Relaxed Locally Correctable Codes in Computationally Bounded Channels*

Jeremiah Blocki, Venkata Gandikota, Elena Grigorescu, and Samson Zhou

1Department of Computer Science, Purdue University. {elena-g, jblocki}@purdue.edu
2College of Information and Computer Science, University of Massachusetts. gandikota.venkata@gmail.com
3School of Informatics, Computing, and Engineering, Indiana University. samsonzhou@gmail.com

Abstract—Error-correcting codes that admit local decoding and correcting algorithms have been the focus of much recent research due to their numerous applications. An important goal is to obtain the best possible tradeoffs between the number of symbols of the codeword that the local decoding algorithm must examine (the locality of the task), and the amount of redundancy in the encoding (the information rate).

In Hamming’s classical adversarial channel model, the current tradeoffs are dramatic, allowing either small locality, but superpolynomial blocklength, or small blocklength, but high locality. However, in the computationally bounded, adversarial channel model, proposed by Lipton (STACS 1994), constructions of locally decodable codes suddenly exhibit small locality and small blocklength, but these constructions require strong trusted setup assumptions e.g., Ostrovsky, Pandey and Sahai (ICALP 2007) construct private locally decodable codes in the setting where the sender and receiver already share a symmetric key.

We study variants of locally decodable and locally correctable codes in computationally bounded, adversarial channels, in a setting with no public-key or private-key cryptographic setup. The only setup assumption we require is the selection of the public parameters (seed) for a collision-resistant hash function. Specifically, we provide constructions of relaxed locally correctable and relaxed locally decodable codes over the binary alphabet, with constant information rate, and poly-logarithmic locality.

Our constructions, which compare favorably with their classical analogs, crucially employ collision-resistant hash functions and local expander graphs, extending ideas from recent cryptographic constructions of memory-hard functions.

I. INTRODUCTION

Classically, an error-correcting code is a tuple \((\text{Enc}, \text{Dec})\) of encoding and decoding algorithms employed by a sender to encode messages, and by a receiver to decode them, after potential corruption by a noisy channel during transmission. Specifically, the sender encodes a message \(m\) of \(k\) symbols from an alphabet \(\Sigma\) into a codeword \(c\) of block-length \(n\) consisting of symbols over the same alphabet, via \(\text{Enc} : \Sigma^k \rightarrow \Sigma^n\). The receiver uses \(\text{Dec} : \Sigma^n \rightarrow \Sigma^k\) to recover the message from a received word \(w\) in \(\Sigma^n\), a corrupted version of some \(\text{Enc}(m)\). Codes over the binary alphabet \(\Sigma = \{0,1\}\) are preferred in practice. The quantities of interest in designing classical codes are the information rate, defined as \(k/n\), and the error rate, which is the tolerable fraction of errors in the received word. Codes with both large information rate and large error rate are most desirable.

In modern uses of error-correcting codes, one may only need to recover small portions of the message, such as a single bit. In such settings, the decoder may not need to read the entire received word \(w\) in \(\Sigma^n\), but only read a few bits of it. Given an index \(i \in [n]\) and oracle access to \(w_i\), a local decoder must make only \(q = o(n)\) queries into \(w_i\), and output the bit \(m_i\). Codes that admit such fast decoders are called locally decodable codes (LDCs) [1], [2]. The parameter \(q\) is called the locality of the decoder. A related notion is that of locally correctable codes (LCCs). LCCs are codes for which the local decoder with oracle access to \(w\) must output bits of the codeword \(c\), instead of bits of the message \(m\). LDCs and LCCs have widespread applications in many areas of theoretical computer science, including private information retrieval, probabilistically checkable proofs, self-correction, fault-tolerant circuits, hardness amplification, and data structures (e.g., [3]–[9] and surveys [10], [11]). However, constructions of such codes suffer from apparently irreconcilable tension between locality and rate: existing codes with constant locality have slightly subexponential blocklength [12]–[14], and

* A full version of this paper is accessible at: https://arxiv.org/abs/1803.05652

This work was done in part while V.G. was a postdoc at the Department of Computer Science, JHU; and S.Z. was a graduate student at the Department of Computer Science, Purdue University. Research of E.G. and S.Z. was supported in part by NSF CCF-1649515 and by a grant from the Purdue Research Foundation. Research of J.B. was supported in part by NSF CNS-1755708.
codes of linear blocklength have slightly subpolynomial query complexity [15]. For example, see surveys by Yekhanin [16] and by Kopparty and Saraf [17].

