22] Ronen Shaltiel and Emanuele Viola. On hardness assumptions needed for "extreme high-end" PRGs and fast derandomization. In *Proceedings of the 13th Innovations in Theoretical Computer Science Conference (ITCS)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[SZ99] Aravind Srinivasan and David Zuckerman. Computing with very weak random sources. *SIAM Journal on Computing*, 28(4):1433–1459, 1999.

[TUZ07] Amnon Ta-Shma, Christopher Umans, and David Zuckerman. Lossless condensers, unbalanced expanders, and extractors. *Combinatorica*, 27:213–240, 2007.

[TV00] Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2000)*, pages 32–42. IEEE, 2000.

[Vad12] Salil Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014.

[Zuc90] David Zuckerman. General weak random sources. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 534–543. IEEE, 1990.

[Zuc96] David Zuckerman. Simulating BPP using a general weak random source. *Algorithmica*, 16(4):367–391, 1996.

[Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of Max Clique and Chromatic Number. *Theory of Computing*, 3:103–128, 2007.

A Deferred Proofs

A.1 Condensing from Suffix Friendliness and Any Rate

This section provides details on the construction in and proof of Theorem 5.4.

The construction is nearly identical to that in the non suffix-friendly case, with a slight modification of some parameters. Namely, we slightly adjust the size of both the small and big graph, along with some other parameters, in order to facilitate the slightly different analysis. We list out all parameters again here for completeness.

Let $c^\star$ be the global constant from Corollary 2.8. For any constant δ, suppose we have a $(\gamma, \lambda, \Lambda)$-suffix-friendly-almost δ-CG source $X_1, \dots, X_t$ with each $X_i \sim \{0,1\}^d$. Again let $e = 100$ and suppose that

- $d \geq \max\left\{\frac{10^3}{\delta^2}, \frac{2c^\star}{\delta}\right\}$,
- $\gamma \leq 2^{-100/\delta}$, and,
- $\lambda \leq \frac{1}{10^8 e^2}\delta^2$.

Our construction again consists of an optimal constant sized lossless conductor, and the lossless conductor from [CRVW02].

The Small Graph (Suffix Friendly) The parameters here are nearly identical to before. Except, we utilize the fact that $\sqrt{\lambda} \leq .01\delta$ to replace the term $(0.9\delta - \sqrt{\lambda})$ with 0.89δ. We'll need to assume this worst case upper bound for technical reasons in the averaging argument of Lemma A.1.[30] Namely, we set $\beta = 1 - \delta/2$, $\Delta = \delta/3$. We set the epoch length $b = \frac{10^6 d^3 e}{\delta}$, and

$$d_{\mathsf{CRVW}} = \log D_{\mathsf{CRVW}} = 0.89 \cdot \delta d b + \frac{2\beta}{\log(1/\gamma)} d^2 + 2d + \log c^\star.$$

We let $H = ([D_{\mathsf{CRVW}}], [D_{\mathsf{CRVW}}], E)$ be the D-regular bipartite graph that is a $\left(K = \frac{D_{\mathsf{CRVW}}}{c^\star D^2}, \varepsilon = \frac{1}{D^\beta}\right)$-lossless expander guaranteed to us by by Corollary 2.8.

The Big Graph (Suffix Friendly) We change the size N (and thus indirectly K) to facilitate the conditions of Theorem 4.15 rather than Theorem 4.13. We also modify δ_{CRVW} to reflect the entropy rate in each Z_i when assuming the inequality $\sqrt{\lambda} \leq 0.01\delta$ is tight. Again, let $\beta_{\mathsf{CRVW}} = 1/6$, $\Delta_{\mathsf{CRVW}} = 1/12$, $\gamma_{\mathsf{CRVW}} = \eta = 2^{-100}$, and

$$\delta_{\mathsf{CRVW}} = \frac{k - 0.01\delta d b - 2\beta d^2 - 2d \log\frac{1}{\eta}}{d_{\mathsf{CRVW}}} \geq \frac{11}{12}.$$

Let $G = ([N], [N], E)$ be the $\left(K_{\mathsf{CRVW}} = \Omega\left(\frac{N}{D^2_{\mathsf{CRVW}}}\right), \varepsilon_{\mathsf{CRVW}} = \frac{1}{D^{1/6}_{\mathsf{CRVW}}}\right)$-expander guaranteed to us by Theorem 2.5 with

$$k_{\mathsf{CRVW}} = \left(0.9\delta_{\mathsf{CRVW}} - 2e\sqrt{\lambda}\right) d_{\mathsf{CRVW}} \frac{t}{b} - 2 \cdot \frac{\delta \Lambda e}{10^6} d_{\mathsf{CRVW}} + \frac{2e\beta_{\mathsf{CRVW}}}{\log(1/\gamma_{\mathsf{CRVW}})} d^2_{\mathsf{CRVW}}.$$

Note again that $n = \log N = k_{\mathsf{CRVW}} + O(d_{\mathsf{CRVW}})$.

The Analysis

The construction given the above graphs is the same as in the non-suffix-friendly case. We divide the random walk into t/b blocks of length b, and we let Z_j be the distribution on H after a random walk using the j-th block as instructions.

