Radio Network Coding Requires Logarithmic Overhead

Klim Efremenko
Ben-Gurion University
klimefrem@gmail.com

Gillat Kol
Princeton University
gillat.kol@gmail.com

Raghuvansh R. Saxena
Princeton University
rrsaxena@princeton.edu

***Abstract*—We consider the celebrated *radio network* model for abstracting communication in wireless networks. In this model, in any round, each node in the network may broadcast a message to all its neighbors. However, a node is able to hear a message broadcast by a neighbor only if no *collision* occurred, meaning that it was the only neighbor broadcasting.**

While the (noiseless) radio network model received a lot of attention over the last few decades, the effect of noise on radio networks is still not well understood. In this paper, we take a step forward and show that making radio network protocols resilient to noise may require a substantial performance overhead. Specifically, we construct a multi-hop network and a communication protocol over this network that works in T rounds when there is no noise. We prove that any scheme that simulates our protocol and is resilient to stochastic noise, requires at least $cT \log(n)$ rounds, for some constant c. This stands in contrast to our previous result (STOC, 2018), showing that protocols over the single-hop (clique) network can be made noise resilient with only a constant overhead. Our result also settles a recent conjecture by Censor-Hillel, Haeupler, Hershkowitz, Zuzic (2018).

We complement the above result by giving a scheme to simulate any protocol with a fixed order of transmissions with only an $O(\log(n))$ overhead.

***Index Terms*—Interactive Coding; Wireless Broadcast; Lower Bounds; Communication Complexity;**

I. Introduction

We consider the extensively studied radio network model for abstracting communication in wireless networks [CK85]. In this model, a set of agents represented as nodes in a graph, want to disseminate their privately stored data throughout the network, or, more generally, perform a joint computation that involves all their data. The nodes are each equipped with a wireless device that, in any time step, can either act as a transmitter or as a receiver. We think of the neighbors of a transmitting node as being within its range of transmission, which is necessary, but not sufficient for receiving the transmission. If a node v transmits at time t, then, a neighbor u of the node v receives v's transmission if and only if u decides to receive at time t and v is the only neighbor of u that transmits at time t.

The (noiseless) radio network model received a lot of attention over the last few decades, and has an ever growing number of wireless and distributed applications. On the other hand, the effect of noise on radio networks is not well understood, although wireless communication errors are extremely common in practice. In our previous paper [EKS18], we took a step forward and initiated the study of interactive coding for the single-hop (clique) network, following the great work of [Gam87], [Gal88], [GKS08]. We proved that the single-hop network admits constant rate coding. That is, we showed how to convert any protocol that works over the noiseless single-hop network, to a noise resilient protocol of similar length, up to a multiplicative constant.

In our current paper, we consider noise in general multi-hop networks, and show that, in contrast to the single-hop case, noise resilience may require a substantial performance overhead. Roughly, our main result proves that there exists a network topology G and a communication protocol Π over G with T rounds that works in the noiseless setting, such that every noise resilient protocol Π' that 'simulates' Π requires $\Omega(T \log(n))$ rounds. We complement our result with a scheme for converting any protocol Π with a fixed order of transmissions over any network G, to a noise resilient protocol Π', while only incurring a multiplicative $\mathcal{O}(\log n)$ overhead.

A. Modeling Collisions and Noise

To be able to give precise statements of our results, we need to define the assumed radio network model. We chose to present our results in the model recently suggested by [CHHZ18], described below. We mention, however, that our results are proved for more general

2575-8454/19/$31.00 ©2019 IEEE
DOI 10.1109/FOCS.2019.00030

models (see more about that in subsection I-B and section II).

To define the assumed (noiseless) radio network model, called here RADIO, and the assumed noisy radio network model, called here NOISY RADIO, one is asked to make two modeling decisions. The first decision is about the way the radio network deals with *collisions*. As explained above, when two or more neighbors of a node v transmit in the same round, v may not receive (any of) their messages. What does v receive then? The most commonly studied collision model in the literature, and the one we are using in RADIO and NOISY RADIO, is called *"collision-as-silence"*, and it asserts that v simply does not get anything. Thus, v is unable to distinguish a collision from the background noise occurring when no neighbor transmits. The main alternatives to collision-as-silence, studied by prior works (somewhat to lesser extent), are the *collision detection* model, where v is notified that a collision occurred, and the *flaky collision* model, where v may receive one of the transmitted messages. For an excellent survey, see [Pel07].

