Online Condensing of Unpredictable Sources via Random Walks

Abstract

A natural model of a source of randomness consists of a long stream of symbols $X = X_1 \circ \ldots \circ X_t$, with some guarantee on the entropy of X_i conditioned on the outcome of the prefix $x_1, \ldots, x_{i-1}$. We study *unpredictable* sources, a generalization of the almost Chor–Goldreich (CG) sources considered in [DMOZ23]. In an unpredictable source X, for a typical draw of $x \sim X$, for most i-s, the element x_i has a low probability of occurring given $x_1, \ldots, x_{i-1}$. Such a model relaxes the often unrealistic assumption of a CG source that for *every* i, and *every* $x_1, \ldots, x_{i-1}$, the next symbol X_i has sufficiently large entropy. Unpredictable sources subsume all previously considered notions of almost CG sources, including notions that [DMOZ23] failed to analyze, and including those that are equivalent to general sources with high min entropy.

For a lossless expander $G = (V, E)$ with $m = \log |V|$, we consider a random walk $V_0, V_1, \ldots, V_t$ on G using unpredictable instructions that have sufficient entropy with respect to m. Our main theorem is that for almost all the steps $t/2 \leq i \leq t$ in the walk, the vertex V_i is close to a distribution with min-entropy at least $m - O(1)$.

As a result, we obtain seeded *online* condensers with constant entropy gap, and seedless (deterministic) condensers outputting a constant fraction of the entropy. In particular, our condensers run in space comparable to the output entropy, as opposed to the size of the stream, and even when the length t of the stream is not known ahead of time. As another corollary, we obtain a new extractor based on expander random walks handling lower entropy than the classic expander based construction relying on spectral techniques [Gil98].

As our main technical tool, we provide a novel analysis covering a key case of adversarial random walks on lossless expanders that [DMOZ23] fails to address. As part of the analysis, we provide a "chain rule for vertex probabilities". The standard chain rule states that for every $x \sim X$ and i, $\Pr(x_1, \ldots, x_i) = \Pr[X_i = x_i | X_{[1,i-1]} = x_1, \ldots, x_{i-1}] \cdot \Pr(x_1, \ldots, x_{i-1})$. If $W(x_1, \ldots, x_i)$ is the vertex reached using $x_1, \ldots, x_i$, then the chain rule for vertex probabilities essentially states that the same phenomena occurs for a typical x:

$$\Pr[V_i = W(x_1, \ldots, x_i)] \lesssim \Pr[X_i = x_i | X_{[1,i-1]} = x_1, \ldots, x_{i-1}] \cdot \Pr[V_{i-1} = W(x_1, \ldots, x_{i-1})],$$

where V_i is the vertex distribution of the random walk at step i using X.

ISSN 1433-8092

Contents

1 Introduction

Randomness is an extremely useful and ubiquitous tool in computer science. Algorithms, protocols and reductions often assume access to uniformly distributed bits. An inherent question is what kind of randomness we can reasonably obtain from nature (or engineering), and whether we can make such randomness as useful as uniform bits. This has spawned a long line of research with many deep and interesting results.

Random walks and their analysis are also an essential tool in computer science, as well as in mathematics and physics. It is natural, therefore, to ask

> How do random walks behave when the instructions for each step are not truly uniform and independent?

Such a scenario occurs when the instructions come from a weak source of randomness. A common assumption about a weak source is that overall, it has min-entropy.[1] The hope is that the quality of the vertex distribution of the random walk is much better, and more useful, than that of the original source.

Gillman's Chernoff bound for random walks on expanders [Gil98] implies that most nodes on an expander random walk are close to uniform for any source with entropy rate very close to 1 (see, e.g., [Zuc07]). However, random walks cannot mix for general rate $1/2$ sources, since an adversary that controls half the steps can have the even numbered steps undo the odd numbered steps. It's therefore interesting to ask if any structure in the source can enable random walks to mix for lower entropy rates. The first paper to address this question was [DMOZ23], who showed that for certain low-rate sources random walks do mix well.

Successful analyses of such random walks give clean constructions of extractors and condensers, which purify the randomness in a weak source. Specifically, an *extractor* is a function $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ that uses an independent and uniform ℓ-bit seed Y to convert X into a distribution $\mathsf{Ext}(X,Y)$ that is statistically close to uniform. A *condenser* is slightly weaker: It's a function $\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ that converts X into a distribution with high entropy rate. If $\mathsf{Cond}(X,Y)$ is close to a k'-source, then a successful condensing means that $k'/m \gg k/n$. The ℓ-bit string in both cases is called the *seed*. In certain cases, seedless (deterministic) extraction and condensing, where $\ell = 0$, is also possible.

A natural family of weak sources is one where X is a long stream of short symbols, $X = X_1 \circ \ldots \circ X_t \sim \{0,1\}^{dt=n}$, with each symbol being revealed one at a time. Indeed, historically, some of the first definitions of weak sources [SV86, CG88] were streaming models. Similarly, a very natural question is how well a random walk mixes when one uses the stream of short symbols as instructions. The streaming model of randomness also corresponds to common sources of randomness in practice. Probably the most popular sources of entropy involve the exact timing of interrupts from mouse movements, keyboard strokes, disk I/O, receiving network packets, and other unpredictable events. Other sources include thermal noise and repeatedly looking at the last few digits of a clock timed according to an independent clock.

To model such streaming sources, one needs some property that implies that each $X_i \sim \{0,1\}^d$ has some entropy, even conditioned on the previously observed $x_1, \ldots, x_{i-1}$ (we often abbreviate this as $x_{[1,i-1]}$). Commonly studied notions such as sequences of independent sources,

[1] We say $X \sim \{0,1\}^n$ has min-entropy k, $H_\infty(X) \geq k$, if $\max_{x \sim X} \Pr[X = x] \leq 2^{-k}$. We call such an X a k-source and k/n the entropy rate.

Santha–Vazirani (SV) sources [SV86], and Chor–Goldreich (CG) sources [CG88] indeed have these properties. A δ-CG source is a sequence of random blocks $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, such that for any i and any prefix $a \in \{0,1\}^{d(i-1)}$, it holds that $H_\infty(X_i | X_{[1,i-1]} = a) \geq \delta d$. A previous work [DMOZ23] shows that random walks using CG sources can in fact mix, and obtains excellent deterministic condensers as a result.

One distinct advantage of random walk based extractors and condensers is that they are readily *online*. That is, it is not necessary to know ahead of time how long the stream $X_1, \ldots, X_t$ is, and nevertheless, the procedure can utilize most of the total entropy k within, by processing each symbol of the stream sequentially in a read-once fashion, and in space comparable to the amount of entropy m we need, as opposed to the length of the stream t. We emphasize that the notion of online extracting and condensing goes hand in hand with streaming models of randomness such as CG sources, which in turn goes hand in hand with random walks. Moreover, aside from a few works [DGSX21a, DGSX21b, DMOZ23], online constructions for various types of sources are scarce.

Unfortunately, the assumptions of a CG source are quite strong, and may be unrealistic in practice. In a CG source, *no matter the outcome of the previous symbols* $x_1, \ldots, x_i$, it must be the case that the next symbol X_i has high entropy. In some sense, such a definition asserts that the randomness stream can contain no "errors" of a certain type. Although [DMOZ23] effectively analyzes certain kinds of errors, for others, it completely fails.

Our work continues the line of inquiry in [DMOZ23] by generalizing the notion of CG sources and asking whether online condensing of randomness streams is possible even in the presence of errors. We give a novel analysis of random walks using a stream of symbols, which may be of independent interest, showing that mixing is possible even in the presence of a very general notion of errors. Thus, we give online extractors and condensers for a more general and practical class of sources. The class of randomness streams we consider are what we call *unpredictable* sources.

Definition 1.1 (unpredictable source, simplified). *We say that $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, is a (δ, ρ) (simplified) unpredictable source, if, for every i, with probability at least $1 - \rho$ over $\sigma \sim X$, it holds that $\Pr[X_i = \sigma_i | X_{[1,i-1]} = \sigma_{[1,i-1]}] \leq 2^{-\delta d}$.*

In words, in an unpredictable source, for every i, with high probability over a sample $\sigma \sim X$, the next symbol σ_i is unlikely, conditioned on its prefix. This notion forgoes the often demanding assumption of CG sources: It is no longer true that X_i has high entropy regardless of the outcome of the previous symbols. (For another simplified instantiation of unpredictable sources, see Definition 1.5.)

We think of ρ as the error parameter of the source. Intuitively, it represents the probability of seeing a low-entropy step. As we discuss later on, the analysis of [DMOZ23] completely fails on unpredictable sources, and so online condensing of such sources is unknown. Our results even hold for a more general notion of unpredictable source, where only the average "error" over i needs to be small.

Definition 1.2 (unpredictable source). *We say that $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, is a (δ, ρ) unpredictable source, if, when defining*

$$\rho_i(X, \delta) = \Pr_{x \sim X}\left[\Pr\left[X_i = x_i \mid X_{[1,i-1]} = x_{[1,i-1]}\right] > 2^{-\delta d}\right],$$

it holds that $\mathbb{E}_{i \sim [t]}[\rho_i(X, \delta)] \leq \rho$.

We present and study this definition for two main reasons. The first is that it seems natural. In particular, it elegantly generalizes all previous notions of almost CG source considered in [DMOZ23] (see Definition 1.5). The second is that, perhaps surprisingly, this notion is weak enough to capture arbitrary high-entropy sources. Indeed, not only are unpredictable sources close to having high min-entropy, but the converse is true as well: Arbitrary $(1-\rho)n$-sources are $(1/2, 2\rho)$ unpredictable sources! (see Proposition 6.4).

The crux of our result is an analysis that shows that random walks using unpredictable sources can "mix" sufficiently well.

Theorem 1 (main (informal); see Theorem 3.14 and Corollary 3.16)**.** *Let $\delta, \rho > 0$ and $D = 2^d > 1$ be constants. Let $X = X_1 \circ \cdots \circ X_t$, $X_i \sim \{0,1\}^d$, be a (δ, ρ) unpredictable source. Let G be a sufficiently good D-regular lossless expander on an appropriately chosen number of vertices $M = 2^m$.*[2]

Suppose that the X_i-s are used as instructions for a random walk on G from an arbitrary starting vertex. Given $x \sim X$, let $W_i(x)$ denote the vertex reached in the i-th step using x as instructions, and let $Z_i(x) = -\log \Pr_{x\sim X}[W_i(X) = W_i(x)]$. Then,

$$\Pr_{i\in[t/2,t], x\sim X}[Z_i(x) < m - O(1)] \leq O(\rho)$$

In words, for most steps in the second half of the random walk, most of the vertices reached are unlikely. Thus, for a random i, $W_i(X)$ is close to an extremely high entropy distribution, namely one with constant entropy gap! We emphasize again that the analysis requires a new technique, since the analysis in [DMOZ23] fails to give anything useful for unpredictable sources. Moreover, this works for low entropy rate, unlike the high-entropy result that follows from Gillman's Chernoff bound for random walks. Constructions of online condensers and extractors follow as corollaries to this analysis, and we discuss them next.

1.1 Online Extracting and Condensing

Due to the streaming nature of our source X, one would like a condenser (or extractor) for such sources to process each symbol sequentially as it is received. The notion of online condensing achieves exactly this. As in the model from Dodis, Guo, Stephens-Davidowitz, and Xie [DGSX21a, DGSX21b], in (deterministic) online condensing, the function Cond is implemented by a procedure that starts in a state S_0, and makes a sequence of calls to an update procedure

$$S_{i+1} \leftarrow \mathsf{Update}(S_i, X_{i+1}).$$

The length of each state S_i should be not much larger than the final output length m. The procedure may then output the final state $S_t \in \{0,1\}^m$ (or perhaps some function of S_t).

We clarify that the question of whether online condensing is possible is entirely orthogonal to the question of whether *seeded* condensing is possible. Indeed, a natural question to ask is whether (and how well) can general weak sources be condensed in an online manner. In the case of seeded online condensing (and in some regimes, we provably must use a seed), one may consider an update procedure that also takes as input a seed $Y \in \{0,1\}^\ell$ that is independent of the stream X,

[2] See Section 1.3 for a discussion about lossless expanders. $M = M(d, t, \delta, \rho)$ is chosen to be up to roughly 2^k, for k being the (smooth) min-entropy of X.

and computes $S_{i+1} \leftarrow \mathsf{Update}(S_i, X_{i+1}, Y)$.[3] The guarantee of such a Cond should be that at any point in the stream i, the state S_i should contain most of the entropy seen so far in $X_1, \ldots, X_i$ (or about $|S_i|$, if this length is smaller), and the length of Y relative to i should be small. The hope is that the update procedure accumulates the additional entropy from X_i in each step into the state S_i, and thus the final state S_t contains most of the entropy in X (or about the length of $|S_t|$, if this length is smaller). As noted earlier, the advantages of such an online model of condensing are that it allows one to utilize the entropy that was overall contained in the entire stream X, even if one does not know the length of the stream t ahead of time. Moreover, when one only wishes to get m bits of entropy out of a very long stream of length $t \gg m$. An online construction would use space m rather than t.

1.2 Online Condensing via Random Walks

In [DMOZ23], they showed that a natural way to condense a randomness stream is to use its symbols as instructions for a random walk over an expander G, starting from an arbitrary fixed vertex (similarly to Theorem 1). The intuition is that if a step in a random walk makes progress towards mixing, then that step accumulates the entropy from the instruction into the vertex distribution. Thus, using the current vertex in the walk as our "state" yields a (deterministic) condenser with output length $m = \log M$, where M is the number of vertices of G.[4]

In this work, we are primarily focused on sequences $X_1, \ldots, X_t$ that may not mix at *every* step, as is the case for unpredictable sources. We now give broad intuition on how such erroneous steps affect condensing. Suppose that a symbol X_i is highly correlated with the previous instructions $x_1, \ldots, x_{i-1}$. Also, assume that the vertex distribution at step $i-1$ is uniform on some set S of size K. If G is a D-regular graph, then an adversarially chosen X_i may cause the walk to "consolidate" the vertices of S into groups of size D. This would result in a vertex distribution that is uniform on a set of size K/D, and hence $d = \log D$ bits of entropy were lost. In general, one can show that this is the worst that can happen, and so if there are very few bad steps overall, then overwhelmingly, mixing, and thus condensing, indeed occurs. Realizing this intuition for unpredictable sources poses several challenges, that we discuss further in Section 1.5.

1.3 CG-Sources, Lossless Expanders, and the [DMOZ23] Condenser

Towards discussing unpredictable sources, let us first review in more detail the previous work, [DMOZ23], on condensers via random walks. Both here, and in [DMOZ23], we study random walks on *lossless expanders*. A degree-D (K, ε)-lossless expander on M vertices is an undirected graph G such that for any set $S \subseteq V$, $|S| \leq K$, the size of the neighborhood $\Gamma(S)$ satisfies $|\Gamma(S)| \geq$

[3] Since the seed length ℓ typically depends on t, when the length of the stream t is not known in advance, one can model the use of a uniform seed by also viewing it as a stream of uniform and independent bits. For example, the seed can be initialized to the empty string (or some constant length uniform string), and Cond may call, in conjunction with each update, an additional procedure, $Y \leftarrow \mathsf{ExtendSeed}(S_i, Y)$, that may choose to increase the length of Y by one or several bits, depending on the current state S_i. Generally, the choice to extend Y will depend on how many symbols Cond has seen so far, which would be stored in the state.

[4] One may notice that committing to a graph of size M may be problematic when the length of the stream is not known ahead of time, as one would like m to be comparable to t. This can be fixed with a trick of repeatedly increasing the size of the graph at regular intervals. We discuss this more in Section 1.5.1, and for most of the introduction, we will assume that t is known ahead of time.

$(1-\varepsilon)D|S|$.[5] [DMOZ23] proved that a random walk using X, starting from an arbitrary fixed vertex, on a sufficiently good lossless expander, accumulates entropy and thus yields a deterministic condenser.

Theorem 1.3 ([DMOZ23], informal)**.** *Let δ, η be constants. Suppose that for every $M = 2^m$, there exists an explicitly computable $D = 2^d$-regular (K, ε)-lossless expander on M vertices, with $\varepsilon \ll D^{-(1-\delta)}$, and $K = M/\mathrm{poly}(D)$. Then, 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 a δ-CG source X, $\mathsf{Cond}(X)$ is η-close to an $m - O(\log(1/\eta))$-source. Moreover, Cond can be computed in an online manner.[6]

That is, Cond condenses X to within a *constant entropy gap*, where we say that the entropy gap (with error ε) of a distribution $Z \sim \{0,1\}^m$ is Δ if Z is ε-close to an $(m-\Delta)$-source.

But in practice, it may be unreasonable to assert that for *every* i and *every* prefix $x_1, \ldots, x_i$, the next symbol X_i is highly unpredictable. To address this, [DMOZ23] does consider generalized versions of CG sources, although the guarantee about $\mathsf{Cond}(X)$ from Theorem 1.3 for such sources is much weaker there. In particular, the situation (before this work) becomes quite bleak when introducing the generalization coined ρ-error in [DMOZ23].

Definition 1.4 (ρ-almost CG source)**.** *A ρ-almost δ-CG source is a sequence of random variables $X = X_1 \circ \ldots \circ X_t$ with $X_i \sim \{0,1\}^d$, such that for each $i \in [t]$, we have that for at least probability $1-\rho$ over the prefix $a \in \{0,1\}^{d(i-1)} \sim X_1 \circ \ldots \circ X_{i-1}$, it holds that $H_\infty(X_i | X_{[1,i-1]} = a) \geq \delta d$.*

It turns out that in the presence of ρ-error (together with other error types discussed shortly), general min-entropy sources are almost CG sources in some regime of parameters (see [DMOZ23, Section 8]). Thus, in general, deterministic condensing to within a constant entropy gap of such sources is impossible. Moreover, before this work, it was unknown whether or not a random walk using ρ-almost CG sources mix well in any sense at all. In this paper we show that it is indeed the case that *random walks mix, even under the more general notion of unpredictable sources*, which captures all previously considered generalizations of CG sources.

1.4 Unpredictable Sources

As discussed previously, the notion of CG sources is quite strong – it assumes that *every* prefix leads to a high entropy distribution is quite strong. An unpredictable source does not make such an assumption. Indeed, notice that in the definition of an unpredictable source, we do not directly insist on any guarantee on the (smooth) min-entropy of the distribution $X_i | X_{[1,i-1]} = x_{[1,i-1]}$, only that usually, the next symbol x_i is unlikely conditioned on $x_{[1,i-1]}$. We give unpredictable sources their name as it closely resembles the intermediate objects of the same name that show up in pseudorandom constructions such as *reconstructive extractors* [Tre01, SU05, TZS06] (for the precise definition of reconstructive extractors, see, e.g., [TU06]).

When talking about (δ, ρ) unpredictable sources, we informally refer to δ as the "entropy rate" of the unpredictable source, and ρ as its "error rate".[7] It is easy to see that this definition captures

[5] More technically, we consider bipartite graphs $([M],[M],E)$, as these are what the known explicit constructions yield. However, at a high level, this distinction is not necessary.

[6] [DMOZ23], also handled *almost* δ-CG sources. In this definition, instead of each X_i being a δd-source for every prefix, each X_i is only γ-close to being a δd-source.