Lemma A.1. *The sequence $Z_1, \ldots, Z_{t/b}$ is an $\left(\gamma_{\mathsf{CRVW}}, \sqrt{\lambda}, \frac{\delta\Lambda}{10^6}\right)$-suffix-friendly almost δ_{CRVW}-CG source.*

[30] We note that there is no place in this work where it is *necessary* for us to consider the more fine grained fact that the entropy loss is $0.9\delta - \lambda$ rather than $0.9\delta - 0.01\delta$. In other words, we always "think" of λ as large as possible in terms of δ. However, we've kept the entropy loss as accurate as possible whenever we can.

Proof: Call the j-th epoch of the original CG source, $X_{(j-1)b+1}, \ldots, X_{(j-1)b+b}$ "good" if less than 0.01δ fraction of the X_i-s in the epoch are bad steps. We'll show that for any suffix of the Z_j-s, there are at most $\frac{\delta\Lambda}{10^6} + \sqrt{\lambda}(t-j)$ bad epochs. Consider any suffix of length s of the Z_j-s. There are at most $\Lambda + \lambda bs$ bad steps in X in this suffix. By an averaging argument, for any a, there are at most

$$\frac{\left(\frac{\Lambda}{s} + \lambda b\right) s}{a}$$

epochs with more than a bad steps in them. Setting $a = \sqrt{\lambda}b + 0.005\delta b \leq 0.01\delta b$ tells us that the number of bad epochs in the suffix is at most

$$\frac{\Lambda}{\sqrt{\lambda}b + 0.005\delta b} + \frac{\lambda b}{\sqrt{\lambda}b + 0.005\delta b} s \leq \frac{200\Lambda}{\delta b} + \sqrt{\lambda}s \leq \frac{\delta\Lambda}{10^6} + \sqrt{\lambda}s,$$

where the last inequality comes from our choice of b and our lower bound on d. Now, as before one can observe that the all conditions for Theorem 4.13 are met and therefore the distribution on Z_j conditioned on any prefix is η-close to a source with

$$k - 0.01\delta db - 2\beta d^2 - 2d\log\frac{1}{\eta} = \delta_{\mathsf{CRVW}} \cdot d_{\mathsf{CRVW}}$$

min-entropy. ■

Lemma A.2. *For any $\eta_{\mathsf{CRVW}} > 0$, the distribution of $W = \mathsf{RW}(G, Z_1, \ldots, Z_{t/b})$ is η_{CRVW}-close to a $(k - \mathrm{poly}(d, 1/\delta) \cdot (\Lambda + \log 1/\eta_{\mathsf{CRVW}}))$-source*

Proof: We verify all conditions of Theorem 4.15 are met to apply it with G and the Z_i-s:

- $1 > \beta_{\mathsf{CRVW}} = 1/6 > \Delta_{\mathsf{CRVW}} = 1/12$,
- $\delta_{\mathsf{CRVW}} > 11/12 = 1 - \beta_{\mathsf{CRVW}} + \Delta_{\mathsf{CRVW}}$,
- $d_{\mathsf{CRVW}} \geq 10^6 e \geq 80 \cdot 12 \cdot \frac{12}{11} \cdot e \geq \frac{80e}{\delta_{\mathsf{CRVW}}\Delta_{\mathsf{CRVW}}}$,
- $\gamma_{\mathsf{CRVW}} = \eta \leq 2^{-200e} \leq 2^{-100e/\delta_{\mathsf{CRVW}}}$,
- $\sqrt{\lambda} \leq \frac{\delta}{10^4 e} \leq \frac{\delta}{6e}$, and,
- $(0.9\delta_{\mathsf{CRVW}} - 2\sqrt{\lambda})d_{\mathsf{CRVW}}\frac{t}{b} - 2 \cdot \frac{\delta\Lambda e}{10^6} d_{\mathsf{CRVW}} = k_{\mathsf{CRVW}} - \frac{2\beta_{\mathsf{CRVW}}}{\log 1/\gamma_{\mathsf{CRVW}}} d^2_{\mathsf{CRVW}}$.

Therefore, for any $\eta_{\mathsf{CRVW}} > 0$, W is η_{CRVW}-close to a source with min-entropy:

$$k_{\mathsf{CRVW}} - 2\beta e d^2_{\mathsf{CRVW}} - \left(\frac{\Lambda e^2}{10^5} + 2e\log(1/\eta_{\mathsf{CRVW}})\right) d_{\mathsf{CRVW}}.$$

Notice that all the terms after k_{CRVW} in the above expression are $\mathrm{poly}(d, 1/\delta) \cdot (\Lambda + \log(1/\eta_{\mathsf{CRVW}}))$. ■

Using the fact that $n = k_{\mathsf{CRVW}} + O(1)$, and that $k_{\mathsf{CRVW}} = \Omega(\delta dt)$, yields Theorem 5.4.

A.2 The Nisan–Zuckerman (NZ) Construction

The overall idea is as follows. Given a CG source $X = X_1 \circ \ldots \circ X_t$, we use a constant-sized Y to extract a uniform Z_1 from a $X_{[t-a+1,t]}$ where $a = O(1)$. Then, we use Z_1 as a seed to extract from $X_{[t-b+1,t-a]}$ to get Z_2, where the latter interval is roughly twice as long as the former. Continuing this way for $s-1 = O(\log t)$ times, we finally use Z_{s-1} as a seed to extract from a perfix of X of length $\Omega(dt)$, which is our final output. Formally, Theorem 1.2 follows from the following, more general, theorem.