Ben-Sasson et al. [18] propose the notion of relaxed locally decodable codes (RLDCs) that remedies the dramatic tradeoffs of classical LDCs. In this notion the decoding algorithm is allowed to output ⊥ sometimes, to signal that it does not know the correct value; however, it should not output an incorrect value too often. More formally, given $i \in [k]$ and oracle access to the received word $w$ assumed to be relatively close to some codeword $c = \text{Enc}(m) \in \Sigma^n$, the local decoder (1) outputs $m_i$ if $w = c$; (2) otherwise, with probability $2/3$ outputs either $m_i$ or ⊥; and (3) the set of indices $i$ such that the decoder outputs $m_i$ (the correct value) with probability $2/3$ is of size $\geq \rho \cdot k$ for some constant $\rho > 0$. The relaxed definition allows them to achieve RLDCs with constant query complexity and blocklength $n = k^{1+\epsilon}$.

Recently, Gur et al. [19] introduce the analogous notion of relaxed locally correctable codes (RLCCs). In particular, upon receiving a word $w \in \Sigma^n$ assumed to be close to some codeword $c$, the decoder: (1) outputs $c_i$ if $w = c$; (2) outputs either $c_i$ or ⊥ with probability $2/3$, otherwise; and (3) the set of indices $i$ such that the decoder outputs $c_i$ with probability $2/3$ is of size $\rho \cdot n$, for some $\rho > 0$. In fact, [19] omits condition (3) in their definition, since the first two conditions imply the 3rd, for codes with constant locality that can withstand a constant fraction of error [18]. The reduction from [18], however, does not maintain the asymptotic error rate, and in particular, in the non-constant query complexity regime, the error rate becomes subconstant. Since our results work in the $\omega(1)$-query regime, we will build codes that achieve the 3rd condition as well (for constant error rate). The results in [19] obtain significantly better parameters for RLCCs than for classical LCCs; namely, they construct RLCCs with constant query complexity, polynomial block length, and constant error rate, and RLCCs with quasipolynomial query complexity, linear blocklength (constant rate), with the caveat that the error rate is subconstant. These results immediately extend to RLDCs, since their codes are systematic, meaning that the initial part of the encoding consists of the message itself.

In this work we study RLDCs and RLCCs in the more restricted, yet natural, computationally bounded adversarial channel, introduced by Lipton [20]. All the above constructions of local codes assume a channel that may introduce a bounded number of adversarial errors, and the channel has as much time as it needs to decide what positions to corrupt. Lipton argued that one can always reasonably assume that an adversarial channel is computationally bounded and can be modeled as polynomial time probabilistic (PPT) algorithms. Variants of this model have been initially studied for classical error-correcting codes [20]–[24] to show better error rate capabilities than in the Hamming model. See full version for further details on related work [25].

Other work has focused on the construction of locally decodable codes when the sender and receiver have already exchanged cryptographic keys. Ostrovsky et al. [26] construct a “private key” locally decodable codes with constant rate and a small (superconstant) locality against computationally bounded channels. Similarly, Hemenway and Ostrovsky [27] and Hemenway et al. [28] construct public-key LDGs i.e., the encoding algorithm uses a secret key, but the decoder only needs to know the sender’s public key. A crucial difference is that in our constructions of RLDCs and RLCCs, the encoding algorithm is completely public and does not require the sender to use any secret key. Instead our constructions only rely on the existence of a collision-resilient hash function $H$, which will be used in both the encoding and decoding algorithms. In practice, one could instantiate $H$ with SHA3 and no further setup assumptions are needed.

