

# Round-Vs-Resilience Tradeoffs for Binary Feedback Channels

Mark Braverman  

Princeton University, NJ, USA

Klim Efremenko  

Ben-Gurion University, Beer-Sheva, Israel

Gillat Kol  

Princeton University, NJ, USA

Raghuvansh R. Saxena 

Tata Institute of Fundamental Research, Mumbai, India

Zhijun Zhang  

Princeton University, NJ, USA

---

## Abstract

---

In a celebrated result from the 60's, Berlekamp showed that feedback can be used to increase the maximum fraction of adversarial noise that can be tolerated by binary error correcting codes from  $\frac{1}{4}$  to  $\frac{1}{3}$ . However, his result relies on the assumption that feedback is "continuous", i.e., after every utilization of the channel, the sender gets the symbol received by the receiver. While this assumption is natural in some settings, in other settings it may be unreasonable or too costly to maintain.

In this work, we initiate the study of *round-restricted feedback channels*, where the number  $r$  of feedback rounds is possibly much smaller than the number of utilizations of the channel. *Error correcting codes* for such channels are *protocols* where the sender can ask for feedback at most  $r$  times, and, upon a feedback request, it obtains all the symbols received since its last feedback request. We design such error correcting protocols for both the adversarial binary *erasure* channel and for the adversarial binary *corruption* (bit flip) channel. For the erasure channel, we give an exact characterization of the round-vs-resilience tradeoff by designing a (constant rate) protocol with  $r$  feedback rounds, for every  $r$ , and proving that its noise resilience is optimal.

Designing such error correcting protocols for the *corruption* channel is substantially more involved. We show that obtaining the optimal resilience, even with one feedback round ( $r = 1$ ), requires settling (proving or disproving) a new, seemingly unrelated, "clean" *combinatorial conjecture*, about the maximum cut in weighted graphs versus the "imbalance" of an average cut. Specifically, we prove an upper bound on the optimal resilience (impossibility result), and show that the existence of a matching lower bound (a protocol) is *equivalent* to the correctness of our conjecture.

**2012 ACM Subject Classification** Theory of computation → Error-correcting codes

**Keywords and phrases** Round-restricted feedback channel, error correcting code, noise resilience

**Digital Object Identifier** 10.4230/LIPIcs.ITCS.2025.22

**Related Version** Full Version: <https://eccc.weizmann.ac.il/report/2022/179> [8]

**Funding** *Mark Braverman*: Supported in part by the NSF Alan T. Waterman Award, Grant No. 1933331, a Packard Fellowship in Science and Engineering, and the Simons Collaboration on Algorithms and Geometry.

*Klim Efremenko*: Supported by the Israel Science Foundation (ISF) through grant No. 1456/18 and European Research Council Grant number: 949707.

*Gillat Kol*: Supported by a National Science Foundation CAREER award CCF-1750443 and by a BSF grant No. 2018325.

*Raghuvansh R. Saxena*: Supported by the Department of Atomic Energy, Government of India, under project no. RTI4001.

*Zhijun Zhang*: Supported by a National Science Foundation CAREER award CCF-1750443.

**Acknowledgements** We thank Noga Alon for discussions regarding Conjecture 3.



© Mark Braverman, Klim Efremenko, Gillat Kol, Raghuvansh R. Saxena, and Zhijun Zhang;

licensed under Creative Commons License CC-BY 4.0

16th Innovations in Theoretical Computer Science Conference (ITCS 2025).

Editor: Raghu Meka; Article No. 22; pp. 22:1–22:23



Leibniz International Proceedings in Informatics

LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

## 1 Introduction

**Cybernetics.** Consider the following two scenarios. Scenario one: a steersperson wishes to steer a longship to shore. She maintains a steady course in a changing environment (wind, waves, storms, currents, tides, etc.) by adjusting her steering in continual response to the effect it is observed as having. Scenario two: a teacher has a semester-worth of topics he wishes to teach to his class. He schedules exams throughout the semester to help him adapt his pace and determine what material should be repeated.

The above two scenarios are examples of *cybernetics*, a field that studies *self-regulating* processes. A core concept in cybernetics is *circular causality*, which is typically implemented using *feedback* mechanisms, where the observed outcomes of actions are taken as inputs for further actions. This is the case for, e.g., spacecraft navigators, artificial limbs, and our bodies' regulation of hormone and blood sugar levels. The term Cybernetics<sup>1</sup> was coined in 1948 by the mathematician and philosopher Norbert Wiener for “the science of control and communication in the animal and the machine” [37], following exchanges between numerous fields during the 1940s, including anthropology, mathematics, neuroscience, psychology, and engineering.

**Feedback in information theory.** Cybernetics grew alongside and built on Claude Shannon’s information theory, that was developed to improve the transmission of information and introduced the notion of *error correcting codes*. Shannon was interested in knowing whether the existence of a “feedback link” in the channel, where after every utilization of the channel, the (possibly incorrect) symbol obtained by the receiver is also given to the sender, allows for better codes. A discouraging early result by Shannon showed that feedback does not improve the capacity of memoryless channels [31]. It would be another decade or so before Berlekamp proves that feedback can, in fact, increase the maximum fraction of *adversarial* errors that can be tolerated. Specifically, Berlekamp showed that the *maximum noise resilience* of the (adversarial) binary channel increases from  $\frac{1}{4}$  to  $\frac{1}{3}$  given feedback [4, 5] (also see [39, 35, 1]).

A key property of the feedback channel exploited by Berlekamp’s result, as well as by follow up work, is that it supports “*continuous*” feedback – after *every* communication round, the sender gets the symbol received by the receiver. This assumption is natural in some settings, e.g., in scenario one, the steersperson continuously watches the ship’s motion as she steers. However, this assumption may be unreasonable or too costly to maintain in other settings, e.g., in scenario two, the teacher may not want to continuously quiz his students.

**This work: round-restricted feedback.** Motivated by such examples, in this work, we initiate the study of *round-restricted feedback channels*, where the number of feedback rounds is possibly much smaller than the number of communication rounds. Specifically, we wish to design protocols with optimal noise resilience that allow the sender (Alice) to transmit a message to the receiver (Bob), where during the execution of the protocol, the sender can ask for feedback at most  $r$  times. Upon such a request, the sender obtains all the bits received by the receiver from the last time feedback was solicited.

One can consider two models for scheduling the feedback rounds: the *adaptive* and the *non-adaptive* models. In the non-adaptive model, the sender decides ahead of time (before the protocol is run and before the input is known) when to schedule the  $r$  feedback rounds, while in the adaptive model, the timing of each feedback request may depend on the previously received feedback. In the second scenario, for example, the non-adaptive setting

---

<sup>1</sup> Interestingly, Cybernetics comes from the Greek word “Kubernetes”, which means steersperson.

corresponds to scheduling all exams at the beginning of the semester, while the adaptive setting corresponds to scheduling the next exam after the previous one was given. While our techniques hold for both the adaptive and non-adaptive settings, we choose to present our results for the non-adaptive setting. See Section 1.3 and Section 2.1 for a discussion of the implication of our techniques for the adaptive setting.

We consider such message transmission protocols with  $r$  feedback rounds over both the (adversarial) binary *erasure channel*, that erases some of the sent bits (those bits are received as “ $\perp$ ”), and over the (adversarial) binary *corruption channel*, that flips some of the sent bits. As was mentioned before, classical results in information theory show that with no feedback the maximum noise resilience of the binary corruption channel is  $\frac{1}{4}$  [25], while with continuous feedback, the maximum resilience improves to  $\frac{1}{3}$  [4, 5]. For the binary erasure channel, it is known that with no feedback the maximum resilience is  $\frac{1}{2}$ , and it is easy to see that with continuous feedback it approaches 1: the sender re-transmits each symbol until the receiver receives it.

We mention that rounds (or passes) are often considered to be a scarce resource and that round-restricted algorithms are extensively studied in other communication settings, e.g., communication complexity, distributed computing, streaming algorithms, and cryptographic protocols, and that we draw inspiration from these settings.

## 1.1 Our Results and Conjecture

Due to page limit, sections containing formal discussion on the corruption case and all proofs are deferred to the full version [8]. Here, we focus on the erasure case and only provide high level ideas for the corruption case.

### 1.1.1 The (Adversarial) Binary Erasure Channel

As discussed above, the maximum resilience of the erasure channel is known for the extreme cases of no feedback and of continuous feedback. Our first result is an optimal *rounds-vs-resilience tradeoff* for the erasure channel with any number of non-adaptive feedback rounds.

► **Theorem 1.** *The maximum noise resilience of the (adversarial) binary erasure channel with  $r$  rounds of feedback is  $\frac{5}{7}$  if  $r = 1$  and  $1 - \frac{7}{12(r+1)}$  if  $r > 1$ . Furthermore, the maximum noise resilience can be obtained by a deterministic, constant-rate protocol.*

Theorem 1 can be viewed as a “hierarchy theorem”, showing that more feedback rounds allow for strictly better resilience. On the other hand, Theorem 1 also shows that a constant number  $O_\epsilon(1)$  of feedback rounds already suffices to get a noise resilience of  $1 - \epsilon$  for the erasure channel.

**Techniques.** The main ingredient in our proof of Theorem 1 is the construction of a *list decodable* code for the binary erasure channel with  $m$  codewords, for *all* (not necessarily asymptotic) values of  $m$ . Our code is optimal in the sense that it achieves the maximum error resilience for every list size simultaneously. We emphasize that for our protocols, we need such a code for all possible  $m$ , which corresponds to all possible “block sizes”. We call codes with small  $m$ ’s “*small codes*”. Given these codes, the protocols we use to prove Theorem 1 are rather simple – after every feedback round, Alice and Bob agree on a (smaller, unless there was a lot of noise) set  $\Gamma$  of candidate inputs  $x$  and Alice encodes  $x$  with our optimal list decodable code with  $m = |\Gamma|$  codewords. On the analysis front, we are able to

argue that, unless the adversary erases many of the sent bits, the size of the candidate set  $\Gamma$  shrinks substantially between feedback rounds, and measure this shrinkage exactly. See Section 2.1 for a detailed overview.

### 1.1.2 The (Adversarial) Binary Corruption Channel

Theorem 1 gives a complete characterization of the noise resilience of the erasure feedback channel as a function of the number of feedback rounds. However, as will be explained next, the case of corruptions is much more involved, and we will focus on protocols with one round of feedback. We mention that since the adaptive and non-adaptive models are the same for protocols with one feedback round, the results in this section hold for both the adaptive and non-adaptive settings. Our next theorem gives an upper bound on the noise resilience of such one-round protocols.

