On the Error Parameter of Dispersers

Ronen Gradwohl * Guy Kindler † Omer Reingold ‡ Amnon Ta-Shma §

Abstract

Optimal dispersers have better dependence on the error than optimal extractors. In this paper we give the first *explicit* disperser constructions that beat in some parameters the best possible extractors. Our constructions are not strong, and we show that having such explicit *strong* constructions implies a solution to the Ramsey graph construction problem.

1 Introduction

Extractors [19, 9] and dispersers [14] are combinatorial structures with many random-like properties [1]. Extractors and dispersers have found many applications, including simulation with weak sources, deterministic amplification, construction of depth-two superconcentrators, hardness of approximating clique, and much more [8]. For nearly all applications, *explicit* constructions are required.

Extractors and dispersers have several parameters: the input length n, the seed length d, the output length m, the amount of required entropy in the source k and the error parameter ϵ. Two parameters that are of central interest for this paper are the seed length d and the entropy loss that is defined as $\Lambda = k+d-m$ and measures how much randomness is lost by an application of the disperser/extractor.

The currently best explicit extractor construction of Lu, Reingold, Vadhan and Wigderson [7] has seed length $d = O(\log \frac{n}{\epsilon})$ and outputs $m = \Omega(k)$, bits and thus has large entropy loss (isn't this true only for some range of ϵ??). The best *non-explicit* extractor construction has seed length $d = \log(n-k) + 2\log\frac{1}{\epsilon} + \Theta(1)$ and entropy loss $\Lambda = \log\frac{1}{\epsilon} + \Theta(1)$, and this was shown to be tight by [11].

The situation for optimal dispersers is different, though. It can be shown that *non-explicitly* there exist dispersers with seed length $d = \log(n-k) + \log\frac{1}{\epsilon} + \Theta(1)$ and entropy loss $\Lambda = \log\log\frac{1}{\epsilon} + \Theta(1)$, and, again, this was shown to be tight by [11]. In particular we see that optimal dispersers can have shorter seed length and significantly smaller entropy loss

*Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, 76100 Israel. E-mail: `ronen.gradwohl@weizmann.ac.il`.

†Institute for Advanced Study, Princeton, New Jersey. E-mail: `gkindler@ias.edu`.

‡Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, 76100 Israel. E-mail: `omer.reingold@weizmann.ac.il`. Research supported by US-Israel Binational Science Foundation Grant 2002246.

§School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel. E-mail: `amnon@tau.ac.il`.

[1]For a definition, see Section ??

than the best possible extractors. However, up to this work no such explicit constructions were known, and getting such explicit constructions is the goal of this paper .

1.1 Our results

1.1.1 A good disperser for high min-entropies

Our goal is to construct a disperser for very high min-entropies ($k = n - 1$) but with very small error error ϵ, the correct entropy loss $\log\log(\frac{1}{\epsilon})$ and seed is length $1\log(\frac{1}{\epsilon})$.

One such disperser associates the input string with a vertex of an expander graph (see Definition 2.10), and the output with the vertex reached after taking one step on the graph [4]. This disperser has the optimal $d = \log\frac{1}{\epsilon}$, but its entropy loss $\Lambda = d = \log\frac{1}{\epsilon}$ is exponentially larger than the required $\log\log\frac{1}{\epsilon}$.

A better disperser construction was given in Reingold, Vadhan, and Wigderson [13], where constructions of high min-entropy dispersers in which the degree and entropy loss are nearly optimal (what's the seed length?) are given. Their constructions are based on the extractors obtained by the Zig-Zag graph product, but the construction times include factors of $\mathrm{poly}(\frac{1}{\epsilon})$ or even $2^{1/\epsilon}$, so they are inefficient for super-polynomially small error. The reason they incur this cost is that they view their extractors as bipartite graphs, and then define their dispersers as the same graph, but with the roles of the left-hand-side and right-hand-side reversed. This inversion of an extractor is nontrivial, and thus requires much computation.

Our constructions are a new twist on the old theme of using random walks on expander graphs to reduce the error. Formally, we get,

Theorem 1.1. *For every $\epsilon = epsilon(n) > 0$, there exists an explicit $(n-1, \epsilon)$-disperser $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ with $d = (2 + o(1))\log(\frac{1}{\epsilon})$ and entropy loss $\Lambda = (1 + o(1))\log\log(\frac{1}{\epsilon})$.*

1.1.2 error reduction for dispersers

Error reduction takes a disperser with, say, constant error, and converts it to a disperser with the designated (small) error ϵ. One way for achieving error reduction for dispersers was suggested by Ta-Shma and Zuckerman [18], and Reingold, Vadhan, and Wigderson [13], and is obtained by first applying the constant error disperser, and then applying a disperser with only ϵ error, that works well for sources of very high min-entropy (say, $k = n - 1$). Using the disperser of Theorem 1.1 we get:

Theorem 1.2. *(error reduction for dispersers) Suppose there exists an explicit $(k, \frac{1}{2})$-disperser $\mathrm{DSP}_1 : \{0,1\}^n \times \{0,1\}^{d_1} \mapsto \{0,1\}^{m_1}$ with entropy loss Λ_1. Then for every $\epsilon = \epsilon(n) > 0$ there exists an explicit (k, ϵ)-disperser $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ with entropy loss $\Lambda = \Lambda_1 + (1 + o(1))\log\log\frac{1}{\epsilon}$ and $d = d_1 + (2 + o(1))\log\frac{1}{\epsilon}$).*

In particular, if we take the best explicit disperser construction for constant error (stated in Theorem 2.9) we get:

Corollary 1.3. *There exists an explicit (k, ϵ)-disperser $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ with seed length $d = O(\log\frac{n}{\epsilon})$ and entropy loss $\Lambda = O(\log n) + \log\log\frac{1}{\epsilon}$.*

1.1.3 A disperser for low entropies

Finally, we construct a disperser for low min-entropies. The seed length of this disperser is $d = O(k) + (1 + o(1)) \log \frac{1}{\epsilon}$ and the entropy loss is $\Lambda = O(k + \log\log \frac{1}{\epsilon})$. Formally,

Theorem 1.4. *There exists an efficiently constructible* (k, ϵ)*-disperser* DSP : $\{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *with* $d = O(k) + (1 + o(1)) \log \frac{1}{\epsilon}$ *and* $\Lambda = O(k + \log\log \frac{1}{\epsilon})$ *entropy loss.*

Notice that when ϵ is small (say $2^{-n^{2/3}}$) and k is small (say, $O(\log(n))$) both the entropy loss and the seed length are almost optimal up to the constant (the seed length is $1 \cdot \log(\frac{1}{\epsilon})$ rather than $2\log(\frac{1}{\epsilon})$ in the previous construction).

We now shortly discuss what optimality means. At first sight it looks that improving the entropy loss from $O(\log(\frac{1}{\epsilon}))$ to $O(\log\log(\frac{1}{\epsilon}))$ is a huge *exponential* improvement, while improving the seed length from $O(\log(n)) + 2\log(\frac{1}{\epsilon})$ to $O(\log(n)) + 1 \cdot \log(\frac{1}{\epsilon})$ is a small constant factor improvement. However, we believe this approach is false, as we explain below.

Say DSP : $\{0,1\}^{n_1} \times \{0,1\}^{d_1} \to \{0,1\}^{n_2}$ is a (k_1, ϵ_1) disperser. We can view DSP as a bipartite graph with $N_1 = 2^{n_1}$ vertices on the left, $N_2 = 2^{n_2}$ vertices on the right, having an edge (v_1, v_2) in the bipartite graph iff $\mathrm{DSP}(v_1, y) = v_2$ for some $y \in \{0,1\}^{d_1}$. The disperser property then translates to the property that any set $A_1 \subseteq \{0,1\}^{n_1}$ of size $K_1 = 2^{k_1}$, and any set $A_2 \subseteq \{0,1\}^{n_2}$ of size $\epsilon_1 N_2$ have an edge between them. This property is symmetric, and so every disperser DSP from $[N_1]$ to $[N_2]$ can equivalently be thought of as a disperser from $[N_2]$ to $[N_1]$, which we call the em inverse disperser.

It then turns out that having a disperser with seed length with dependence $1 \cdot \log \varepsilon$ on the error, is *equivalent* to the inverse disperser having the the entropy loss dependence on the error being $O(\log\log(\frac{1}{\epsilon}))$ [2]. It therefore follows that having seed length $1 \cdot \log(\frac{1}{\epsilon})$ is an equivalent property to having entropy loss $O(\log\log(\frac{1}{\epsilon}))$, and what looks to us as a huge improvement vs. a small improvement is due to change of scales.

1.1.4 The connection to Ramsey graphs

Our last result shows that constructing a *strong* disperser with $O(\log\log(\frac{1}{\epsilon}))$ entropy loss and $1\log(\frac{1}{\epsilon})$ seed length dependence on the error, implies a solution to the Ramsey graph problem. Formally,

2 Preliminaries

2.1 Dispersers and strong dispersers

Dispersers are formally defined as follows:

Definition 2.1 (disperser). *A bipartite graph* $G = (L, R, E)$ *is a* (k, ϵ)-disperser *if for every* $S \subset L$ *with* $|S| \geq 2^k$, $|\Gamma(S)| \geq (1 - \epsilon)|R|$, *where* $\Gamma(S)$ *denotes the set of neighbors of the vertices in* S.

[2] We remark that the entropy loss lower bound in [11] is obtained by proving a seed length lower bound, and then using this equivalence.

As noted earlier, we will denote $|L| = 2^n$ and $|R| = M = 2^m$. Also, the left-degree of the disperser will be denoted by $D = 2^d$.

We can describe dispersers as functions $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$. DSP is a disperser if, when choosing x uniformly at random from S and r uniformly at random from $\{0,1\}^d$, the distribution of $\mathrm{DSP}(x,r)$ has support of size $(1-\epsilon)M$.

We will also be interested in strong dispersers. Loosely speaking, this means that they have a large support for most seeds. Equivalently, if we were to concatenate the seed to the output of the disperser, then this extended output would have a large support. Here we actually give a slightly more general definition, to include the case of almost-strong dispersers, which have a large support when "most" of the seed is concatenated to the output.