### II. Our Contributions

We introduce the notions of a computationally relaxed locally correctable codes (CRLCCs) which is analogous to RLCCs when the channel is computationally bounded. We construct CRLCCs with constant information and error rates as well as polylogarithmic locality, improving on [19] for bounded channels. Our codes are systematic, and therefore give computationally relaxed locally decodable codes (CRLDCs).

Since these codes interact with an adversarial channel, their strength is not only measured in their error correction and locality capabilities (as is the case for RLCCs/RLDCs in the Hamming channel), but also in the security they provide against the channel. We present these codes while describing how they interact with the channel, in order to make the analogy with the classical setting. We use the notation $\text{Enc}$ and $\text{Dec}$ to denote encoding and decoding algorithms.

**Definition 2.1:** A local code is a tuple $(\text{Gen}, \text{Enc}, \text{Dec})$ of probabilistic algorithms such that

- $\text{Gen}(\lambda)$ takes as input security parameter $\lambda$ and generates a public seed $s \in \{0, 1\}^\lambda$. This public seed $s$ is fixed once and for all.
• Enc takes as input the public seed $s$ and a message $x \in \Sigma^n$ and outputs a codeword $c = \text{Enc}(s, x)$ with $c \in \Sigma^r$.

• Dec takes as input the public seed $s$, an index $i \in [n]$, and is given oracle access to a word $w \in \Sigma^n$. Dec$(s, i)$ outputs a symbol $b \in \Sigma$ (which is supposed to be the value at position $i$ of the closest codeword to $w$).

We say that the code is efficient if Gen, Enc, Dec are all probabilistic polynomial time (PPT) algorithms, and we say that the (information) rate of the code is $k/n$.

Definition 2.2: A computational adversarial channel $\mathcal{A}$ with error rate $\tau$ is an algorithm that interacts with a local code $(\text{Gen}, \text{Enc}, \text{Dec})$ in rounds, as follows. In each round of the execution, given a security parameter $\lambda$,

1. Generate $s \leftarrow \text{Gen}(1^\lambda)$; $s$ is public, so Enc, Dec, and $\mathcal{A}$ have access to $s$. This public seed $s$ is fixed once and for all.
2. The channel $\mathcal{A}$ on input $s$ hands a message $x$ to the sender.
3. The sender computes $c = \text{Enc}(s, x)$ and hands it back to the channel (in fact the channel can compute $c$ without this interaction).
4. The channel $\mathcal{A}$ corrupts at most $\tau n$ entries of $c$ to obtain a word $w \in \Sigma^n$ and selects a challenge index $i \in [n]$; $w$ is given to the receiver’s $\text{Dec}$ with query access along with the challenge index $i$.
5. The receiver outputs $b \leftarrow \text{Dec}^w(s, i)$.
6. We define $\mathcal{A}(s)$’s probability of fooling $\text{Dec}$ on this round to be $p_{\mathcal{A}, s} = \Pr[b \not\in \{\perp, c_i\}]$, where the probability is taken only over the randomness of the Dec$^w(s, i)$. We say that $\mathcal{A}(s)$ is $\gamma$-successful at fooling Dec if $p_{\mathcal{A}, s} > \gamma$. We say that $\mathcal{A}(s)$ is $\rho$-successful at limiting Dec if $\Pr[\text{Dec}^w(s, j) = c_j] > \frac{\rho}{3}$. We use Fool$\mathcal{A}, s(\gamma, \tau, \lambda)$ (resp. Limit$\mathcal{A}, s(\rho, \tau, \lambda)$) to denote the event that the attacker was $\gamma$-successful at fooling Dec (resp. $\rho$-successful at limiting Dec) on this round.

We now define our secure RLCC codes against computational adversarial channels.