► **Theorem 2.** *The maximum noise resilience of the (adversarial) binary corruption channel with one round of feedback is at most  $\frac{7}{23}$ .*

We conjecture that the upper bound of  $\frac{7}{23}$  on the noise resilience in Theorem 2 is tight, and that it can be achieved by a constant-rate protocol. Perhaps surprisingly, proving this is equivalent to showing the following combinatorial conjecture about the existence of large cuts in graphs.

► **Conjecture 3.** *Let  $G$  be a graph with  $n$  vertices and non-negative edge weights summing up to 1. Let  $\text{wt}(S)$  be the sum of weights of all the edges with both endpoints in the subset of vertices  $S$ , and let  $\text{Max-Cut}(G)$  be the maximum total weight of all the edges across any cut in  $G$ . Then,<sup>2</sup>*

$$\text{Max-Cut}(G) \geq \frac{2}{3} - \frac{16}{15} \cdot \mathbb{E}_{S \subseteq [n]} [\min(\text{wt}(S), \text{wt}(\bar{S}))]. \quad (1)$$

We prove Conjecture 3 for (large enough) graphs where all edges have equal weight, i.e., “unweighted” graphs. However, the case for general weighted graphs seems much harder, and, despite our best effort, we were unable to prove (or disprove) it. We also mention that Conjecture 3 is tight for some graphs (e.g., cliques of size 3 and 5 with edges of equal weight), and related bounds on Max-Cut were studied in other contexts, e.g., [26, 2, 22].

The next theorem gives the equivalence between Conjecture 3 and the tightness of Theorem 2.

► **Theorem 4.** *Theorem 2 is tight if and only if Conjecture 3 holds. Furthermore, Conjecture 3 implies a constant rate protocol achieving the maximum noise resilience.*

In essence, Theorem 4 connects the problem of designing optimal error correcting protocols with one round of feedback to a combinatorial question about graphs. As we discuss later in Section 2.2, our techniques can also be used to connect the problem of designing optimal error correcting protocols with multiple rounds of feedback to similar questions about graphs.

---

<sup>2</sup> As the expectations of  $\text{wt}(S)$  and  $\text{wt}(\bar{S})$ , for a uniformly random  $S$ , are  $\frac{1}{4}$ , Equation (1) can be equivalently written as  $\text{Max-Cut}(G) \geq \frac{6}{15} + \frac{8}{15} \cdot \mathbb{E}_{S \subseteq [n]} [\lvert \text{wt}(S) - \text{wt}(\bar{S}) \rvert]$ , where the term inside the expectation is the “imbalance” of a random cut.

**Techniques.** The proof of Theorem 4 is technically involved and a detailed overview can be found in Section 2.2. At a high level, the main ingredient in designing our protocol is the construction of a special type of “weighted” codes, called **dc-codes**. A **dc**-code  $C$  is parameterized by a “distance contribution function”  $\text{dc}$  that assigns a value in  $[0, 1]$  to each possible message  $x \in \{0, 1\}^k$ . We require that for all  $x \neq x' \in \{0, 1\}^k$ , the codewords  $C(x)$  and  $C(x')$  are at least (relative) Hamming distance  $\text{dc}(x) + \text{dc}(x')$  apart. Equivalently, we ask that the balls of radii  $\text{dc}(x)$  around  $C(x)$  are all disjoint.<sup>3</sup> We note that unlike traditional error correcting codes that have only one distance guarantee for all pairs of codewords (i.e., the minimum distance), the distance guarantees for different pairs of codewords in a **dc**-code are different. In fact, traditional codes can be viewed as **dc**-codes for a constant **dc** function.

**dc**-codes for non-constant **dc** functions are useful for our protocol as if the adversary already used up many of its corruptions before the feedback round, Alice knows she can afford to send her message  $x$  encoded with an error correcting code that does not guarantee a large distance between  $C(x)$  and the other codewords. Geometrically, designing a **dc**-code is a sphere packing problem where we need to pack spheres of different radii  $\text{dc}(x)$ . As for some  $x$ ’s a small radius  $\text{dc}(x)$  suffices, some of the spheres are small, which allows the other spheres being packed to be larger.

The proof of Theorem 4 shows that Conjecture 3 implies the existence of **dc**-codes that are needed for our protocol to work. We assume that Alice uses a uniformly random code to encode her message before the feedback. The codeword sent by Alice can be corrupted by the channel in many ways, and each such way would imply a function **dc** such that Alice would like to use a **dc**-code to encode her message after the feedback. We denote by  $Q$  the set of **dc** functions for which the corresponding **dc**-codes are needed by our protocol. We also denote by  $P$  the set of **dc** functions for which **dc**-codes exist. We wish to show  $Q \subseteq P$ . To this end, we show that both  $P$  and  $Q$  are closed and convex, and that in every direction  $z$ , the extremal point of  $P$  in direction  $z$  is “farther” than the extremal point of  $Q$  in direction  $z$ . We then recast this geometric problem as a combinatorial problem by interpreting the direction vector  $z$  as a weighted graph  $G$ , and show that the extremal point of  $P$  in direction  $z$  corresponds to a **Max-Cut** in  $G$  (as in the left hand side of Conjecture 3), while the extremal point of  $Q$  in direction  $z$  corresponds to the right hand side of Conjecture 3.

For the converse direction of Theorem 4, we show that the arguments in the above paragraph are actually equivalences, except for the assumption that Alice uses a randomly sampled code to encode her message before the feedback. At a high level, we use *Ramsey theory* to show that the assumption that this code is a random code is, at least in some sense, without loss of generality (see Section 2.2.2 for a more precise statement).

## 1.2 Related Work

Feedback channels were studied since the early days of information theory and are still actively studied [31, 24, 14, 5, 9, 27, 32, 13, 33, 34, to cite a few]. While feedback does not increase the capacity of discrete memoryless channels with vanishing error, there are settings where feedback is known to allow improvement, like in the 0-error capacity case [31], and under variable decision time [9].

**Partial feedback.** Haeupler, Kamath, and Velingker [23] considered the setting where the feedback is partial, and showed that even if Alice receives feedback bits from Bob for an arbitrarily small constant fraction of her transmissions, resilience close to (the optimal

---

<sup>3</sup> We mention that **dc**-codes are an example of non-equally spaced codes defined in [11].

resilience of  $\frac{1}{3}$  is possible using a randomized protocol. However, the number of feedback rounds their protocol needs grows linearly with  $n$ , the length of Alice's input. See [36] for a subsequent result.

Independently and concurrently to our work, [15] improved [23] and showed a deterministic protocol that uses  $\mathcal{O}(\log n)$  feedback bits over  $\mathcal{O}(1)$  feedback rounds to get resilience approaching  $\frac{1}{3}$ , along with a similar result for the erasure channel showing that the resilience approaches 1 for this channel. The main difference between [15] and the current work is that we focus on finding the optimal resilience for any given number  $r$  of feedback rounds whereas [15] focuses on showing that the resilience approaches the optimal value as the constant  $r$  increases. Additionally, their work measures both the number of feedback rounds and the number of feedback bits, while we only focus on the number of rounds.

**Two-way codes and interactive codes.** As discussed above, feedback is also known to increase the noise resilience of the adversarial binary corruption channel [4, 5], and this result played a big role in recent work in *interactive coding* [10, 18, 17] and *two-way coding* [16, 19, 11]. In interactive coding [28, 29, 30], we wish to simulate a communication protocol  $\Pi$  that was designed to work over the noiseless channel, by a protocol  $\Pi'$  that works over a noisy channel. In the setting of two-way codes, like in the setting of traditional error correcting codes, Alice wishes to transmit a message  $x$  to Bob over a noisy channel. However, unlike the case of traditional codes, where Alice is the only party that can transmit messages, in two-way codes Bob can also use the (noisy) channel to transmit messages back to Alice.

Observe that since Bob has no input, any two-way code can be run over the feedback channel and thus two-way error correcting codes can be viewed as protocols over a *noisy feedback* channel. In particular, since the noise tolerance of the binary corruption channel is only  $\frac{1}{3}$ , the noise resilience of binary two-way codes over the binary corruption channel is at most  $\frac{1}{3}$ . In the same way, results for the bounded round feedback channel give upper bounds on the noise resilience of the corresponding two-way channels.

Gupta, Kalai, and Zhang [16, 19] studied two-way error correcting codes over the binary *erasure channel*. Their main result is a code that is resilient to a  $\frac{3}{5}$  fraction of adversarial errors, improving on the noise tolerance of the one-way binary erasure channel that is known to be  $\frac{1}{2}$ . We mention that the two-way coding schemes of [16, 19] exchange (almost) linear number of messages. The work of [16] also gives an upper bound of  $\frac{2}{3}$  on the maximum tolerance of the two-way binary erasure channel, and an upper bound of  $\frac{2}{7}$  on the maximum tolerance of the two-way binary corruption channel. Given those upper bounds, a corollary of our results is that even a single round of noiseless feedback allows for a better error tolerance than any number of noisy feedback rounds over both the erasure and corruption channels<sup>4</sup>.

The recent work of Efremenko, Kol, Saxena, and Zhang [11] shows that the maximum noise resilience of two-way error correcting codes for the binary corruption channel is strictly better than the noise resilience of traditional error correcting codes for this channel, which is

---

<sup>4</sup> To see why, observe that if Bob's messages are noiseless, we can assume without loss of generality that Bob's messages are much shorter, say at most an  $\epsilon$  fraction, of Alice's messages. Indeed, if not, consider a modified protocol where all messages from Alice are repeated  $k$  times, for some large  $k$ . For the erasure channel, either all the repetitions of a bit from Alice are erased or Bob knows the bit exactly. Thus, his communication does not grow with  $k$ . For the corruption channel, it suffices for Bob to say how many of the repetitions were received as 1, which can be done using  $\log k \ll k$  bits. Moreover, we mention that for this claim, we do not need to rely on Conjecture 3, as a lower bound slightly smaller than  $\frac{7}{23}$  (but greater than  $\frac{2}{7}$ ) on the maximum error resilience of protocols with one feedback round over the binary corruption channel can be obtained unconditionally using our techniques (but is not included in the current work).

known to be  $\frac{1}{4}$  [25]. At a very high level, those results for two-way codes are obtained by implementing a (weak) feedback mechanism over channels with no built-in feedback. Related ideas were used in [10, 18, 17] to give interactive binary error correcting codes with high noise resilience.