[7] An (δ, ρ) unpredictable source is indeed, roughly, ρ-close to an δn sources.

almost-CG sources with all previously considered error parameters.

Definition 1.5 (almost CG source with all error parameters)**.** *We say that $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, is a $(\delta, \gamma, \rho, \lambda)$ almost CG source, if, for at least $(1-\lambda)$ fraction of $i \in [t]$, the following holds:*

$$\Pr_{x \sim X}\left[H_\infty^\gamma\left(X_i | \left\{X_{[1,i-1]} = x_{[1,i-1]}\right\}\right) < \delta d\right] \leq \rho.$$

Indeed, a $(\delta, \gamma, \rho, \lambda)$ almost CG source is a $(\delta, \gamma + \rho + \lambda)$ unpredictable source.[8] Additionally, as was the case for almost CG sources with all error parameters, every general source with $(1-\rho)n$ min entropy is an unpredictable source, with a much more straightforward argument, with fewer constraints on ρ, and with less loss in converting ρ into the error parameters of an almost CG source (see Proposition 6.4). In this work, we give the following result for unpredictable sources, analogous to Theorem 1.3:

Theorem 2 (seeded condensing (informal); see Corollary 3.16)**.** *Let δ, ρ be constants, and let D and ε be constants that satisfy $\varepsilon < D^{-(1-\delta)}$. Suppose that for every M, there exists an explicitly computable $D = 2^d$-regular (K, ε)-lossless expander on $\Theta(M)$ vertices, with $K = M/\mathrm{poly}(D)$. Then, for any positive integer t, there exists an explicit function*

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

such that given a (δ, ρ) unpredictable source X and an independent and uniform Y, $\mathsf{Cond}(X,Y)$ is $O\left(\frac{1}{\delta}(\varepsilon D^{1-\delta} + \rho)\right)$-close to an $m - O(\log(1/\varepsilon))$ source. Moreover, Cond can be computed in an online manner.[9]

As suggested by prior discussion, the construction is again to simply use X as instructions for a random walk on a lossless expander, with the random seed indicating the stopping time. In order to show that such a walk mixes, we develop a new analysis, different than that of [DMOZ23], which we discuss in detail in Section 1.5.

Let us briefly discuss the error term in the theorem's statement. First, roughly speaking, the term $\varepsilon D^{1-\delta}$ is the probability that a set S "does not expand" in some sense. For example, $\varepsilon D^{1-\delta}$ is the probability over a uniformly chosen vertex $v \in S$ and uniform neighbor $\Gamma(v)$, that $\Gamma(v)$ has another neighbor in S. Broadly speaking, events such as this are undesirable: they represent "collisions" of paths in the random walk. Thus we expect such an error term, as it corresponds to the (inherent) error of the expander. We should also expect the error rate ρ to appear for the same reason: this corresponds to the probability that "expansion does not occur" due to low quality randomness (as opposed to the expander's error). Indeed, if one considers the $(1, \rho)$ unpredictable distribution X that is 0^n with probability ρ, and uniform otherwise, one can see that at every step of a random walk using X, some vertex will have probability mass at least ρ.

Finally, we comment on the relationship between these parameters. In general, as ρ is an error probability over the entire space of X, it is possible for it to be sub-constant. While ρ can be subconstant, the constant-ρ regime is more interesting, and our theorem handles the case when ρ is in fact fairly large, for example $\rho \geq D^{-\delta}$. Thus, the error term can be thought of as $O(\rho/\delta)$, and

[8] It is also true that the converse holds via several averaging arguments, although with a large loss in parameters. We prefer to study and phrase our results for unpredictable sources, as the statements are clean, and with minimal artifacts of analysis.

[9] See Appendix B for a more detailed description of the online version.

in fact it is necessary that $\rho < \delta$. Intuitively, using a (δ, ρ) unpredictable source, in a typical run of the random walk, one expects there to be roughly t steps each accumulating δd bits of entropy (for a total of δdt entropy gained), and roughly ρt steps that lose d bits of entropy (for a total of ρdt entropy lost). Thus overall, one should not expect anything good to happen when the entropy rate of the unpredictable source is smaller than the error rate.

We can also show that even without a random stopping time, we can use our new analysis of random walks to get a result about deterministic condensing, although, as expected, the entropy gap is not constant.

Theorem 3 (seedless condensing (informal); see Corollary 3.17)**.** *Let δ, ρ be constants and let D and ε be constants that satisfy $\varepsilon < D^{-(1-\delta)}$. Suppose that for every M, there exists an explicitly computable $D = 2^d$-regular (K, ε)-lossless expander on $\Theta(M)$ vertices, with $K = M/\text{poly}(D)$. Then, 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 a (δ, ρ) unpredictable source X, $\mathsf{Cond}(X)$ is $O\left(\sqrt{\varepsilon D^{1-\delta} + \rho}\right)$-close to a $(1-\beta)m$-source, where $\beta = O\left(\frac{1}{\delta} \cdot \sqrt{\varepsilon D^{1-\delta} + \rho}\right)$. Moreover, Cond can be computed in an online manner.

Overall, Theorem 2 and Theorem 3 indicate that it is indeed possible to condense a very general class of sources in an online manner.

On Extracting from Unpredictable Sources. As a final note for this section, recall that we cannot hope to extract from arbitrary unpredictable sources without a seed. However, even if one only cares about extracting, rather than condensing, from unpredictable sources, our work is the *first to do so in an online manner*: Indeed, one can use Theorem 2 to within constant entropy gap, and then apply a known online construction for constant entropy gap (see Theorem 2.12). We stress that known constructions for arbitrary weak sources with linear entropy rate, such as [Zuc97, Zuc07], are not online, and thus are unsatisfying for *streams* of randomness.

1.4.1 A Two-Stage Construction, and Recent Developments in Lossless Expanders

An expert reader may notice that the statements of Theorem 1.3, Theorem 2, and Theorem 3 are slightly weaker than what is actually achievable. In each of these theorems, we require explicit expanders with $\varepsilon \ll D^{-(1-\delta)}$. In other words, as the entropy rate δ of the source gets smaller, the error of the expander that we use must improve. Optimal, non-explicit expanders (as well as random ones) can achieve a dependence of $\varepsilon \approx 1/D$ and would thus allow us to handle any constant entropy rate δ. However, the [CRVW02] explicit construction only achieves $\varepsilon \approx D^{1/6}$, and more recent works can improve this to $\varepsilon \approx D^{1/2}$ [CRT23, Gol24]. Thus, even considering recent improvements, Theorem 1.3, Theorem 2, and Theorem 3 can only support entropy rate $\delta > 1/2$.

Fortunately, explicit optimal constructions are not necessary. A trick from [DMOZ23], the *two-stage construction*, utilizes constant-sized optimal expanders (found by brute force) to condense small blocks of the stream $X_1, \ldots, X_t$ into a higher entropy rate $\delta' > \delta$ larger blocks, that is high enough to use the known (suboptimal) explicit constructions. For details of the construction for unpredictable sources, see Section 4.

Nevertheless, we choose to present our results and phrase our theorems assuming optimal expanders for several reasons. The first is that the parameters are better, aesthetically simpler, and easier to analyze when no two stage construction is required. Moreover, we wish to highlight that the novelty of this work is the analysis of random walks on a single expander, without the trick of the two-stage construction. Finally, because of the two-stage construction, the current lack of better constructions of explicit expanders is not an inherent barrier to the plausibility of explicit condensing for smaller δ.

We give instantiations of Theorem 2, and Theorem 3 for any entropy rate $\delta > 0$ in Theorem 5.3 and Theorem 5.5, using currently known explicit expanders and the two-stage construction. As the statement is relevant for the next discussion, we give an informal version of the latter here, which considers deterministic condensing.

Theorem 4 (informal; see Theorem 5.5)**.** *Let $\delta > 0$ be any constant, let $d \geq \mathrm{poly}(1/\delta)$, and $\rho \leq \mathrm{poly}(\delta)$. Then, 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 (δ, ρ) unpredictable source $X = X_1 \circ \cdots \circ X_t$ with each $X_i \sim \{0,1\}^d$, $\mathsf{Cond}(X)$ is $\approx \rho^{1/C}$ close to a $(1-\beta)m$, source, where $\beta \approx \rho^{1/C}$, for some universal constant C.

1.4.2 Perspective: Condensing from Unpredictable Sources vs. General Sources

Having presented our main results about *seeded* condensing to within constant entropy gap, and *deterministic* condensing outputting a constant fraction of the entropy, we provide a few observations about the nature of unpredictable sources.

First, we know that general sources are unpredictable sources in the high entropy regime (see Proposition 6.4). Indeed, if $H(X) \geq (1-\rho)n$, then X is already a $(\delta = 0.99, 100\rho)$ unpredictable source. Thus, essentially for any δ, nontrivial condensing requires seed, and we provide a simple such condenser that is even online.

However, this does *not* imply that for every δ, an unpredictable source is as hard to deal with as a general source of the same entropy rate. Indeed, Theorem 4 shows that for small entropy rate δ, deterministic condensing is possible to with output entropy rate roughly $1 - \rho^{1/C}$. Thus, one can deterministically condense unpredictable sources from entropy rate 0.01 to entropy rate 0.99. Such a feat is not possible for general sources! Indeed, if X is a general source with entropy rate, say, 0.6, a simple argument shows that it cannot be deterministically condensed to entropy rate, say, 0.7 (see Claim A.1). This suggests an interesting property about unpredictable sources: Deterministic condensing "past the entropy rate" of such sources is easy, while the hard part is condensing "past the error rate." In particular, although our analysis only achieves an output entropy rate of $1 - \rho^{1/C}$, we suspect that deterministic condensing to an entropy rate of $1 - O(\rho)$ is possible. Moreover, there is good reason to believe that there is a barrier to condensing past this entropy rate: When considering general $(1-\rho)n$ sources, the entropy gap ρ becomes the error rate when thinking of it as an unpredictable source.

We end the discussion by leaving as an open line of inquiry to determine the exact threshold of the output entropy rate of deterministic condensers.

1.5 Technical Overview: Random Walks Using Unpredictable Sources

We are now ready to present a technical overview of how we analyze random walks via unpredictable sources. Since unpredictable sources subsume CG sources, we'll start with discussing the challenges inherent to both sources, and the previous solution for CG sources. An initial observation is that for both types of sources, spectral analysis fails, and for two main reasons. The first one is that spectral expanders may not be lossless, and even the best spectral expanders may only be $(K, \varepsilon = 1/2)$-lossless. Such expansion is insufficient for us, as it intuitively means that for a distribution on a set of vertices S, at least half of all edges leaving S may lead to collisions (with perhaps D other nodes in S). Since, as discussed before, collisions imply a loss in entropy, even the good high entropy steps fail to mix, unless the steps were almost uniform.

The second reason is that random walks using CG sources and unpredictable sources are *non-Markovian*: The distribution of the next step depends on the entire history of the walk up until that point. Therefore, we cannot analyze the evolution of the vertex distribution at each step by repeatedly applying a transition matrix and bounding the norm of the corresponding probability vector. Moreover, the distribution of the next instruction X_{i+1}, given a prefix $x_1, \dots, x_i$, can be *adversarial*. That is, whatever the vertex distribution p_i may be for each i, X_{i+1} could be the worst possible edge distribution that yields the least amount of improvement for p_{i+1} (while still satisfying the overall conditions on the source X).

Nevertheless, [DMOZ23] provides a direct analysis that shows that the norm of the vertex distribution does evolve favorably over time. Specifically, they show that for the q-norm, setting $q = 1 + \alpha$, if p_i is the vertex distribution of the random walk at step i, then $\|p_{i+1}\|_q^q \le \frac{1}{D^{\delta\alpha}} \|p_i\|_q^q$. More concretely they prove:

Theorem 1.6 (informal; see [DMOZ23], Theorem 5)**.** *Let $G = (U = [M], V = [M], E)$ be a sufficiently good lossless expander, and let $q = 1 + \alpha$ for some sufficiently small constant α.*

Fix any i, and let r_u, for each $u \in \mathrm{Supp}(p_i)$, 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. By definition, p_{i+1} is defined as $p_{i+1}(v) = \sum_{u \in \Gamma(v)} r_u(u, v) p_i(u)$. Then,

$$\|p_{i+1}\|_q^q \le O\left(\frac{1}{D^{\delta\alpha}}\right) \cdot \|p_i\|_q^q ,$$

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

This essentially implies that the entropy of the vertex distribution increases by at least δd. Thus, inductively, the final distribution will have $\|p_t\|_q^q \le \frac{1}{K^\alpha}$, implying that it has min entropy roughly k.

For ease of exposition, for the remainder of the section, we consider unpredictable sources where for every i, $\rho_i = \rho$ for some constant ρ. This case captures most of the intuition and difficulty at a high level.

The q-norm analysis fails. Unfortunately, in the case of unpredictable sources, the q-norm analysis cannot give a good bound on the norm of the final vertex distribution, $\|p_t\|_q$. To see this, consider a distribution X that is 0^n with probability ρ, and is a δ-CG source otherwise. This is both a CG source with ρ-error and a (δ, ρ)-unpredictable source. However, the norm of the final vertex distribution will always be at least $\|p_t\|_q \ge \rho$. Thus, the q-norm will not help us to establish that the final (smoothed) min entropy is large. But note that it *is* true in this example, where clearly we are ρ-close to having high min-entropy. It has been an open question since [DMOZ23] to give an analysis that shows this is always the case.

Beating the union bound, once again. Naively, one might try to fix the q-norm analysis as follows. For each i, condition on the event that X_{i+1} has "high entropy" given $x_1, \dots, x_i$. The original q-norm analysis could then work on this conditional distribution, and since the probability that this event does not happen is at most ρ, we can conclude that in the i-th step, the q-norm decreases "except with error ρ." Unfortunately, it is not clear how to chain such an argument multiple times over all steps t, without using a union bound which would require $\rho < 1/t$.

We remark that originally, in the case of CG sources, [DMOZ23] uses the q-norm analysis in part to beat the union bound over the *expander error* ε. As we've seen from the discussion above, the q-norm analysis does not allow you to do the same for ρ. Thus, beating the union bound over ρ is yet another challenge to overcome.

Probability evolution, not distribution evolution: a "chain rule" for vertex probabilities. The issue with the approaches above is that they attempt to make a statement about the quality of the vertex distribution at every step i. As discussed, it is not clear how to make any such statement. This leads us to search for an alternative approach. Denote by $W(x_1, \dots, x_i)$ as the vertex reached when taking the instructions $x_1, \dots, x_i$. In an unpredictable source, given a typical $x \sim X$, one expects to see roughly $(1-\rho)t$ "good" steps i in which

$$\Pr[X_i = x_i | X_{[1,\dots,i-1]} = x_{[1,\dots,i-1]}] \le D^{-\delta},$$

and ρt "bad" steps in which $\Pr[X_i = x_i | X_{[1,\dots,i-1]} = x_{[1,\dots,i-1]}] > D^{-\delta}$. One would like to argue that this directly translates to good steps and bad steps in the random walk. In other words, for every good step,

$$p_i(W(x_1, \dots, x_i)) \le \frac{1}{D^\delta} \cdot p_i(W(x_1, \dots, x_{i-1})),$$

and for every bad step, $p_i(W(x_1, \dots, x_i)) \le D \cdot p_i(W(x_1, \dots, x_{i-1}))$.

Notice that such an approach does not directly make a statement about the distribution at each step: we do not claim that for each i, the overall entropy of p_i increases. Rather, we say that individually, each path of vertices that the random walk takes is on its own journey of ups and and downs *in individual probability*, and typically there are few downs. We are able to make this approach concrete with the following key lemma.

Theorem 5 (chain rule for vertex probabilities, see Theorem 3.6, Corollary 3.7)**.** *Let G be a D-biregular (K, ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, and fix some $0 < \delta \le 1$. Then, for any $i \in [t]$, there is a subset $S_i \subseteq \{0,1\}^{n=dt}$ with $\Pr[X \in S_i] \ge 1 - 4\varepsilon D^{1-\delta} - 2\rho$, such that for every $x \in S_i$,*

$$p_i(W(x_1, \dots, x_i)) \le \max\left(\frac{2}{D^\delta} \cdot p_{i-1}(W(x_1, \dots, x_{i-1})), \frac{D^{O(\log 1/\varepsilon)}}{K}\right).$$

We believe that this chain rule for vertex probabilities is interesting in its own right. The standard chain rule for probability states that for every $x \in \mathrm{Supp}(X)$, and every i the probability of $x_1, \dots, x_i$ decreases from the probability of $x_1, \dots, x_{i-1}$ by a factor of $\Pr[X_i = x_i | X_{[1,i-1]} = x_{1,i-1}]$. The chain rule for vertex probabilities states that the same evolution of probabilities occurs when considering $W(x_1, \ldots, x_i)$ and $W(x_1, \ldots, x_{i-1})$, as long as the conditional probability of x_{i+1} "has entropy" (as is needed for expansion), and accounting for the probability of a collision due to the inherent error of the expander or the probability the next step has no entropy.[10]

Analyzing a full random walk. So far, we've shown that at every step, there is a high probability over $x \sim X$ that the corresponding vertex probabilities decrease. Notice we have made no assertion yet about how drastically the vertex probability might *increase* when the event S_i does not occur. However, it is not too hard to show that it is extremely unlikely for the probability to increase drastically. Overall, we can argue that in expectation over $x \sim X$, there are roughly $(1-\rho)t$ steps for which the vertex probability goes down by a factor of roughly $D^{-\delta}$, more accurately, $p_i(W(x_1, \ldots, x_i) \leq \frac{2}{D^\delta} \cdot p_{i-1}(W(x_1, \ldots, x_{i-1}))$, and the *total factor increase* from the remaining ρt steps, is roughly $D^{\rho t}$.

This argument so far is essentially all we need to obtain Theorem 3: When ρ is small, for a typical x, the number of good steps is overwhelmingly large comparing to the number of bad steps, and therefore one expects most runs of the random walk to end up at a vertex that has probability at most $p_t(W(x_1, \ldots, x_t)) \leq D^{-(\delta t - \rho t)}$.

Using a random stopping time. To obtain a seeded condenser with constant entropy gap, as in Theorem 2, we must characterize a bit more accurately how a typical run of the random walk behaves. In reality, Theorem 5 states that if i is a good step for $x_1, \ldots, x_t$ (that is, $x \in S_i$), then the probability of the vertex reached at step i decreases by $\approx \frac{1}{D}$ *as long as the vertex probability has not already reached the "capacity"* $D^{O(\log 1/\varepsilon)}/K$, which is a constant factor smaller than $1/M$, for M being the number of vertices in the expander.[11] If we can prove that over a random $x \sim X$, and a random stopping time i, that the vertex probability is at capacity with high probability, then we have proven Theorem 2.

Suppose we choose M to be noticeably less than $D^{\delta t}$, say, $D^{(\delta/2)t}$. Then, we expect that in a typical run of the random walk, the vertex probability reaches the capacity (or is close to it) after $t/2$ steps. We can assume for simplicity that the vertex probability is exactly at capacity after $t/2$ steps. Now, let us consider what happens in the *last* $t/2$ steps, under this assumption. There are only $\rho t < t/2$ steps for which the vertex probability can increase, each of which increases it by a factor of roughly D. Thus, most of the other $t/2$ steps either keep the vertex probability at capacity, or "repairs" a deficit from capacity by a factor of $1/D^\delta$. Overall, this means that over a random stopping time in the last $t/2$ steps, the probability of not being at capacity (meaning the walk has recently taken one of the ρt bad steps, or one of the $(\rho/\delta)t$ "repairing" good steps), is roughly $\rho + \rho/\delta = O(\rho/\delta)$.