Definition 2.2 (strong disperser). *Denote by $x_{[0..t-1]}$ the first t bits of a string x. A (k,ϵ)-disperser $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ is* strong in t bits *if for every $S \subseteq L$ with $|S| \geq 2^k$, $|\{(\mathrm{DSP}(x,r), r_{[0..t-1]}) : x \in S, r \in \{0,1\}^d\}| \geq (1-\epsilon)2^{m+t}$. DSP is a* strong disperser *if it is strong in all d bits.*

We also need the following proposition, which is implicit in [13].

Proposition 2.3 ([13]). *Let*

- $\mathrm{DSP}_1 : \{0,1\}^{n_1} \times \{0,1\}^{d_1} \mapsto \{0,1\}^{m_1}$ *be a (k,ϵ_1)-disperser with entropy loss Λ_1, and let*
- $\mathrm{DSP}_2 : \{0,1\}^{m_1} \times \{0,1\}^{d_2} \mapsto \{0,1\}^{m_2}$ *be an $(m_1 - \log\frac{1}{1-\epsilon_1}, \epsilon_2)$-disperser with entropy loss Λ_2.*

Then $\mathrm{DSP} : \{0,1\}^{n_1} \times \{0,1\}^{d_1+d_2} \mapsto \{0,1\}^{m_2}$, where

$$\mathrm{DSP}(x, r_1, r_2) = \mathrm{DSP}_2(\mathrm{DSP}_1(x, r_1), r_2),$$

is a (k,ϵ_2)-disperser with entropy loss $\Lambda = \Lambda_1 + \Lambda_2$.

2.2 Extractors

To formally describe extractors, we first need a couple of other definitions:

Definition 2.4 (statistical difference). *For two distributions X and Y over some finite domain, denote the* statistical difference *between them by $\Delta(X,Y)$, where:*

$$\Delta(X,Y) = \frac{1}{2}\sum_{i\in\mathrm{supp}(X\cup Y)} |\Pr[X=i] - \Pr[Y=i]|.$$

X and Y are ε-close *if $\Delta(X,Y) \leq \varepsilon$.*

We also need a measure of the randomness of a distribution.

Definition 2.5 (min-entropy). *The* min-entropy *of a distribution X, denoted by $H_\infty(X)$, is defined as*

$$H_\infty(X) = \min_{i\in\mathrm{supp}(X)} \log\frac{1}{\Pr[X=i]}.$$

Definition 2.6 (extractor). $\mathrm{EXT} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *is a* (k,ϵ)*-extractor if, for any distribution* X *with* $H_\infty(X) \geq k$*, when choosing* x *according to* X *and* r *uniformly at random from* $\{0,1\}^d$*, the distribution of* $\mathrm{EXT}(x,r)$ *is* ϵ*-close to uniform.*

As with dispersers, we also consider strong extractors:

Definition 2.7 (strong extractor). $\mathrm{EXT} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *is a* (k,ϵ)*-extractor if, for any distribution* X *with* $H_\infty(X) \geq k$*, when choosing* x *according to* X *and* r *uniformly at random from* $\{0,1\}^d$*, the distribution of* $(\mathrm{EXT}(x,r), r)$ *is* ϵ*-close to uniform.*

Finally, for both dispersers and extractors, and $S \subseteq (n)$ and $T \subseteq (d)$, denote by $\mathrm{EXT}(S,T) = \{\mathrm{EXT}(s,t) : s \in S, t \in T\}$ and $\mathrm{DSP}(S,T) = \{\mathrm{DSP}(s,t) : s \in S, t \in T\}$.

2.3 Previous explicit constructions

In our construction, we will use the following extractor of Srinivasan and Zuckerman [15] and disperser of Ta-Shma, Umans and Zuckerman [17]:

Theorem 2.8 ([15]). *There exists an efficiently constructible strong* (k,ϵ)*-extractor* $\mathrm{EXT}_{SZ} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *with* $d = O(m + \log \frac{n}{\epsilon})$ *and* $\Lambda = 2\log\frac{1}{\epsilon} + O(1)$*.*

Theorem 2.9 ([17]). *For every* $k < n$ *and any constant* $\epsilon > 0$*, there exists an efficiently constructible disperser* $\mathrm{DSP}_{TUZ} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *with* $d = O(\log n)$ *and* $m = k - 3\log n - O(1)$*.*

2.4 Expander Graphs

A main tool we use in our construction is an expander graph.

Definition 2.10 (expander). *Let* $G = (V,E)$ *be a regular graph with normalized adjacency matrix* A *(the adjacency matrix divided by the degree), and denote by* λ *the second largest eigenvalue (in absolute value) of* A*. Then* G *is an* (N,D,α)-expander *if* G *is* D*-regular,* $|V| = N$*, and* $\lambda \leq \alpha$*.*

An expander G is explicit if there exists an algorithm that, given any vertex $v \in V$ and index $i \in \{0,1,\ldots,D-1\}$, computes the i'th neighbor of v in time $\mathrm{poly}(\log|V|)$. Intuitively, the smaller the value of α, the better the expander, implying better parameters for our construction. Thus, we wish to use graphs with the optimal $\alpha \leq \frac{2}{\sqrt{D}}$, called Ramanujan Graphs.