**List decodable codes.** List decodable codes were introduced in the 50's [12, 38] and have been studied over numerous papers and found many applications since then. We next list the works most related to ours. Most of the work on list decoding was done in the asymptotic regime, where the number of codewords goes to infinity. In this work, we are interested in the optimal list decodable codes for any (potentially small) number of codewords. However, as an ingredient in our proof, we use the asymptotic results of [3] (see also [6, 21, 7]) for optimal list decoding of the corruption channel (see Lemma 11). The list decoding question was also considered for other channels, for example, over the corruption channel with feedback [32] and the erasure channel [20].

### 1.3 Open Problems

Our work suggests the study of feedback channels through a new lens, namely, their feedback round complexity. We next list some suggestions for future work in this direction.

**Graph-theoretic conjectures.** The most immediate question we leave open is proving Conjecture 3 for all weighted graphs. We also propose the following potentially related conjecture, which is tight for all odd cliques with edges of equal weight.

► **Conjecture 5.** *Let  $G$  be a graph with  $n$  vertices and non-negative edge weights summing up to 1. Let  $\text{wt}(i)$  be the sum of weights of all edges incident on vertex  $i$ . Then,*

$$\text{Max-Cut}(G) \geq \frac{1}{2} + \frac{1}{8} \cdot \sum_{i \in [n]} \text{wt}(i)^2.$$

**Round-vs-resilience tradeoff for other channels.** Proving Conjecture 3 would imply that our protocol in Theorem 4 has optimal noise resilience among protocols with one round of feedback over the corruption channel. Obtaining a general round-vs-resilience tradeoff for any number of feedback rounds  $r$  for the corruption channel and for other well-studied channels (e.g., the binary insertion-deletion channel<sup>5</sup>, the binary deletion-only channel, and non-binary channels), would be interesting.

**Adaptive corruptions over the erasure channel.** Theorem 1 considers the case of *non-adaptive* feedback rounds, where Alice decides ahead of time when to ask for feedback. It can be shown that the case of *adaptive* feedback rounds, where Alice chooses when to ask for another round of feedback after seeing the previous feedback, allows for (strictly) better round-vs-resilience tradeoffs. Our techniques can be used to write a recursive formula for the noise resilience in the adaptive case, and finding a “clean”, closed-form formula for this setting (if one exists) is left open (see Section 2.1).

---

<sup>5</sup> We note that this first requires a suitable definition for the insertion-deletion channel with constant number of rounds.

## 2 Proof Overview

In this section, we overview the proofs of Theorems 1 and 4, starting with the relatively easier Theorem 1.

### 2.1 Result for the Erasure Channel – Theorem 1

The defining feature of the erasure channel is that the receiver (Bob) either receives the bit sent by Alice or receives a special erasure symbol  $\perp$ . This means that in any round where Bob receives  $\perp$ , he is certain that this is due to the erasures in the channel, while if he receives a symbol different from  $\perp$ , he is certain that the symbol must be what Alice sent in that round. In turn, this means that Bob knows exactly the amount of erasures introduced by the channel and also means that Bob can (recall that he is trying to determine Alice's input) remove from consideration any candidate input that is “inconsistent” and would make Alice send a different symbol in any such round.

**The general format of a protocol.** The above observation implies that protocols for the erasure channel with  $r$  rounds of feedback (and therefore  $r + 1$  messages from Alice) have the following format: Alice starts with an input  $x \in \Gamma_0 = \{0, 1\}^n$ . For her first message, she takes a code<sup>6</sup>  $C_0 : \Gamma_0 \rightarrow \{0, 1\}^*$  and sends  $C_0(x)$  to Bob. Some of the bits of  $C_0(x)$  are received correctly by Bob, while the remaining bits are erased and replaced with  $\perp$ . Using the bits he received correctly, Bob can calculate the number of erasures  $N_1$  introduced by the channel in this round and can identify a subset  $\Gamma_1 \subseteq \Gamma_0$  of inputs for Alice that are consistent with the message he received. Note that Alice's input  $x$  must be in  $\Gamma_1$ .

Then, a feedback round takes place, and as Alice learns all the received symbols, she can also compute  $N_1$  and  $\Gamma_1$ . As both parties now know these values, they can now “forget” this round and “reduce”<sup>7</sup> to a smaller problem where Alice wants to transmit an element  $x \in \Gamma_1$  to Bob using a protocol with  $r - 1$  rounds of feedback and the maximum number of erasures the channel can insert is  $N_1$  lower than what it was before. Continuing this way, the goal of the parties is to reduce to a problem with 0 rounds of feedback, and set of inputs  $\Gamma_r$  such that there exists a (standard) error correcting code for elements in  $\Gamma_r$  resilient to the number of erasures that the channel can insert in the last round.

**List-decodable small codes.** It is readily seen that the protocol format described above does not care about the exact strings in the sets  $\Gamma_0, \dots, \Gamma_r$ , as long as their sizes stay the same. Thus, the question of whether or not the above protocol format can be instantiated to get a protocol that is resilient to  $\theta$  fraction of adversarial erasures, for some  $\theta \in [0, 1]$ , reduces to determining when to schedule the feedback rounds, and given two feedback rounds, determining the codes  $C_i$  to be used by Alice between these rounds. The codes  $C_i$  should be such that, given an initial set size  $m = |\Gamma_i|$  and a target set size<sup>8</sup>  $k = |\Gamma_{i+1}|$ , the number

<sup>6</sup> At this point, it may be helpful to view this as a function instead of a code. We explain why we are calling it a code later. Also, a more precise way to state this would be to say that there exists an  $L > 0$  such that  $C_0 : \Gamma_0 \rightarrow \{0, 1\}^L$ , as all codewords need to be of the same length to avoid the parties from signaling through the length of the codeword. Nonetheless, we stick with statements like  $C_0 : \Gamma_0 \rightarrow \{0, 1\}^*$  throughout this sketch for simplicity.

<sup>7</sup> We elaborate what this means exactly in the paragraph on adaptive feedback rounds below.

<sup>8</sup> Note that  $k$  is not known to the parties in advance, and thus it will be ideal if the code used is optimal for all  $k$  simultaneously.

of erasures required to reduce the set size from  $m$  to  $k$  is the highest. Using such codes, Alice ensures that unless the adversary invests many erasures, the set of candidates shrinks substantially between feedback rounds. We first focus on designing such codes.

Codes like the above are known as list decodable codes, and have been well studied in the asymptotic regime, where  $m$  tends to infinity, and exact answers are known (see, e.g., [12, 38, 6, 20, 7, 3, 32] and Lemma 11). However, for our purposes, we need the exact answer for smaller values of  $m$  as well. Codes with small  $m$ , i.e., “small codes” or codes with few codewords, have recently received a lot of attention and have proven to be useful in designing binary protocols with high error resilience in several contexts [10, 11, 16, 18, 19]. In the current paper, we provide a complete analysis of the list-decodability of these codes for the erasure channel, giving a function  $d(m, k)$  that characterizes exactly the minimum amount of erasure noise needed such that for *any* code  $C : [m] \rightarrow \{0, 1\}^*$ , one can erase  $d(m, k)$  fraction of the bits and ensure that Bob gets a list of candidates of length strictly smaller than  $k$ .

The formula for  $d(m, k)$  is given in Equation (7). Proving that this formula is correct requires showing both a construction (of codes with resilience approaching  $d(m, k)$ ) and an impossibility result. Our construction has the nice property that the same code is tight simultaneously for all values of  $k$ . Roughly speaking, our code achieves this optimal erasure noise resilience by ensuring that every coordinate is as differentiating as possible, i.e., we ensure that for all coordinates  $j$ , exactly  $\lfloor \frac{m}{2} \rfloor$  (uniformly chosen) codewords have 0 in that coordinate, while the remaining  $\lceil \frac{m}{2} \rceil$  codewords have 1 (see Lemmas 9 and 10). This is as opposed to randomly sampled codes where, e.g., a  $\frac{1}{2^m}$  fraction (which is large for small  $m$ ) of the coordinates are expected to be 0 for all the codewords, and therefore not differentiate between any pair of codewords.

**Scheduling the feedback rounds.** Even with an exact formula for  $d(m, k)$  in hand, it still remains to schedule the feedback round correctly in order to maximize the overall noise resilience of the obtained protocol. The fact that our constructed code is tight simultaneously for all values of  $k$  is of great help for this part, as the actual value of  $k$  is determined by the erasures inserted by the channel and not in our control. This means that in order to schedule the feedback rounds optimally, one needs to go over all possible values of  $k$  (across all rounds) that may happen over the channel and maximize the corresponding error resilience. This requires a careful analysis of the obtained formula for  $d(m, k)$ .

**Adaptive feedback rounds.** We finish this section by briefly discussing the extension of our result to adaptive feedback rounds, as hinted in Section 1.3. Recall our reduction above from  $r$  to  $r - 1$  feedback rounds, and note that this reduction is not perfect in the following sense: the erasures inserted by the adversary in Alice’s first message in the  $r$ -round protocol dictate the set  $\Gamma_1$  of candidates and the budget of the  $(r - 1)$ -round protocol. Observe that the  $(r - 1)$ -round protocol with maximal noise resilience for transmitting a message depends on the size of the set of candidates and on the erasure budget. Now, since our  $r$ -round protocol is non-adaptive, meaning that the timing of all feedback rounds is fixed in advance and cannot be recalculated given the erasures in the first round, our  $r$ -round protocol may reduce to a sub-optimal  $(r - 1)$ -round protocol. Therefore, when scheduling the feedback rounds for our  $r$ -round protocol, one needs to consider the values of  $k$  that are possible across all rounds in order to get the optimal schedule.

On the other hand, if the feedback rounds can be scheduled adaptively, the reduction is indeed perfect. In this case, one just needs to schedule the first feedback round beforehand based on the possible values of  $k = |\Gamma_1|$  for this round alone, and then, upon seeing the  $N_1$

and  $\Gamma_1$  values, one can take the  $(r - 1)$ -feedback round protocol with the maximum error resilience (when Alice's input is from  $\Gamma_1$  and the number of erasures is  $N_1$  lower) and schedule the remaining feedback rounds according to this protocol. Thus, our techniques also lead to a tight recursive formula for the maximum error resilience in the case of adaptive feedback rounds, but converting it to a “clean” closed form formula (if at all possible) is left open.

## 2.2 Result for the Corruption Channel – Theorem 4

Compared to the erasure channel, where Bob knows exactly the amount of noise inserted and can safely eliminate many candidate inputs for Alice, the corruption channel is much harder. Here, upon receiving a message from Alice, all Bob can compute is, given a candidate input  $y$  for Alice, what is the number  $N(y)$  of corruptions the channel inserted assuming Alice's input was indeed  $y$ . Crucially, this value of  $N(y)$  may be very different for different  $y$ , and unless it exceeds the maximum possible number of corruptions in the channel (which can only happen when the protocol is quite far advanced), it can never have Bob eliminate  $y$  from consideration entirely.