1.5.1 Making Our Condensers Fully Online

A random walks based condenser is online if each symbol X_i is processed and used to update the state sequentially in a "read-once" fashion. However, there is an issue when the length of the stream

[10] An expert reader might ask how the Theorem 5 compares to the standard statement about lossless expanders as lossless conductors. In the standard case, when $\delta = 1$ and $\rho = 0$, the property of lossless conductors states that if p_{i-1} is a source with min entropy $k' < k = \log K$, then p_i is ε-close to a $k+d$-source. Theorem 5 on the other hand, requires no assumption on the entropy of the input source p_{i-1}, and is still able to conclude that the distribution "improves" in one step.

[11] In general, for constant degree lossless expanders, $K = M/\text{poly}(D)$.

t is not known ahead of time. Indeed, if one must settle on a graph of size M ahead of time, then one cannot hope for a final output entropy larger than m. This is problematic if the length of the stream t is not known ahead of time, and ends up being much larger than m, as we will miss out on most of the entropy of the stream. Broadly speaking, the workaround to this issue is as follows: If we know how much entropy we expect overall in each $X_1, \ldots, X_i$, then we can repeatedly increase the size of the graph at regular intervals, to accommodate the additional entropy expected. This allows us to maintain the guarantee that the entropy of the vertex distribution is close to m for all (or most) steps.

In a bit more detail, for simplicity, assume that the total length of the stream is a power of two (although still unknown). We begin the random walk from a fixed vertex on a small D-regular graph of size 2^C for some constant C. If we see more than $\approx C$ symbols from the stream, we embed the current vertex into a D-regular graph of size 2^{2C} and walk for another C steps. If a node in the smaller graph is represented by $v \in \{0,1\}^C$, then one can embed it in the larger graph as $v \circ 0^C$. Such an embedding provides a one-to-one mapping from the vertex distribution in the small graph to one with the same entropy in the large graph. We repeat this embed-and-walk process until the stream ends. This idea essentially suffices to implement the deterministic condenser of Theorem 3 in an online fashion.

To implement the condenser from Theorem 2, we use the same embed-and-walk process, but we must take care to implement the random stopping time in an online fashion as well. Once again, for simplicity, assume that t is a power of two. We once again begin the walk on a constant sized graph of size 2^C, and initialize a seed of length $\approx c = \log C$ to pick a random stopping time in case $t \leq c$. When the random stopping time is reached, we save the resulting vertex additionally in the state, and we continue the random walk until time c. If the stream ends at time c, output the saved vertex. Otherwise, we embed the current vertex into a graph of size 2^{2C}, and use $\mathsf{ExtendSeed}$ to add one more bit to the seed. Now, Y represents a random stopping time between 1 and $2c$, and we can repeat this process until the stream ends. Ultimately, this shows that for every $i \in [t]$ that is a power of two, the distribution obtained from using $X_1, \ldots, X_i$ as a random walk (with a random stopping time) contains most of the entropy k seen so far, within a graph whose size M is not too much larger than $K = 2^k$, all the while only needing to generate $\log i$ bits of seed.

As the details of such an online implementation are mostly minor alterations to the main results of our work, we defer a more detailed explanation to Appendix B.

1.6 Improved Random Walk-Based Extractors for High Min-Entropy Sources

So far, the main takeaway from our results is that lossless expanders can handle unpredictable sources with low entropy rate. On the other hand, spectral expanders can handle unpredictable sources with high entropy rate, as can be seen by looking at the classic random walks based extractor.

However, even for the case of sources with high entropy, the lossless expander random walk yields a quantitatively better result. In the classic expander random walk extractor (or sampler), based on the expander Chernoff bound, in order to achieve an error of ρ in the output distribution, it is necessary for the entropy of the input source X to be at least $(1 - \rho^2/C)n$ for some constant C.

Theorem 1.7 (standard RW-based extractor; see Theorem 6.2)**.** *There exists a universal constant C such that the following holds. For every positive integer n, and any $\rho > 0$, there exists an explicit (k, ρ) extractor*

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

for any $k \geq (1 - \rho^2/C)n + \log(1/\rho)$.

The ρ^2 factor is inherent in the use of the expander Chernoff bound (see Theorem 6.1). On the other hand, if one uses our new lossless expander random walk to condense to constant entropy gap (and then apply known constructions of extractors for sources with constant entropy gap with short seed length), one only needs the input source to have entropy $(1 - \rho/C')n$ for some constant C' in order to obtain final output error ρ. In addition, all of this can be implemented in a fully online manner, even when n is not known ahead of time.

Theorem 6 (new RW-based extractor; see Theorem 6.6)**.** *There exist universal constants $\rho_0 \in (0, 1)$ and $C > 1$ such that the following holds. For every positive integer n, and any constant $\rho \in (0, \rho_0)$, there exists an explicit (k, ρ) extractor*

$$\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^{\ell = \log n + O(\log(1/\rho))} \to \{0,1\}^{m=\Omega(k)},$$

for any $k \geq (1 - \rho/C)n$.

1.7 Related Work

Before introducing ρ-almost CG sources, [DMOZ23] first generalizes CG sources by introducing what is coined λ-error.

Definition 1.8 (λ-almost CG source)**.** *A λ-almost δ-CG source is a sequence of random variables $X = X_1 \circ \ldots \circ X_t$ with $X_i \sim \{0,1\}^d$, such that for at least $(1 - \lambda)t$ of $i \in [t]$, we have that for any prefix $a \in \{0,1\}^{d(i-1)}$, it holds that $H_\infty(X_i | X_{[1,i-1]} = a) \geq \delta d$.*

Unlike ρ-error, for λ-error, [DMOZ23] is still able to construct condensers by running a random walk using X, although not with a constant entropy gap. Instead, the gap is roughly λm, for reasons inherent to the random walk construction itself. Intuitively, for the λ fraction of bad indices i, X_i could be completely determined (and adversarially chosen) based on x_{i-1}. Therefore, whatever edge x_{i-1} instructs the walk to take, x_i could instruct to return via the same edge, effectively wiping out the progress made from x_{i-1}. Overall, when all the bad indices are at the end, it can wipe out λt steps of entropy accumulation, leaving an entropy gap of λt.[12] The case of λ-error is interesting in its own right. In fact, a recent work of Chattopadhyay, Gurumukhani, and Ringach [CGR24b], shows that deterministic condensing of λ-almost δ-CG sources is impossible, even with large entropy gap, in the regime where $\lambda \geq \frac{1}{2}$.

Goodman, Li, and Zuckerman [GLZ24] showed how to condense CG sources even when the blocks are long, and the entropy rate is subconstant. However, their constructions are not online, and they don't address the case of almost CG sources.[13]

Previous works that directly consider (deterministic) online extraction [DGSX21a, DGSX21b] assume a strong notion of unpredictability, wherein the X_i-s are independent (but with some min-entropy). In their model, they assume that for every i, $|S_i| = |X_i| = n$, with the length of the stream t sufficiently long that the total entropy k of X is at least n. Recall that in our work, we

[12] When $\lambda > 0$ but 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 *can* regain constant entropy gap. See [DMOZ23, Section 3.1], where this property is called *suffix friendliness*.

[13] However, their constructions work for suffix-friendly CG sources.

generally think of each $X_i \sim \{0,1\}^d$ for some *constant* d, $n = dt$, and $|S_t| \approx k \ll n$. More specifically, [DGSX21a] considers how entropy accumulates for specific update functions that are based off of practical random number generation. They show that entropy accumulates when the X_i-s are independent draws from certain classes of distributions known as 2-monotone distributions. [DGSX21b], considers *linear* update functions and shows that entropy accumulates when the X_i-s are independent k-sources.

Other previously studied notions of sequential sources include Somewhere Honest Entropy Look Ahead (SHELA) sources [AOR+20], also known as online Non-Oblivious Symbol Fixing (oNOSF) sources [CGR24b]. Such sources are essentially the λ-error CG sources discussed above, except for two distinctions. The first is that the good steps are all high entropy distributions that are *independent* from each other. The second is that each bad step only depends on previous blocks.[14] Aggarwal et a. [AOR+20] shows that extracting from oNOSF sources is impossible, however one can convert oNOSF sources into *uniform* oNOSF sources. Chattopadhyay, Gurumukhani, and Ringach explored the limits of online condensing of such sources [CGR24b], and later achieved constructions with essentially optimal parameters [CGR24a].

A recent work by Xun and Zuckerman [XZ24] provides constructions of strong *offline* extractors whose seed length has *nearly* optimal dependence on n and ε: for any desired $\alpha > 0$, their construction gives an extractor with seed length $(1+\alpha)\log(n-k) + (2+\alpha)\log 1/\varepsilon + O(1)$, as long as the entropy rate k/n is sufficiently close to 1 (depending on α). To compare, our results discussed in Section 1.6 provide *online* extractors with seed length $\log n + O(\log 1/\varepsilon) + O(1)$ when $k/n \geqslant 1 - \Theta(\varepsilon)$.

1.8 Organization

Section 2 gives preliminary definitions and results needed for our work. Section 3 provides our novel analysis of random walks using unpredictable sources on lossless expanders. Section 4 gives the analysis using the idea of a two-stage construction from [DMOZ23], while Section 5 plugs in known explicit constructions of lossless expanders to give explicit (with and without two-stage) constructions of online condensers. Finally, Section 6 compares and contrasts the extractors we obtain from classic random walks expanders to the ones we get from our lossless expanders.

2 Preliminaries

For integers denoted by lowercase letters such as n, m, k, d, we typically use capital letters to denote their power $N = 2^n, M = 2^m, K = 2^k, D = 2^d$.

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 $|X - Y| \leq \varepsilon$ we say that X is ε-close to Y and denote it by $X \approx_\varepsilon Y$. We denote by U_n the random

[14] The latter property is why those sources are called "online" – an adversary corrupts the bad blocks while only knowing the history. They do not give online condensers for such sources.

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 $|f(X) - f(Y)| \le |X - Y|$, 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 $H_\infty(X) \le H(X)$. 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, $H_\infty(X) \ge \log \frac{1}{\varepsilon}$.*

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.

Claim 2.2. *Let $X \sim \{0,1\}^d$ be a random variable, then the following hold:*

- *If $H_\infty^\varepsilon(X) \ge \delta d$, then $\Pr_{x \sim X}[\Pr[X = x] \ge 2D^{-\delta}] \le 2\varepsilon$.*
- *Suppose $\delta d \le d - 2$. If $\Pr_{x \sim X}[\Pr[X = x] \ge D^{-\delta}] \le \varepsilon$ then $H_\infty^\varepsilon(X) \ge \delta d$*

Proof: For the first bullet, let H be the set of x-s such that $\Pr[X = x] \ge 2D^{-\delta}$. On the one hand, $\Pr[X \in H] \le \varepsilon + \frac{|H|}{D^\delta}$. On the other hand, $\Pr[X \in H] \ge |H| \cdot 2D^{-\delta}$. Thus $|H| \le \varepsilon D^\delta$. So $\Pr[X \in H] \le 2\varepsilon$.

For the second bullet, let p be the distribution of x (i.e. $p(x) = \Pr[X = x]$). Let H_1 be the set of x-s such that $p(x) > D^{-\delta}$ and let H_2 be the set of x-s such that $\frac{1}{2}D^{-\delta} \le (x) \le D^{-\delta}$. Consider the following probability distribution r.

$$r(x) = \begin{cases} 0 & \text{if } x \in H_1, \\ p(x) & \text{if } x \in H_2, \\ p(x) + \frac{\sum_{y \in B_1} p(y)}{|\{0,1\}^n \setminus (H_1 \cup H_2)|} & \text{otherwise.} \end{cases}$$

By construction, r is ε-close to p. So it suffices to show that r is a δd source. Observe that:

$$|\,\{0,1\}^d \setminus (H_1 \cup H_2)| \ge 2^d - 2^{\delta d+1} \ge 2^{\delta d+2} - 2^{\delta d+1} = 2^{\delta d+1}.$$

Therefore, for any $x \in \{0,1\}^d \setminus (H_1 \cup H_2)$, $p(x) \le D^{-\delta}$. ■

Claim 2.3. *Suppose that $X = (1-\rho)X' + \rho X''$, where X' is γ-close to a δd source. Then, X is $(\gamma+\rho)$-close to a δd source.*

Proof: Let $X^\star$ be the δd source that X' is γ-close to. Then,

$$\sum_{x \in \mathrm{Supp}(X)} |\Pr[X = x] - \Pr[X^\star = x]| = \sum_{x \in \mathrm{Supp}(X)} \left|(1-\rho)\Pr[X' = x] + \rho \Pr[X'' = x] - \Pr[X^\star = x]\right|$$
$$\leq \rho + \sum_{x \in \mathrm{Supp}(X)} \left|\Pr[X' = x] - \Pr[X^\star = x]\right| \leq \rho + \gamma.$$

■

Claim 2.4. *Let X, Y be random variables with $X \sim \{0,1\}^d$. Assume that $\delta d \leq d-2$, and that*

$$\Pr_{(x,y)\sim(X,Y)} [\Pr[X = x | Y = y] > D^{-\delta}] \leq \rho.$$

Then, $\Pr_{x \sim X} \left[\Pr[X = x] > 2D^{-\delta}\right] \leq 4\sqrt{\rho}$.

Proof: Let $A(x,y)$ be the indicator random variable for the event that $\Pr[X = x | Y = y] > D^{-\delta}$. Then $\mathbb{E}_{(x,y)\sim(X,Y)}[A(X,Y)] = \mathbb{E}_{y\sim Y} E_{x \sim X|Y=y}[A(x,y)] \leq \rho$. By Markov's, with probability at least $1-\sqrt{\rho}$ over y, $\mathbb{E}_{x\sim X|Y=y}[A(x,y)] \leq \sqrt{\rho}$. Therefore, by the second bullet of Claim 2.2, with probability at least $1-\sqrt{\rho}$ over y, the conditional distribution $X|Y=y$ is $\sqrt{\rho}$-close to a δd-source. Call such fixed y-s "good", and "bad" otherwise. Then:

$$X = (X | y \text{ is good}) \Pr[y \text{ is good}] + (X | y \text{ is bad}) \Pr[y \text{ is bad}]$$

However, the distribution $(X|y \text{ is good})$ is convex combination of distributions that are $\sqrt{\rho}$-close to a δd source, and is thus itself a δd source. By Claim 2.3, we see that X is $2\sqrt{\rho}$-close to a δd source. Finally, the first bullet of Claim 2.2 yields the result. ■

Lemma 2.5. *Let X, Y be random variables with $X \sim \{0,1\}^d$ and Y. Assume that $\delta d \leq d-2$, and that*

$$\Pr_{(x,y)\sim(X,Y)} \left[\Pr[X = x|Y = y] > D^{-\delta}\right] \leq \rho.$$

Then, for any deterministic function $f \colon \mathrm{Supp}(Y) \to \Omega$, for some finite range Ω, it holds that

$$\Pr_{(x,\omega)\sim(X,f(Y))} \left[\Pr[X = x | f(y) = \omega] > 2D^{-\delta}\right] \leq 5\rho^{1/4}.$$

Proof: We have:

$$\rho \geq \Pr_{(x,y)\sim(X,Y)} \left[\Pr[X = x|Y = y] > D^{-\delta}\right]$$
$$= \sum_{\omega \in \Omega} \Pr[f(Y) = \omega] \Pr_{(x,y)\sim(X,Y)} \left[\Pr[X = x|Y = y] > D^{-\delta} \middle| f(Y) = \omega\right].$$

Therefore, by Markov's, with probability at least $1-\sqrt{\rho}$ over $\omega \sim f(Y)$, we have that

$$\Pr_{(x,y)\sim(X,Y)} \left[\Pr[X = x|Y = y] > D^{-\delta} \middle| f(Y) = \omega\right] \leq \sqrt{\rho}.$$

Therefore, by Claim 2.4, for every such good ω, we have:

$$\Pr_{x \sim X}\left[\Pr[X = x] > 2D^{-\delta} \middle| f(Y) = w\right] \le 4\rho^{1/4}$$

Finally we can conclude that

$$\Pr_{(x,\omega)\sim(X,f(Y))}[\Pr[X = x|f(y) = \omega] > 2D^{-\delta}] \le 4\rho^{1/4} + \rho^{1/2} \le 5\rho^{1/4}.$$

■

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 $(i_1, \dots, i_t) \in [D]^t$ is the sequence $(v_0, v_1, \dots, v_t)$, where v_j is the i_j-th neighbor of v_{j-1}.

Definition 2.6 (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*.[15] The expanders we work with will general be balanced: $N = M$, and biregular, so every node in the bipartite graph has exactly D neighbors.

Theorem 2.7 (nonexplicit lossless expanders). *There exists a universal constant $c^\star$ such that for every positive integers N and D, there exists a D-biregular bipartite graph $G = ([N], [N], E)$ that is a (K, ε) lossless expander for $\varepsilon \le \frac{c^\star}{D}$ and $K = \frac{N}{c^\star D^2}$. By brute-force, such an expander can be found deterministically in time $N^{O(ND)}$.*

We make use of recent constructions of explicit biregular expanders, that simplify the seminal [CRVW02] construction, and improve upon its dependence between ε and D.

Theorem 2.8 ([CRT23, Gol24]). *There exists a universal constant $c^\star$ such that for every positive integers N and D, there exists an explicit D-regular bipartite graph $G = ([N], [N], E)$ that is a $(K, \varepsilon = \frac{c^\star}{D^{1/3}})$ expander and $K = \frac{N}{D^{c^\star}}$.*

Remark 2.9. *The actual dependence between ε and D in the new constructions is even better, roughly $D = O\left(\frac{\log 1/\varepsilon}{\varepsilon^2}\right)$. However, in order for us to state a simple theorem for all D-s, we use the weaker bound $\varepsilon = O(1/D^{1/3})$. We note that the fact that the new constructions are biregular greatly simplify the analysis. On the other hand, the improvement in dependence between ε and D from [CRVW02] only marginally improves upon the entropy rate δ that we can handle without a two stage construction.*

Remark 2.10. *For simplicity, for the remainder of the paper, we consider the global $c^\star$ from both of the above theorems the same.*

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

2.2.1 The Expander's Error for Entropy Rate δ

Given a (K, ε)-expander $G = (U, V, E)$, it is a well known fact that for any set of size $|S| \le K$, there are at least $(1 - 2\varepsilon)D|S|$ unique neighbors of S. Assume for simplicity that each vertex $u \in S$ has exactly $(1 - 2\varepsilon)D$ unique neighbors. This mean that the probability over a choice of $u \in S$ (using any distribution over S), and a uniform random neighbor $v \in \Gamma(u)$, that $p(v) \le \frac{1}{D}p(u)$ is at least $1 - 2\varepsilon$.