A minor technicality involving Ramanujan Graphs is that they are explicitly constructible only for certain values of N and D. However, for any given N and D, it is possible to find a description of the Ramanujan Graphs of Lubotzky, Phillips and Sarnak [6] that are sufficient for our needs. These graphs satisfy all the requirements, with $|V| \in [N, (1+\delta)\cdot N]$ and degree D, and can be found in time $\mathrm{poly}(n, \frac{1}{\delta})$ [1, 4].

Since the Ramanujan Graphs of [6] are Cayley graphs, they also have the useful property of being consistently labelled: in any graph $G = (V,E)$, the label of an edge $(u,v) \in E$ is i if following edge i leaving u leads to v. The graph is consistently labelled if for all vertices

$u, v, w \in V$, if $(u, v) \in E$ with label i and $(w, v) \in E$ with label j, then $i \neq j$. Loosely speaking, if we take two walks on a consistently labelled graph following the same list of labels but starting at different vertices, then the walks will not converge into one walk.

The reason expander graphs will be useful is the Expander Path Lemma of Kahale [5]:

Lemma 2.11 ([5]). *Let G be an (N, D, α)-expander, and let $B \subset V$ with density $\beta = \frac{|B|}{|V|}$. Choose $X_0 \in V$ uniformly at random, and let $X_0, \ldots, X_t$ be a random walk on G starting at X_0. Then*

$$\Pr[\forall i, X_i \in B] \leq (\beta + \alpha)^t .$$

3 A high min-entropy disperser

In this section we prove Theorem 1.1. Our technique is motivated by the construction of Cohen and Wigderson [3]. In the context of deterministic amplification, Cohen and Wigderson [3] used a random walk on an expander graph to reduce the error from polynomially small to exponentially small. Translating their work into the current notion of a disperser, their construction has the following parameters: $d = (2 + \alpha) \log(1/\epsilon)$ and $\Lambda = \log \log(1/\epsilon)$, for some constant α. Cohen and Wigderson used this disperser as a sampling procedure, and they were only interested in the case $k \geq n - \frac{1}{\mathrm{poly}(n)}$. In fact, their construction does not attain the above parameters for smaller entropy thresholds, such as $k = n - 1$.

3.1 The basic construction

Lemma 3.1. *For all positive w, t, and δ there exists an $(n-c, \epsilon)$-disperser* $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *computable in time* $\mathrm{poly}(n, \frac{1}{\delta})$ *with:*

- $\epsilon \leq \left(1 - \frac{1}{(1+\delta)2^c} + \frac{2}{2^{w/2}}\right)^t$
- $d = wt + \log t$
- $\Lambda \leq \log t$

Proof. Consider a Ramanujan Graph $G = (V, E)$ with $|V| \in [N, (1 + \delta)N]$ and degree 2^w, which can be constructed in time $\mathrm{poly}(\frac{1}{\delta})$ (see Section 2.4). Associate with an arbitrary set of N vertices of G the strings (n), which in turn correspond to the vertices on the left side of the disperser. Again, for every string $s \in (d)$, we think of $s = r \circ i$, where $r \in (wt)$ and $i \in (\log t)$. The disperser is defined in the same manner as before, namely $\mathrm{DSP}(x, r \circ i) = (y, r)$, where $y \in V$ is the vertex of G reached after taking a walk consisting of the last i steps encoded by r, starting at the vertex associated with x. Let $S \subset (n)$, with $|S| \geq 2^{n-c}$.

For a pair (y, r), denote by $\overleftarrow{r_y}$ the walk r ending at y, backwards. By backwards we mean that if the i'th step of the walk is from v to u, then the $(t-i)$'th step of $\overleftarrow{r_y}$ has the label of the edge from u to v. If (y, r) is bad (i.e. for no $x \in S$ and $i \in (\log t)$, $\mathrm{DSP}(x, r \circ i) = (y, r)$), then the walk $\overleftarrow{r_y}$ starting at vertex y in G never hits a vertex $x \in S$. This implies that in all t steps, the walk stays in $\overline{S}$. Note that because the graph is consistently labelled, there is a bijection from pairs (y, r) to pairs $(y, \overleftarrow{r_y})$. Thus, to bound ϵ, it suffices to bound the probability that a random walk in G stays in $\overline{S}$. We use the Expander Path Lemma (Lemma 2.11)to bound this probability by

$$\epsilon \leq \left(\frac{|\overline{S}|}{|V|} + \lambda\right)^t,$$

where λ is the second largest eigenvalue (in absolute value) of G. Since $|S| \geq \frac{N}{2^c}$,

$$\frac{|\overline{S}|}{|V|} \leq \frac{|V| - \frac{N}{2^c}}{|V|} \leq \frac{(1+\delta)N - \frac{N}{2^c}}{(1+\delta)N} = 1 - \frac{1}{(1+\delta)2^c}$$