Theorem A.3 (follows from [NZ96])**.** *For any positive integers t, d, any $\varepsilon, \delta > 0$, and any $\alpha > 1$, the following holds. For some positive integer a, let $\{\mathsf{Ext}_i\}_i$ be a family of explicit extractors, each $\mathsf{Ext}_i \colon \{0,1\}^{\alpha^{i-1}ad} \times \{0,1\}^{b_i} \to \{0,1\}^{m_i}$ being a $(k_i = \delta\alpha^{i-1}ad, \varepsilon_i)$ extractor, where $m_i \geq b_{i+1}$.*

Setting $s = \lceil \log_\alpha(1 + (\alpha-1)\frac{t}{a}) \rceil$, there exists an explicit extractor

$$\mathsf{ExactCGExt} \colon \left(\{0,1\}^d\right)^t \times \{0,1\}^{b_1} \to \{0,1\}^{m_s}$$

for δ-CG sources with error $\varepsilon = \sum_{i \in [s]} \varepsilon_i$.

Proof: We are given a δ-CG source $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, and $\alpha > 1$. We define a sequence of exponentially-growing intervals as follows. Set $I_1 = [t-a+1, t]$, and for all $i > 1$, given $I_{i-1} = [\ell_{i-1}, \ell_i - 1]$, we set $I_i = [\ell_{i-1} - \alpha(\ell_i - \ell_{i-1}), \ell_{i-1} - 1]$. Note that $|I_i| = \alpha \cdot |I_{i-1}|$, and that I_s starts at 1 when $s = \log_\alpha(1 + (\alpha-1)\frac{t}{a})$.[31] The parameter s stands for the number of extraction steps we perform, and note that $s = O(\log t)$ whenever a and α are constants.

Recall that b_1 is the seed length of $\mathsf{Ext}_1 \colon \{0,1\}^{ad} \times \{0,1\}^{b_1} \to \{0,1\}^{m_1}$, and let $Y \sim \{0,1\}^{b_1}$ be uniform and independent of X. We define the following sequence of random variables.

- $Z_1 = \mathsf{Ext}_1(X_{I_1}, Y)$, and,
- For all $i \in [2, s]$, denote $Z_i = \mathsf{Ext}_i(X_{I_i}, Z_{i-1})$, where $\mathsf{Ext}_i \colon \{0,1\}^{\alpha^{i-1}ad} \times \{0,1\}^{b_i} \to \{0,1\}^{m_i}$.

Recall that we indeed require $m_i \geq b_{i+1}$. (If $m_i > b_{i+1}$, we only use the first b_{i+1} bits of Z_i to extract from $X_{I_{i+1}}$.) The δ-CG source extractor is given by

$$\mathsf{ExactCGExt}(X, Y) = Z_s.$$

Clearly, $\mathsf{ExactCGExt}$ is explicit since the Ext_i-s are explicit. For simplicity, we also denote $Z_0 = Y$ and $m_0 = b_1$. The correctness will follow from the following lemma.

Lemma A.4. *For all $i \in [s]$ it holds that*

$$\left(X_{I_s}, \ldots, X_{I_{i+1}}, Z_i\right) \approx_{\varepsilon(i)} \left(X_{I_s}, \ldots, X_{I_{i+1}}\right) \times U_{m_i},$$

where $\varepsilon(i) = \sum_{j=1}^{i} \varepsilon_j$.

[31] We assume s and $\alpha(\ell_i - \ell_{i-1})$ are integers, and otherwise the construction can be adjusted without significant loss in parameters.

Proof: The proof is by induction on i. For $i = 1$ the lemma readily holds, since Y is uniform and independent of X, and X_{I_1} is a δ-source conditioned on any fixing of $(X_{I_s}, \dots, X_{I_2})$. The latter readily follows from the fact that X is a δ-CG source. Next, fix $i > 1$ and assume the claim holds for $i - 1$. By the induction's hypothesis,

$$\left(X_{I_s}, \dots, X_{I_{i+1}}, X_{I_i}, Z_{i-1}\right) \approx_{\varepsilon(i-1)} \left(X_{I_s}, \dots, X_{I_{i+1}}, X_{I_i}\right) \times U_{m_{i-1}}.$$

Conditioned on any value of $X_{I_s}, \dots, X_{I_{i+1}}$, the random variable X_{I_i} is a δ-source over $\alpha^{i-1}ad$ bits. Thus, recalling that $Z_i = \mathsf{Ext}_i(X_{I_i}, Z_{i-1})$, it follows that

$$\begin{aligned}\left(X_{I_s}, \dots, X_{I_{i+1}}, Z_i\right) &\approx_{\varepsilon(i-1)} \left(X_{I_s}, \dots, X_{I_{i+1}}\right) \times \mathsf{Ext}_i(X_{I_i}, U_{m_{i-1}}) \\ &\approx_{\varepsilon(i-1)+\varepsilon_i} \left(X_{I_s}, \dots, X_{I_{i+1}}\right) \times U_{m_i},\end{aligned}$$

as required. ■

Applying Lemma A.4 with $i = s$ we get that $\mathsf{ExactCGExt}(X, Y) = Z_s \approx_\varepsilon U_{m_s}$, concluding the proof. Note that if the extractors are strong, one can increase the number of output bits by considering $Z_i = \mathsf{Ext}_i(X_{I_i}, Z_{i-1}) \circ Z_{i-1}$ rather than $Z_i = \mathsf{Ext}_i(X_{I_i}, Z_{i-1})$. ■