In the case when the distribution over neighbors is no longer uniform, but a flat source over D^{δ}, the new probability of a "successful dampening", i.e. that $p(v) \le \frac{1}{D^{\delta}}p(u)$, is at least $1 - 2\varepsilon D^{1-\delta}$. Under these simplifying assumptions, we can consider the term $\varepsilon D^{1-\delta}$ as the error of the expander for entropy rate δ. Indeed we will be able to show that the intuition above holds in the general case, with a similar error term appearing in the full analysis. Notice that when $\delta = 1$, the error of the lossless expander is essentially equivalent to the error of the "conductor" corresponding to G (see [CRVW02]).

As a final remark, note that the expanders from Theorem 2.7 and Theorem 2.8 yield errors roughly $D^{-\delta}$ and $D^{1/2-\delta}$ for entropy rate δ respectively. Thus, while an optimal lossless expander can yield a nontrivial error for any entropy rate $\delta > 0$, current explicit constructions can only do so for sufficiently large error rates $\delta > 1/2$.

2.3 Seeded Extractors, Condensers, and Samplers

Definition 2.11 (extractor). *A function*

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

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 $\{0,1\}^{\ell}$ and is independent of X. We say Ext is strong *if $(\mathsf{Ext}(X, Y), Y) \approx U_m \times Y$.*

We'll make use of extractors that work well when X has a small entropy gap.

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 $\mathsf{Ext}_{\mathsf{GW}}\colon \{0,1\}^n \times \{0,1\}^{\ell} \to \{0,1\}^m$, where $\ell = O(\Delta + \log \frac{1}{\varepsilon})$ 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*

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

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 $H_{\infty}^{\varepsilon}(\mathsf{Cond}(X, Y)) \ge k'$, where Y is uniformly distributed over $\{0,1\}^{\ell}$ and is independent of X. When $\ell = 0$, we say that $\mathcal{X}$ admits deterministic condensing.

We next define *density samplers*.

Definition 2.14 (sampler). *Let $\Gamma\colon [N] \times [D] \to [M]$.*

- *We say $x \in [N]$ is ε-*bad for *$B \subseteq [M]$ if*

$$\left| \Pr_{y \sim U_{[D]}} [\Gamma(x,y) \in B] - \mu(B) \right| > \varepsilon.$$

- *We say Γ is a (δ, ε) sampler if for every $B \subseteq [M]$ we have that*

$$|\{x \in [N] : x \text{ is } \varepsilon\text{-bad for } B\}| < \delta N.$$

Lemma 2.15 ([Zuc97])**.** *Let $\mathsf{Ext} \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a (k, ε) extractor. Then, Ext is also a $(\delta = 2^{k-n+1}, \varepsilon)$ sampler.*

Conversely:

Lemma 2.16. *Let $\Gamma \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a (δ, ε) sampler. Then, Γ is also a $(k = n - \log 1/\delta + \log 1/\varepsilon, 2\varepsilon)$ extractor.*

3 Condensing Unpredictable Sources via Random Walks

In this section we demonstrate our novel analysis of random walks via (δ, ρ) unpredictable sources, on (K, ε)-lossless expanders. We state our theorems in full generality, for any δ, ρ, and ε, without regard to the quality of these parameters. Naturally, the theorems will only give interesting results (and we will only apply them) with sufficiently large δ, and sufficiently small ε and ρ.

3.1 The Expander Framework

We use the following framework, which was also used in [DMOZ23].

Definition 3.1 (weight function)**.** *Let $G = (U, V, E)$ be a bipartite graph. Let $w \colon U \to \mathbb{R}_{\geq 0}$ be a weight function. Let $|w| = \sum_{u \in U} w(u)$. We define a weight function on $\mathcal{N}(w) \colon V \to \mathbb{R}_{\geq 0}$ as:*

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

We denote u_v as the node that achieves $\max_{u \in \Gamma(v)} w(u)$.

Theorem 3.2 (weight expansion, see Lemma 4.2 in [DMOZ23])**.** *Let G be a (K, ε)-expander. Suppose that w is a weight function supported on K nodes. Then, $|\mathcal{N}(w)| \geq (1 - \varepsilon) D|w|$.*

The following corollary is immediate.

Corollary 3.3. *Let $G = (U, V, E)$ be a (K, ε)-expander. Suppose that w is a weight function supported on K nodes. Then,*

$$\sum_{v \in V} \sum_{u \in \Gamma(v) \setminus u_v} w(u) \leq \varepsilon D|w|.$$

Proof: Observe that

$$D|w| = \sum_{v \in V} \sum_{u \in \Gamma(v)} w(u) = \sum_{v \in V} w(u_v) + \sum_{v \in V} \sum_{u \in \Gamma(v) \setminus u_v} w(u),$$

and then apply Theorem 3.2. ■

3.2 A Warmup

To help illustrate the ideas of proving our main technical result, we first reprove a generalized notion of unique neighbor expansion (see also [DMOZ23, Appendix D]).

Theorem 3.4. *Let $G = (U, V)$ be a D-biregular (K, ε) lossless expander. Let p be a probability distribution supported on at most K nodes of V. Then,*

$$\sum_{v \in V} \left(\frac{1}{D} p(u_v) - \sum_{u \in \Gamma(v), u \neq u_v} \frac{1}{D} p(u) \right) \geq 1 - 2\varepsilon. \tag{1}$$

Proof: Observe that

$$1 = \sum_{v \in V} \left(\frac{1}{D} p(u_v) + \sum_{u \in \Gamma(v), u \neq u_v} \frac{1}{D} p(u) \right).$$

By Corollary 3.3, $\sum_{v \in V} \sum_{u \in \Gamma(v), u \neq u_v} \frac{1}{D} p(u) \leq \varepsilon$. Thus, subtracting twice this quantity from both sides yields the result. ■

We now turn to interpreting the above theorem. Let $V^+ \subseteq V$ be the set of vertices v for which $\frac{1}{D} p(u_v) - \sum_{u \in \Gamma(v), u \neq u_v} \frac{1}{D} p(u) \geq 0$. Or in other words, $p(v) = \sum_{u \in \Gamma(v)} \frac{1}{D} p(u) \leq \frac{2}{D} p(u_v)$. It must still be the case that

$$\sum_{v \in V^+} \left(\frac{1}{D} p(u_v) - \sum_{u \in \Gamma(v), u \neq u_v} \frac{1}{D} p(u) \right) \geq 1 - 2\varepsilon,$$

and also

$$\frac{1}{D} \sum_{v \in V^+} p(u_v) \geq 1 - 2\varepsilon. \tag{2}$$

Observe that the quantity on the left-hand side of Equation (2) measures the total probability (over a choice of u from p, and a uniformly random neighbor of u), that a step in the graph leads to a $v \in V^+$ from its heaviest weight neighbor u_v. By the definition of V^+, this means that with probability at least $1 - 2\varepsilon$, a vertex u moves from having a probability of $p(u)$, to having a probability of $\frac{2}{D} p(u)$.

So far, we have shown that given any initial probability distribution $p(u)$, with support size at most K, a uniform random step on the expander will send u to a neighbor with probability at most $2/D$ times smaller than the original $p(u)$. In order to handle the general case, we must generalize the argument to work when $p(u)$ is supported on more than K vertices, and when the next step is not uniform, but only usually (with probability at least $1 - \rho$ over the choice of u and a next step) a low probability outcome. Intuitively, the latter adds an extra $-\rho$ to the right-hand side of Equation (1). To handle the former, considering only nodes with probability at least $1/K$ according to p yields a support of size at most K. We hope to argue that the the contribution of nodes with probability smaller than $1/K$ to the overall probability is small and can also be subtracted from the right-hand side of Equation (1). For technical reasons, we will need to instead consider only nodes with probability at least D^a/K for some parameter a.

3.3 Analyzing a Single Step in the Random Walk

Definition 3.5. *Let $X = X_1 \circ \cdots \circ X_t$ be a sequence of random variables, each $X_i \sim \{0,1\}^d$. Let $G = (U, V, E)$ be a balanced bipartite D-regular graph. For $i \in [t]$, we define the random variable $W_i = W_i(X)$ to be the vertex in U reached upon a walk over G for i steps using the instructions $x_1, \dots, x_i \sim X_{[1,i]}$.*[16]

Let p_0 be a distribution concentrated entirely on an arbitrary start vertex. Define $p_i \colon U \to \mathbb{R}$ as the probability distribution over the vertices in U after those i steps. We also overload this notation, by defining $p_i \colon X \to \mathbb{R}$ as

$$p_i(x_1, \dots, x_t) = p_i(W_i(x_1, \dots, x_t)).$$

Furthermore, let $\ell_G \colon E \to [D]$ be a labelling function for G's edges such that for every u, every edge leaving u is assigned a unique label in $[D]$. When considering a single step on G from distribution $p_{i-1} = p_U$ to $p_i = p_V$, we define $r_u(u,v) = \Pr[X_i = \ell_G(u,v) | W_{i-1} = u]$.

We are now ready to give our main technical result. In order to use the ideas presented in Section 3.2, we will decompose the total probability weight into "good events", such as the one measured in the left-hand side of Equation (2), and "bad events" whose measure we will bound.

Theorem 3.6. *Let G be a D-biregular (K, ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, and fix some $0 < \delta \leq 1$. For every $i \in [t]$, we abbreviate $\rho_i = \rho_i(X, \delta)$ (see Definition 1.2).*

Then, for any $i \in [t]$ and any parameter $a > 2$, there is a subset $S_i \subseteq \{0,1\}^{n=dt}$ with $\Pr[X \in S_i] \geq 1 - 2\varepsilon D^{1-\delta} - 2\rho_i - 2D^{-(a-2)}$, such that for every $x \in S_i$,

$$p_i(x) \leq \max\left(\frac{2}{D^\delta} \cdot p_{i-1}(x), \frac{D^{a+1}}{K}\right).$$

Before we prove the theorem we note that we will commonly instantiate the theorem so that the term $D^{-(a-2)}$ is absorbed by the $\varepsilon D^{1-\delta}$ term, by setting $a = \log(1/\varepsilon)/d + 2$, so that $D^{-(a-2)} \leq \varepsilon \leq \varepsilon D^{1-\delta}$.

Corollary 3.7. *Keeping the notation above, for every $i \in [t]$ there is a subset $S_i \subseteq \{0,1\}^{n=dt}$ with $\Pr[X \in S_i] \geq 1 - 4\varepsilon D^{1-\delta} - 2\rho_i$, such that for every $x \in S_i$,*

$$p_i(x) \leq \max\left(\frac{2}{D^\delta} \cdot p_{i-1}(x), \frac{D^{3+\log(1/\varepsilon)/d}}{K}\right).$$

For a (K, ε)-expander G, we'll often refer to the quantity

$$k_{\text{capacity}} = k - \log(1/\varepsilon) - 3d$$

as the *capacity* of G, as this is intuitively the highest the "entropy" can accumulate to in the graph.

Proof (of Theorem 3.6): Fix $i \in [t]$, and denote $p_U = p_{i-1}$, $p_V = p_i$, and $\rho_i = \rho$. Note that

$$p_V(v) = \sum_{u \in \Gamma(v)} \Pr_X[X_i = \ell_G(u,v) | W_{i-1} = u] \cdot p_U(u) = \sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u).$$

[16] Formally, along an $(i+1)$-partite graph with a copy of G between each two layers.

For any vertex u, define $B_u \subseteq \operatorname{Supp}(X)$ as

$$B_u = \left\{ x : W_{i-1}(x) = u \wedge \Pr\left[X_i = x_i | X_{[1,i-1]} = x_{[1,i-1]}\right] > D^{-\delta} \right\}. \tag{3}$$

That is, B_u is the set of paths reaching u at step i, but whose next instruction has high probability of occurring. We can then write $r_u(u,v)$ conditioned on being in B_u or not as follows:

$$\begin{aligned} r_u(u,v) &= \Pr[x \in B_u | W_{i-1} = u] \Pr[X_i = \ell_G(u,v) | x \in B_u, W_{i-1} = u] \\ &\quad + \Pr[x \notin B_u | W_{i-1} = u] \Pr[X_i = \ell_G(u,v) | x \notin B_u, W_{i-1} = u] \\ &\triangleq \theta_{u,b} \cdot r_{u,b}(u,v) + \theta_{u,g} \cdot r_{u,g}(u,v). \end{aligned}$$

Notice that:

$$\sum_{v \in V} \sum_{u \in \Gamma(v)} \theta_{u,b} \cdot r_{u,b}(u,v) \cdot p_U(u) = \rho. \tag{4}$$

Notice also that by definition, for any u, v, $r_{u,g}(u,v) \leq D^{-\delta}$. The idea will be to decompose $\sum_{v \in V} \sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u)$ into different parts. First, we'll separate those v which will definitely have sufficiently small probability from those that might not. To this end, let H_a for a parameter a be the set of $u \in U$ with $p_U(u) \geq \frac{D^a}{K}$, and let V_{H_a} be the set of vertices $v \in V$ with at least one neighbor into H_a. In other words, $V_{H_a} = \Gamma_G(H_a)$. Define also $H = H_0$ as the set of nodes with probability at least $1/K$. Then,

$$\begin{aligned} \sum_{v \in V} \sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u) = & \sum_{v \notin V_{H_a}} \left(\sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u) \right) \\ & + \sum_{v \in V_{H_a}} \left(\sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u) \right). \end{aligned}$$

For any $v \notin V_{H_a}$, we know that $p_V(v) \leq D^{a+1}/K$, and therefore is already sufficiently small probability. Hence ultimately, the term $\sum_{v \notin V_{H_a}}$ measures a "good event." Next, for nodes $v \in V_{H_a}$, we can separate the contribution to $p_V(v)$ (the inner sum) from "good" next steps (those with low probability) and the "bad" ones.

$$\begin{aligned} \sum_{v \in V} \sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u) = & \sum_{v \notin V_{H_a}} \left(\sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u) \right) \\ & + \sum_{v \in V_{H_a}} \sum_{u \in \Gamma(v)} \theta_{u,g} \cdot r_{u,g}(u,v) \cdot p_U(u) \qquad (5) \\ & + \sum_{v \in V_{H_a}} \sum_{u \in \Gamma(v)} \theta_{u,b} \cdot r_{u,b}(u,v) \cdot p_U(u). \qquad (6) \end{aligned}$$

Notice we can bound the summand in (6) by ρ, whereas we hope to work with (5) as in the warmup, as this summand corresponds to "high entropy steps." Specifically, for summand (5), we can partition the neighbors $u \in \Gamma(v)$ into three categories:

- The u that is u_v. Note that u_v is always in $H_a \subset H$.

- The u-s in H (but not u_v).
- The u-s not in H.

Therefore, we can rewrite once again according to these classifications:

$$\begin{aligned}\sum_{v\in V}\sum_{u\in\Gamma(v)} r_u(u,v)\cdot p_U(u) = &\sum_{v\notin V_{H_a}}\left(\sum_{u\in\Gamma(v)} r_u(u,v)\cdot p_U(u)\right)\\ &+\sum_{v\in V_{H_a}}\theta_{u_v,g}\cdot r_{u_v,g}(u_v,v)\cdot p_U(u_v)\\ &+\sum_{v\in V_{H_a}}\sum_{u\in(\Gamma(v)\setminus u_v)\cap H}\theta_{u,g}\cdot r_{u,g}(u,v)\cdot p_U(u) \qquad (7)\\ &+\sum_{v\in V_{H_a}}\sum_{u\in(\Gamma(v)\setminus u_v)\cap \overline{H}}\theta_{u,g}\cdot r_{u,g}(u,v)\cdot p_U(u) \qquad (8)\\ &+\sum_{v\in V_{H_a}}\sum_{u\in\Gamma(v)}\theta_{u,b}\cdot r_{u,b}(u,v)\cdot p_U(u).\end{aligned}$$

The following claim bounds the summand in (7).

Claim 3.8.

$$\sum_{v\in V_{H_a}}\sum_{u\in(\Gamma(v)\setminus u_v)\cap H}\theta_{u,g}\cdot r_{u,g}(u,v)\cdot p_U(u)\le\varepsilon D^{1-\delta}.$$

Proof: Let $w\colon U\to\mathbb{R}$ be the weight function that is $w(u)=0$ if $u\notin H$ and $w(u)=p_U(u)$ otherwise. Notice that $|w|=\Pr[x\in H]$. We make use of the fact that $r_{u,g}\le D^{-\delta}$:

$$\begin{aligned}\sum_{v\in V_{H_a}}\sum_{u\in(\Gamma(v)\setminus u_v)\cap H}\theta_{u,g}\cdot r_{u,g}(u,v)\cdot p_U(u) &= \sum_{v\in V_{H_a}}\sum_{u\in\Gamma(v)\setminus u_v}\theta_{u,g}\cdot r_{u,g}(u,v)\cdot w(u)\\ &\le\frac{1}{D^{\delta}}\sum_{v\in V}\sum_{u\in\Gamma(v)\setminus u_v}w(u)\le\frac{1}{D^{\delta}}\cdot\varepsilon D|w|\le\varepsilon D^{1-\delta},\end{aligned}$$

where the penultimate inequality follows from Corollary 3.3. ■

The next claim now bounds summand (8):

Claim 3.9. *It holds that*

$$\sum_{v\in V_{H_a}}\sum_{u\in(\Gamma(v)\setminus u_v)\cap\overline{H}}\theta_{u,g}\cdot r_{u,g}(u,v)\cdot p_U(u)\le\frac{1}{D^{a-2}}.$$

Proof: First,

$$\sum_{v\in V_{H_a}}\sum_{u\in(\Gamma(v)\setminus u_v)\cap\overline{H}} r_{u,g}(u,v)\cdot p_U(u)\le\sum_{v\in V_{H_a}}\sum_{u\in(\Gamma(v)\setminus u_v)\cap\overline{H}}\frac{1}{K},$$

where we used the fact that $p_U(u)\le 1/K$ for every $u\notin H$. We observe that the number of terms in the double sum is at most the number of edges touching V_{H_a}. H_a has size at most K/D^a, and therefore there are at most K/D^{a-2} edges overall that touch V_{H_a} (assuming the right degree is also D). Thus overall, the sum is at most $(1/K)\cdot K/D^{a-2}=1/D^{a-2}$. ■

Using these claims, and the bound on ρ from (4), we can conclude that

$$\begin{aligned}
1 - 2\varepsilon D^{1-\delta} - 2\rho - 2D^{-(a-2)} \leq &\sum_{v \notin V_{H_a}} \left(\sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u) \right) \\
&+ \sum_{v \in V_{H_a}} \theta_{u_v,g} \cdot r_{u_v,g}(u_v,v) \cdot p_U(u_v) \\
&- \sum_{v \in V_{H_a}} \sum_{u \in (\Gamma(v) \setminus u_v) \cap H} \theta_{u,g} \cdot r_{u,g}(u,v) \cdot p_U(u) \\
&- \sum_{v \in V_{H_a}} \sum_{u \in (\Gamma(v) \setminus u_v) \cap \overline{H}} \theta_{u,g} \cdot r_{u,g}(u,v) \cdot p_U(u) \\
&- \sum_{v \in V_{H_a}} \sum_{u \in \Gamma(v)} \theta_{u,b} \cdot r_{u,b}(u,v) \cdot p_U(u).
\end{aligned}$$

Now, let $V_+ \subseteq V_{H_a}$ be the set of v-s such that

$$\begin{aligned}
&\theta_{u_v,g} \cdot r_{u_v,g}(u_v,v) \cdot p_U(u_v) \\
&- \sum_{u \in (\Gamma(v) \setminus u_v) \cap H} \theta_{u,g} \cdot r_{u,g}(u,v) \cdot p_U(u) \\
&- \sum_{u \in (\Gamma(v) \setminus u_v) \cap \overline{H}} \theta_{u,g} \cdot r_{u,g}(u,v) \cdot p_U(u) \\
&- \sum_{u \in \Gamma(v)} \theta_{u,b} \cdot r_{u,b}(u,v) \cdot p_U(u) \geq 0.
\end{aligned}$$

It follows that

$$\begin{aligned}
1 - 2\varepsilon D^{1-\delta} - 2\rho - 2D^{-(a-2)} \leq &\sum_{v \notin V_{H_a}} \left(\sum_{u \in \Gamma(v)} r_u(u,v) \cdot p_U(u) \right) &\quad (9) \\
&+ \sum_{v \in V_+} \theta_{u_v,g} \cdot r_{u_v,g}(u_v,v) \cdot p_U(u_v). &\quad (10)
\end{aligned}$$

We observe that the above inequality suggests a subset of $x \in \mathrm{Supp}(X)$ (of large density) for which good things happen:

- The term (9) measures the probability of the set of x-s that lead to a vertex in $v \in V_H$, but no neighbor of this v has probability weight $p_U(u) > D^a/K$. Such vertices have weight at most $\frac{D^{a+1}}{K}$.
- The term (10) measures the probability of x-s that lead to each $v \in V_+$ from u_v.