Thus,

$$\epsilon \leq \left(1 - \frac{1}{(1+\delta)2^c} + \lambda\right)^t \leq \left(1 - \frac{1}{(1+\delta)2^c} + \frac{2}{2^{w/2}}\right)^t$$

for a Ramanujan Graph. It is clear that $d = wt + \log t$. Since the length of the output is at least $n + wt$, the entropy loss $\Lambda \leq \log t$. □

3.2 Parameters

By using Lemma 3.1 with $w = 6$, $t = \frac{5}{2}$, and $\delta = \frac{1}{\text{poly}(n)}$ we get:

Corollary 3.2. *There exists an explicit* $(n-1, \epsilon)$*-disperser* $\text{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *with* $d = 15 \log \frac{1}{\epsilon}$ *and entropy loss* $\Lambda = \log\log \frac{1}{\epsilon} + O(1)$*.*

For our high min-entropy disperser, we wish to reduce the coefficient in the seed length. Thus, we will compose two different dispersers guaranteed by Lemma 3.1. The effect of composing two dispersers is captured by Proposition 2.3.

The first disperser will have entropy requirement $n - 1$, and error $\epsilon_1 = \frac{2}{2^{w/2}}$. This is chosen in order to satisfy $\epsilon_1 = \lambda$, the second eigenvalue of the expanders used in the disperser constructions, and make the analysis simpler. The second disperser will take a subset of $\{0,1\}^{n_2}$ with error $\epsilon_1 = \frac{2}{2^{w/2}}$ (or, equivalently, entropy requirement $n_2 - \log \frac{1}{1-\epsilon_1}$), and have error ϵ, as desired. Both dispersers will be as guaranteed by Lemma 3.1, with the same w. This means that both expander graphs will have the same degree, and thus the same bound on λ, but their sizes will be different. Note that for the first expander, we do not care too much about the size of $|V|$, and so we choose δ to be some constant, to yield a graph with $|V| \in [N, \frac{3}{2}N]$. For the second expander, however, this size is relevant, so we must choose a smaller δ.

We now prove Theorem 1.1.

Proof. Our disperser DSP will consist of the composition of two other dispersers, an $(n-1, \epsilon_1)$-disperser $\text{DSP}_1 : \{0,1\}^n \times \{0,1\}^{d_1} \mapsto \{0,1\}^{m_1}$ and a $(n - \log \frac{1}{1-\epsilon_1}, \epsilon)$-disperser $\text{DSP}_2 : \{0,1\}^{m_1} \times \{0,1\}^{d_2} \mapsto \{0,1\}^m$. These dispersers will be as guaranteed by Lemma 3.1, with parameters $t_1, w_1, \delta_1 = \frac{1}{2}$ and t_2, w_2, δ_2 respectively. Recall that t_i indicates the length of the walk, w_i is the number of bits needed to specify a step in the expander graph, and δ_i is related to the size of the expander graph. We will have

$$\mathrm{DSP}(x, r_1 \circ r_2 \circ i_1 \circ i_2) = \mathrm{DSP}_2(\mathrm{DSP}_1(x, r_1 \circ i_1), r_2 \circ i_2).$$

Let $w = w_1 = w_2$, a parameter we will choose later. For DSP_1, we wish its error to be $\epsilon_1 \leq \frac{2}{2^{w/2}}$. By Lemma 3.1, $\left(1 - \frac{2}{3} \cdot \frac{1}{2} + \frac{2}{2^{w/2}}\right)^{t_1} = \left(\frac{2}{3} + \frac{2}{2^{w/2}}\right)^{t_1} \leq \frac{2}{2^{w/2}}$. If $\lambda < \frac{1}{6}$ (which it will be), then it suffices to satisfy $(\frac{3}{4})^{t_1} \leq \frac{2}{2^{w/2}}$. This occurs whenever $t_1 \geq \frac{3}{2}w$, so set $t_1 = \frac{3}{2}w$. Thus, DSP_1 will output m_1 bits, with error $\epsilon_1 \leq \frac{2}{2^{w/2}}$. Furthermore, the required seed length $d_1 = t_1 w = \frac{3}{2}w^2$.

We now apply DSP_2 on these strings, with $\delta_2 = \frac{1}{1-\epsilon_1} - 1$. The construction time is $\mathrm{poly}(\frac{1}{\delta_2}) = 2^{O(w)}$. By Lemma 3.1, the error is

$$\begin{aligned}\epsilon &\leq \left(1 - \frac{1-\epsilon_1}{1+\delta_2} + \lambda\right)^{t_2} \\ &< (2\epsilon_1 + \lambda)^{t_2} = \left(\frac{6}{2^{w/2}}\right)^{t_2}\end{aligned}$$

This implies that

$$\begin{aligned}t_2 &= \frac{\log \epsilon}{\log \frac{6}{2^{w/2}}} = \log \frac{1}{\epsilon} \left(\frac{1}{\log \frac{2^{w/2}}{6}}\right) \\ &< \log \frac{1}{\epsilon} \left(\frac{2}{w-6}\right).\end{aligned}$$

The required seed length is

$$d_2 = t_2 w = \frac{2w}{w-6} \log \frac{1}{\epsilon} = (2 + \frac{12}{w-6}) \log \frac{1}{\epsilon}.$$