The second modeling decision we face is regarding the *noise*. We choose to model the noise in NOISY RADIO as stochastic erasures. More formally, let v and u be neighbors such that v is the only neighbor of u transmitting in round t. Then, for $\epsilon > 0$, in the NOISY RADIO$_\epsilon$ model, in round t, u receives v's message with probability $1-\epsilon$ (independently of any other event), and otherwise witnesses silence. We selected stochastic erasures as they are "weaker" than other models of noise studied in the literature (*e.g.,* bit flips, insertion-deletion errors, adversarial errors), and thus, make our lower bound stronger.

B. Our Results

a) Lower bound: The following theorem is the main contribution of our paper. It lower bounds the overhead, in terms of rounds, required in order to make a protocol noise resilient. A proof sketch is given in section II, and the full proof is given in section III and section V. As far as we know, this is the first lower bound that also rules out *adaptive* simulations. Such lower bounds are tricky, as when a message is lost due to an erasure, it may not only cause the parties to change their future messages, but can also cause them to completely change their broadcast/receive patterns. Thus, the parties can dynamically allocate more rounds to the parties that were erased the most, decreasing the length of the simulation (see [GHS14] to learn more about this phenomenon).

Theorem I.1 (main). *Let Γ be a set such that $|\Gamma| > 1$, $n \in \mathbb{N}$ be sufficiently large, and $\epsilon > 0$ be a constant. There exists a network G with n nodes and a (deterministic and non-adaptive) protocol Π of length $T(n)$ in the* RADIO *model over G with message set Γ, such that any protocol Π' that simulates Π in the* NOISY RADIO$_\epsilon$ *model over G with message set Γ, requires $\Omega(T(n)\log n)$ rounds.*

The protocol Π we construct in order to prove Theorem I.1 is a protocol for a routing task (*a.k.a.* a *multicommodity multicast* task), called TribeTransfer. In multicommodity multicasts, we are given a set of source-sink pairs so that each source has an input (*a.k.a.* rumor) that must be transmitted to the corresponding sink. Such tasks are a generalization of the extensively studied broadcast task (one node is a source and all the others are sinks) and gossip task (all pairs of nodes are sink-source pairs). Our result can be interpreted as showing that even protocols for "simple" tasks, such as routing tasks, that do not involve any joint computation, can become expensive in the presence of noise. We note that it is an interesting open problem to get a near-logarithmic lower bound, like the one in Theorem I.1, for the case where the length of the protocol T can approach infinity (independently of the number of nodes n).

The noiseless protocol for TribeTransfer is also "simple" in the sense that it is: (1) deterministic; (2) the set of nodes communicating in every round is known in advance, and is independent of the nodes' inputs, the messages received in previous rounds, and the noise. Protocols admitting the second property are called *non-adaptive* (*a.k.a. static* and *oblivious*). Interestingly, despite the obvious advantages of adaptivity, for many useful tasks, the state of the art algorithms are non-adaptive. For such non-adaptive protocols, we prove that our $\Omega(\log n)$ lower bound is in fact tight. Getting a similar upper bound for general (adaptive) protocols is an interesting problem.

b) Upper bound: For completeness, we also give a way to convert any non-adaptive noiseless protocol Π, such as the one used in the proof of our main result, to a noise resilient protocol with an $\mathcal{O}(\log n)$ blowup to the length. We note that the claim is trivial when the length of Π is $\mathrm{poly}(n)$, as each message in Π can be repeated $\mathcal{O}(\log n)$ times. This ensures that even in the presence of noise, all nodes get the messages

'correctly' with high probability. To prove that the noise can be dealt with even for longer protocols, we use the (by now standard) recursive "rewind-if-error" interactive coding mechanism (see, for example, [Sch92], [EKS18]). Our simulation protocol and its analysis are given in Appendix section B.

Theorem I.2. *Let Γ be a set such that $|\Gamma| > 1$, $n, T \in \mathbb{N}$ be sufficiently large, and $\epsilon \geq 0$ be constant. Let G be a graph with n vertices, and let Π be a non-adaptive protocol of length T in the* RADIO *model over G with message set Γ. Then, there exists a protocol Π' that simulates Π in the* NOISY RADIO$_\epsilon$ *model over G and only requires $\mathcal{O}(T \log n)$ rounds.*

We mention that while, for simplicity of exposition, the statements of our results use the RADIO and NOISY RADIO models, we actually prove stronger statements. Our lower bound is proved assuming an even weaker model for the simulating protocol, where in round t, *all* nodes get *all* messages that were successfully received by at least one node in round t (see subsection IV-C for a formal definition). In addition, our upper bound gives a scheme that is resilient against the very strong noise model where, when a collision occurs, an adversary is controlling the received massages (see [EKS18], [Hae14]).