Consider now a protocol over the corruption channel with one round of feedback (and therefore, two messages from Alice). Suppose that Alice's input  $x$  comes from the set  $\{0, 1\}^n$ . As explained above, after receiving the first message from Alice, Bob knows  $N(y)$  for all  $y \in \{0, 1\}^n$ . By subtracting  $N(y)$  from the maximum possible number of corruptions, Bob can compute, for all  $y \in \{0, 1\}^n$ , a number  $dc(y)$  which is the leftover corruptions, or, equivalently, the degree to which the second message of Alice can be corrupted, assuming her input is  $y$ . As Alice receives feedback from Bob, she can also compute the values  $dc(y)$  for all  $y \in \{0, 1\}^n$ . In the remainder of this sketch, we normalize  $dc(y)$  by dividing it by the length of Alice's second message. This will result in a value in  $[0, 1]$ .

**dc-codes.** Using this feedback, Alice's goal in her second message is to allow Bob to uniquely identify her input. If  $C : \{0, 1\}^n \rightarrow \{0, 1\}^*$  is the code used by Alice in her second message, the only way Bob can uniquely decode Alice's input is if for all  $y \neq y' \in \{0, 1\}^n$ , the codewords  $C(y)$  and  $C(y')$  are at least (relative) Hamming distance  $dc(y) + dc(y')$  apart. The reason is that if  $y$  is Alice's input, then the adversary has fractional budget  $dc(y)$  that it can use to corrupt  $C(y)$ , and thus the codeword received by Bob can be any string of (relative) Hamming distance at most  $dc(y)$  from  $C(y)$ . Similarly, if  $y'$  is Alice's input, then the codeword received by Bob can be any string of Hamming distance at most  $dc(y')$  from  $C(y')$ . Note that the adversary cannot arrange for the received encodings to be the same if and only if  $C(y)$  and  $C(y')$  are at least (relative) Hamming distance  $dc(y) + dc(y')$  apart. We call a code that satisfies this (relative) Hamming distance property a dc-code and mention that the values  $dc(y)$  can equivalently be seen as the “distance contributed” by  $y$  in such a code.

We note that unlike traditional error correcting codes that have only one distance guarantee for all pairs of codewords (i.e., the minimum distance), for dc-codes, the distance between a pair of codewords may be different depending on the “compatibility” of the messages they encode. Specifically, we think of each codeword as having a different “radius” and the code needs to “pack” all the induced balls of different radii. We point out that dc-codes are an example of non-equally spaced codes defined in [11].

We also observe that the small code used in our protocol for erasures can be viewed as a dc-code where  $dc(y) = 0$  for all inputs  $y$  that Bob has ruled out (and therefore, do not need any distance guarantees), and  $dc(y) = c$  for all inputs  $y$  that he has not ruled out, where  $c$  is

the best possible constant ( $c$  is determined by the  $d(m, k)$  function). We mention that for the erasure channel, our protocol also needed list-decoding guarantees that are not needed here as we are only attempting to get a one feedback round protocol.

The discussion so far shows that the existence of a protocol with a given error resilience amounts to determining whether or not it holds that for all functions  $dc(\cdot)$  that can be induced by the corruptions inserted in Alice's first message, there exists a  $dc$ -code that Alice can use to compute her second message. Curiously, we show in the next subsection that this question is equivalent to our seemingly unrelated combinatorial conjecture (Conjecture 3) about the existence of large cuts in graphs.

**Towards multiple rounds of feedback.** The above approach of designing  $dc$ -codes (that have no rounds of feedback) to construct protocols with one round of feedback can be generalized. One can similarly argue that, for any  $r \geq 0$ ,  $dc$ -codes with  $r$  rounds of feedback can be used to construct protocols with  $r + 1$  rounds of feedback. Analogously to the above, the “extra” round is the first round, and dictates which  $dc$ -code is used in the rest of the protocol. Moreover, questions about constructing  $dc$ -codes with  $r$  rounds of feedback can be translated to questions about graphs. The  $r = 0$  case is explained next, but similar ideas may be used for general  $r$ , with appropriate changes in the definitions of the set  $P$  and  $Q$  (see below).

### 2.2.1 Conjecture 3 Implies a Tight Protocol

We first show why Conjecture 3 implies the existence of a tight protocol. In fact, we shall show the existence of a protocol where Alice's message in the first round is simply the encoding of her input  $x$  using a randomly sampled code. Let  $m = 2^n$ . A distance function is a function  $dist : \binom{[m]}{2} \rightarrow \mathbb{R}$ , where  $\binom{[m]}{2}$  is the set of all subsets of  $[m]$  of size 2. For a code  $C : [m] \rightarrow \{0, 1\}^*$ , we denote by  $dist_C$  the distance function induced by  $C$ , i.e.,  $dist_C(i, i')$  is the (relative) Hamming distance between  $C(i)$  and  $C(i')$ . For a distance contribution function  $dc$ , we denote by  $dist_{dc}$  the distance function induced by  $dc$ , i.e.,  $dist_{dc}(i, i') = dc(i) + dc(i')$ . For simplicity, throughout this overview we assume that  $dc(y) = 1 - N(y)$  (recall that  $dc(y)$  is actually the normalized leftover corruption count, but in this sketch we will ignore the exact multiplicative and additive constants in this function).

**Recasting as a geometric problem.** We denote by  $P$  the set of all distance functions  $dist_C$  that are induced by codes  $C : [m] \rightarrow \{0, 1\}^*$ . We denote by  $Q$  the set of all distance functions  $dist_{dc}$  induced by  $dc$  functions that can be obtained by the corruptions inserted in Alice's first message (recall that  $dc$  depends on  $N$ , which is a function of the corruptions inserted in Alice's first message). In other words,  $P$  is the set of distance functions that can be realized and  $Q$  is the set of distance functions required by our protocol. We wish to prove  $Q \subseteq P$ .

We view distance functions  $dist$  as  $\binom{m}{2}$ -dimensional vectors. We observe that both  $P$  and  $Q$  are closed and convex and that the set  $P$  is “downwards-closed”, meaning that if  $dist \in P$  then any  $dist'$  that is coordinate wise at most  $dist$  is also in  $P$ . This means that showing  $Q \subseteq P$  is equivalent to showing that for all  $\binom{m}{2}$ -dimensional non-negative hyperplanes  $z$ , it holds that:

$$\max_{dist \in P} \langle z, dist \rangle \geq \max_{dist \in Q} \langle z, dist \rangle, \quad (2)$$

**Recasting as a combinatorial problem.** By scaling, we can assume that the entries of  $z$  sum to 1 and view them as the weights on the edges of an  $m$ -vertex graph  $G_z$  as in Conjecture 3. As both  $P$  and  $Q$  are closed and convex, both the maximums are attained at one of their vertices.

To reason about Equation (2), it will be useful to represent a code  $C : [m] \rightarrow \{0, 1\}^L$  as a sequence of  $L$  one-bit functions  $b : [m] \rightarrow \{0, 1\}$  (the first one-bit function corresponds to the first coordinate of  $C(i)$ , etc.). Observe that for the code  $b : [m] \rightarrow \{0, 1\}$  (i.e.,  $L = 1$ ), it holds that  $\text{dist}_b$  is a boolean function with  $\text{dist}_b(i, i') = 1$  if and only if  $b(i) \neq b(i')$ .

**The LHS of Equation (2).** Since a general code  $C$  is a sequence of one-bit functions, it can be shown that the function  $\text{dist}_C$  is a convex combination of the functions  $\text{dist}_b$  that are induced by one-bit functions  $b$ . In particular, this means that the vertices of  $P$  are distance functions induced by one-bit functions. Using the expression above for  $\text{dist}_b$  for one-bit function  $b : [m] \rightarrow \{0, 1\}$ , the value of  $\langle z, \text{dist}_b \rangle$  is the value of the cut in the graph  $G_z$  indicated by  $b$ :

$$\langle z, \text{dist}_b \rangle = \sum_{(i, i')} z_{i, i'} \cdot \text{dist}_b(i, i') = \sum_{(i, i'): b(i) \neq b(i')} z_{i, i'}. \quad (3)$$

Thus, the left hand side of Equation (2) is the maximum cut in  $G_z$ , as in Conjecture 3.

**The RHS of Equation (2).** We view the code used by Alice in her first message as a sequence of one-bit functions. Since in our protocol this code is randomly sampled, each of the  $2^m$  one-bit functions is expected to appear equally often in Alice's message<sup>9</sup>. As the channel can corrupt each of these one-bit functions independently of all the others, we get that a distance function  $\text{dist}$  can be induced by the corruptions inserted in Alice's first message (i.e.,  $\text{dist} \in Q$ ) if and only if it is the expectation (under the uniform distribution over one-bit functions) of the distance functions that can be induced by corrupting one-bit functions.

Now, if Alice is sending a one-bit function  $b : [m] \rightarrow \{0, 1\}$ , there are only two possibilities for Bob: either he receives a 0 or he receives a 1. Let  $\text{dc}_b$  be the distance contribution function dictated by Bob's received bit. We next show that in the former case, where Bob receives 0, the value of  $\langle z, \text{dist}_{\text{dc}_b} \rangle$  is the value of the cut in  $G_z$  indicated by  $b$  plus twice the weight of all edges such that  $b(\cdot) = 0$  on both its endpoints. To see that, recall that  $\text{dist}_{\text{dc}_b}(i, i') = \text{dc}_b(i) + \text{dc}_b(i')$  and that we assume  $\text{dc}_b(y) = 1 - N(y)$ . In our case,  $\text{dc}_b(i) = 1 - 0 = 1$  if  $b(i) = 0$  (Alice's bit was not corrupted) and  $\text{dc}_b(i) = 0$  if  $b(i) = 1$  (Alice's bit was corrupted). This implies that  $\text{dist}_{\text{dc}_b}(i, i') = 0$  if  $b(i) = b(i') = 1$ , and that  $\text{dist}_{\text{dc}_b}(i, i') = 1$  if  $b(i) \neq b(i')$ , and that  $\text{dist}_{\text{dc}_b}(i, i') = 2$  if  $b(i) = b(i') = 0$ . Therefore,

$$\max_{\text{dist} \in Q} \langle z, \text{dist}_{\text{dc}_b} \rangle = \sum_{(i, i')} z_{i, i'} \cdot \text{dist}_{\text{dc}_b}(i, i') = \sum_{(i, i'): b(i) \neq b(i')} z_{i, i'} + \sum_{(i, i'): b(i) = b(i') = 0} 2 \cdot z_{i, i'}. \quad (4)$$

Similarly, it can be shown that in the latter case, where Bob gets 1, the value of  $\langle z, \text{dist}_{\text{dc}_b} \rangle$  is the value of the cut in  $G_z$  indicated by  $b$  plus twice the weight of all edges such that  $b(\cdot) = 1$  on both its endpoints.