But for any $v \in V_+$,

$$\begin{aligned} p_V(v) = \theta_{u_v,g} \cdot r_{u_v,g}(u_v, v) \cdot p_U(u_v) \\ + \sum_{u \in (\Gamma(v) \setminus u_v) \cap H} \theta_{u,g} \cdot r_{u,g}(u,v) \cdot p_U(u) \\ + \sum_{u \in (\Gamma(v) \setminus u_v) \cap \overline{H}} \theta_{u,g} \cdot r_{u,g}(u,v) \cdot p_U(u) \\ + \sum_{u \in \Gamma(v)} \theta_{u,b} \cdot r_{u,b}(u,v) \cdot p_U(u) \\ \le 2 \cdot \theta_{u_v,g} \cdot r_{u_v,g}(u_v, v) \cdot p_U(u_v) \le \frac{2}{D^{\delta}} \cdot p_U(u_v). \end{aligned}$$

Overall, we can consider the corresponding events:

$$E_1 = \left\{ x \in \mathrm{Supp}(X) : W_i(x) \in \overline{V}_{H_a} \right\}$$
$$E_2 = \left\{ x \in \mathrm{Supp}(X) : W_i(x) \in V_+, \ \ W_{i-1}(x) = u_{W_i(x)}, \ \ \Pr\left(X_i = x_i | X_{[1\dots i-1]} = x_{[1\dots i-1]} \right) \le D^{-\delta} \right\},$$

and set $S_i = E_1 \cup E_2$. Recalling that $V_+ \subseteq V_{H_a}$, we see that the two events are disjoint. Thus, indeed, we have our lower bound on $\Pr[X \in S_i]$, and for each $x \in S_i$, either $p_V(x) \le \frac{D^{a+1}}{K}$ or $p_V(x) \le \frac{2}{D^{\delta}} \cdot p_U(x)$. ■

The next lemma states that upon walking over any graph G (not necessarily an expander), and for any next-step distribution, the probability of a node increasing by a large amount cannot be too large.

Lemma 3.10. *Let $G = (U, V, E)$ be a balanced D-biregular graph. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$. Then, for every $i \in [t]$, and any parameter $a > 2$, there is a subset $S_i \subseteq \{0,1\}^{n=dt}$ with $\Pr[X \in S_i] \ge 1 - D^{-(a-2)}$, such that for every $x \in S_i$,*

$$p_i(x_1, \dots, x_t) \le D^{a+1} \cdot p_{i-1}(x_1, \dots, x_t).$$

Proof: As before, fix an i, and let $p_U = p_{i-1}$ and $p_V = p_i$. For every $v \in V$, define

$$T_v = \left\{ u \in \Gamma(v) : p_U(u) \ge \frac{1}{D^a} p_U(u_v) \right\},$$

and recall that u_v is the neighbor of v with the heaviest probability under p_U. Then, we have that

$$1 = \sum_{v \in V} \left(\sum_{u \in T_v} r_u(u,v) \cdot p_U(u) + \sum_{u \in \Gamma(v) \setminus T_v} r_u(u,v) \cdot p_U(u) \right),$$

however,

$$\sum_{v \in V} \sum_{u \in \Gamma(v) \setminus T_v} r_u(u,v) \cdot p_U(u) \le \frac{1}{D^a} \sum_{v \in V} \sum_{u \in \Gamma(v) \setminus T_v} r_u(u,v) \cdot p_U(u_v) \le \frac{D^2}{D^a}.$$

Therefore:

$$1 - D^{-(a-2)} \leq \sum_{v \in V} \sum_{u \in T_v} r_u(u,v) \cdot p_U(u).$$

We conclude by observing that the RHS measures the density (under X) of a set of x-s for which

$$p_V(x_1, \ldots, x_t) \leq D^{a+1} \cdot p_U(x_1, \ldots, x_t).$$

This is because, for every v, and any $u \in T_v$, $p_U(u) \geq 1/D^a \cdot p_U(u_v)$ and $p_V(v) \leq D \cdot p_U(u_v)$. The set in question is $S_i = \{x \in \mathrm{Supp}(X) : W_{i-1}(x) \in T_{W_{i(x)}}\}$. ■

3.4 Analyzing the Entire Random Walk

In this section we analyze how the vertex probabilities evolve (on average) over multiple steps on the graph. The next lemma says that if many steps add a lot of entropy, and the total entropy lost in "bad" steps is low, then many time steps must have high entropy. We use the notation t' below instead of t, because we will eventually use the lemma on a suffix (of length t') of a sequence of length t.

Lemma 3.11. *Let $a, b, t', k_{\text{capacity}}, k_{\text{start}}$ be positive integers, with $k_{\text{start}} \leq k_{\text{capacity}}$. Let $y_1, \ldots, y_{t'}$ be a sequence of real numbers. Suppose that $\sum_{i:y_i<0} y_i \geq -b$, and that $y_i \geq a$ whenever $y_i \geq 0$.*

Define the sequence $z_0 = k_{\text{start}}$, and

$$z_i = \begin{cases} \min(z_{i-1} + y_i, k_{\text{capacity}}) & y_i \geq 0, \\ \max(z_{i-1} + y_i, 0) & y_i < 0. \end{cases}$$

Then, the number of z_i-s that are smaller than k_{capacity} is at most

$$2 \cdot |\{i : y_i < 0\}| + \frac{k_{\text{capacity}} - k_{\text{start}} + 2b}{a}.$$

Proof: Call z_i *good* if $z_i = k_{\text{capacity}}$. Call z_i *bad* otherwise. Call a step y_i *bad* if $y_i < a$.

First, there must be some good $z_{i^\star}$ in the first

$$\varphi = |\{i : y_i < 0\}| + \frac{k_{\text{capacity}} - k_{\text{start}} + b}{a}$$

of the i-s. This is because even in the worst case, if all bad steps are within the first φ i-s, the remaining $\frac{k_{\text{capacity}} - k_{\text{start}} + b}{a}$ good steps will bring the value of z_i back to capacity from the worst case deficit of $k_{\text{capacity}} - k_{\text{start}} + b$. We thus add φ to our upper bound on the number of z_i-s that are bad. We can then assume that $z_0 = k_{\text{capacity}}$, and count the number of bad z_i-s in this special case. To do so, we write the sequence $y_1, \ldots, y_{t'}$ as $W_0, Y_1, W_1, \ldots, Y_s, W_s$, where:

- Every W_j is a contiguous sequence of i-s such that $y_i \geq a$.
- Every Y_j starts with some $i^\star$ such that $y_{i^\star} < 0$, and ends at the first $i^\star + \ell$ such that

$$\sum_{i \in [i^\star, i^\star+\ell] : y_i < 0} -y_i \leq \sum_{i \in [i^\star, i^\star+\ell] : y_i \geq 0} y_i.$$

- The last Y_s may end prematurely.

Observe that it suffices to count the length of each of the sequences Y_j, as this will be the number of times z_i is bad. Towards this end, denote $b_j = \sum_{i:y_i<a, y_i \in Y_j} -y_i$. Then, the length of each Y_j is

$$|\{i \in Y_j : y_i < 0\}| + |\{i \in Y_j : y_i \geq 0\}| \leq |\{i \in Y_j \mid y_i < 0\}| + \frac{b_j}{a}.$$

Therefore, the sum of the lengths of the Y_i-s is at most $|\{i \mid y_i < 0\}| + \frac{b}{a}$. Therefore, overall, the number of bad z_i-s is at most $\varphi + |\{i : y_i < 0\}| + \frac{b}{a}$. ■

The following lemma is similar to the previous one, and states that if there are many good steps, and small entropy loss overall, then after a large number of steps, the entropy should be large.

Lemma 3.12. *Let $b, t, k_{\text{capacity}}$ be positive integers. Let $Y_1, \dots, Y_t$ be a sequence of real valued random variables over the domain $\{0,1\}^n$. Suppose that $\mathbb{E}\left[\sum_{i=1}^{t} -\mathbf{1}_{Y_i<0} \cdot Y_i\right] \leq b$. Define the random variables $Z_0 = 0$, and*

$$Z_i = \begin{cases} \min(Z_{i-1} + Y_i, k_{\text{capacity}}) & Y_i \geq 0, \\ \max(Z_{i-1} + Z_i, 0) & Y_i < 0. \end{cases}$$

Suppose further that for some ℓ, it holds that $\mathbb{E}\left[\sum_{i=1}^{\ell} \mathbf{1}_{Y_i \geq 0} \cdot Y_i + \sum_{i=1}^{t} \mathbf{1}_{Y_i<0} \cdot Y_i\right] \geq k_{\text{capacity}}$. Then, it also holds that

$$\mathbb{E}[Z_\ell] \geq k_{\text{capacity}} - b.$$

Proof: Define the event $C \subseteq \{0,1\}^n$ to be the set of x-s such that Z_i is at capacity for *some* $i \in [\ell]$. Notice that if $x \in C$ then $Z_\ell(x) \geq k_{\text{capacity}} + \sum_{i=1}^{\ell} \mathbf{1}_{Y_i<0} \cdot Y_i$. If $x \notin C$ then $Z_\ell(x) \geq \sum_{i=1}^{\ell} Y_i(x) \geq \sum_{i=1}^{\ell} \mathbf{1}_{Y_i \geq 0}(x) \cdot Y_i(x) + \sum_{i=1}^{t} \mathbf{1}_{Y_i<0}(x) \cdot Y_i(x)$. So we have

$$\begin{aligned} \mathbb{E}[Z_\ell] &= (1 - \Pr[C]) \cdot \mathbb{E}[Z_\ell \mid \overline{C}] + \Pr[C] \cdot \mathbb{E}[Z_\ell \mid C] \\ &\geq (1 - \Pr[C]) \cdot \mathbb{E}\left[\sum_{i=1}^{\ell} \mathbf{1}_{Y_i \geq 0} \cdot Y_i + \sum_{i=1}^{t} \mathbf{1}_{Y_i<0} \cdot Y_i\right] + \Pr[C] \cdot \mathbb{E}\left[k_{\text{capacity}} + \sum_{i=1}^{\ell} \mathbf{1}_{Y_i<0} \cdot Y_i\right] \\ &\geq k_{\text{capacity}} + \mathbb{E}\left[\sum_{i=1}^{t} \mathbf{1}_{Y_i<0} \cdot Y_i\right] \geq k_{\text{capacity}} - b. \end{aligned}$$

■

We can now state a result that tells us how for a typical $x \sim X$, the "surprise" $h(x_1, \dots, x_i)$ changes as i increases from 1 to t. Since the true surprises (which we denote Z_i) may behave differently than the "worst case" bounds developed in Theorem 3.6, we define a proxy sequence, Z'_i, that captures the worst case behaviour for those surprises.

Lemma 3.13. *Let G be a D-biregular (K, ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, and fix some $0 < \delta \leq 1$. For every $i \in [t]$, recall that we defined*

$$\rho_i = \Pr_{x \sim X}\left[X_i = x_i | \left\{X_{[1,i-1]} = x_{[1,i-1]}\right\}\right) > D^{-\delta}].$$

For every i, let S_i be defined as in Corollary 3.7, and let $k_{\text{capacity}} = k - \log(1/\varepsilon) - 3d$, recalling that $k = \log K$ and $d = \log D$. Define $Z_i(x) = -\log p_i(x)$ for $i \in \{0, \dots, t\}$. Define $Y_i(x) = Z_i(x) - Z_{i-1}(x)$ for $i \in [t]$. Also for $i \in [t]$, let:

$$Y_i'(x) = \begin{cases} \delta d - 1 & x \in S_i \\ -4d & x \notin S_i,\ -4d \le Y_i(x) < \delta d - 1 \\ Y_i(x) & x \notin S_i,\ Y_i(x) < -4d \end{cases}$$

and

$$Z_i'(x) = \begin{cases} \min(Z_{i-1}'(x) + Y_i'(x), k_{\text{capacity}}) & Y_i'(x) \ge 0 \\ \max(Z_{i-1}'(x) + Y_i'(x), 0) & Y_i'(x) < 0 \end{cases}$$

with $Z_0'(x) = 0$. Then, the following holds.

1. *For all $i \in [t]$ and $x \in \mathrm{Supp}(X)$, $Z_i(x) \ge Z_i'(x)$.*
2. *$\mathbb{E}_X[|\{i : Y_i' < 0\}|] \le 4\varepsilon D^{1-\delta} t + 2\sum_{i\in[t]} \rho_i$.*
3. *$\mathbb{E}_X[\sum_i -\mathbf{1}_{Y_i'<0} \cdot Y_i'] \le (16\varepsilon D^{1-\delta} + 20D^{-1})dt + 8d \cdot \sum_{i\in[t]} \rho_i$.*

Proof: We prove Item 1 by fixing any x and inducting on i. As a base case, we have that $Z_0(x) = Z_0'(x) = 0$. For the inductive step, first suppose that $x \in S_i$. Notice that by Corollary 3.7, for each $i \in [t]$, if $x \in S_i$ we have $Z_{i+1}(x) \ge \min\left(Z_i(x) + \delta d - 1, k_{\text{capacity}}\right)$. Then:

$$Z_i(x) \ge \min\left(Z_{i-1}(x) + \delta d - 1, k_{\text{capacity}}\right) \ge \min\left(Z_{i-1}'(x) + Y_i'(x), k_{\text{capacity}}\right) = Z_i'(x).$$

Suppose now that $x \notin S_i$. Then:

$$Z_i(x) = Z_{i-1}(x) + Y_i(x) = \max\left(Z_{i-1}(x) + Y_i(x), 0\right) \ge \max\left(Z_{i-1}'(x) + Y_i'(x), 0\right) = Z_i'(x).$$

The second equality here uses the fact that the Z_i-s are, by definition, always nonnegative. The inequality follows from the inductive hypothesis, and the fact that if $x \notin S_i$ then $Y_i(x) \ge Y_i'(x)$. For Item 2, we simply have

$$\mathop{\mathbb{E}}_X\left[|\{i : Y_i' < 0\}|\right] \le \sum_i \mathop{\mathbb{E}}_X\left[\mathbf{1}_{\overline{S_i}}\right] \le 4\varepsilon D^{1-\delta} t + 2\sum_i \rho_i.$$

For Item 3, we bound $\mathbb{E}_X\left[-\mathbf{1}_{Y_i'<0} \cdot Y_i'\right]$ for each i:

$$\begin{aligned}
&\mathop{\mathbb{E}}_X\left[-\mathbf{1}_{Y_i'<0} \cdot Y_i'\right] \\
&\le 4d \cdot \Pr[x \notin S_i, -4d \le Y_i' < \delta d - 1] + \sum_{a=3}^{\infty} (a+2)d \cdot \Pr[x \notin S_i, -(a+2)d \le Y_i' < -(a+1)d] \\
&\le 4d \cdot \Pr[x \notin S_i] + \sum_{a=3}^{\infty} (a+2)d \cdot \Pr[Y_i' < -(a+1)d] \\
&\le 4d \cdot \Pr[x \notin S_i] + \sum_{a=3}^{\infty} (a+2)d \cdot D^{-(a-2)} \\
&\le 4d \cdot (4\varepsilon D^{1-\delta} + 2\rho_i) + \frac{5D-4}{(D-1)^2} d \le (16\varepsilon D^{1-\delta} + 8\rho_i)d + \frac{20}{D} d.
\end{aligned}$$

The third inequality is from Lemma 3.10 since $\Pr[Y_i' < -(a+1)d] = \Pr[Y_i < -(a+1)d]$ by definition, when $a \geq 3$. The result then follows by linearity of expectation. ■

3.5 Using a Random Stopping Time

With the framework to analyze the behavior of the vertex probabilities over the entire random walk, we can now show what kind of distributions over the vertices we can obtain.

We can first give a weak result, without the need of random stopping time, to give a deterministic condenser for unpredictable sources with a linear entropy gap. This result will also be useful for the two stage construction in Section 4.

Theorem 3.14 (deterministic condensing). *Let G be a D-biregular (K,ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, be a (δ,ρ) unpredictable source for some $0 < \delta \leq 1$. Suppose that $k = \log K$ is such that*

$$k_{\text{capacity}} \leq (\delta d - 1)\left(t - 4\varepsilon D^{1-\delta} t - 2t\rho\right) - (16\varepsilon D^{1-\delta} + 20 D^{-1})dt - 8td\rho,$$

where

$$k_{\text{capacity}} = k - \log(1/\varepsilon) - 3d.$$

Then, $W_t(X_1,\ldots,X_t)$ is η-close to a $k_{\text{capacity}} - \eta dt$ source, where

$$\eta = \sqrt{36\varepsilon D^{1-\delta} + 8\rho}.$$

Proof: First, we have that $\mathbb{E}[Z_t] \geq \mathbb{E}[Z_t']$. By the condition on k_{capacity}, we can apply Lemma 3.12 and get

$$\mathbb{E}[Z_t] \geq \mathbb{E}[Z_t'] \geq k_{\text{capacity}} - (16\varepsilon D^{1-\delta} - 20D^{-1})dt - 8d\sum_i \rho_i,$$

where again, we abbreviate $\rho_i = \rho_i(X,\delta)$. By an averaging argument, with probability at most η over $x \sim X$, we have:

$$\frac{k_{\text{capacity}} - Z_t'(x)}{dt} > \eta.$$

Therefore with probability at most η, we have that $Z_t < k_{\text{capacity}} - \eta dt$. ■

Finally, a random stopping time allows us to get close to capacity.