Definition 2.3 (Computational) Relaxed Locally Correctable Codes (CRLCC): A local code $(\text{Gen}, \text{Enc}, \text{Dec})$ is a $(\gamma, \tau, \rho, \gamma(\cdot), \mu(\cdot))$-CRLCC against a class $\mathcal{A}$ of adversaries if Dec$^w$ makes at most $q$ queries to $w$ and satisfies the following:

1. For all public seeds $s$ if $w \leftarrow \text{Enc}(s, x)$ then Dec$^w(s, i)$ outputs $b = (\text{Enc}(s, x))_i$.
2. For all $\mathcal{A} \in \mathcal{A}$ we have $\Pr[\text{Fool}_{\mathcal{A}, s}(\gamma(\lambda), \tau, \lambda)] \leq \mu(\lambda)$, where the randomness is taken over the selection of $s \leftarrow \text{Gen}(1^\lambda)$ as well as $\mathcal{A}$’s random coins.
3. For all $\mathcal{A} \in \mathcal{A}$ we have $\Pr[\text{Limit}_{\mathcal{A}, s}(\rho, \tau, \lambda)] \leq \mu(\lambda)$, where the randomness is taken over the selection of $s \leftarrow \text{Gen}(1^\lambda)$ as well as $\mathcal{A}$’s random coins.

When $\mu(\lambda) = 0$, $\gamma(\lambda) = \frac{1}{2}$ is a constant and $\mathcal{A}$ is the set of all (computationally unbounded) channels, we say that the code is a $(\gamma, \tau, \rho, \gamma)$-CRLCC. When $\mu(\cdot)$ is a negligible function and $\mathcal{A}$ is restricted to the set of all probabilistic polynomial time (PPT) attackers, we say that the code is a $(\gamma, \tau, \rho, \gamma)$-CRLCC (computational relaxed locally correctable code).

We say that a code that satisfies conditions 1 and 2 is a Weak CRLCC, while a code that satisfies conditions 1, 2 and 3 is a Strong CRLCC code.

We construct Weak and Strong CRLCCs against PPT adversaries, under the assumption that Collision-Resistant Hash Functions (CRHF) exist. Briefly, a CRHF function is a pair $(\text{Gen}H, H)$ of probabilistic polynomial time (PPT) algorithms, where $\text{Gen}H$ takes as input a security parameter $\lambda$ and outputs a public seed $s \in \{0, 1\}^*$; the function $H : \{0, 1\}^* \times \Sigma^* \rightarrow \Sigma^n(\lambda)$, takes as input the seed $s$ and a long enough input that is hashed into a string of length $\ell(\lambda)$. We note that $H$ is deterministic upon fixing $s$. The value $\ell(\lambda)$ is the length of the hash function. $(\text{Gen}H, H)$ is said to be collision-resistant if for all PPT adversaries that take as input the seed $s$ generated by $\text{Gen}(1^\lambda)$, the probability that a collision pair $(x, x')$ is produced, i.e. such that $H(s, x) = H(s, x')$ and $x \neq x'$, is negligible in $1^\lambda$.

Theorem 2.4, our main result, states that it is possible to construct a constant rate Strong CRLDC with polylog locality that withstands a constant error rate $\tau$. Furthermore, the information rate $r(\tau)$ approaches 1 as $\tau$ approaches 0. By contrast, the classical RLCCs of [19] achieve constant information rate, but subconstant error rate and $(\log n)^{O(\log \log n)}$ query complexity in the Hamming channel.

Theorem 2.4: Assuming the existence of a collision-resistant hash function $(\text{Gen}H, H)$ with length $\ell(\lambda)$, there exists a constant $0 < \tau' < 1$ and negligible functions $\mu(\cdot), \gamma(\cdot)$ such that for all $\tau \leq \tau'$ there exist constants $0 < r(\tau), \rho(\tau) < 1$ such that there exists a $(\ell(\lambda) \cdot \log \log n, \tau, \rho(\tau), \gamma(\lambda), \mu(\cdot))$-Strong CRLCC of blocklength $n$ over the binary alphabet with rate $r(\tau)$ where $r(\cdot)$ and $\rho(\cdot)$ have the property that $\lim_{\tau \rightarrow 0} r(\tau) = 1$ and $\lim_{\tau \rightarrow 0} \rho(\tau) = 1$. In particular,
that the node is graph $G$.