---

<sup>9</sup> We mention that this is only in expectation and the length of Alice's message need not depend exponentially on  $m$ .

Recall that the bit function  $b$  is a uniformly random bit-function. Taking an expectation over one bit functions  $b$ , the value of the cut in  $G_z$  indicated by  $b$  is exactly the constant  $\frac{1}{2}$  and the other terms on the right hand side of Equation (3) and Equation (4) are exactly as on the right hand side of Conjecture 3, where the maximum becomes minimum because of the constants involved. Equation (2) now directly follows from Conjecture 3.

### 2.2.2 A Tight Protocol Implies Conjecture 3

We now finish this sketch by arguing why a tight protocol implies Conjecture 3. For this, we note that all the arguments in Section 2.2.1 were actually equivalences, except two, one of which was explicitly stated and one was not. The explicit one was our assumption that Alice's first message is simply the encoding of her input using a randomly sampled code. The second one was that Alice gets feedback from Bob at round  $\frac{8T}{23}$ , where  $T$  is the total number of rounds of the protocol. The constant  $\frac{8}{23}$  may seem arbitrary, but it is the constant one gets when one tries to match the constants obtained in the analysis in Section 2.2.1 with the constants in Conjecture 3.

Both these assumptions are actually without loss of generality. We start by arguing this for the second one, again ignoring the actual constants and only stating the high level idea. Roughly speaking, the second assumption is without loss of generality as Conjecture 3 is tight for cliques of size 3 and 5, and if the constant is anything other than  $\frac{8}{23}$ , Equation (2) will fail to hold for  $z$  corresponding to one of these cliques.

It remains to show why the first assumption is without loss of generality. For this, our approach is to take an arbitrary code  $C : [m] \rightarrow \{0, 1\}^*$  that Alice may use for her first message, and in several steps, convert it to a code that looks more and more like a random code, at the cost of a smaller  $m$ . In each step  $k$ , we convert  $C$  to a code that is  $k$ -random, in the sense that any set of  $k$  codewords of the new code looks like  $k$  codewords from a randomly sampled code.

For  $k = 1$ , this means that we have to show that each codeword has an equal number of 0s and 1s, and this can be easily achieved by concatenating all codewords with their negations (which preserves the distance properties). We now show how to get a 2-random code from a 1-random code, noting that similar (but technically more involved) ideas allow us to get a  $(k + 1)$ -random code from a  $k$ -random code, for any  $k \geq 1$ . To show that a code is 2-random, we need to show that it is 1-random and that the fractional distance between any pair of codewords is (roughly)  $\frac{1}{2}$ .

For this, let  $\epsilon > 0$  be an error parameter and construct a complete graph with the  $m$  codewords as the vertices, and color the edge between codewords  $i$  and  $i'$  as (1) red, if the fractional distance between them is smaller than  $\frac{1}{2} - \epsilon$ , (2) blue, if the fractional distance between them is between  $\frac{1}{2} - \epsilon$  and  $\frac{1}{2} + \epsilon$ , (3) green, if the fractional distance between them is larger than  $\frac{1}{2} + \epsilon$ . As  $m$  gets larger and larger, Ramsey theory tells us that there must exist a large (going to infinity with  $m$ ) monochromatic clique in this graph. This clique cannot be red, as we show that a large number of pairwise close codewords can be used to break the protocol. It also cannot be green, as that would violate known distance bounds for error correcting codes. Thus, it must be blue, implying that restricting attention to this clique gives us our desired 2-random code.

### 3 Model and Preliminaries

#### 3.1 Notation and Preliminaries

For  $x \in \mathbb{R}$ , let  $\vec{x}$  be the vector (of appropriate dimension inferred from context) with all its coordinates being  $x$ . Throughout, all inequalities between vectors are coordinate-wise. For  $k \geq 0$ ,  $\Delta^k = \{(x_0, \dots, x_k) \in \mathbb{R}^{k+1} \mid \sum_{i=0}^k x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i \in [0, k]\}$  denotes the  $k$ -dimensional standard simplex. For  $x \in \mathbb{R}$  and  $k \geq 0$ , we write  $x^k$  as a shorthand for falling factorial  $\prod_{i=0}^{k-1} (x - i)$ .<sup>10</sup> For a set  $S$  and  $k \geq 0$ , let  $\binom{S}{k}$  be the collection of all subsets of  $S$  of size  $k$ . For a function  $f : X \rightarrow Y$  and subset  $X' \subseteq X$ ,  $f|_{X'}$  denotes the restriction of  $f$  onto  $X'$ . For  $x, y \geq 1$ ,  $R(x, y)$  is the (two-color) Ramsey number for  $x, y$ , which is well-known to be finite. For  $k \geq 1$  and two bit strings  $x, y \in \{0, 1\}^k$ , their Hamming distance is  $\Delta(x, y) = \sum_{i=1}^k \mathbb{1}[x_i \neq y_i]$ .

#### 3.2 Our Model: Round-Restricted Binary Feedback Channels

We now define (deterministic, binary) *protocols* with (non-adaptive) round-restricted feedback for the message transfer task, where Alice has an input and Bob's goal is to learn this input. Such a protocol is defined by a tuple:

$$\Pi = \left( n, r, \{L_i\}_{i \in [r+1]}, \{f_i\}_{i \in [r+1]}, \text{out} \right), \quad (5)$$

where (1)  $\{0, 1\}^n$  is the set of all possible inputs for Alice; (2)  $r$  is the number of feedback rounds. Equivalently, we can say that Alice speaks in  $r + 1$  rounds; (3) For all  $i \in [r + 1]$ ,  $L_i$  is the length of Alice's message in the  $i$ -th round. Throughout, we use  $L = \sum_{i=1}^{r+1} L_i$ ; (4) For all  $i \in [r + 1]$ ,  $f_i : \{0, 1\}^n \times \{0, 1, \perp\}^{L_1} \times \dots \times \{0, 1, \perp\}^{L_{i-1}} \rightarrow \{0, 1\}^{L_i}$  is the message function Alice uses in the  $i$ -th round; (5)  $\text{out} : \{0, 1, \perp\}^{L_1} \times \dots \times \{0, 1, \perp\}^{L_{r+1}} \rightarrow \{0, 1\}^n$  is the function Bob uses to compute the output.

**Execution of a protocol.** Let  $\Pi$  be a protocol as above. An adversary for  $\Pi$  is defined by a function  $\text{Adv} : \{0, 1\}^n \rightarrow \{0, 1, \perp\}^{L_1} \times \dots \times \{0, 1, \perp\}^{L_{r+1}}$ . For  $i \in [r + 1]$ , we will use  $\text{Adv}_i(\cdot)$  to denote the function that outputs the  $i$ -th coordinate of  $\text{Adv}(\cdot)$ . We next define an execution of  $\Pi$  in the presence of an adversary  $\text{Adv}$  for  $\Pi$ : At the beginning of the execution, Alice starts with an input  $x \in \{0, 1\}^n$ . The execution consists of  $r + 1$  rounds and before the  $i$ -th round, for  $i \in [r + 1]$ , Alice and Bob have the (same) transcript  $\tau_{<i} \in \{0, 1, \perp\}^{L_1} \times \dots \times \{0, 1, \perp\}^{L_{i-1}}$ . In round  $i$ , Alice computes the message  $f_i(x, \tau_{<i}) \in \{0, 1, \perp\}^{L_i}$  and sends it to Bob bit by bit, while Bob receives the string  $\tau_i = \text{Adv}_i(x)$ . As we assume a feedback channel, if  $i \leq r$ , Alice also receives the string  $\tau_i$  and both the parties add  $\tau_i$  to  $\tau_{<i}$  and continue executing the protocol.

If  $i = r + 1$ , the execution of the protocol terminates and Bob outputs  $\text{out}(\tau_{\leq r+1})$ . Observe that this execution is completely determined by  $x$ ,  $\Pi$ , and  $\text{Adv}$ . We denote the output of  $\Pi$  on input  $x$  in the presence of adversary  $\text{Adv}$  by  $\text{out}_{\Pi, \text{Adv}}(x)$ .

**Counting the noise.** Let  $\Pi$  be a protocol as above and  $\text{Adv}$  be an adversary for  $\Pi$ . For  $x \in \{0, 1\}^n$ , the amount of noise added by  $\text{Adv}$  in  $\Pi$  on input  $x$  is the number of times Bob's received bit is different from the bit Alice sent. Formally, we have:

$$\text{noise}_{\Pi, \text{Adv}}(x) = \sum_{i=1}^{r+1} \Delta(\text{Adv}_i(x), f_i(x, \text{Adv}_{<i}(x))). \quad (6)$$

---

<sup>10</sup>See also Falling factorials (Wikipedia) for the notation.

For  $\theta \in [0, 1]$ , we say that an adversary  $\text{Adv}$  has budget  $\theta$  if we have

$$\max_{x \in \{0,1\}^n} \text{noise}_{\Pi, \text{Adv}}(x) \leq \theta L.$$

**Types of Adversaries.** Let  $\Pi$  be a protocol as above and  $\text{Adv}$  be an adversary for  $\Pi$ . We say that  $\text{Adv}$  is a *corruption adversary* if it never outputs the symbol  $\perp$ , i.e., for all  $x \in \{0, 1\}^n$  and all  $i \in [r + 1]$ , we have  $\text{Adv}_i(x) \in \{0, 1\}^{L_i}$ . We say that  $\text{Adv}$  is an *erasure adversary* if it only “erases” the symbols sent by Alice. More precisely, we say that  $\text{Adv}$  is an erasure adversary if for all  $x \in \{0, 1\}^n$ , all  $i \in [r + 1]$ , and all  $j \in [L_i]$ , if  $(\text{Adv}_i(x))_j \neq \perp$ , then we have  $(\text{Adv}_i(x))_j = (f_i(x, \text{Adv}_{<i}(x)))_j$ .

**Resilience of a protocol.** Let  $\Pi$  be a protocol as above and  $\theta \in [0, 1]$ . We say that  $\Pi$  has *resilience*  $\theta$  over the binary erasure channel if for all erasure adversaries with budget  $\theta$  and all  $x \in \{0, 1\}^n$ , it holds that  $\text{out}_{\Pi, \text{Adv}}(x) = x$ . Resilience over the binary corruption channel is defined analogously.