Theorem 3.15 (random stopping time). *Let G be a D-biregular (K,ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, be a (δ,ρ) unpredictable source for some $0 < \delta \leq 1$. For any given ℓ, suppose that $k = \log K$ is such that*

$$k_{\text{capacity}} \leq (\delta d - 1)\left(\ell - 4\varepsilon D^{1-\delta} t - 2t\rho\right) - (16\varepsilon D^{1-\delta} + 20 D^{-1})dt - 8td\rho,$$

where

$$k_{\text{capacity}} = k - \log(1/\varepsilon) - 3d.$$

Then, the probability over random stopping time $\ell \leq i \leq t$, and a random $x \sim X$, that $Z_i(x)$ is not at capacity (i.e., $Z_i(x) < k_{\text{capacity}}$) is at most

$$\frac{224}{\delta} \cdot \frac{\varepsilon D^{1-\delta} t + \sum_i \rho_i}{t - \ell}.$$

Before proving the theorem, we note that we will commonly use the theorem in the case where $0 < \delta < 1$ and $\frac{1}{t}\sum_i \rho_i$ are constant, and ε is sufficiently is small so that $\varepsilon D^{1-\delta} \approx D^{-\delta}$. In such a scenario, setting, say, $\ell = t/2$ tells us that the error is $O\left(\frac{1}{\delta t}\sum_i \rho_i\right)$.

Proof: The probability over $i \in [\ell, t]$ and $x \sim X$ that $Z_i(x)$ is not at capacity is:

$$\begin{aligned}\sum_x \Pr_X[X = x] \Pr_i[Z_i(x) \text{ is not at capacity}] &\leq \sum_x \Pr_X[X = x] \Pr_i[Z_i'(x) \text{ is not at capacity}] \\ &= \frac{1}{t-\ell} \cdot \mathop{\mathbb{E}}_{x\sim X}\left[\sum_{i=\ell}^{t} \mathbf{1}_{Z_i'(x)<k_{\text{capacity}}}\right],\end{aligned}$$

where we used the notation of Lemma 3.13, and the fact that for every i and x, $Z_i'(x) \leq Z_i(x)$. For any fixed x we can use Lemma 3.11 to bound the number of $Z_i'(x)$-s not at capacity and so we can bound the above expectation by

$$\mathbb{E}\left[\sum_{i=\ell}^{t} \mathbf{1}_{Z_i'(x)<k_{\text{capacity}}}\right] < \mathbb{E}\left[2\left|\{i : Y_i'(x) < 0\}\right| + \frac{k_{\text{capacity}} - Z_\ell'(x) + 2\sum_i -\mathbf{1}_{Y_i'(x)<0} \cdot Y_i'(x)}{\delta d - 1}\right].$$

We can upper bound $\mathbb{E}\left[|\{i : Y_i' < 0\}|\right]$ and $\mathbb{E}\left[\sum_i -\mathbf{1}_{Y_i'<0} \cdot Y_i'\right]$ via Lemma 3.13. We lower bound $\mathbb{E}[Z_\ell']$ via Lemma 3.12, noting that we can apply it because the assumed upper bound on k_{capacity} implies that it is at most $k_{\text{capacity}} \leq \mathbb{E}\left[\sum_{i=1}^{\ell} \mathbf{1}_{Y_i \geq 0} \cdot Y_i + \sum_{i=1}^{t} \mathbf{1}_{Y_i<0} \cdot Y_i\right]$. Therefore,

$$\mathbb{E}[Z_\ell'] \geq k_{\text{capacity}} - (16\varepsilon D^{1-\delta} + 20D^{-1})dt - 8\left(\sum_i \rho_i\right)d.$$

Overall, this gives us:

$$\begin{aligned}\Pr_{i,x}[Z_i(x) \text{ is not at capacity}] &\leq \frac{1}{t-\ell}\left(2\left(4\varepsilon D^{1-\delta}t + 2\sum_i \rho_i\right) + 3 \cdot \frac{(16\varepsilon D^{1-\delta} + 20D^{-1})dt + 8d\sum_i \rho_i}{\delta d - 1}\right) \\ &\leq \frac{1}{t-\ell}\left(2\left(4\varepsilon D^{1-\delta}t + 2\sum_i \rho_i\right) + \frac{6}{\delta} \cdot \left((16\varepsilon D^{1-\delta} + 20D^{-1})t + 8\sum_i \rho_i\right)\right) \\ &\leq \frac{1}{\delta}\frac{224\varepsilon D^{1-\delta}t + 52\sum_i \rho_i}{t-\ell}.\end{aligned}$$

■

We next give a corollary with simpler parameters by setting $\ell = t/2$ and setting ρ sufficiently small. For simplicity, the corollary makes no attempt to optimize the output entropy of the random walk relative to the entropy of the source, which is roughly δdt.

Corollary 3.16. *Let G be a D-biregular (K, ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, be a (δ, ρ) unpredictable source for some $0 < \delta \leq 1$. Let $\varepsilon D^{1-\delta} = D^{-\alpha}$ for some α, and suppose that $d \geq \frac{1000}{\delta \cdot \alpha}$ and $\rho \leq \frac{\delta}{1000}$. Also, suppose that $k = \log K$ is such that*

$$k_{\text{capacity}} \leq (\delta/4)dt,$$

where

$$k_{\text{capacity}} = k - \log(1/\varepsilon) - 3d.$$

Then, the probability over random stopping time $t/2 \le i \le t$, and a random $x \sim X$, that $Z_i(x)$ is not at capacity (i.e., $Z_i(x) < k_{\text{capacity}}$) is at most

$$\frac{500}{\delta}(D^{-\alpha} + \rho).$$

Proof: By the constraints on d and ρ, one can verify that $\delta d - 1 \ge 0.99\delta d$, and every other summand except the first term, $(\delta d - 1)(t/2)$, in

$$(\delta d - 1)\left(t/2 - 4\varepsilon D^{1-\delta}t - 2t\rho\right) - (16\varepsilon D^{1-\delta} + 20D^{-1})dt - 8td\rho$$

is at most $.01\delta dt$. And so overall, the expression is at least $(\delta/4)dt$. Therefore, applying Theorem 3.15 with $\ell = t/2$ yields the result. ■

We also provide a simplified analogue of Theorem 3.14 too, using the same argument, and the fact that $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$

Corollary 3.17. *Let G be a D-biregular (K, ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, be a (δ, ρ) unpredictable source for some $0 < \delta \le 1$. Let $\varepsilon D^{1-\delta} = D^{-\alpha}$ for some α, and suppose that $d \ge \frac{1000}{\delta \cdot \alpha}$ and $\rho \le \frac{\delta}{1000}$. Also, suppose that $k = \log K$ is such that*

$$k_{\text{capacity}} \le (\delta/2)dt,$$

where

$$k_{\text{capacity}} = k - \log(1/\varepsilon) - 3d.$$

Then, $W_t(X_1, \ldots, X_t)$ is η-close to a $k_{\text{capacity}} - \eta dt$ source, where

$$\eta = 6(D^{-\alpha/2} + \rho^{1/2}).$$

Proof: Again, by the constraints, every term except δdt in

$$(\delta d - 1)\left(t - 4\varepsilon D^{1-\delta}t - 2t\rho\right) - (16\varepsilon D^{1-\delta} + 20D^{-1})dt - 8td\rho$$

■

is at most $0.01\delta dt$ and so overall the expression is at least $(\delta/2)dt$. Thus applying Theorem 3.14 yields the result.

4 The Two Stage Construction

So far we have shown our novel analysis that unpredictable sources mix well on an lossless expander with error ε that is sufficiently small relative to the sources entropy rate δ. Unfortunately, due to the limitations of currently known constructions of lossless expanders, one cannot immediately apply our analysis to obtain condensers for unpredictable sources of any constant entropy rate. To handle any constant entropy rate $\delta > 0$, we utilize the idea from [DMOZ23] of a two stage construction, and generalize its analysis to the case of unpredictable sources. We make no attempt to optimize parameters, and only use this section to serve as a proof of concept that the lack of sufficiently good lossless expanders is not an inherent barrier to condensing from any rate. We first recall the construction here.

Definition 4.1. *Let $X = X_1 \circ \cdots \circ X_t$, with each $X_i \in \{0,1\}^d$. We define the new sequence,*

$$X' = X'_1, \ldots, X'_{t/b}$$

with each $X'_i = X_{[(i-1)b+1,ib]} \in \{0,1\}^{db}$. We often refer to b as the epoch or block length.

We can then define the big graph, the small graph and the two level random walk between them.

Definition 4.2. *Let H be a degree $D = D_H$-biregular (K_H, ε_H)-expander on D_G vertices. Let G be a degree D_G-biregular, (K_G, ε_G)-expander on M vertices. Given instructions $x_1, \ldots, x_t$, with $x_i \in \{0,1\}^d$, denote $W_H(x_1, \ldots, x_t)$ as the vertex reached on H after using $x_1, \ldots, x_t$ as instructions. Also, define $X''_i = W_H(X'_i)$, and $W_{i,G}(x''_1, \ldots, x''_{t/b})$ as the vertex reached at step i on G using the x''-s as instructions. We often refer to X''_i as the unpredictable source of the second stage.*

First, we give a general lemma that we use to show that no matter whether the previous small walks $W_H(X'_1), \ldots, W_H(X'_{i-1})$ mixed or not, the next steps will still be unpredictable. We state the theorem with the assumption that the entropy rate $\delta < 0.99$. This assumption simplifies some arguments, and we note that no two stage construction is necessary for such high entropy rates anyways.

Lemma 4.3. *Let $X = X_1, \ldots, X_t$, with $X_i \in \{0,1\}^d$ be a (δ, ρ) unpredictable source, and let $D^{-\alpha_H} = \varepsilon_H D^{1-\delta}$. Let $b > 0$ be any epoch length, and assume that*

- $\delta \leq 0.99$,
- $d \geq \frac{1000}{\delta \cdot \alpha_H}$,
- $\rho \leq \frac{\delta^{16}}{10^{64}}$,

and

$$k_{\text{capacity},H} \leq (\delta/4)db,$$

where

$$k_{\text{capacity},H} = k_H - \log(1/\varepsilon_H) - 3d.$$

Let

$$\eta = 12(D^{-\alpha_H/2} + \rho^{1/32}).$$

Then, $X'' = X''_1, \ldots, X''_{t/b}$ is a $\left(\delta' = \frac{k_{\text{capacity},H} - \eta db}{d_G}, \rho' = 3\eta\right)$ unpredictable source.

Proof: Since $\delta \leq 0.99$, and $d \geq 200$, we have that $\delta d \leq d - 2$. For every $i \in [t]$, define $f_i \colon \{0,1\}^{d(i-1)}$ as follows. On input

$$X_1, \ldots, X_{i-1} = X'_1, \ldots, X'_{\lfloor (i-1)/b \rfloor}, X_{\lfloor (i-1)/b \rfloor \cdot b+1}, \ldots, X_{i-1},$$

f_i outputs

$$f_i(X_1, \ldots, X_{i-1}) = X''_1, \ldots, X''_{\lfloor (i-1)/b \rfloor}, X_{\lfloor (i-1)/b \rfloor \cdot b+1}, \ldots, X_{i-1}.$$

In words, f_i outputs the result of the random walks on H in all the epochs before the epoch containing i, concatenated with the X_j-s in the epoch containing index i, up to but not including i. Now let $A(i,x)$ be the indicator random variable for whether

$$\Pr_{(x_i,\omega)\sim(X_i,f(X_1,\dots,X_{i-1}))}[X_i = x_i | f(X_1,\dots,X_{i-1}) = \omega] \geq 2D^{-\delta}.$$

By Lemma 2.5, for every i, $\mathbb{E}_x[A(i,x)] \leq 5\rho_i^{1/4}$. This implies that $\mathbb{E}_i \mathbb{E}_x[A(i,x)] \leq 5\rho^{1/4}$ (using Jensen's). Thus:

$$\begin{aligned} 5\rho^{1/4}b &= \frac{1}{t/b}\sum_{i=1}^{t}\mathop{\mathbb{E}}_x[A(i,x)] \\ &= \frac{1}{t/b}\sum_{i=1}^{t/b}\sum_{j=1}^{b}\mathop{\mathbb{E}}_x[A((i-1)b+j,x)] \\ &= \frac{1}{t/b}\sum_{i=1}^{t/b}\mathop{\mathbb{E}}_x\left[\sum_{j=1}^{b}A((i-1)b+j,x)\right]. \end{aligned}$$

By Markov's inequality, for at least $1-3\rho^{1/8}$ fraction of $i \in [t/b]$, $\mathbb{E}_x\left[\sum_{j=1}^{b}A((i-1)b+j,x)\right] \leq 3\rho^{1/8}b$, and therefore for such good i

$$\mathop{\mathbb{E}}_{x\sim X}\mathop{\mathbb{E}}_{j\in[b]}[A((i-1)b+j,x)] \leq 3\rho^{1/8}.$$

Now fix any such good block $i \in [t/b]$. By yet another Markov's we see that with probability at least $1-2\rho^{1/16}$ over a fixing of $x''_1,\dots,x''_{i-1} \sim X''_1,\dots,X''_{i-1}$, we have that:

$$\mathop{\mathbb{E}}_{x'\sim X'_i|X''_1,\dots,X''_{i-1}=x''_1,\dots,x''_{i-1}}\mathop{\mathbb{E}}_{j\in[b]}[A((i-1)b+j,x)] \leq 2\rho^{1/16}.$$

In other words, in a good block $i \in [t/b]$, with high probability over result of the outcome of the previous small random walks $X''_1,\dots,X''_{i-1}$, the next block of instructions X'_i is a $(\delta/2, 2\rho^{1/16})$ unpredictable source (using the fact that the true rate $\delta - 1/d > \delta/2$ by the constraints on d). Applying Corollary 3.17 (using the constraints on d and ρ in the statement of this lemma to satisfy the required constraints), we see that conditioned on such a good fixing of the previous small random walks, X''_i is η-close to a $k_{\text{capacity},H} - \eta db$ source.

Thus the probability over a random i, and random X''_i that X''_i is too likely conditioned on its prefix is $\eta + 2\rho^{1/16} + 3\rho^{1/8} \leq 3\eta$. ■

Using Lemma 4.3, we can finally apply Corollary 3.16 to our two-stage construction.

Theorem 4.4. *Let H be a $D = D_H$-biregular (K_H, ε_H)-expander on D_G vertices. Let G be a degree D_G-biregular, (K_G, ε_G)-expander on M vertices. Define:*

- $k_{\text{capacity},H} = k_H - \log(1/\varepsilon_H) - 3d$*, the capacity of H.*
- $k_{\text{capacity},G} = k_G - \log(1/\varepsilon_G) - 3d_{G,}$*, the capacity of G.*

- $D^{-\alpha_H} = \varepsilon_H D^{1-\delta}$, *the expander error for entropy rate* δ.
- $\eta = 12(D^{-\alpha_H/2} + \rho^{1/32})$, *the error of the second stage unpredictable source, according to* Lemma 4.3.
- $\delta' = \frac{k_{\text{capacity},H} - \eta db}{d_G}$, *the designated entropy rate of the second stage unpredictable source with block length* b, *according to* Lemma 4.3.
- $D_G^{-\alpha_G} = \varepsilon_G D_G^{1-\delta'}$, *the expander error for entropy rate* δ.

Let $X = X_1 \circ \cdots \circ X_t$, *each* $X_i \sim \{0,1\}^d$, *be a* (δ, ρ) *unpredictable source. Suppose that:*

- $\delta \leq 0.99$,
- $d \geq \frac{1000}{\delta \cdot \alpha_H}$,
- $\rho \leq \frac{\delta^{16}}{10^{64}}$,
- $d_G \geq \frac{1000}{\delta' \cdot \alpha_G}$,
- $3\eta \leq \frac{\delta'}{1000}$,
- $k_{\text{capacity},H} \leq (\delta/4)db$, *and,*
- $k_{\text{capacity},G} \leq (\delta'/4)d(t/b)$.

Then, the probability over random stopping time $t/(2b) \leq i \leq t/b$, *and a random* $x \sim X$, *that*

$$\Pr[W_{i,G}(X_1'', \ldots, X_{t/b}'') = W_{i,G}(x_1'', \ldots, x_{t/b}'')] > 2^{-k_{\text{capacity},G}},$$

is at most

$$\frac{500}{\delta'} \cdot \left(\varepsilon_G D_G^{1-\delta'} + 2\eta\right) \leq \frac{10^5}{\delta'}(D_G^{-\alpha_G} + D^{-\alpha_H/2} + \rho^{1/32}).$$

We also give the deterministic version without a random stopping time, obtained by applying Corollary 3.17 to the second stage unpredictable source instead.

Theorem 4.5. *Assume the same premise as in* Theorem 4.5. *Then,* $W_{t/b,G}(X_1'', \ldots, X_{t/b}'')$ *is* η'*-close to a* $k_{\text{capacity},G} - \eta' dt$*-source, for*

$$\eta' = D_G^{\alpha_G/2} + 6\left(D^{-\alpha_H/4} + \rho^{1/64}\right)$$

5 Using Explicit Expanders

We now use known constructions of explicit lossless expanders to give our explicit condensers results. We first give a theorem that does not use a two-stage construction, and only requires applying Corollary 3.16 using the expanders from Theorem 2.8.

Theorem 5.1. *Let* $d \in \mathbb{N}$, $\delta, \rho \in (0,1)$ *be constants that satisfy the following constraints:*

- $\delta > \frac{5}{6}$,
- $d > \frac{10^4 c^\star}{\delta} \geq 10^5 c^\star$, *where* $c^\star$ *is the constant from* Theorem 2.8, *and*

- $\rho \leq \frac{1}{2000} \leq \frac{\delta}{1000}$.

For any positive integer t, there exists an explicit function

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=\log(t)-1} \to \{0,1\}^m$$

with $m = \Omega(\delta dt)$, such that for any (δ,ρ) unpredictable source $X = X_1 \circ \cdots \circ X_t$ with each $X_i \sim \{0,1\}^d$, $\mathsf{Cond}(X)$ is $O(D^{-1/6}+\rho)$-close to an $m - O(d)$ source (recalling that $D = 2^d$).

Proof: We utilize, from Theorem 2.8, the fact that for any M and D, there exists an explicit $\left(K = \frac{M}{D^{c^\star}}, \varepsilon = \frac{c^\star}{D^{1/3}}\right)$ expander. Notice that $\varepsilon D^{1-\delta} = c^\star D^{-1/3} = D^{-1/3+\log(c^\star)/d} \triangleq D^{-\alpha} \leq D^{-1/6}$. We use this explicit construction to construct a G on M vertices such that $M = 2^m$ satisfies

$$\frac{\delta}{4} \cdot dt = k_{\text{capacity}} \triangleq k - \log(1/\varepsilon) - 3d = m - O(d).$$

Or in other words, so that $m = (\delta/4)dt + O(d)$. Then, we can apply Corollary 3.16 observing that the constraints on d and ρ are satisfied (we use the fact that $\alpha \geq 1/6$, so that $d > 10^5 \geq \frac{1000}{\delta\alpha}$).

This implies that the function Cond taking input the unpredictable source X, and an independent and uniform $Y \sim [t/2, t]$, and outputs the Y-th vertex in the random walk on G using X, is the desired condenser. ■

As a corollary, we record the result for almost CG sources, using the fact that a $(\delta,\gamma,\rho,\lambda)$ almost CG source is a $(\delta, \gamma+\rho+\lambda)$ unpredictable source.