Our constructions are systematic and imply the existence of a Strong CRLDC with the same parameters.

III. TECHNICAL INGREDIENTS

At a technical level our construction uses two main building blocks: local expander graphs and collision resistant hash functions. Below we describe how we use these tools to construct Weak CRLCCs. We will also briefly sketch how the construction can be modified to obtain Strong CRLCCs though we defer a complete treatment to the full version of the paper [25].

Weak CRLCCs (Encoding): Our construction of Weak CRLCCs involves labeling a $k$-node directed acyclic graph $G$ with a collision resistant hash function $H(s,\cdot)$. We require that $G$ satisfies key combinatorial property called $\delta$-local expansion described later. Given an input word $x = (x_1 \ldots x_k)$ (broken up into bit strings of length $\ell(\lambda)$) and a $k$ node local expander graph $G$, the label of node $v$ with parents $v_1, \ldots, v_d$ is computed as $\ell_{v,s} = H(s, x_v \circ \ell_{v_1,s} \circ \ldots \circ \ell_{v_d,s}) \in \{0,1\}^{\ell(\lambda)}$, where $\ell_{v_1,s}, \ldots, \ell_{v_d,s}$ are the labels of the parents of node $v$ (if any exist), and $\circ$ denotes string concatenation. When $\ell(\lambda) \in \Theta(\text{polylog } n)$ we will select $\lambda \in \text{polylog}(n)$ to ensure that $\ell(\lambda) \in \Theta(\text{polylog } n)$. We use the notation Enc and Dec for the encoding and decoding of our construction, while we use ECC and ECCD to denote the efficient encoding and decoding algorithms for a good binary code with constant rate and relative distance (e.g., [29], [30]), which can decode efficiently from some constant fractions of errors.

We first apply ECC to $x_1, \ldots, x_k$ to obtain codewords $c_1, \ldots, c_k \in \{0,1\}^{\ell(\lambda)}$ where $c_i = \text{ECC}(x_i)$. Also, for $v \in [k]$ we let $c_{v+k} = \text{ECC}(\ell_{v,s})$ which is the encoding of the label corresponding to the node $v$ in $G$. The final output is $c = (c_1 \circ \ldots \circ c_{2k-1} \circ c_{2k} \circ c_{2k+1} \circ \ldots \circ c_{3k})$ where $c_{2k+1} = \ldots = c_{3k} = c_{2k}$ consists of $k$ copies of the last codeword $c_{2k}$. The final word is an $n$ bit message with $n = \Theta(k \ell(\lambda))$. By repeating this last codeword $k$ times we ensure that it is not possible for the attacker to irreparably corrupt the final label $\ell_{k,s}$.

Weak CRLCCs (Local Decoder): Given a (possibly corrupted) codeword $c'$ produced by a PPT attacker $\mathcal{A}$ we let $x' = (x'_1 \circ \ldots \circ x'_k)$ with $x'_i = \text{ECCD}(c'_i)$ (possibly $\bot$) and we let $\ell'_{v,s} = \text{ECCD}(c'_{v+k})$ for $v \in [k]$ and $\ell'_{v,s,j} = \text{ECCD}(c'_{2k+j})$ for each $j \in [k]$. We say that a node $v$ is green if it is locally consistent i.e., $\ell'_{v,s} = H(s, x'_v \circ \ell'_{v_1,s} \circ \ldots \circ \ell'_{v_d,s})$, otherwise, we say that the node is red. Because $H(s,\cdot)$ is collision resistant it is not too difficult to prove that if a green node has the correct label $\ell'_{v,s} = \ell_{v,s}$ then it must be the case that $x'_v = x_v$ and $\ell'_{v,i,s} = \ell_{v_i,s}$ for each of $v$’s parents $v_1, \ldots, v_d$ — otherwise $\mathcal{A}$ would have found a hash collision! If a graph contains too many red nodes then this is easily detectable by random sampling and our weak local decoder is allowed to output $\bot$ if it detects any red nodes.