In [NZ96], they take the family of seeded extractors to be a universal family of hash functions (see Theorem 2.11). Instantiating Theorem A.3 with $\mathsf{Ext}_{\mathsf{ILL}}(x, y) = h_y(x) \circ \bar{y}$, for $\bar{y}$ being a suitable prefix of y, we get the following explicit extractor for δ-CG sources.[32]

Corollary A.5. *For any positive integers t, d, any $\varepsilon > 0$ and a constant $\delta > 0$, the following holds. For any constant $\beta > 0$ there exists an explicit function*

$$\mathsf{ExactCGExt}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell} \to \{0,1\}^m$$

where $m = (1-\beta)\delta n$ and $\ell = d + O_{\beta,\delta}(\log \frac{1}{\varepsilon})$, such that given a δ-CG source $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, and an independent uniform $Y \sim \{0,1\}^{\ell}$, it holds that $\mathsf{ExactCGExt}(X, Y) \approx_\varepsilon U_m$.

Proof: The extractor from Theorem 2.11 is strong, so we can append to $h_y(x)$ any prefix of y. We choose the prefix so that $m_i = b_{i+1}$. In particular, $b_{i+1} = \alpha^i ad$, and the length of $h_y(x)$ is $k_i - 2\log\frac{1}{\varepsilon_i}$, so we take $\bar{y}$ to be of length $\alpha^i ad - k_i + 2\log\frac{1}{\varepsilon_i}$, and we need it to be at most $|y| = \alpha^{i-1}ad$. Towards this end, we choose

- $\alpha = 1 + (1-\beta)\delta$,
- a to be the smallest integer larger than $\frac{4}{\beta(1-\beta)\delta^2} \cdot \frac{\log(1/\varepsilon)}{d}$, and ,
- For all $i \in [s]$, $\varepsilon_i = \frac{(\alpha-1)\varepsilon}{\alpha^i}$.

Under those choices:

- One can verify that indeed $\alpha^i ad - k_i + 2\log\frac{1}{\varepsilon_i} \le \alpha^{i-1}ad$, and also,
- The errors satisfy $\sum_{i\in[s]} \varepsilon_i = (\alpha-1)\varepsilon \sum_{i\in[s]} \alpha^{-i} \le \varepsilon$, as required.

Finally, note that the seed length is $b_1 = ad = d + O(\log(1/\varepsilon))$, and the output length is

$$m_s = b_{s+1} = \alpha^s ad \ge \left(1 + (\alpha-1)\frac{t}{a}\right) ad \ge (\alpha-1)td = (1-\beta)\delta dt.$$

■

[32] Choosing to work with the extractor of [RVW02] instead of $\mathsf{Ext}_{\mathsf{ILL}}$ would improve the seed length ℓ in the high ε regime. See the discussion at the end of Section 7.

A.3 The NZ Construction for Almost CG Sources

We show that the above construction can be applied to $(\gamma, 0)$-almost δ-CG sources as well, thereby establishing Theorem 1.4. The idea is natural once we saw, in Lemma 3.3, that grouping into blocks improves γ at the expense of slightly worse smooth entropy rate per block: Each time we group consecutive blocks for the next extraction step, we make sure we group enough blocks so that we can union-bound over the smoothness error. For brevity, we use the notation γ-almost δ-CG source in place of $(\gamma, 0)$-almost δ-CG source.

Theorem A.6. *For any positive integers t, d, any $\varepsilon, \delta, \gamma \in (0,1)$, and any $\alpha > 1$, the following holds. For some positive integer a, let $\{\mathsf{Ext}_i\}_i$ be a family of explicit extractors, each $\mathsf{Ext}_i \colon \{0,1\}^{\alpha^{i-1}ad} \times \{0,1\}^{b_i} \to \{0,1\}^{m_i}$ being a $(k_i = (1-2\gamma)\delta\alpha^{i-1}ad, \varepsilon_i)$ extractor, where $m_i \geq b_{i+1}$ and $a\alpha^{i-1} \geq \frac{8}{(1-\gamma)^2} \ln \frac{1}{\varepsilon_i}$.*

Setting $s = \lceil \log_\alpha(1 + (\alpha - 1)\frac{t}{a}) \rceil$, there exists an explicit extractor

$$\mathsf{AlmostCGExt} \colon \left(\{0,1\}^d\right)^t \times \{0,1\}^{b_1} \to \{0,1\}^{m_s}$$

for δ-CG sources with error $\varepsilon = 2\sum_{i \in [s]} \varepsilon_i$.

Proof: The proof closely follows Theorem A.3, where the only difference is that we need to make sure α and a are large enough so that each X_{I_i} is *sufficiently* close to having high min-entropy. Towards that end, we extend Lemma A.4 as follows, wherein we use the same construction and notation of Theorem A.3.

Lemma A.7. *For all $i \in [s]$ it holds that*

$$\left(X_{I_s}, \ldots, X_{I_{i+1}}, Z_i\right) \approx_{\varepsilon(i)} \left(X_{I_s}, \ldots, X_{I_{i+1}}\right) \times U_{m_i},$$

where $\varepsilon(i) = 2\sum_{j=1}^{i} \varepsilon_j$.