c) Simulation: Both our lower and upper bounds consider the simulation of a protocol Π by another protocol Π'. But, what does it mean to simulate? As is usually the case in the field of interactive coding, in our upper bound result, Π simulates Π' in a very strong sense: given a transcript for Π' one can deduce the transcript for Π with the same randomness and noise. However, our lower bound rules out much weaker simulations. Namely, recall that the protocol Π is deterministic, thus, for any set of inputs, each node has a single correct output. We show that for any protocol Π' with fewer than $c \cdot T \log n$ rounds (for some fixed constant $c > 0$), there exists a set containing at least $4/5$ fraction of the sink nodes, such that for every node in the set, the probability (over the selection of inputs, the randomness used by the simulation, and the noise) that it outputs an incorrect value is greater than $3/4$ (see Theorem V.1 for an exact statement)[1].

C. Related Work

The field of coding for interactive communication was introduced in seminal papers of Schulman [Sch92], [Sch93], [Sch96]. Various aspects of two-party interactive coding (such as computational efficiency, interactive channel capacity, noise tolerance, list decoding, different channel types, *etc.*) were considered in recent years, with most works focusing on simulation protocols [GMS11], [BR11], [BKN14], [Bra12], [MS14], [BE14], [GMS14], [GH14], [GHK+16], [EGH16], [BGMO17], [EKS18], to cite a few.

Few strong interactive coding lower bounds are known, and these entail substantial technical difficulties [GKS08], [KR13], [BEGH16], [GK17]. Out of these, the one that is the most relevant to our work is [BEGH16], which proves that coding over the point-to-point star network, requires almost-logarithmic blowup to the number of rounds, matching the beautiful scheme of [RS94]. Observe that the point-to-point model allows for more freedom for designing a protocol, as each node can transmit different messages to each of its neighbors over a private channel, and, more importantly, one does not need to worry about collisions. Yet, the problems of coding over point-to-point networks and over radio networks are incomparable, as in the former case, both the faultless and the noise resilient protocols use point-to-point channels. Due to the fundamental differences between the models, as is evident by the fact that the star topology does admit constant rate coding in the NOISY RADIO model [EKS18], we were not able to use the techniques developed by [BEGH16] (see section II for a more detailed comparisons).

While, as mentioned above, no lower bounds for adaptive simulations were previously known, several papers give strong upper bounds. The two-party setting is considered by the very interesting works [Hae14], [GHS14], [AGS16], where it is shown that adaptive coding schemes can allow for strictly better rate and

[1]We note that the weak statement that, with probability $3/4$, there exists at least one node whose output is incorrect, has an easy proof: consider the star topology where the center wants to broadcast its input. This statement is weaker than ours in two respect: (1) the size of the incorrect set is small; (2) the order of quantifiers is different, as we show that the *same set* of nodes is incorrect in many executions. Our order of quantifier is the standard one for interactive coding. It parallels the notion of randomized algorithms, where it's ok if for every random string, the computation fails for a (different) set of inputs, but the computation for any input should not fail for many random strings.

error tolerance[2]. In [EKS18], we construct a constant rate adaptive coding scheme for the single-hop network, when it was known that the best rate achievable by a non-adaptive scheme is $\mathcal{O}(1/\log\log n)$ [GKS08].

Noise in radio networks was considered by several prior works [Gam87], [Gal88], [FK00], [KM05], [New04], [GKS08], [EKS18], [CHHZ17], [GHM18], [CHHZ18]. The related dual graph model, where some of the links are unreliable, is studied by [KLN+10], [GHLN12], [CGK+14], [GLN13], and [KKP01], [KPP10] consider radio networks with faulty nodes.

a) Relation to [CHHZ18]: The most relevant to our work, aside from [EKS18], is the very recent work [CHHZ18] defining the model $\texttt{NOISY RADIO}_\epsilon$ we use in this paper. They show that every non-adaptive protocol can be made noise resilient with only a blowup of $\text{poly}(\log\Delta, \log\log n)$ in length, where Δ is the maximum degree of the graph. They also prove that, under certain assumptions (*e.g.*, the noise resilient protocol is non-coding), an $\Omega(\log(\Delta))$ multiplicative round overhead is required. The authors further conjecture that such a statement can be proved assumptions-free for some specific protocol and topology which is described in the paper.