## 4 Optimal List-Decodable Small Codes

In this section, we construct the codes used by our protocol.

### 4.1 Definitions of List Decodability

**Codes for erasures.** We start by defining list decodability for erasures.

► **Definition 6.** Let  $m, k, L \geq 1$  and  $d \in [0, 1]$ . We say that a code  $C : [m] \rightarrow \{0, 1\}^L$  is less-than- $k$ -list decodable for erasures up to radius  $d$  if for all subsets  $\Gamma \in \binom{[m]}{k}$ , we have  $\text{ns}_C(\Gamma) > d$ , where:

$$\text{ns}_C(\Gamma) = 1 - \frac{1}{L} \cdot \sum_{j=1}^L \mathbf{1}[\exists b \in \{0, 1\} \forall i \in \Gamma : C_j(i) = b].$$

To get the intuition behind the definition of  $\text{ns}$ , observe that  $\text{ns}_C(\Gamma)$  is the minimum fraction  $e$  of erasures for which there exists  $\tau \in \{0, 1, \perp\}^L$  such that for all  $i \in \Gamma$ , it is possible to erase  $e \cdot L$  symbols from  $C(i)$  and get  $\tau$ . Observe that this is equal to the fraction of coordinates where the encodings  $\{C(i)\}_{i \in \Gamma}$  are *not* all the same ( $\text{ns} = \text{not same}$ ).

For  $m, k \geq 1$ , we define  $\mathbf{d}_{\text{erase}}(m, k)$  to be the supremum of all values  $d \in [0, 1]$  for which there exists  $L \geq 1$  and a code  $C : [m] \rightarrow \{0, 1\}^L$  that is less-than- $k$ -list decodable for erasures up to radius  $d$ .

**Codes for corruptions.** Next, we define list decodability for corruptions:

► **Definition 7.** Let  $m, k, L \geq 1$  and  $d \in [0, 1]$ . We say that a code  $C : [m] \rightarrow \{0, 1\}^L$  is less-than- $k$ -list decodable for corruptions up to radius  $d$  if for all  $\tilde{x} \in \{0, 1\}^L$ , we have

$$|\{i \in [m] : \Delta(C(i), \tilde{x}) < dL\}| < k.$$

Analogous to  $\mathbf{d}_{\text{erase}}$ , for  $m, k \geq 1$ , we define  $\mathbf{d}_{\text{corr}}(m, k)$  to be the supremum of all values  $d \in [0, 1]$  for which there exists  $L \geq 1$  and a code  $C : [m] \rightarrow \{0, 1\}^L$  that is less-than- $k$ -list decodable for corruptions up to radius  $d$ .

## 4.2 Lemmas about $d_{\text{erase}}$ and $d_{\text{corr}}$

In this section, we show the results we need about  $d_{\text{erase}}$  and  $d_{\text{corr}}$ . First, we define a helper function  $d(\cdot, \cdot)$ :

$$d(m, k) = 1 - \frac{\binom{\lfloor m/2 \rfloor}{k} + \binom{\lceil m/2 \rceil}{k}}{\binom{m}{k}}. \quad (7)$$

We now show some useful properties of the function  $d$  defined in Equation (7).

► **Claim 8.** The following hold:

1. For all  $m \geq k \geq 1$ , it holds that

$$d(m, k) = 1 - \frac{(\lceil m/2 \rceil - 1)^{k-1}}{(2\lceil m/2 \rceil - 1)^{k-1}}.$$

2. For all  $m_2 \geq m_1 \geq k \geq 1$ , it holds that  $d(m_2, k) \leq d(m_1, k)$ , and moreover,

$$\lim_{m \rightarrow \infty} d(m, k) = 1 - \frac{1}{2^{k-1}}.$$

It follows that  $d(m, k) \geq 1 - \frac{1}{2^{k-1}}$  for all  $m \geq k \geq 1$ .

3. For all  $m \geq k_2 \geq k_1 \geq 1$ , it holds that  $d(m, k_2) \geq d(m, k_1)$ .

### 4.2.1 Lemmas about $d_{\text{erase}}$

We now show that the functions  $d_{\text{erase}}$  and  $d$  are the exact same. Owing to this lemma, we omit writing  $\text{erase}$  in the subscript in the rest of this text.

► **Lemma 9.** For all  $m, k \geq 1$ , we have:

$$d_{\text{erase}}(m, k) = d(m, k).$$

► **Lemma 10.** For all  $\epsilon > 0$ , there exists a constant  $K$  such that for all  $K' \geq K$  and for all  $m \geq 1$ , there exists a code  $C : [m] \rightarrow \{0, 1\}^{K' \log m}$  such that for all  $k \in [m]$ , the code  $C$  is less-than- $k$ -list decodable up to radius  $d(m, k) - \epsilon$ .

### 4.2.2 Lemmas about $d_{\text{corr}}$

Using the results of Section 4.2.1, we show the following lemma:

► **Lemma 11.** For all  $m, k \geq 2$ , we have:

$$d_{\text{corr}}(m, 2) = \frac{d(m, 2)}{2} \quad \text{and} \quad \lim_{m \rightarrow \infty} d_{\text{corr}}(m, k) = \frac{1}{2} - \frac{\binom{k-1}{\lceil k/2 \rceil - 1}}{2^k}.$$

## 5 Protocols Against Erasures

In this section, we show one direction of Theorem 1, as formalized below. Later, in Section 6, we prove the other direction.

► **Theorem 12.** For all  $\epsilon > 0$  and  $r, n \in \mathbb{N}$ , there exists a constant-rate (polynomial in  $\epsilon$ ) protocol for message transfer with  $r$  rounds of feedback, input length  $n$ , and the following resilience over the binary erasure channel:

$$\begin{cases} \frac{5}{7} - \epsilon, & \text{if } r = 1 \\ 1 - \frac{7}{12(r+1)} - \epsilon, & \text{if } r > 1 \end{cases}.$$

We prove Theorem 12 in the rest of this section. Throughout, we fix  $\epsilon > 0$  and  $r, n \in \mathbb{N}$ . We assume  $r < \frac{10}{\epsilon}$ . This is without loss of generality as a protocol for large  $r$  follows from a protocol for smaller  $r$ .

## 5.1 Our Protocol

Let  $K$  be the constant from Lemma 10 for  $\epsilon$ . For all  $K' \geq K$  and all  $m \geq 1$ , let  $C_{m,K'} : [m] \rightarrow \{0,1\}^{K' \log m}$  be as promised by Lemma 10. We will omit  $K'$  when it is clear from context. For a set  $\Gamma$  of size  $m$ , we will also view  $C_m$  as a code  $C_\Gamma : \Gamma \rightarrow \{0,1\}^{K' \log m}$ . Our protocol is given in Algorithm 1, where the lengths of the rounds are given as follows:

$$L_i = \begin{cases} \frac{4}{3} \cdot Kn, & \text{if } i = r = 1 \\ Kn, & \text{otherwise} \end{cases} \quad (8)$$

■ **Algorithm 1** Message transfer protocol over the erasure channel with  $r \geq 1$  feedback rounds.

**Input:** Alice has input  $x \in \Gamma_0 = \{0,1\}^n$ .

**Output:** Bob outputs  $y \in \{0,1\}^n$ .

- 1: **for**  $i = 1, \dots, r + 1$  **do**
- 2:   Alice sends  $C_{\Gamma_{i-1}}(x) \in \{0,1\}^{L_i}$  bit by bit.
- 3:   Bob receives  $\tau_i \in \{0,1,\perp\}^{L_i}$  and sends  $\tau_i$  via the noiseless feedback channel.
- 4:   Bob computes

$$\Gamma_i = \{x' \in \Gamma_{i-1} \mid \forall j \in [L_i] : \tau_{i,j} \in \{C_{\Gamma_{i-1},j}(x'), \perp\}\}.$$

- 5:   If  $i \leq r$ , Alice receives  $\tau_i$  as feedback and also computes  $\Gamma_i$  as above.

6: **end for**

- 7: Bob outputs the lexicographically first element in  $\Gamma_{r+1}$ , aborting if  $\Gamma_{r+1} = \emptyset$ .

## 5.2 Analysis

We now analyze Algorithm 1 and finish proving Theorem 12. That the protocol is constant rate is clear from Algorithm 1. It remains to show that it has the claimed noise resilience. For this, we fix an input  $x$  for Alice and an erasure adversary  $\text{Adv}$  for the protocol with desired budget as in Theorem 12. Observe that fixing  $x$  and  $\text{Adv}$  fixes the value of all the variables in the execution of Algorithm 1. For the analysis, we first show that:

► **Lemma 13.** *For all  $i \in [0, r + 1]$ , we have  $x \in \Gamma_i$ .*

► **Lemma 14.** *For all  $m \geq k' \geq k \geq 2$  such that  $(k', k) \neq (3, 2)$ , it holds that*

$$d(m, k') + d(k', k) \geq 1 + d(m, k).$$

At a high level, Lemma 14 will be applied as follows: Consider an adversary that shrinks the set  $\Gamma$  from size  $m$  to size  $k'$  in a given round and from size  $k'$  to size  $k$  in the next round. Lemma 14 shows that, if the two rounds are of equal length (recall from Equation (8) that the round lengths are always the same except when  $i = r = 1$ ), then it is always better for the adversary to erase one of the rounds completely and shrink from size  $m$  to size  $k$  directly in the other round.

We now divide the proof into two cases based on whether or not  $r = 1$ .

### 5.2.1 Proof of Theorem 12 When $r = 1$

Let  $k_i = |\Gamma_i|$  for  $i \in [0, 2]$ . We prove the theorem by showing that  $k_2 \leq 1$ . Together with Lemma 13 and Algorithm 1, this shows the correctness of Algorithm 1.