Proof: The proof is by induction on i. For $i = 1$, conditioned on any value of $X_{I_s}, \ldots, X_{I_2}$, X_{I_1} comprises a blocks that form a γ-almost δ-CG source. By the second item of Lemma 3.3, X_{I_1} is γ_1-close to having δ' entropy rate, where $\gamma_1 = e^{-\frac{\gamma^2}{2}a}$ and $\delta' = (1-2\gamma)\delta$.[33] Under each such conditioning, recalling that Y is independent of X, we have that $Z_1 \approx_{\varepsilon_1 + \gamma_1} U_{m_1}$. This is also true on average over $X_{I_s}, \ldots, X_{I_2}$. By our assumption on a, we have that $\gamma_1 \leq \varepsilon_1$, which established the claim for $i = 1$.

Next, fix $i > 1$ and assume the claim holds for $i - 1$. By the induction's hypothesis,

$$\left(X_{I_s}, \ldots, X_{I_{i+1}}, X_{I_i}, Z_{i-1}\right) \approx_{\varepsilon(i-1)} \left(X_{I_s}, \ldots, X_{I_{i+1}}, X_{I_i}\right) \times U_{m_{i-1}}.$$

Conditioned on any value of $X_{I_s}, \ldots, X_{I_{i+1}}$, the random variable X_{I_i} comprises $\alpha^{i-1}a$ blocks that form a γ-almost δ-CG source. Again, by Lemma 3.3, each conditioning gives rise to a source over $\alpha^{i-1}ad$ bits that is γ_i-close to a source with δ' entropy rate, for $\gamma_i = e^{-\frac{-\gamma^2}{2}\alpha^{i-1}a}$. By our assumption on a and α, we have that $\gamma_i \leq \varepsilon_i$. Thus, similarly to Lemma A.4 we get

$$\left(X_{I_s}, \ldots, X_{I_{i+1}}, Z_i\right) \approx_{\varepsilon(i-1)+2\varepsilon_i} \left(X_{I_s}, \ldots, X_{I_{i+1}}\right) \times U_{m_i},$$

as desired. ■

[33] Technically, Lemma 3.3 only states a result for an almost CG source divided into evenly sized blocks. However, it's easy to see such a result also holds for an almost CG source divided into any collection of contiguous blocks of arbitrary size: each block will have appropriate entropy rate, while the smoothness error will be exponentially small in the length of the block.

As before, we apply the above lemma with $i = s$ to conclude the proof. ■

Similar to what we did in Appendix A.2, we instantiate the above theorem with a universal family of hash functions as extractors, and set the parameters accordingly.

Corollary A.8. *For any positive integers t, d, any $\varepsilon > 0$ and any constants $\delta, \gamma > 0$ the following holds. For any constant $\beta \geq \frac{\gamma^2}{d\delta \ln 2}$, there exists an explicit function*

$$\mathsf{AlmostCGExt}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell} \to \{0,1\}^m$$

where $m = (1-2\gamma-\beta)\delta n$ and $\ell = O_{\beta,\delta,\gamma}(d \log \frac{1}{\varepsilon})$, such that given a γ-almost δ-CG source $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, and an independent uniform $Y \sim \{0,1\}^{\ell}$, it holds that $\mathsf{AlmostCGExt}(X, Y) \approx_{\varepsilon} U_m$.

Proof: As in Corollary A.5, we choose $m_i = b_{i+1}$ and append the seed accordingly. We choose:

- a to be the smallest integer larger than $\max\left\{\frac{2\ln 2}{\gamma^2} + \frac{2}{\gamma^2}\ln\frac{1}{\varepsilon}, \frac{2\ln 2}{(1-2\gamma-\beta)\delta\gamma^2}\right\}$,
- $\alpha = 1 + \Delta$ for $\Delta = \frac{2\ln 2}{\gamma^2 a}$, and,
- For all $i \in [s]$, $\varepsilon_i = e^{-\frac{\gamma^2}{2}a\alpha^{i-1}}$. This matches the requirement in Theorem A.6.

The parameters are set so that for all $i \in [s]$, $\varepsilon_i \leq \frac{\varepsilon}{2^{i+1}}$. To see this, note that $a \geq \frac{2\ln 2}{\gamma^2} + \frac{2}{\gamma^2}\ln\frac{1}{\varepsilon}$ and $\alpha - 1 \geq \frac{2\ln 2}{\gamma^2 a}$. The latter inequality implies that $\alpha^i - \alpha^{i-1} \geq \frac{2\ln 2}{\gamma^2 a}$ for all i, and so $a\alpha^{i-1} \geq \frac{2\ln 2}{\gamma^2}(i+1) + \frac{2}{\gamma^2}\ln\frac{1}{\varepsilon}$, which in turn implies that $\varepsilon_i = e^{-\frac{\gamma^2}{2}a\alpha^{i-1}} \leq \frac{\varepsilon}{2^{i+1}}$. Thus, $\sum_{i\in[s]} 2\varepsilon_i \leq \varepsilon$.