Corollary 5.2. *Let $d \in \mathbb{N}$, $\delta,\gamma,\rho,\lambda \in (0,1)$ be constants that satisfy the following constraints:*

- $\delta > \frac{5}{6}$,
- $d > 10^5 c^\star$, *and,*
- $\gamma+\rho+\lambda \leq \frac{1}{2000}$.

For any positive integer t, there exists an explicit function

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=\log(t)-1} \to \{0,1\}^m$$

with $m = \Omega(\delta dt)$ such that for any $(\delta,\gamma,\rho,\lambda)$ almost CG source $X = X_1 \circ \cdots \circ X_t$ with each $X_i \sim \{0,1\}^d$, $\mathsf{Cond}(X)$ is $O(D^{-1/6}+\gamma+\rho+\lambda)$-close to an $m - O(d)$ source.

To remove the constraint on the entropy rate δ, we use the two stage construction from Theorem 4.4. Again, we make no attempt here (or in any previous stage of the analysis) to optimize the required dependence between δ and ρ, or the dependence on ρ in the final error. We give the following theorems as a proof of concept that the two stage construction still works, and we note again that with sufficiently good lossless expanders, no two stage construction is necessary, and the dependence on ρ in both senses above is much more natural.

Theorem 5.3 (two-stage timed RWs with unpredictable sources)**.** *Let $d \in \mathbb{N}$, $\delta,\rho \in (0, 0.99)$ be constants such that $\rho \leq \frac{\delta^{32}}{10^{300}}$ and $d \geq \frac{10^6 c^\star}{\delta^2}$, where $c^\star$ is the constant from Theorem 2.8. For any positive integer t, there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=\log t - O(1)} \to \{0,1\}^m$$

with $m = \Omega(\delta dt)$ such that for any (δ,ρ) unpredictable source $X = X_1 \circ \cdots \circ X_t$ with each $X_i \sim \{0,1\}^d$, $\mathsf{Cond}(X,Y)$ is $O\left(D^{-\delta/4}+\rho^{1/32}\right)$-close to an $m - O(d)$ source.

Proof: We utilize Theorem 2.7 to construct H, a size-D_G D-biregular $\left(K_H = \frac{D_G}{c^\star D^2}, \varepsilon_H = \frac{c^\star}{D}\right)$ expander. We will choose the constant D_G later. For the large graph, we again follow Theorem 2.8, that for any M and D, guarantees the existence of an explicit $\left(K = \frac{M}{D^{c^*}}, \varepsilon = \frac{c^\star}{D^{1/3}}\right)$ expander, for M to be chosen soon.

Notice we have

$$\varepsilon_H D^{1-\delta} = c^\star D^{-\delta} = D^{-\delta+\log(c^*)/d} \triangleq D^{-\alpha_H} \le D^{-\delta/2}$$

and set $\eta = 12(D^{-\delta/2} + \rho^{1/32})$, which is the same η defined in Theorem 4.4. In order to apply Theorem 4.4, we wish to choose the parameter b, the number of instructions in each block, appropriately, as this determines $k_{\text{capacity},H} = (\delta/8)db$, which in turn determines D_G and δ'. The main concern is to ensure that the error term $D_G^{-\alpha_G} = \varepsilon_G D_G^{1-\delta'}$ is non trivial and the constraint $d_G \ge \frac{1000}{\delta' \cdot \alpha_G}$ is satisfied. For any b, if we choose the size of H, D_G, such that $k_{\text{capacity},H} = (\delta/4)db$, then:

$$d_G = (\delta/4)db + \log(1/\varepsilon_H) + 5d + \log c^\star \le (\delta/4)db + 6d + \log c^\star.$$

By choosing $b = \frac{10^5 c^*}{\delta}$, we get that $d_G \le 1.01(\delta/4)db$. Hence,

$$\delta' = \frac{k_{\text{capacity},H} - \eta db}{d_G} \ge \frac{(\delta/4)db - \eta db}{1.01(\delta/4)db} \ge \frac{1 - \frac{4\eta}{\delta}}{1.01} \ge 0.99$$

Where we used the constraints on d and ρ to show that $8\eta/\delta < 0.0001$. We can next compute

$$\varepsilon_G D_G^{1-\delta'} \le c^\star D_G^{-1/3+0.01} \le D_G^{-1/3+0.01+\log c^\star/d} = D_G^{-\alpha_G} \le D_G^{-1/6}.$$

Thus, we can verify for our choice of b, that

$$d_G \ge (\delta/4)db \ge 250d \ge 10^6 \ge \frac{1000}{\delta' \cdot \alpha_G}.$$

As before, we choose M, the size of G such that:

$$(\delta'/4)d(t/b) = k_{\text{capacity},G} = k_G - \log(1/\varepsilon_G) - 3d_G = m - O(d_G) = m - O(d).$$

We finally verify that every constraint required to apply Theorem 4.4 is satisfied:

- $d \ge \frac{10^6 c^*}{\delta^2} \ge \frac{1000}{\delta \cdot \alpha_H}$ (since $\alpha_H \ge \delta/2$).
- $\rho \le \frac{\delta^{16}}{10^{64}}$.
- $d_G \ge \frac{1000}{\delta' \cdot \alpha_G}$, as computed above.
- $3\eta = 36(D^{-\alpha_H/2} + \rho^{1/32}) \le \frac{1}{2000} \le \frac{0.99}{1000} = \frac{\delta'}{1000}$, by the constraints on d and ρ in the theorem statement.
- $k_{\text{capacity},H} = (\delta/4)db$ by construction.
- $k_{\text{capacity},G} = (\delta'/4)d(t/b)$ by construction.

Thus, we can conclude that the two-stage construction yields the desired Cond. With error $O\left(D_G^{-1/6} + D^{-\delta/4} + \rho^{1/32}\right)$. Finally, using our choice of b, and that $d_G \leq 1.01(\delta/4)db$, we get that $D_G^{-1/6} \leq D^{-1000} \leq D^{-\delta/4}$. ■

Corollary 5.4 (timed RWs with almost CG sources). *Let $d \in \mathbb{N}$, $\delta, \gamma, \rho, \lambda \in (0,1)$ be constants such that $\rho + \gamma + \lambda \leq \frac{\delta^{32}}{10^{300}}$ and $d > \frac{10^6 c^\star}{\delta^2}$, where $c^\star$ is the constant from Theorem 2.8. For any positive integer t, there exists an explicit function*

$$\mathsf{Cond}\colon \{0,1\}^{n=dt} \times \{0,1\}^{\ell=\log t - O(1)} \to \{0,1\}^m$$

with $m = \Omega(\delta dt)$ such that for any $(\delta, \gamma, \rho, \lambda)$ almost CG source $X = X_1 \circ \cdots \circ X_t$ with each $X_i \sim \{0,1\}^d$, $\mathsf{Cond}(X)$ is $O\left(D^{-1/6} + D^{-\delta/4} + (\rho+\gamma+\lambda)^{1/32}\right)$-close to an $m - O(d)$ source.

Finally, we give a result about deterministic condensing from any rate using the two-stage construction. Utilizing the same constructions of explicit expanders, with the same parameters, but with Theorem 4.5 instead of Theorem 4.4.

Theorem 5.5 (two-stage untimed RWs with unpredictable sources). *Let $d \in \mathbb{N}$, $\delta, \rho \in (0, 0.99)$ be constants such that $\rho \leq \frac{\delta^{32}}{10^{300}}$ and $d \geq \frac{10^6 c^\star}{\delta^2}$, where $c^\star$ is the constant from Theorem 2.8. 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 (δ, ρ) unpredictable source $X = X_1 \circ \cdots \circ X_t$ with each $X_i \sim \{0,1\}^d$, $\mathsf{Cond}(X)$ is β-close to an $m - O(\beta dt) = (1 - O(\beta))m$ source, where $\beta = O\left(D^{-\delta/8} + \rho^{1/64}\right)$.

6 New Random Walks Based Extractors for High Min-Entropy Sources

In this section, we compare the extractor we can obtain using our new random walk analysis with the standard random walk extractor.

First, we record the result for the classical random walk based extractor, that is implied by the expander Chernoff bound, and the sampler it implies. We reproduce the proof in order to illustrate the necessity of having at least $\approx (1-\rho^2)n$ min-entropy, when one aims for error ρ. Indeed, the bottleneck comes from the use of the expander Chernoff bound.

Theorem 6.1 (see, e.g., [Gil98] or Theorem 4.22 in [Vad12]). *Let $G = (V, E)$ be a D-regular λ-spectral expander[17] on M. Let $X = V_1 \circ X_1 \circ \cdots \circ X_{t-1} \sim \{0,1\}^n$ be a uniformly distributed random variable, with $V_1 \sim [M] \equiv \{0,1\}^m$, and $X_i \sim [D] \equiv \{0,1\}^d$. Let $V_1, V_2, \ldots, V_t$ be the sequence of vertices obtained by a random walk that starts at V_0 and uses the instructions $X_1, \ldots, X_{t-1}$. Then, for every set $A \subset \{0,1\}^m$, it holds that*

$$\Pr_{x \sim X}\left[\left|\frac{1}{t}\sum_{i=1}^{t} \mathbf{1}_{V_i \in A} - \frac{|A|}{2^m}\right| \geq \rho\right] \leq e^{-b(1-\lambda)\rho^2 t}$$

for some universal constant $b < 1$.

[17] Letting $\lambda_n \leq \ldots \leq \lambda_1 = 1$ be the eigenvalues of the normalized adjacency matrix of G, we say that G is a λ-spectral expander if $\max\{\lambda_2, -\lambda_n\} \leq \lambda$.

We phrase the statement of the expander Chernoff based extractor, discussed in the introduction, as first fixing a desired error ρ, and asking how large the entropy k of the source must be to achieve that error.

Theorem 6.2 (standard RW-based extractor)**.** *There exists a universal constant C such that the following holds. For every positive integer n, and any $\rho > 0$, there exists an explicit (k, ρ) extractor $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^{\ell = \log n - O(1)} \to \{0,1\}^{m=\Omega(k)}$ for any $k \geq (1 - \rho^2/C)n + \log(1/\rho)$.*

Proof: For M to be chosen later, let $G = (V = [M], E)$ be a D-regular λ-spectral expander, where $\lambda = \frac{1}{2}$, and such explicit expanders are known with $D = O(1)$. The extractor interprets its input x as a length-t random walk on G, namely $x = v_1, x_1, \ldots, x_{t-1}$. Letting $v_1, \ldots, v_t$ be the vertices resulting from that random walk, the extractor, on a seed $i \in [t]$, outputs v_i. For concreteness (and to guarantee a large enough m), we'll choose $t = \frac{n}{4d}$, but the specific constant will not change our theorem's statement. Since $n = m + d(t-1)$, this implies that $m \geq \frac{3n}{4}$, and $\ell = \log n - O(1)$.

To prove that the construction is an (k, ρ) extractor for some sufficiently large k, fix any $A \subseteq \{0,1\}^m$, and let B be the set of $x \in \{0,1\}^n$ such that

$$\left| \frac{1}{t} \sum_{i=1}^{t} \mathbf{1}_{V_i(x) \in A} - \frac{|A|}{2^m} \right| \geq \frac{\rho}{2}.$$

Fix any flat k-source X, and let $Y \sim \{0,1\}^d$ be uniform and independent from X. Conditioning on whether $X \in B$ or not, and then using Theorem 6.1, we have that:

$$\begin{aligned}\left| \Pr_{x \sim X y \sim Y}[\mathsf{Ext}(x,y) \in A] - \Pr_{u \sim U}[u \in A] \right| &\leq \frac{\rho}{2} + \Pr[X \in B] \leq \frac{\rho}{2} + |B| \cdot 2^{-k} \\ &\leq \frac{\rho}{2} + 2^n \cdot 2^{-b(1-\lambda)\rho^2 t} \cdot 2^{-k} = \frac{\rho}{2} + 2^{n-k} \cdot 2^{-(b/2)\rho^2 n}.\end{aligned}$$

To make the second term at most $\rho/2$ (for sufficiently large n), we see that we need $n - k \leq (b/2)\rho^2 n - \log(1/\rho) - 1$, or $k \geq (1 - \rho^2/C)n + \log(1/\rho)$. ■

We can rephrase the result in the language of samplers using Lemma 2.15.

Corollary 6.3. *There exists a universal constant C such that the following holds. For every positive integer n, and any $\rho > 0$, there exists a $\left(2^{-\frac{\rho^2}{C}n}, \rho\right)$ sampler $\Gamma\colon \{0,1\}^n \times \{0,1\}^{\ell = \log n - O(1)} \to \{0,1\}^{m = \Omega(k)}$.*

That is, the confidence parameter's dependence on ρ is exponential in $-\rho^2 n$. Next, we give our new random walk extractor and sampler, utilizing the random walk condenser achieved in the previous sections. But before that, we give the following characterization of high entropy sources as unpredictable sources.

Proposition 6.4. *Let $X \sim \{0,1\}^n$ be a random variable with $H(X) \geq (1-\rho)n$. For any positive integers t, d with $n = dt$, write $X = X_1 \circ \ldots \circ X_t$, with each $X_i \sim \{0,1\}^d$. Then, for any constant $c \in (0,1)$, X is a $(\delta = 1 - \frac{1}{c}, c \cdot \rho)$ unpredictable source.*

Proof: Define $h_i(x) = -\log \Pr[X_i = x_i | X_{[1,i-1]} = x_{[1,i-1]}]$. For every x, we have $h(x) = \sum_i h_i(x) \geq (1-\rho)n$. Therefore, by Markov's inequality, for every x, for at most $c \cdot \rho$ fraction of the i-s, we have that $h_i(x) \leq (1-1/c)d$. Defining the indicator random variable $A(x,i)$ as 1 if $\Pr[X_i = x_i | X_{[1,i-1]} = x_{[1,i-1]}] > D^{-\delta}$, we get:

$$c\rho \geq \mathop{\mathbb{E}}_x \left[\mathop{\mathbb{E}}_i [A(x,i)] \right] = \mathop{\mathbb{E}}_i \left[\mathop{\mathbb{E}}_x [A(x,i)] \right] = \mathop{\mathbb{E}}_i [\rho_i(X,\delta)] \, .$$

$\blacksquare$

We give the following instantiation of our random walk based condenser.

Lemma 6.5. *For every positive integer n, and any $\rho \in (0, \frac{1}{2000})$, there exists a (k, k', ρ) condenser*

$$\mathsf{Cond} \colon \{0,1\}^n \times \{0,1\}^{\ell = \log n - \log\log(1/\rho) - O(1)} \to \{0,1\}^{m = \Omega(k)}$$

for any $k \geq (1-\rho/C)n$, where C is some universal constant, and $k' = m - O(\log(1/\rho))$.

Proof: Let $X \sim \{0,1\}^n$ be a $k = (1-\phi/200)n$ source for some $\phi \leq \rho$ to be chosen later. We divide X into blocks of length $d = \max(10^5 c^\star, 12\log(1/\rho))$ where c^* is the constant from Theorem 2.8. By Proposition 6.4 set with $c = 100$, $X = X_1 \circ \ldots \circ X_t$, each $X_i \sim \{0,1\}^d$, is a $(\frac{1}{100}, \phi/2)$ unpredictable source. Finally, by Theorem 5.1, there is an explicit $\mathsf{Cond} \colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ such that $\mathsf{Cond}(X)$ is $O(D^{-1/6} + \phi/2) = O(\phi)$ close to an $m - O(d)$ source. We choose $\phi = \rho/b$ for some universal constant b such that the $O(\phi)$ term is at most ρ, and finally we set $C = 200b$, so that $k = (1-\rho/C)n$. $\blacksquare$

We recall that Cond is our simple random walk over a lossless expander condenser, since the entropy rate δ is large enough.

For our new random walk based extractor, we will use the high min-entropy extractor from Theorem 2.12. We can compose our condenser with this extractor for a final statement comparable to Theorem 6.2, using $\Delta = \log(1/\rho)$.

Theorem 6.6 (new RW-based extractor)**.** *There exists universal constants $\rho_0 \in (0,1)$ and $C > 1$ such that the following holds. For every positive integer n, and any $\rho \in (0, \rho_0)$, there exists an explicit (k,ρ) extractor*

$$\mathsf{Ext} \colon \{0,1\}^n \times \{0,1\}^{\ell = \log n + O(\log(1/\rho))} \to \{0,1\}^{m = \Omega(k)} \, ,$$

for any $k \geq (1-\rho/C)n$. Moreover, the computation of Ext *involves only walks over suitably chosen expanders.*

We remark that one can implement the extractor in Theorem 2.12 as a function $\mathsf{Ext}(x,y)$ that associates $x \in \{0,1\}^n$ with a vertex in a suitable N-vertex spectral expander, and uses y to take a single step (or multiple independent steps) from x. Thus overall, we obtain a purely random-walks based extractor that first uses a seed Y_1 to pick a random stopping time for the random walk using the original source X on a *lossless expander* on M vertices. This results in a vertex $v \in \{0,1\}^m$. We then continue the random walk starting from v, but now on a suitable *spectral expander* on M vertices, using an additional seed Y_2 to take one additional uniform step (or multiple ones, depending on the degree), on the new expander. Similar to before, the extractor readily implies a random walk based sampler.

Corollary 6.7. *There exists universal constants $\rho_0 \in (0,1)$ and $C > 1$ such that the following holds. For every positive integer n, and any $\rho \in (0,\rho_0)$, there exists a $\left(2^{-\frac{\rho}{C}n}, \rho\right)$ sampler $\Gamma \colon \{0,1\}^n \times \{0,1\}^{\ell = \log n + O(\log(1/\rho))} \to \{0,1\}^{m = \Omega(k)}$.*

References

[AOR$^+$20] Divesh Aggarwal, Maciej Obremski, João Ribeiro, Luisa Siniscalchi, and Ivan Visconti. How to extract useful randomness from unreliable sources. In *Advances in Cryptology–EUROCRYPT 2020: 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 343–372. Springer, 2020.

[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.

[CGR24a] Eshan Chattopadhyay, Mohit Gurumukhani, and Noam Ringach. Condensing against online adversaries. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2024.

[CGR24b] Eshan Chattopadhyay, Mohit Gurumukhani, and Noam Ringach. On the existence of seedless condensers: Exploring the terrain. In *Proceedings of the 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1451–1469. IEEE, 2024.

[CRT23] Itay Cohen, Roy Roth, and Amnon Ta-Shma. HDX condensers. In *Proceedings of the 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1649–1664. IEEE, 2023.

[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.

[DGSX21a] Yevgeniy Dodis, Siyao Guo, Noah Stephens-Davidowitz, and Zhiye Xie. No time to hash: On super-efficient entropy accumulation. In *Advances in Cryptology–CRYPTO 2021: 41st Annual International Cryptology Conference*, 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.

[DMOZ23] Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. Almost Chor–Goldreich sources and adversarial random walks. In *Proceedings of the 55th Annual Symposium on Theory of Computing (STOC)*, pages 1–9. ACM, 2023.

[Gil98] David Gillman. A Chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[GLZ24] Jesse Goodman, Xin Li, and David Zuckerman. Improved condensers for Chor-Goldreich sources. In *Proceedings of the 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1513–1549. IEEE, 2024.

[Gol24] Louis Golowich. New explicit constant-degree lossless expanders. In *Proceedings of the 35th Annual Symposium on Discrete Algorithms (SODA)*, pages 4963–4971. ACM-SIAM, 2024.