Observe that at the beginning of the  $i$ -th round, Alice and Bob agree on  $\Gamma_{i-1}$ , the subset of all remaining possibilities for  $x$  from the perspective of Bob, that are still consistent with the partial transcript  $\tau_1, \dots, \tau_{i-1}$  so far. In order to keep Bob confused among  $\Gamma_i$ , the adversary has to erase at least a  $\mathbf{d}(k_{i-1}, k_i) - \epsilon$  fraction of Alice's  $i$ -th message, due to Lemma 10. As this holds for all rounds, the overall fraction of erasures is lower bounded by

$$\frac{4}{7}(\mathbf{d}(k_0, k_1) - \epsilon) + \frac{3}{7}(\mathbf{d}(k_1, k_2) - \epsilon) = \frac{4\mathbf{d}(k_0, k_1) + 3\mathbf{d}(k_1, k_2)}{7} - \epsilon.$$

Now suppose  $k_2 \geq 2$ . It is sufficient to show  $4\mathbf{d}(k_0, k_1) + 3\mathbf{d}(k_1, k_2) \geq 5$  for a contradiction. Without loss of generality, we assume  $k_2 = 2$  since  $\mathbf{d}(k_1, k_2)$  only decrease as  $k_2$  becomes smaller by Item 3 of Claim 8. If  $k_1 = 3$ , we have

$$4\mathbf{d}(k_0, k_1) + 3\mathbf{d}(k_1, k_2) = 4\mathbf{d}(k_0, 3) + 3\mathbf{d}(3, 2) \geq 4 \cdot \frac{3}{4} + 3 \cdot \frac{2}{3} = 5$$

by Item 2 of Claim 8. Otherwise, by Lemma 14, we also get

$$\begin{aligned} 4\mathbf{d}(k_0, k_1) + 3\mathbf{d}(k_1, k_2) &= 4(\mathbf{d}(k_0, k_1) + \mathbf{d}(k_1, 2)) - \mathbf{d}(k_1, 2) \\ &\geq 4(1 + \mathbf{d}(k_0, 2)) - \mathbf{d}(k_1, 2) \\ &\geq 3 + 4\mathbf{d}(k_0, 2) \quad \text{(as } \mathbf{d}(\cdot, \cdot) \text{ is always upper bounded by 1)} \\ &\geq 3 + 4 \cdot \frac{1}{2} \quad \text{(by Item 2 of Claim 8)} \\ &= 5. \end{aligned}$$

### 5.2.2 Proof of Theorem 12 When $r > 1$

Let  $k_i = |\Gamma_i|$  for  $i \in [0, r+1]$ . Similarly to the proof in Section 5.2.1, we show that  $k_{r+1} \leq 1$ . This ensures Bob outputs the correct  $y = x$  because of Lemma 13 and Algorithm 1. Using a similar argument to Section 5.2.1, we have that the overall fraction of erasures is lower bounded by

$$\frac{1}{r+1} \cdot \sum_{i=1}^{r+1} \mathbf{d}(k_{i-1}, k_i) - \epsilon.$$

Now for the purpose of contradiction, suppose that  $k_{r+1} \geq 2$ . It is sufficient to show

$$\frac{1}{r+1} \cdot \sum_{i=1}^{r+1} \mathbf{d}(k_{i-1}, k_i) \geq 1 - \frac{7}{12(r+1)}.$$

In the following, we again assume without loss of generality that  $k_{r+1} = 2$  since  $\mathbf{d}(k_r, k_{r+1})$  decreases as  $k_{r+1}$  becomes smaller by Item 3 of Claim 8. Let  $j = \min\{t \in [r+1] \mid k_t \leq 3\}$ . By repeatedly applying Lemma 14, we have

$$\begin{aligned}
& \frac{1}{r+1} \cdot \sum_{i=1}^{r+1} d(k_{i-1}, k_i) \\
& \geq \frac{1}{r+1} \cdot \left( 1 + d(k_0, k_2) + \sum_{i=3}^{r+1} d(k_{i-1}, k_i) \right) \\
& \quad \vdots \\
& \geq \frac{1}{r+1} \cdot \left( j-1 + d(k_0, k_j) + \sum_{i=j+1}^{r+1} d(k_{i-1}, k_i) \right).
\end{aligned}$$

Since  $k_0 \geq \dots \geq k_{r+1} = 2$  by definition of  $\Gamma_i$ , either  $k_j = 2$  or  $k_j = 3$ .

In the former case where  $k_j = 2$ , we also have  $k_{j+1} = \dots = k_{r+1} = 2$  and thus

$$\begin{aligned}
& \frac{1}{r+1} \cdot \sum_{i=1}^{r+1} d(k_{i-1}, k_i) \\
& \geq \frac{1}{r+1} \cdot \left( j-1 + d(k_0, k_j) + \sum_{i=j+1}^{r+1} d(k_{i-1}, k_i) \right) \\
& = \frac{1}{r+1} \cdot (j-1 + d(k_0, 2) + (r+1-j) \cdot d(2, 2)) \\
& \geq \frac{1}{r+1} \cdot \left( j-1 + \frac{1}{2} + r+1-j \right) \quad (\text{by Item 2 of Claim 8}) \\
& = 1 - \frac{1}{2(r+1)} \\
& \geq 1 - \frac{7}{12(r+1)}.
\end{aligned}$$

In the latter case where  $k_j = 3$ , let  $j' = \min\{t \in [r+1] \mid k_t = 2\}$ . Then we have

$$\begin{aligned}
& \frac{1}{r+1} \cdot \sum_{i=1}^{r+1} d(k_{i-1}, k_i) \\
& \geq \frac{1}{r+1} \cdot \left( j-1 + d(k_0, k_j) + \sum_{i=j+1}^{r+1} d(k_{i-1}, k_i) \right) \\
& = \frac{1}{r+1} \cdot (j-1 + d(k_0, 3) + (j'-1-j) \cdot d(3, 3) + d(3, 2) + (r+1-j') \cdot d(2, 2)) \\
& \geq \frac{1}{r+1} \cdot \left( j-1 + \frac{3}{4} + j'-1-j + \frac{2}{3} + r+1-j' \right) \quad (\text{by Item 2 of Claim 8}) \\
& = 1 - \frac{7}{12(r+1)}.
\end{aligned}$$

This concludes the proof.

## 6 Impossibility Result for Erasures

In this section, we show the other direction of Theorem 1, as formalized below.

► **Theorem 15.** *For all  $r \in \mathbb{N}$ , there exists an  $n \in \mathbb{N}$  such that the resilience of any protocol for message transfer with  $r$  rounds of feedback and input length  $n$  over the binary erasure channel is at most:*

$$\begin{cases} \frac{5}{7}, & \text{if } r = 1 \\ 1 - \frac{7}{12(r+1)}, & \text{if } r > 1 \end{cases}.$$

We prove Theorem 15 in the rest of this section. Throughout, we work with a fixed  $r \in \mathbb{N}$  and define  $n$  to be large enough for asymptotic inequalities to hold. We now divide the proof into two cases based on whether or not  $r = 1$ .

### 6.1 Proof of Theorem 15 When $r = 1$

Fix a protocol  $\Pi$  with input length  $n$  and one round of feedback. Recall Equation (5) and let  $L_1, L_2$  be the lengths of Alice's messages sent in the two rounds, and  $f_1 : \{0, 1\}^n \rightarrow \{0, 1\}^{L_1}, f_2 : \{0, 1\}^n \times \{0, 1, \perp\}^{L_1} \rightarrow \{0, 1\}^{L_2}$  be the two message functions Alice uses in the two rounds.

First suppose that  $L_1 \geq \frac{4}{7}(L_1 + L_2)$ . By Lemma 9, there exists a subset  $\Gamma = \{x_1, x_2\} \in \binom{\{0, 1\}^n}{2}$  such that  $\text{ns}_{f_1}(\Gamma) \leq d(2^n, 2)$ . This implies the adversary is able to erase a  $d(2^n, 2)$  fraction of Alice's first message so that Bob's view when Alice's input is  $x_1$  is identical to Bob's view when Alice's input is  $x_2$ , and therefore Bob is forced to send the same feedback  $\tau_1 \in \{0, 1, \perp\}^{L_1}$  in both cases. Now the adversary simply erases Alice's second message entirely implying that Bob can never output the correct answer. By Item 2 of Claim 8, the overall fraction of erasures is upper bounded as

$$d(2^n, 2) \cdot \frac{L_1}{L_1 + L_2} + 1 \cdot \frac{L_2}{L_1 + L_2} \leq \frac{4 \cdot d(2^n, 2) + 3}{7} \xrightarrow{n \rightarrow \infty} \frac{5}{7}.$$

Now consider the other case where  $L_1 \leq \frac{4}{7}(L_1 + L_2)$ . Again by Lemma 9, there exists a subset  $\Gamma = \{x_1, x_2, x_3\} \in \binom{\{0, 1\}^n}{3}$  such that  $\text{ns}_{f_1}(\Gamma) \leq d(2^n, 3)$ . In this case, the adversary erases a  $d(2^n, 3)$  fraction of Alice's first message so that Bob's view is the same when Alice's input is any of  $x_1, x_2, x_3$ . Bob must send the same feedback  $\tau_1 \in \{0, 1\}^{L_1}$  in all three cases. Note that  $f_2(\cdot, \tau_1)$  can also be viewed as a valid code and thus Lemma 9 still applies. In particular, it is always possible to erase a  $d(3, 2) = \frac{2}{3}$  fraction of Alice's second message so that for at least two of  $x_1, x_2, x_3$ , Bob's view remains the same at the end of the protocol. This concludes the proof as the overall fraction of erasures is at most

$$d(2^n, 3) \cdot \frac{L_1}{L_1 + L_2} + \frac{2}{3} \cdot \frac{L_2}{L_1 + L_2} \leq \frac{4 \cdot d(2^n, 3) + 3 \cdot \frac{2}{3}}{7} \xrightarrow{n \rightarrow \infty} \frac{5}{7}.$$

### 6.2 Proof of Theorem 15 When $r > 1$

Fix a protocol  $\Pi$  with input length  $n$  and  $r$  rounds of feedback. Recall Equation (5) and for  $t \in [r+1]$ , let  $L_t$  be the length of Alice's message sent in the  $t$ -th round. Let  $L = \sum_{t=1}^{r+1} L_t$ .