We now analyze the parameters we obtained in the composition of DSP_1 and DSP_2. By Proposition 2.3, the total seed length of DSP is $d = (2 + \frac{12}{w-6}) \log \frac{1}{\epsilon} + \frac{3}{2}w^2$. If we pick $w = \omega(1)$, but also $w = o(\log \log \frac{1}{\epsilon})$, then $d = (2 + o(1)) \log \frac{1}{\epsilon}$, as claimed. The entropy loss $\Lambda \leq \log[(2 + \frac{12}{w-6}) \log \frac{1}{\epsilon}] + \log \frac{3}{2}w^2$, and with the above choice for w, we get $\Lambda = (1 + o(1)) \log \log \frac{1}{\epsilon}$. □

3.3 Error reduction for dispersers

With a high min-entropy disperser at hand, we get the error reduction for dispersers of Theorem 1.2.

Proof. This is attained by composing DSP_1 with our high min-entropy disperser, as in Proposition 2.3. □

4 Dispersers and Bipartite Ramsey Graphs

Our construction of low min-entropy dispersers was motivated by the connection of such dispersers to bipartite Ramsey graphs.

Definition 4.1. *A bipartite graph $G = (L, R, E)$ with $|L| = |R| = N$ is (s,t)-Ramsey if for every $S \subseteq L$ with $|S| \geq s$ and every $T \subseteq R$ with $|T| \geq t$, the vertices of S and the vertices of T have at least one edge and at least one non-edge between them in G.*

By the probabilistic method, it can be shown that $(2 \log N, 2 \log N)$-Ramsey graphs exist. However, the best known explicit constructions are of (N^δ, N^δ)-Ramsey graphs, for any constant $\delta > 0$ [2]. We now argue that constructions of low min-entropy dispersers may provide new ways of constructing bipartite Ramsey graphs.

Suppose we have some strong (k, ϵ)-disperser DSP that outputs 1 bit, and has seed length $d = s_n + \log \frac{1}{\epsilon}$. Consider the bipartite graph $G = (L, R, E)$, such that there is an edge between $x \in L$ and $y \in R$ if and only if $\mathrm{DSP}(x,y) = 1$. Then for every $S \subset L$ with $|S| \geq 2^k$, and every $T \subset R$ with $|T| > 2\epsilon|R|$, there must be an edge and a non-edge between S and T. To see this, suppose there was no non-edge. Then this implies that $\mathrm{DSP}(S,t) = 1 \circ t$, for all $t \in T$, and for no $t \in T$ does $\mathrm{DSP}(S,t) = 0 \circ t$. This means that the disperser misses more than an ϵ-fraction of the outputs, a contradiction.

How good of a Ramsey graph does this yield? If we set $k = O(\log n)$ and $\epsilon = 2^{-n-s_n}$, then we get a $2^n \times 2^n$ graph that is $(\mathrm{poly}(n), 2^{s_n} + 1)$-Ramsey. If $s_n = O(\log n)$, then this would yield a $(\mathrm{poly}(n), \mathrm{poly}(n))$-Ramsey graph, which is significantly better than any current construction, even for the non-bipartite case.

The following theorem is a generalization of the above for dispersers in which the seed length $d = s_n + t \log \frac{1}{\epsilon}$. Note that the coefficient t plays a crucial role in the quality of the Ramsey graph. As we saw, if $t = 1$, then it is possible to get extremely good Ramsey graphs. On the other hand, if the seed length is $d > 2 \log \frac{1}{\epsilon}$, then the corresponding graph is worse than $(2^k, \sqrt{2^n})$-Ramsey.

Theorem 4.2. *Let $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^1$ be a strong (k,ϵ)-disperser with $d = s_n + t \log \frac{1}{\epsilon}$, where s_n is only a function of n. Let $\epsilon = 2^{-\frac{n-s_n}{t}}$, implying that $d = n$. Define the bipartite graph $G = (L, R, E)$ with $|L| = |R| = 2^n$, such that there is an edge between $x \in L$ and $y \in R$ if and only if $\mathrm{DSP}(x,y) = 1$. Then G is $(2^k, 2^{\frac{t-1}{t}n + \frac{s_n}{t} + 1} + 1)$-Ramsey.*

Proof. Suppose towards a contradiction that G is not $(2^k, 2^{\frac{t-1}{t}n + \frac{s_n}{t} + 1})$-Ramsey, i.e. there exist some $S \subseteq L$ and $T \subseteq R$ with $|S| = 2^k$ and $|T| > 2^{\frac{t-1}{t}n + \frac{s_n}{t} + 1}$, such that all vertices of S are connected (or disconnected) to all vertices of T. Without loss of generality, assume they are connected. This implies that $\mathrm{DSP}(S,T) = 1 \circ x$ for $x \in T$. In particular, this means that for no $x \in T$ is the string $0 \circ x$ in the output of $\mathrm{DSP}(S,(d))$. But $|\{0 \circ x : x \in T\}| = |T| > 2^{\frac{t-1}{t}n + \frac{s_n}{t} + 1} = \epsilon 2^{n+1}$, and $\mathrm{DSP}(S,(d))$ can not miss more than an ϵ-fraction of the 2^{n+1} outputs, so this is a contradiction. $\square$