Our local decoder will first obtain the final label $\ell_{k,s}$ by random sampling some labels from $\ell'_{k,s,1}, \ldots, \ell'_{k,s,k}$ and checking to make sure each of these labels is equal to $\ell'_{k,s}$. If this check passes then with high probability we must have $\ell'_{k,s} = \ell_{k,s}$ since the attacker cannot corrupt too many of these labels. The local decoder also checks that the node $k$ is green i.e., $\ell'_{k,s}$ is locally consistent. If $\ell'_{k,s}$ is correct and node $k$ is green then this node serves as an anchor point. In particular, if node $v$ and node $k$ are connected by a green path then we can conclude that this part of the codeword has not been corrupted i.e., $\ell'_{v,s} = \ell_{v,s}$ and that $x_v = x'_v$. Thus, the remaining challenge is to design a test with small locality that outputs $0$ (whp) whenever there is no green path connecting $v$ to $k$ and always outputs $1$ when there are no red nodes. We show how to design such a test whenever $G$ is a local expander.

Local Expander Graphs: Given a graph $G = (V,E)$ and distinguished subsets $A, B \subseteq V$ of nodes such that $A \cap B = \emptyset$ and $|A| \leq |B|$ we say that the pair $(A,B)$ contains a $\delta$-expander if for all $X \subseteq A$ and $Y \subseteq B$ with $|X| > \delta|B|$ and $|Y| > \delta|B|$ there is an edge connecting $X$ and $Y$. We say that a DAG $G$ is a $\delta$-local expander around a node $v$ if for any radius $r > 0$ and any node $v \geq 2r$, the contiguous pair $A = \{v-2r+1, \ldots, v-r\}$ and $B = \{v-r, \ldots, v\}$ contain a $\delta$-expander and the contiguous pair $C = \{v, v+1, \ldots, v+r-1\}$ and $D = \{v+r-1, \ldots, v+2r-1\}$ contain a $\delta$-expander. When this property holds for every node $v \in V(G) = \{1, \ldots, n\}$ we simply say that the DAG $G$ is a $\delta$-local expander. For any constant $\delta > 0$ it is possible to (explicitly) construct a $\delta$-local expander with the property that $\text{indeg}(G), \text{outdeg}(G) \in \Theta(\log n)$ [31], [32]. Local expanders have several nice properties that have been recently exploited in cryptography in the design and analysis of data-independent memory hard functions (iMHFs) [32]–[36], proofs of sequential work [37] and proofs of space [38].

If $G$ is a $\delta$-local expander (for a suitably small constant $\delta > 0$) and we delete a large number of nodes $S \subseteq V$, then we can still prove that (1) at least $n - 2|S|$ of the nodes have the property that they are “$\alpha$-good” with respect to the deleted set $S$, and (2) any pair of $\alpha$-good
nodes \( u \) and \( v \) are connected by a directed path. Here, a node \( v \) is \( \alpha \)-good with respect to \( S \) if, for any radius \( r < v \), we have at most \( \alpha v \) nodes in \( S \cap [v - r + 1, v] \) and for any radius \( r \leq n - v + 1 \), we have at most \( \alpha v \) nodes in \( S \cap [v, v + r - 1] \).

Recall that given a (potentially corrupted) codeword, we defined a node \( v \) to be green if the hash label was locally consistent. Now if we let \( S \) denote the set of red nodes then we can develop an efficient randomized testing algorithm to check if a particular node \( v \) is \( \alpha \)-good or not. (We defer to the full version [25] the details of this test and of the full construction of Strong CRLCCs using the ideas developed above.)

REFERENCES