[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.

[SU05] Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *Journal of the ACM*, 52(2):172–216, 2005.

[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.

[Tre01] Luca Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–879, 2001.

[TU06] Amnon Ta-Shma and Christopher Umans. Better lossless condensers through derandomized curve samplers. In *Proceedings of the 47th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 177–186. IEEE, 2006.

[TZS06] Amnon Ta-Shma, David Zuckerman, and Shmuel Safra. Extractors from Reed–Muller codes. *Journal of Computer and System Sciences*, 72:786–812, 2006.

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

[XZ24] Zhiyang Xun and David Zuckerman. Near-optimal averaging samplers. *Electron. Colloquium Comput. Complex.*, TR24-097, 2024.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11(4):345–367, 1997.

[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

We give a theorem that deterministic condensing for general sources is impossible.

Claim A.1. *Let $\alpha > 0$ be constant, and let $n > m$ be (sufficiently large) positive integers. Then, there is no (deterministic) function $f\colon \{0,1\}^n \to \{0,1\}^m$ that is a $((1-\alpha)n, (1-\alpha)m+1, \varepsilon)$-condenser, for any $\varepsilon < 1/2$.*

Proof: First, suppose that $m \le \alpha n$. In this case, there exists a $v \in \{0,1\}^m$ such that $|f^{-1}(v)| \ge 2^{(1-\alpha)n}$. Thus, an X that is flat on $f^{-1}(v)$ is a $(1-\alpha)n$ source for which $f(X)$ has no entropy.

Next, suppose that $m > \alpha n$. Let $A = 2^{m-\alpha n}$ and let S a set of size A of $v \in \{0,1\}^m$ such that $f^{-1}(v)$ is largest (that is, the top A elements $v \in \{0,1\}^m$ when sorted according to $|f^{-1}(v)|$). We first claim that $\sum_{v \in S} |f^{-1}(v)| \ge 2^{(1-\alpha)n}$. Indeed, if not, then the $v \in S$ with the smallest preimage must satisfy $|f^{-1}(v)| < \frac{2^{(1-\alpha)n}}{A} = 2^{n-m}$. Therefore, the size of any preimage $f^{-1}(u)$ for $u \notin S$ must also be smaller than 2^{n-m}, and overall, this means the total number of preimages is smaller than $(2^m - A)2^{n-m} + 2^{(1-\alpha)n} = 2^n$, a contradiction.

Now, consider a distribution X that is flat on S. Such a distribution has min-entropy $(1-\alpha)n$, and thus entropy rate $1-\alpha$. However, $f(X)$ is contained in a set of size $A = 2^{m-\alpha n}$. Thus $f(X)$ must be at least $1/2$-far from any $(1-\alpha)m+1$ source, since such a source must have at least half of its probability mass outside of $\mathrm{Supp}(f(X))$. ■

B Making the Random Walk Fully Online

We now demonstrate how the construction of Corollary 3.16 can be made online. Recall that the main trick is to repeatedly embed the current vertex distribution (when it is expected that the distribution on the current graph is "saturated") into a larger graph, so that entropy can continue to accumulate. The embedding procedure is to simply pad the representation of the current vertex with the appropriate amount of 0-s. Thus, for example, if $v \in \{0,1\}^m$ is a vertex in the small graph, and the larger graph has vertices associated with $\{0,1\}^{2m}$, then we simply pad v to $v \circ 0^m$. It is clear such an embedding provides a one-to-one mapping between the vertex distribution on the small graph, and an "isomorphic" vertex distribution on the large graph. Also, it is indeed true that we can continue to apply Theorem 3.6 at each step after embedding, and so we can conclude that the vertex probabilities continue to decrease in expectation at every step. Therefore, it would be nice to say that after walking on the small graph, embedding in a larger graph, and then continuing to walk on the larger graph, **we get a resulting distribution that is identical to the one we would get if we had always just been walking on the larger graph to begin with**. If that were the case, we can readily apply the analysis from the main portion of the paper.

Unfortunately, this is not quite true. The issue is that while walking on the small graph, some paths may have reached the capacity of the small graph quite early. Thus, the vertex probabilities on this path would have stagnated at the capacity of the small graph. On the other hand, the same path, had it been used to walk solely on the larger graph, would not have such stagnation. Therefore, we expect to accumulate less entropy overall when using the embedding trick rather than not.

This issue also affects the choice of random stopping time. Recall that in Section 3.5 we pick a random stopping time between $t/2$ and t, with the understanding that after $t/2$ steps, the vertex probabilities are expected to be close to the capacity of the graph. However, in the streaming case, after $t/2$ steps, the vertex probabilities are only expected to be close to the capacity of the small graph. Therefore, we do not expect a random stopping time between $t/2$ and t to be at the capacity of the large graph with high probability. The solution is to instead pick the size of the large graph accordingly, so that the steps between $t/2$ and $3t/4$ properly "burn-in" the vertex probabilities to the capacity of the large graph. We can then pick a random stopping time between $3t/4$ and t, with the assumption that the total entropy lost, roughly ρdt, is significantly smaller that the entropy gained in $t/4$ steps of the random walk.

Most of the section here will demonstrate that the entropy loss from the embedding trick is manageable, and in the end we can still obtain an online condenser that accumulates most of the entropy of the source. The idea will be to generalize the analysis in Section 3.4 to the case when Z_0, the random variable representing the initial vertex probability, is not identically 0 (as is the case when starting a walk from a fixed vertex), but instead represents the vertex probability obtained after embedding the final vertex in a walk of the smaller graph into the larger graph (these generalizations can be easily seen by inspection of the proofs). For simplicity, we assume that we take $t/2$ steps in the small graph before embedding in the larger graph and taking the remaining

$t/2$ steps.

We first generalize Corollary 3.7 to this scenario.

Theorem B.1. *Let $G = (V = [M], [M], E)$ be a D-biregular (K, ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, and fix some $0 < \delta \le 1$. For every $i \in [t]$, recall that we defined*

$$\rho_i = \Pr_{x \sim X}\left[\Pr\left[X_i = x_i \mid X_{[1,i-1]} = x_{[1,i-1]}\right] > D^{-\delta}\right],$$

Let $f(x_1, \dots, x_{t/2}) \in V$ be some arbitrary function of the first $t/2$ instructions. Define $p_i(x)$ for $i \in [t/2, t]$, as the probability of the vertex reached starting from vertex $f(x_1, \dots, x_{t/2})$, and then using $x_{t/2+1}, \dots, x_i$ as instructions for a random walk, with $p_{t/2}(x) = \Pr[f(X_1, \dots, X_{t/2}) = f(x_1, \dots, x_{t/2})]$.

Then, for every $i \in [t/2+1, t]$ there is a subset $S_i \subseteq \{0,1\}^{n=dt}$ with $\Pr[X \in S_i] \ge 1 - 4\varepsilon D^{1-\delta} - 2\rho_i$, such that for every $x \in S_i$,

$$p_i(x) \le \max\left(\frac{2}{D^\delta} \cdot p_{i-1}(x), \frac{D^{3+\log(1/\varepsilon)/d}}{K}\right).$$

Next, we do a similar generalization for Lemma 3.13, changing only the definition of the initial Z_0 (now $Z_{t/2}$).

Lemma B.2. *Let $G = (V = [M], [M], E)$ be a D-biregular (K, ε) lossless expander. Let $X = X_1 \circ \cdots \circ X_t$, each $X_i \sim \{0,1\}^d$, and fix some $0 < \delta \le 1$. Let $f(x_1, \dots, x_{t/2}) \in V$ be some arbitrary function of the first $t/2$ instructions. Define $Z_{t/2}(x) = Z'_{t/2}(x) = -\log p_{t/2}(x)$.*

For every i, let S_i be defined as in Theorem B.1, and let $k_{\text{capacity}} = k - \log(1/\varepsilon) - 3d$, recalling that $k = \log K$ and $d = \log D$. Define $Z_i(x) = -\log p_i(x)$ for $i \in \{t/2+1, \dots, t\}$. Define $Y_i(x) = Z_i(x) - Z_{i-1}(x)$ and let:

$$Y'_i(x) = \begin{cases} \delta d - 1 & x \in S_i \\ -4d & x \notin S_i,\ -4d \le Y_i(x) < \delta d - 1 \\ Y_i(x) & x \notin S_i,\ Y_i(x) < -4d, \end{cases}$$

and

$$Z'_i(x) = \begin{cases} \min(Z'_{i-1}(x) + Y'_i(x), k_{\text{capacity}}) & Y'_i(x) \ge 0 \\ \max(Z'_{i-1}(x) + Y'_i(x), 0) & Y'_i(x) < 0. \end{cases}$$

Then, the following holds.

1. *For all $i \in [t/2, t]$ and $x \in \mathrm{Supp}(X)$, $Z_i(x) \ge Z'_i(x)$.*

2. *$\mathbb{E}_X[|\{i \in [t/2+1, t] : Y'_i < 0\}|] \le 4\varepsilon D^{1-\delta}\frac{t}{2} + 2\sum_{i \in [t/2+1,t]} \rho_i$.*

3. *$\mathbb{E}_X[\sum_{i \in [t/2+1,t]} -\mathbf{1}_{Y'_i < 0} \cdot Y'_i] \le (16\varepsilon D^{1-\delta} + 20D^{-1})d\frac{t}{2} + 8d \cdot \sum_{i \in [t/2+1,t]} \rho_i$.*

Finally, we generalize Lemma 3.12

Lemma B.3. *Let $b, t, k_{\text{capacity}}$ be positive integers. Let $Y_1, \dots, Y_t$ be a sequence of real valued random variables over the domain $\{0,1\}^n$. Suppose that $\mathbb{E}\left[\sum_{i=1}^t -\mathbf{1}_{Y_i < 0} \cdot Y_i\right] \le b$. Let Z_0 be a random variable, and define:*

$$Z_i = \begin{cases} \min(Z_{i-1} + Y_i, k_{\text{capacity}}) & Y_i \ge 0, \\ \max(Z_{i-1} + Z_i, 0) & Y_i < 0. \end{cases}$$

Suppose further that for some ℓ, it holds that $\mathbb{E}\left[Z_0 + \sum_{i=1}^{\ell} \mathbf{1}_{Y_i \geq 0} \cdot Y_i + \sum_{i=1}^{t} \mathbf{1}_{Y_i < 0} \cdot Y_i\right] \geq k_{\text{capacity}}$. *Then, it also holds that*

$$\mathbb{E}[Z_\ell] \geq k_{\text{capacity}} - b.$$

We now demonstrate a setting of parameters for which one step of the embedding procedure works.

Theorem B.4. *Let $G_1 = (V_1 = [M_1], [M_1], E_1)$ be a D-regular (K_1, ε)-expander on $M_1 = \mathrm{poly}(D) \cdot K_1$ vertices, and let $G_2 = (V_2 = [M_2], [M_2], E_2)$ be a D-regular (K_2, ε)-expander on $M_2 = \mathrm{poly}(D) \cdot K_2$ vertices. Suppose that*

$$k_{1,\text{capacity}} = k_1 - \log(1/\varepsilon) - 3d = (\delta/8)d(t/2) = (\delta/16)dt,$$

and

$$k_{2,\text{capacity}} = k_2 - \log(1/\varepsilon) - 3d = (\delta/8)dt.$$

Let α be such that $\varepsilon D^{1-\delta} = \alpha$. Let $X = X_1 \circ \ldots \circ X_t$, $X_i \sim \{0,1\}^d$, be a (δ, ρ)-unpredictable source for $d \geq \frac{1000}{\delta \cdot \alpha}$ and $\rho \leq \frac{\delta}{1000}$. Consider the process that first uses $x_1, \ldots, x_{t/2}$ as a random walk on G_1 from a fixed start vertex, then embeds the final node into G_2 (by concatenating the appropriate amount of 0), then continues the walk on G using $x_1, \ldots, x_{t/2}$. Define $p_i(x_1, \ldots, x_i)$ as the probability of the vertex reached using $x_1, \ldots, x_i$ (on whichever graph G_1 or G_2 makes sense depending on $i \leq t/2$ or not). Finally, define $Z_i(x) = -\log p_i(x)$ and $Y_i(x) = Z_i(x) - Z_{i-1}(x)$ and define Z_i' and Y_i' as in Lemma B.2.

Assume that $\mathbb{E}[Z_{t/2}(x)] \geq k_{1,\text{capacity}} + \mathbb{E}\left[\sum_{i=1}^{t/2} -\mathbf{1}_{Y_i < 0} \cdot Y_i\right]$. Then the following two properties hold:

- $\mathbb{E}[Z_t] \geq k_{2,\text{capacity}} + \mathbb{E}\left[\sum_{i=1}^{t} \mathbf{1}_{Y_i < 0} \cdot Y_i\right]$
- *The probability over $x \sim X$ and a random stopping time $i \in [3t/4, t]$ that $Z_i(x) < k_{2,\text{capacity}}$ is at most $O\left(\frac{1}{\delta}\left(D^{-\alpha} + \rho\right)\right)$.*

Proof (sketch): To prove both bullet points we use Lemma B.3. For the first point, we apply it on $Z'_{t/2+1}, \ldots, Z'_t$. Indeed we can see that:

$$\begin{aligned}
&\mathbb{E}\left[Z'_{t/2} + \sum_{i=t/2+1}^{t} \mathbf{1}_{Y_i' \geq 0} \cdot Y_i' + \sum_{i=t/2+1}^{t} \mathbf{1}_{Y_i' < 0} \cdot Y_i'\right] \\
&= (\delta/16)dt + \mathbb{E}\left[\sum_{i=t/2+1}^{t} \mathbf{1}_{Y_i' \geq 0} \cdot Y_i'\right] + \mathbb{E}\left[\sum_{i=1}^{t} \mathbf{1}_{Y_i' < 0} \cdot Y_i'\right] \\
&\geq (\delta/16)dt + (\delta/2)d(t/2) + \mathbb{E}\left[\sum_{i=1}^{t} \mathbf{1}_{Y_i' < 0} \cdot Y_i'\right] \geq (\delta/8)dt = k_{2,\text{capacity}}.
\end{aligned}$$

In the last inequality we use the fact that $\mathbb{E}_X[\sum_{i \in [1,t]} -\mathbf{1}_{Y_i' < 0} \cdot Y_i'] \leq (16\varepsilon D^{1-\delta} + 20D^{-1})dt + 8d \cdot \sum_{i \in [1,t]} \rho_i$, and, as before, used the constraints on d and ρ to show each term is at most $0.01\delta dt$. Thus we conclude that $\mathbb{E}[Z_t] \geq \mathbb{E}[Z_t'] \geq k_{2,\text{capacity}} + \mathbb{E}\left[\sum_{i=1}^{t} -\mathbf{1}_{Y_i < 0} \cdot Y_i\right]$.

For the second bullet point, we observe that we can make the same statement using Lemma B.3 up to $3t/4$ instead of t. Indeed, one can verify that:

$$\begin{aligned}
&\mathbb{E}\left[Z'_{t/2} + \sum_{i=t/2+1}^{3t/4} \mathbf{1}_{Y'_i \geq 0} \cdot Y'_i + \sum_{i=t/2+1}^{t} \mathbf{1}_{Y'_i < 0} \cdot Y'_i\right] \\
&\geq (\delta/16)dt + (\delta/2)d(t/4) + \mathbb{E}\left[\sum_{i=1}^{t} \mathbf{1}_{Y'_i < 0} \cdot Y'_i\right] \geq (\delta/8)dt = k_{2,\text{capacity}}.
\end{aligned}$$

And so, $\mathbb{E}[Z_{3t/4}] \geq k_{2,\text{capacity}} + \mathbb{E}\left[\sum_{i=1}^{t} -\mathbf{1}_{Y_i<0} \cdot Y_i\right]$. Thus, we can argue as in Section 3.5 that a random stopping time between $[3t/4, t]$ yields the result. ■

We can now inductively apply the above theorem to give a guarantee in the streaming scenario. We first state the result without the context of streaming.

Theorem B.5. *Let $c > 1$ be some sufficiently large constant. Suppose that $\{G_j\}_{j \in [s]}$ is a sequence of D-regular (K_j, ε)-expanders on $M_j = \text{poly}(D) \cdot K_j$ vertices, such that:*

$$k_{j,\text{capacity}} = k_j - \log(1/\varepsilon) - 3d = (\delta/8)d \cdot 2^{c-1+j}.$$

Let α be such that $\varepsilon D^{1-\delta} = \alpha$. Assume that $d \geq \frac{1000}{\delta \cdot \alpha}$, $\rho \leq \frac{\delta}{1000}$ and that $t = 2^{c+s}$. Let $X = X_1 \circ \ldots \circ X_t$, $X_i \sim \{0,1\}^d$ be such that for every $j \in [s]$, $X_1 \circ \ldots \circ X_{2^{c+j}}$ is a (δ, ρ_j) unpredictable source.

Consider the process that starts with a fixed vertex on G_1, then uses $X_1, \ldots, X_{2^c}$ as a random walk, then embeds the resulting vertex in G_2. In subsequent iterations $j \in [2, s]$ it uses $X_{2^{c+j-1}+1}, \ldots, X_{2^{c+j}}$ to walk on G_j, then embeds the vertex in G_{j+1}. Then, for every $j \in [s]$, using a random stopping time between $\left[\frac{3}{4} \cdot 2^{c+j}, 2^{c+j}\right]$ yields a distribution over the vertices of G_j that is $O\left(\frac{D^{-\alpha}+\rho_j}{\delta}\right)$-close to an $m_j - O(\log 1/\varepsilon)$ source, for $m_j = \Omega(\delta d \cdot 2^{c+j})$.

Proof (sketch): The proof is by induction. In the base case, for sufficiently large c, we'll have $\mathbb{E}[Z_{2^c}(x)] \geq k_{1,\text{capacity}} + \mathbb{E}\left[\sum_{i=1}^{2^c} \mathbf{1}_{Y_i<0} \cdot Y_i\right]$, and also the conclusion will hold true. Inductively, assume that

$$\mathbb{E}[Z_{2^{c+j}}(x)] \geq k_{1,\text{capacity}} + \mathbb{E}\left[\sum_{i=1}^{2^{c+j}} \mathbf{1}_{Y_i<0} \cdot Y_i\right].$$

Then, we can apply Theorem B.4, to see that the above property also holds for $Z_{2^{c+j+1}}$, and that the conclusion of the theorem also holds for $j+1$.[18] ■

Finally, we remark that it is clear that the above process can be implemented in an online fashion, where the state begins with a node in G_1, and is updated by stepping in the expander using instructions from X. When it is time to embed, the length of the state is increased appropriately. Moreover, if a seed of length 2^{c+j-2} is used in the j-th embedding to pick a random stopping time between $\left[\frac{3}{4} \cdot 2^{c+j}, 2^{c+j}\right]$, the seed can be extended by 1, to then represent a random stopping time between $\left[\frac{3}{4} \cdot 2^{c+j+1}, 2^{c+j+1}\right]$ when the $j+1$-th embedding is encountered.

[18] Technically, Theorem B.4 assumes that the walk in the first half is a pure walk with no embedding. However, it is clear that the argument can be adapted when the first half may be a sequence of walks and embeddings. For ease of exposition we ignore this fact in the proof sketch.