5 A Low Min-Entropy Disperser

In this section, we construct a disperser with seed length $d = O(\log n) + (1 + o(1)) \log \frac{1}{\epsilon}$ that is almost strong. The following lemma is identical to Theorem 1.4 except that it also states that the disperser is strong in many of its seed bits:

Lemma 5.1. *There exists an efficiently constructible (k, ϵ)-disperser* $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ *with $d = O(k) + (1 + o(1)) \log \frac{1}{\epsilon}$ that is strong in $d - O(\log k + \log \log \frac{1}{\epsilon})$ bits. The entropy loss of* DSP *is $\Lambda = O(k + \log \log \frac{1}{\epsilon})$.*

Consider the following intuition. Given some source with min-entropy $O(\log n)$, we first apply an extractor with m output bits (and think of m as a constant) and $\frac{1}{2}2^{-m}$ error. The error is small enough relative to the output length to guarantee that half of the seeds are "good" in the sense that all possible output strings are hit, and so the error for them is 0. The error over the seeds, however, is large, because only half of the seeds are good. Our next step is then to use a disperser with error half, to sample a good seed. The main advantage of this approach is that we obtain a disperser with a very low error ϵ, by constructing only objects with relatively high error.

We now formalize the above discussion, and prove Theorem 1.4

Proof. We wish to construct a (k, ϵ)-disperser $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$. We use the following two ingredients:

- $\mathrm{EXT} : \{0,1\}^n \times \{0,1\}^{d_E} \mapsto \{0,1\}^m$ that is a $(k, 2^{-m-1})$-extractor, and,
- $\mathrm{DSP}' : \{0,1\}^{k'+\log(\frac{1}{\epsilon})} \times \{0,1\}^{d'} \mapsto \{0,1\}^{d_E}$ that is a $(k', \frac{1}{2})$-disperser, with $d = d_E + d'$.

where we specify the parameters later. Now, Given $x \in \{0,1\}^n$, $r_1 \in \{0,1\}^{k'+\log(\frac{1}{\epsilon})}$, and $r_2 \in \{0,1\}^{d'}$, define

$$\mathrm{DSP}(x, r_1 \circ r_2) \quad = \quad \mathrm{EXT}(x, \mathrm{DSP}'(r_1, r_2)).$$

Fix $S \subseteq \{0,1\}^n$ with $|S| \geq 2^k$. We say a seed s is "good" for S if $|\mathrm{EXT}(S, s)| = 2^m$, and "bad" otherwise. As EXT is a strong extractor, at least half of all seeds $s \in \{0,1\}^{d_E}$ are good for S.

Claim 5.2. *For all S as above, and all but an ϵ-fraction of the $r_i \in \{0,1\}^{k'+\log \frac{1}{\epsilon}}$, $\mathrm{DSP}'(r_i, \{0,1\}^{d'})$ hits a seed s that is good for S.*

Proof. Suppose towards a contradiction that more than an ϵ-fraction of the r_i do not hit any s that is good for S. This implies that for at least $\epsilon \cdot 2^{k'+\log \frac{1}{\epsilon}} = 2^{k'}$ x's, $\mathrm{DSP}'(x, \{0,1\}^{d'})$ misses more than half of the outputs. Consider the set T of all such x's. Then $|\mathrm{DSP}'(T, \{0,1\}^{d'})| < \frac{1}{2} \cdot 2^{d_E}$, contradicting the fact that DSP' is a $(k', \frac{1}{2})$-disperser. □

Recall our disperser construction $\mathrm{DSP}(x, r_1 \circ r_2) = \mathrm{EXT}(x, \mathrm{DSP}'(r_1, r_2))$, and denote by R the length of r_1. Claim 5.2 above shows that for all but an ϵ-fraction of the r_1's,

$\mathrm{DSP}'(r_1, r_2))$ hits an s that is good for S, and $\mathrm{EXT}(S, s)$ hits all of $\{0,1\}^m$. Thus, for all but an ϵ-fraction of the r_1's, $\mathrm{EXT}(S, \mathrm{DSP}'(r_1, r_2))$ hits all of $\{0,1\}^m$. This implies that DSP is a (k, ϵ)-disperser that is strong in R bits.

To complete the proof, let $\mathrm{EXT} = \mathrm{EXT}_{SZ}$ from Theorem 2.8, and let $\mathrm{DSP}' = \mathrm{DSP}_{TUZ}$ from Theorem 2.9, with $k' = O(k)$ and $m = O(k)$. Then $d' = O(\log(k + \log\frac{1}{\epsilon})) \leq O(\log k + \log\log\frac{1}{\epsilon})$, and $d = R + d' = O(k) + (1 + o(1))\log\frac{1}{\epsilon}$. Thus, DSP is strong in $R = d - d'$ bits, as claimed. Finally, the entropy loss is $O(k)$ from the application of EXT_{SZ}, and an additional $O(\log k + \log\log\frac{1}{\epsilon})$ from the seed of DSP_{TUZ}. Thus, the total entropy loss is $\Lambda = O(k + \log\log\frac{1}{\epsilon})$. □

Plugging $k = O(\log n)$ we get:

Corollary 5.3. *There exists an efficiently constructible* $(O(\log n), \epsilon)$*-disperser* $\mathrm{DSP} : (n) \times (d) \mapsto (m)$ *with* $d = O(\log n) + (1 + o(1))\log\frac{1}{\epsilon}$ *that is strong in* $d - O(\log\log n + \log\log\frac{1}{\epsilon})$ *bits. The entropy loss of* DSP *is* $\Lambda = O(\log n + \log\log\frac{1}{\epsilon})$.