Here too, we also need to make make sure that $\alpha^i ad - k_i + 2\log\frac{1}{\varepsilon_i} \leq \alpha^{i-1}ad$. Plugging-in the expression for ε_i and collecting terms, we get that $\Delta \leq (1-2\gamma)\delta - \frac{\gamma^2}{d\ln 2}$ needs to hold. This indeed holds by our requirement on a. The output length is then

$$m_s = b_{s+1} = \alpha^s ad \geq \left(1 + \Delta \cdot \frac{t}{a}\right) ad \geq \Delta td \geq (1-2\gamma-\beta)\delta dt,$$

where the last inequality follows from our requirement on β. Finally, the seed length is given by $b_1 = ad = O_{\beta,\delta,\gamma}(d\log\frac{1}{\varepsilon})$. ■

We remark that one can also use the first item of Lemma 3.3 instead of the second one. This would allow supporting $\gamma > \frac{1}{2}$ at the expense of worse output length of roughly $m = \frac{1}{2}\delta n$.

A.4 CG Sources from Shannon Sources

Proof of Lemma 8.8: By first, by averaging argument, there are at most $\sqrt{\lambda} \cdot \lfloor t/b \rfloor$ blocks that have more than $\sqrt{\lambda}b$ bad indices for X. Call a block $1 \leq i \leq \lfloor t/b \rfloor$ good if it has less than $\sqrt{\lambda}b$ bad indices.

Fix any good block $1 \leq i \leq \lfloor t/b \rfloor$. For convenience, say $j \in [b]$ is a good step in the block if the step $b(i-1)+j$ is good. We'll show that with high probability over prefixes $X_{[1,\ldots,(i-1)b]}$, the current (good) block $X_{[(i-1)b+1,\ldots,ib]}$ conditioned on the prefix is close to high min-entropy. For convenience, let $X' = X_{[1,\ldots,(i-1)b]}$ and $X'' = X_{[(i-1)b+1,\ldots,ib]}$.

For each good $j \in [b]$, define $\mathbf{G}_j \in \{0,1\}^{db(i-1)+j-1}$ as the set prefixes to step j that are δ-good for X. Additionally, for each good j, define $W_j(x_1, \ldots, x_{ib})$ as the indicator random variable whether $x_1, \ldots, x_{b(i-1)+j-1} \notin \mathbf{G}_j$. Let $W = \sum_{\text{good } j} W_j$. By linearity of expectation:

$$\mathbb{E}_{X',X''}[W] = \mathbb{E}_{X',X''}\left[\sum_{\text{good } j} W_j\right] \leq \rho b$$

Thus by Markov:

$$\Pr_{X',X''}[W \geq \sqrt{\rho} b] \leq \sqrt{\rho}$$

Now define $\mathbf{B} \subset \{0,1\}^{db(i-1)}$ as the set of prefixes a to the current block i such that:

$$\Pr_{X''}\left[W \geq \sqrt{\rho} b \middle| X' = a\right] > \rho^{1/4}$$

By averaging, we know that $\Pr[X' \in \mathbf{B}] < \rho^{1/4}$. Now consider any prefix a outside of $\mathbf{B}$. We show that conditioned on this prefix, the distribution of X'' is close to high min-entropy.

For the rest of this proof, for convenience and brevity, we use X_j to refer to the distribution of $X_{(i-1)b+j}$ conditioned on a good fixed prefix a outside of $\mathbf{B}$. Furthermore, for the rest of this proof, all random variables, expectation statements and probability statements are implicitly conditioned on a.

Let E_1 be the event that $W \geq \sqrt{\rho} b$. Notice that since we're implicitly conditioning on a prefix a outside of $\mathbf{B}$, $\Pr[E_1] \leq \rho^{1/4}$. Let $p(x) = \Pr[X'' = x]$. Now:

$$\Pr_{x \sim X''}[p(x) \geq r] \leq \Pr_{x \sim X''}[p(x) \geq r | \overline{E_1}] + \rho^{1/4}$$

We now bound the probability of $\Pr_{x \sim X''}[p(x) \geq r | \overline{E_1}]$. Again for convenience, all notation beyond this point is implicitly conditioned on $\overline{E_1}$. Conditioned on the fact that $W < \sqrt{\rho} b$, we know there exists at least $(1 - \sqrt{\lambda})b - \sqrt{\rho} b = (1 - \sqrt{\lambda} - \sqrt{\rho})b$ steps j in the block that such that the distribution of X_j conditioned on its respective prefix has Shannon entropy at least δd.

For convenience, denote X_j as the j-th step in the current block. For each $\ell \in [(1 - \sqrt{\lambda} - \sqrt{\rho})b]$, let $J^\ell(x_1, \ldots, x_b) \in [b]$ be the random variable denoting the ℓ-th step in the block j such that $H(X_j | X_{[1,\ldots,j-1]} = x_{[1,\ldots,j-1]}) \geq \delta d$.