We prove the theorem using an approach similar to Section 6.2, i.e., the adversary is always able to erase Alice's messages in such a way that Bob has the same view at the end of the protocol for at least two different inputs. In particular, the adversary erases the entire messages of all rounds except for  $i = \arg \max_{t \in [r+1]} L_t$ , the longest round, and  $j = \arg \max_{t \in [r+1] \setminus \{i\}} L_t$ , the second longest round. Then we have

$$L_i \geq \frac{L}{r+1}, \quad (9)$$

$$L_j \geq \frac{L - L_i}{r}. \quad (10)$$

First consider the case where  $i < j$ . Since the first  $i - 1$  rounds are completely erased, Bob obviously has the same view for all possible inputs at the beginning of the  $i$ -th round. By Lemma 9, the adversary can erase a  $d(2^n, 3)$  fraction of Alice's  $i$ -th message so that Bob's view is the same when Alice's input is any of some subset  $\Gamma = \{x_1, x_2, x_3\} \subseteq \{0, 1\}^n$ . This remains true at the beginning of the  $j$ -th round as all intermediate rounds are completely erased. Now again by Lemma 9, the adversary is able to erase a  $d(3, 2) = \frac{2}{3}$  fraction of Alice's  $j$ -th message so that Bob still has the same view at the end of the  $j$ -th round, for at least two of  $x_1, x_2, x_3$ . As all remaining rounds are also completely erased, Bob can never output the correct answer at the end of the protocol. By Item 2 of Claim 8, the overall fraction of erasures is upper bounded as

$$\begin{aligned} & d(2^n, 3) \cdot \frac{L_i}{L} + \frac{2}{3} \cdot \frac{L_j}{L} + 1 \cdot \frac{L - L_i - L_j}{L} \\ &= 1 - (1 - d(2^n, 3)) \cdot \frac{L_i}{L} - \frac{1}{3} \cdot \frac{L_j}{L} \\ &\leq 1 - (1 - d(2^n, 3)) \cdot \frac{L_i}{L} - \frac{1}{3r} \cdot \frac{L - L_i}{L} \quad (\text{by Equation (10)}) \\ &= 1 - \frac{1}{3r} - \left(1 - d(2^n, 3) - \frac{1}{3r}\right) \cdot \frac{L_i}{L} \\ &\leq 1 - \frac{1}{3r} - \left(1 - d(2^n, 3) - \frac{1}{3r}\right) \cdot \frac{1}{r+1} \\ &\quad (\text{as } 1 - d(2^n, 3) \xrightarrow{n \rightarrow \infty} \frac{1}{4} > \frac{1}{3r} \text{ for } r \geq 2, \text{ and by Equation (9)}) \\ &\xrightarrow{n \rightarrow \infty} 1 - \frac{7}{12(r+1)}. \end{aligned}$$

Now suppose that  $i > j$ . In this case, a similar argument shows the adversary must be able to confuse Bob by erasing a  $d(2^n, 3)$  fraction of Alice's  $j$ -th message as well as a  $d(3, 2) = \frac{2}{3}$  fraction of Alice's  $i$ -th message (in addition to completely erasing all other rounds of messages). Observe that  $L_i \geq L_j$  by definition and that  $d(2^n, 3) \xrightarrow{n \rightarrow \infty} \frac{3}{4} \geq \frac{2}{3}$ . So the overall fraction of erasures is at most

$$d(2^n, 3) \cdot \frac{L_j}{L} + \frac{2}{3} \cdot \frac{L_i}{L} + 1 \cdot \frac{L - L_i - L_j}{L} \leq d(2^n, 3) \cdot \frac{L_i}{L} + \frac{2}{3} \cdot \frac{L_j}{L} + \frac{L - L_i - L_j}{L},$$

which has the desired upper bound as already shown above.

---

## References

---

- 1 Rudolf Ahlswede, Christian Deppe, and Vladimir S. Lebedev. Non-binary error correcting codes with noiseless feedback, localized errors, or both. In *International Symposium on Information Theory (ISIT)*, pages 2486–2487, 2006. doi:10.1109/ISIT.2006.262057.
- 2 Noga Alon. Voting paradoxes and digraphs realizations. *Adv. Appl. Math.*, 29(1):126–135, 2002. doi:10.1016/S0196-8858(02)00007-6.
- 3 Noga Alon, Boris Bukh, and Yury Polyanskiy. List-decodable zero-rate codes. *IEEE Trans. Inf. Theory*, 65(3):1657–1667, 2019. doi:10.1109/TIT.2018.2868957.
- 4 Elwyn R. Berlekamp. *Block coding with noiseless feedback*. PhD thesis, Massachusetts Institute of Technology, USA, 1964. URL: <http://hdl.handle.net/1721.1/14783>.

- 5 Elwyn R Berlekamp. Block coding for the binary symmetric channel with noiseless, delayless feedback. *Error-correcting codes*, pages 61–68, 1968.
- 6 Vladimir Markovich Blinovskii. Bounds for codes in the case of finite-volume list decoding. *Problemy Peredachi Informatsii*, 22(1):11–25, 1986.
- 7 Vladimir M Blinovsky. Plotkin bound generalization to the case of multiple packings. *Problems of Information Transmission*, 45(1):1–4, 2009. doi:10.1134/S0032946009010013.
- 8 Mark Braverman, Klim Efremenko, Gillat Kol, Raghuvansh Saxena, and Zhijun Zhang. Round-vs-resilience tradeoffs for binary feedback channels. *Electron. Colloquium Comput. Complex.*, TR22-179, 2022. URL: <https://eccc.weizmann.ac.il/report/2022/179>, arXiv:TR22-179.
- 9 Marat Valievich Burnashev. Data transmission over a discrete channel with feedback. random transmission time. *Problemy peredachi informatsii*, 12(4):10–30, 1976.
- 10 Klim Efremenko, Gillat Kol, and Raghuvansh R. Saxena. Binary interactive error resilience beyond  $1/8$  (or why  $(1/2)^3 > 1/8$ ). In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 470–481. IEEE, 2020. doi:10.1109/FOCS46700.2020.00051.
- 11 Klim Efremenko, Gillat Kol, Raghuvansh R. Saxena, and Zhijun Zhang. Binary codes with resilience beyond  $1/4$  via interaction. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 1–12. IEEE, 2022. doi:10.1109/FOCS54457.2022.00008.
- 12 Peter Elias. List decoding for noisy channels. Technical report, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1957.
- 13 Krishnan Eswaran, Anand D. Sarwate, Anant Sahai, and Michael Gastpar. Zero-rate feedback can achieve the empirical capacity. *IEEE Transactions on Information Theory*, 56(1):25–39, 2010. doi:10.1109/TIT.2009.2034779.
- 14 David G. Forney. Exponential error bounds for erasure, list, and decision feedback schemes. *IEEE Transactions on Information Theory*, 14(2):206–220, 1968. doi:10.1109/TIT.1968.1054129.
- 15 Meghal Gupta, Venkatesan Guruswami, and Rachel Yun Zhang. Binary error-correcting codes with minimal noiseless feedback. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 1475–1487. ACM, 2023. doi:10.1145/3564246.3585126.
- 16 Meghal Gupta, Yael Tauman Kalai, and Rachel Yun Zhang. Interactive error correcting codes over binary erasure channels resilient to  $> \frac{1}{2}$  adversarial corruption. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 609–622. ACM, 2022. doi:10.1145/3519935.3519980.
- 17 Meghal Gupta and Rachel Yun Zhang. Efficient interactive coding achieving optimal error resilience over the binary channel. *CoRR*, abs/2207.01144, 2022. doi:10.48550/arXiv.2207.01144.
- 18 Meghal Gupta and Rachel Yun Zhang. The optimal error resilience of interactive communication over binary channels. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 948–961. ACM, 2022. doi:10.1145/3519935.3519985.
- 19 Meghal Gupta and Rachel Yun Zhang. Positive rate binary interactive error correcting codes resilient to  $> 1/2$  adversarial erasures. *CoRR*, abs/2201.11929, 2022. arXiv:2201.11929, doi:10.48550/arXiv.2201.11929.
- 20 Venkatesan Guruswami. List decoding from erasures: bounds and code constructions. *IEEE Trans. Inf. Theory*, 49(11):2826–2833, 2003. doi:10.1109/TIT.2003.815776.
- 21 Venkatesan Guruswami and Madhu Sudan. List decoding algorithms for certain concatenated codes. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 181–190, 2000. doi:10.1145/335305.335327.

22 Gregory Z. Gutin and Anders Yeo. Lower bounds for maximum weighted cut. *CoRR*, abs/2104.05536, 2021. doi:10.48550/arXiv.2104.05536.

23 Bernhard Haeupler, Pritish Kamath, and Ameya Velingker. Communication with partial noiseless feedback. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 40 of *LIPICS*, pages 881–897, 2015. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.881.

24 Michael Horstein. Sequential transmission using noiseless feedback. *IEEE Transactions on Information Theory*, 9(3):136–143, 1963. doi:10.1109/TIT.1963.1057832.

25 M. Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6(4):445–450, 1960. doi:10.1109/TIT.1960.1057584.

26 Svatopluk Poljak and Daniel Turzík. A polynomial time heuristic for certain subgraph optimization problems with guaranteed worst case bound. *Discrete Mathematics*, 58(1):99–104, 1986. doi:10.1016/0012-365X(86)90192-5.

27 Anant Sahai. Why do block length and delay behave differently if feedback is present? *IEEE Transactions on Information Theory*, 54(5):1860–1886, 2008. doi:10.1109/TIT.2008.920339.

28 Leonard J Schulman. Communication on noisy channels: A coding theorem for computation. In *Foundations of Computer Science (FOCS)*, pages 724–733, 1992. doi:10.1109/SFCS.1992.267778.

29 Leonard J Schulman. Deterministic coding for interactive communication. In *Symposium on Theory of computing (STOC)*, pages 747–756, 1993. doi:10.1145/167088.167279.

30 Leonard J. Schulman. Coding for interactive communication. *IEEE Transactions on Information Theory*, 42(6):1745–1756, 1996. doi:10.1109/18.556671.

31 Claude E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, 1956. doi:10.1109/TIT.1956.1056798.

32 Ofer Shayevitz. On error correction with feedback under list decoding. In *IEEE International Symposium on Information Theory, ISIT*, pages 1253–1257, 2009. doi:10.1109/ISIT.2009.5205965.

33 Ofer Shayevitz and Meir Feder. Optimal feedback communication via posterior matching. *IEEE Transactions on Information Theory*, 57(3):1186–1222, 2011. doi:10.1109/TIT.2011.2104992.

34 Ofer Shayevitz and Michèle A. Wigger. On the capacity of the discrete memoryless broadcast channel with feedback. *IEEE Transactions on Information Theory*, 59(3):1329–1345, 2013. doi:10.1109/TIT.2012.2227670.

35 Joel Spencer, Peter Winkler, and South St. Three thresholds for a liar. *Combinatorics, Probability and Computing*, 1:81–93, 1992. doi:10.1017/S0963548300000080.

36 Gang Wang, Yanyuan Qin, and Chengjuan Chang. Communication with partial noisy feedback. In *IEEE Symposium on Computers and Communications (ISCC)*, pages 602–607, 2017. doi:10.1109/ISCC.2017.8024594.

37 Norbert Wiener. *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 2019.

38 John M Wozencraft. List decoding. *Quarterly Progress Report*, 48:90–95, 1958.

39 K.Sh. Zigangirov. On the number of correctable errors for transmission over a binary symmetrical channel with feedback. *Problems of Information Transmission*, 12:85–97, 1976.