As we saw above, having such a strong disperser would solve the Ramsey graph construction problem.

6 Discussion

In this work we addressed one aspect of sub-optimality in dispersers, the dependence on ϵ, and constructed min-entropy dispersers in which the entropy loss is optimally dependent on ϵ.

One aspect that follows from this work, is that strong vs. non-strong issue becomes crucial when dealing with objects that are truly close to being optimal.

Another aspect is the following. Say $\mathrm{DSP} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, ϵ) disperser with small error ϵ and close to optimal entropy-loss Λ. For any source $X \subseteq \{0,1\}^n$ of size 2^k, DSP is close to being one-to-one on X (to be more precise, every image is expect to have 2^Λ pre-images, and the disperser property guarantees we are not far from that). In particular every such disperser is also an extractor with some worse error, and if we fix ϵ to be some constant, this disperser would actually be an extractor with constant error (where the constant is greater than $\frac{1}{2}$).

It follows that if we could improve Corollary 1.3 to have entropy loss independent of n, we would get an extractor with constant error, optimal seed length and constant entropy loss! No current extractor construction can attain such entropy loss, regardless of the error, without incurring a large increase in seed length (generally, $d = \log^2 n$ is necessary). So perhaps such dispersers will provide a new means of obtaining better extractors.

Note that if we also had an error reduction for extractors that did not increase the entropy loss too much, then we could get better extractors for any error. Using the techniques of Lu, Reingold, Vadhan and Wigderson [7], such an error reduction would follow from constructions of mergers with minimal entropy loss.

References

[1] N. Alon, J. Bruck, J. Naor, M. Naor and R. Roth. Construction of asymptotically good, low-rate error-correcting codes through pseudo-random graphs. In *Transactions on Information Theory*, 38:509-516, 1992. IEEE.

[2] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. To appear in STOC 2005.

[3] Aviad Cohen and Avi Wigderson. Dispersers, deterministic amplification, and weak random sources. In *Proc. of 30th FOCS*, pages 14-19, 1989.

[4] Oded Goldreich, Avi Wigderson. Tiny families of functions with random properties: a quality-size trade-off for hashing. Random Structures and Algorithms 11:4, pages 315-343, 1997.

[5] Nabil Kahale. Better expansion for Ramanujan graphs. In *Proceedings of 33rd FOCS*, 1992, pages 398-404.

[6] A. Lubotzky, R. Phillips, and P. Sarnak. Explicit expanders and the Ramanujan conjectures. In *Proceedings of 18th STOC*, 1986, pages 240-246.

[7] C. Lu, O. Reingold, S. Vadhan, and A. Wigderson. Extractors: Optimal up to constant factors. In *Proceedings of 35th ACM STOC*, 2003.

[8] Noam Nisan and Amnon Ta-Shma. Extracting randomness: A survey and new constructions. *Journal of Computer and System Sciences*, 58(1):148-173, 1999.

[9] Noam Nisan and David Zuckerman. More deterministic simulation in logspace. In *25th ACM Symposium on Theory of Computing*, 1993, pages 235-244.

[10] Ran Raz. Extractors with weak random seeds. To appear, STOC 2005.

[11] Jaikumar Radhakrishnan and Amnon Ta-Shma. Tight bounds for depth-two superconcentrators. In *38th Annual Symposium on Foundations of Computer Science*, pages 585-594, Miami Beach, Florida, 20-22 October 1997. IEEE.

[12] Ran Raz, Omer Reingold, and Salil Vadhan. Error reduction for extractors. In *40th Annual Symposium on Foundations of Computer Science*, 1999. IEEE.

[13] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proc. 41st FOCS*, pages 3-13, 2000.

[14] Michael Sipser. Expanders, randomness, or time versus space. *Journal of Computer and System Sciences*, 36(3):379-383, June 1988.

[15] Aravind Srinivasan and David Zuckerman. Computing with very weak random sources. In *35th Annual Symposium on Foundations of Computer Science*, 1994, pages 264-275.

[16] Amnon Ta-Shma. Almost optimal dispersers. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 196-202, Dallas, TX, May 1998. ACM.

[17] Amnon Ta-Shma, Christopher Umans, and David Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *Proc. 33rd STOC*, pages 143-152, 2001.

[18] Amnon Ta-Shma and David Zuckerman. Personal communication.

[19] David Zuckerman. General weak random sources. In *31st Annual Symposium on Foundations of Computer Science*, 1990, pages 245-1993.