Define $Y_\ell(x_1, \ldots, x_b)$ as the indicator random variable that is 1 if and only if

$$\Pr[X_{J^\ell} = x_{J^\ell} | X_{[1,\ldots,J^\ell-1]} = x_{[1,\ldots,J^\ell-1]}] \geq \frac{1}{D^{\delta/3}}.$$

Let $Y = \sum_\ell Y_\ell$. First, we observe that:

$$\mathbb{E}[Y] = \sum_\ell E[Y_\ell] \leq \left(1 - \sqrt{\rho} - \sqrt{\lambda}\right) b \cdot E[Y_\ell] \leq (1 - \delta/3)\left(1 - \sqrt{\rho} - \sqrt{\lambda}\right) b \leq \left(1 - \left(\frac{\delta}{3} + \sqrt{\rho} + \sqrt{\lambda}\right)\right) b$$

Where we used Corollary 3.9 for to bound $E[Y_\ell]$. We define the Doob martingale

$$Z_\ell = \mathbb{E}[Y | X_1, \ldots, X_{J^\ell}]$$

with the convention that $Z_0 = \mathbb{E}[Y]$. Note further that $Z_b = Y$. Further, we know that $|Z_j - Z_{j-1}| \leq 1$ for all j. Thus, by the Azuma-Hoeffding inequality, we get

$$\begin{aligned}
&\Pr\left[Y - \left(1 - \left(\frac{\delta}{3} + \sqrt{\rho} + \sqrt{\lambda}\right)\right) b > \frac{1}{2}\left(\frac{\delta}{3} + \sqrt{\rho} + \sqrt{\lambda}\right) b\right] \\
&\leq \Pr\left[Z_b - Z_0 > \frac{1}{2}\left(\frac{\delta}{3} + \sqrt{\rho} + \sqrt{\lambda}\right) b\right] \\
&\leq e^{-\delta^2 b/36}
\end{aligned}$$

Finally, we observe that for any $x_1, \ldots, x_b$ s.t.

$$Y(x_1, \ldots, x_b) \leq \left(1 - \frac{\delta}{6} - \frac{\sqrt{\rho}}{2} - \frac{\sqrt{\lambda}}{2}\right) b$$

has probability at most:

$$\left(D^{-\delta/3}\right)^{(1-\sqrt{\rho}-\sqrt{\lambda})b - \left(1-\frac{\delta}{6}-\frac{\sqrt{\rho}}{2}-\frac{\sqrt{\lambda}}{2}\right)b} = D^{-\delta^2/36+\delta\sqrt{\rho}/6+\delta\sqrt{\lambda}/6}.$$

■

B Maintaining Constant Entropy Gap Throughout the Walk

Here we continue our discussion on a truly online construction that maintains constant entropy gap at all intermediate steps in the case of $\lambda = 0$.

Theorem B.1 (online condensing from almost CG sources). *For any constants $\delta, \varepsilon, \gamma > 0$, and any constant integer $d \geq 1$, there exists a pair of algorithms* $(\mathsf{Update}, \mathsf{Output})$ *such that the following holds.*

- Update *takes as input a state string $s \in \{0,1\}^\star$, and $x \in \{0,1\}^d$, and outputs in time polynomial in $|s|$ a new state string s' with $|s'| \leq |s| + O(1)$.*
- Output *takes as input a state string $s \in \{0,1\}^\star$ and outputs a string* $\mathsf{Output}(s)$ *in time $O(|s|)$.*
- *For any γ-almost δ-CG source $X = X_1 \circ \cdots \circ X_t$ with $X_i \sim \{0,1\}^d$, for the sequence of states $s_i =$* $\mathsf{Update}(s_{i-1}, X_i)$ *(for some fixed s_0), for every i,* $\mathsf{Output}(s_i)$ *is a source on $\Omega(\delta d i)$ bits with constant entropy gap.*

We now briefly discuss why we can support the above theorem given our techniques.

The State. The state s will contain the following information.

- A vertex in a "big graph" G_i from [CRVW02]. G_0 will be of some appropriate constant size. Note that we never store the entire graph, simply an encoding of a vertex within the graph. The vertex is initialized arbitrarily.
- A vertex in a "small graph" H that is an optimal lossless expander of appropriate constant size found via brute force. The vertex is initialized arbitrarily.

- A "buffer" of constant length containing the sequence of the most recently seen X_i-s. The buffer is initialized to be empty.
- A counter of the number of steps taken in H.
- A counter of the number of steps taken in G_i.

The Update Function. Given a current state, the function Update will behave as follows.

- We append the current X_i to the buffer.
- If the sequence of elements in the buffer is long enough to define a step in H, we take a step in H according the sequence, and reset the buffer to empty.
- If the number of (recent) steps in H is enough to ensure that the distribution on H has ample entropy, we use the vertex position of H to take a step in G_i. We reset the position in H to an arbitrary fixed vertex.
- If the number of (recent) steps in G_i is enough to ensure that the distribution on G_i is saturated with entropy, we append an appropriate constant number of 0-s to the encoding of the vertex in G_i. We then treat this as an encoding of a vertex in a [CRVW02] graph G_{i+1} that is a constant times larger than G_i, but with the same degree.

The Output Function. The function Output will simply return the current vertex v in G_i.

The fact that the current vertex distribution always has constant entropy gap follows from the fact that the current graph G_i is never more than a constant times larger than the current entropy. The easiest way to view the embedding process in the last part of the update function is via the lossless conductor view of the [CRVW02] construction. We take a step in the larger G_{i+1} by applying the lossless conductor with the vertex encoding as the source. Clearly the operation of appending 0-s preserves the q-norm of the distribution and so, by Theorem 5, each step in any G_i gradually increases the entropy in the vertex encoding.

C The Construction's Runtime

Our construction is explicit, which is evident by our use of explicit ingredients. But since we also care about fast simualtion via almost CG sources, it is appealing to determine the runtime more accurately.

Claim C.1 (condenser runtime). *Given $t, d, \delta, \gamma, \lambda$, let* $\mathsf{Cond}\colon \{0,1\}^{dt} \to \{0,1\}^m$ *be the condenser from Theorem 5.3. Then, given $x \sim X$, where X is a (γ, λ)-almost δ-CG source over $n = dt$ bits, we can compute* $\mathsf{Cond}(x)$ *in time $\widetilde{O}(n^2)$. (In the TM model.)*

Proof: The construction amounts to taking steps on a constant-sized H (which can be written down in constant time) and the CRVW graph G, over $\{0,1\}^m$, where $m < n$. Thus, we can bound the runtime of Cond by $n \cdot T$, for T being the time it takes to compute Γ_G, the neighborhood function of G.

Inspecting the construction of [CRVW02] for the balanced case, we see that the only non constant-sized object that is being applied is a *permutation conductor*, which can be implemented by taking a step on a constant-degree spectral expander, say a Ramanujan graph Γ over $U = 2^{m-O(1)}$ vertices. For concreteness, we take the LPS Ramanujan graph [LPS88]. Each vertex in Γ is indexed by a 2×2 matrix over a prime field $\mathbb{F}_q$ of cardinality $O(U)$.[34] The graph Γ is a Cayley graph with a set of generators that can be precomputed in linear time. Taking a single step over Γ amounts to matrix multiplication, which then amounts to performing a constant number of field operations. The bit complexity of addition and multiplication in $\mathbb{F}_q$ is $\widetilde{O}(\log q) = \widetilde{O}(n)$. Overall, computing Cond takes $\widetilde{O}(n^2)$ time. ■

Clearly, the same holds for condensing from Shannon entropy (Theorem 6.3) and the suffix-friendly analogues.

To establish the runtime of extraction, we simply need to account for the time it takes to compute the [GW97] extractor.

Claim C.2 (extractor runtime)**.** *The extractors from Section 7, set to extract with constant error $\varepsilon > 0$, run in time $\widetilde{O}(n^2)$, where n is the length of the corresponding source. (In the TM model.)*

Proof: To extract from the condensed output, we apply the [GW97] extractor from Theorem 2.12 on input of length m and constant-length seed. Very roughly, computing the GW extractor amounts to taking a length-$O(\log(1/\varepsilon))$ walk over a spectral expander, followed by an application of a two-universal family of hash functions. As we consider the constant error regime, this can be done in time $\widetilde{O}(n)$. Thus, the condenser's runtime is the dominant factor. We refer the reader to the appendix of [DMOZ20] for a detailed review of the [GW97] construction. ■

Runtime in the RAM Model. The runtime analysis in the above two claims was done in the standard Turing machine model. However, for applications, we often care more about the RAM model, in which we assume that we perform arithmetic operations in $\mathbb{F}_q$ at unit cost, even when q is exponential in n (note that it takes $\log q$ bits to store a field element, and in the RAM model we assume this is the word size). When each field operations takes constant time, it is easy to verify both our condensers and extractors run in time *linear* in n.

D An Analog of Unique Neighbor Expansion

The following definition is a generalization of the notion of "unique neighbor expansion" in expander graphs. Although the notion is not necessary for our result, we believe it is interesting in its own right.

Definition D.1. *Let $G = (U, V, E)$. Let $w \colon U \to [0, \infty)$ be a weight function on U. We define $\mathcal{N}_{\mathsf{diff}}(w) \colon V \to \mathbb{R}$ as:*

$$\mathcal{N}_{\mathsf{diff}}(w)(v) = w(u_v) - \sum_{u \in \Gamma(v) \setminus u_v} w(u)$$

We will also use $|\mathcal{N}_{\mathsf{diff}}(w)|$ to denote $\sum_{v \in V} \mathcal{N}_{\mathsf{diff}}(w)(v)$. Note that here, the meaning of $|\cdot|$ is different than the standard ℓ_1-norm of the weight function. Some terms, and thus even the whole sum, can be negative.

[34] The LPS construction works only for certain primes, but we can handle this without significant loss in parameters.

We can then show that a generalized notion of unique neighbor expansion holds under this definition.

Lemma D.2. *Let $G = (U, V, E)$ be a (K, ε, e)-expander for any e. For all weight functions $w \colon U \to [0, \infty)$ supported on at most K nodes, it holds that*

$$|\mathcal{N}_{\mathsf{diff}}(w)| \geq (1 - 2\varepsilon)D|w|.$$

Proof: Observe that

$$D|w| = \sum_{v \in V} \left(w(u_v) + \sum_{u \in \Gamma(v), u \neq u_v} w(u) \right) = |\mathcal{N}(w)| + \sum_{v \in V} \sum_{u \in \Gamma(v), u \neq u_v} w(u).$$

Thus, we have

$$D|w| - |\mathcal{N}(w)| = \sum_{v \in V} \sum_{u \in \Gamma(v), u \neq u_v} w(u).$$

Since $|\mathcal{N}(w)| \geq (D - \varepsilon D)|w|$, we have

$$\varepsilon D|w| \geq \sum_{v \in V} \sum_{u \in \Gamma(v), u \neq u_v} w(u).$$

Finally,

$$\sum_{v \in V} \mathcal{N}_{\mathsf{diff}}(w)(v) = D|w| - 2 \sum_{v \in V} \sum_{u \in \Gamma(v), u \neq u_v} w(u) \geq (D - 2\varepsilon D)|w|.$$

■