Our main result essentially proves their conjecture with a different topology and protocol: Fix n. Since the degree of the network G in Theorem I.1 with n vertices is $\Delta = \text{poly}(n)$, Theorem I.1 directly proves the conjecture for $\Delta = \text{poly}(n)$. To prove the conjecture for a smaller Δ, we take vertex-disjoint copies of G with $\text{poly}(\Delta)$ vertices and consider the task of simulating Π on all copies of G in parallel.

II. Proof Sketch

The main contribution of this work is an $\Omega(\log n)$ lower bound on the overhead of an interactive coding scheme for a particular multi-hop radio network. We sketch this result in this section.

A. The Star Topology

Consider the star graph on $n+1$ nodes where one 'center' node is connected to n 'leaf' nodes. Suppose that all the leaf nodes have a one bit input that they want to send to the center node. Also, assume that each node may only broadcast one bit in every round. Call this communication task $\mathsf{Transfer}_n$. In the noiseless case, the task $\mathsf{Transfer}_n$ requires n rounds, as no two leaf nodes can broadcast together (otherwise, there is a collision). Clearly, n rounds also suffice, as the nodes can broadcast their inputs one after the other in some pre-determined order.

Consider now the noisy case, where each broadcast from the leaf nodes is erased with a constant probability. We may assume, without loss of generality, that each node is only broadcasting its input bit, as any information it has about the other inputs must have been broadcast to it by the center. We observe that even in the erasure case, the *expected* number of rounds that are needed for the center node to hear the input of one leaf node, assuming it is the only one communicating, is only a constant. If one assumes that when a message reaches the center, all the $n+1$ nodes in the network get *notified* and proceed to sending the next input, an easy Chernoff bound shows that a noise resilient protocol for $\mathsf{Transfer}_n$ with $\mathcal{O}(n)$ rounds exists.

How strong is the assumption that all the nodes get notified of a successful transmission? In the extreme case, when the center node does not broadcast at all, the leaf nodes have no choice but to repeat their input $\Omega(\log n)$ times to allow a union bound over all n inputs. Thus, a noise resilient protocol for $\mathsf{Transfer}_n$ would require $\Omega(n\log n)$ rounds.

However, the center node can give *feedback* to the leaf nodes to reduce the number of rounds in the broadcast protocol. Indeed, in our prior work [EKS18], we (implicitly) implement a sufficiently powerful feedback mechanism for the star topology, reducing the number of rounds all the way down to $\mathcal{O}(n)$. While our focus in [EKS18] was the single-hop (clique) topology, our simulation can be shown to work for topologies satisfying (roughly): (1) there is a 'center' node that can 'reach' all other nodes in a constant number of rounds (*i.e.,* there is a constant round protocol for broadcasting a message from the center to all other nodes); (2) there exists a constant c such that all sets of nodes S that can *broadcast successfully* have $|S| \leq c$. Here, we say that S broadcasts successfully if there exists a node outside S with exactly one neighbor in S. The second property implies that, effectively, at most c nodes are broadcasting in any round. The center is then able to broadcast feedback on their messages, reaching all nodes in constant number of rounds.

The above discussion suggests that while $\Omega(\log n)$ messages per node are required for the center to hear any

[2]We note that the work of [AGS16] considers channels with no collisions.

leaf's input *with high probability*[3], the *expected* number of messages required is only a constant. Our lower bound approach is to find a topology where the feedback from the 'center' node can be somehow controlled, allowing us to "go from the expected cost to the high probability cost".

We mention, as an aside, the work of [BEGH16], showing a near-logarithmic lower bound for the star topology in the *point-to-point* model. Their proof is also "going from the expected cost to the high probability cost", but does so using an inherently different approach. They lower bound the overhead required for simulating a *pointer chasing* protocol in the presence of noise (rather than a message exchange protocol). They exploit the fact that the messages communicated by the nodes in a given round of the protocol crucially depend on all prior messages sent to them, thus all nodes must know all previous messages with high probability. Due to the differences in the models, and as we are mainly interested in protocols for routing information, where the messages communicated by the nodes are typically independent of the history, we are unable to use their ideas.

B. Controlling the Feedback

One easy way to make the feedback redundant is to constrain the leaf nodes to only broadcast in a fixed subset of the rounds, *i.e.*, make the protocol non-adaptive. This subset is decided beforehand and cannot be changed by any feedback from the center node. In this restricted scenario, it is again easy to see an $\Omega(\log n)$ lower bound (see [GKS08] for a related result for the clique network).

Since we aim to prove a lower bound for adaptive protocols, where the nodes can decide whether to receive or transmit based on the communication history and their input, we need other ways to control the feedback. To this end, consider a graph G that has k disconnected copies of the star graph, each with k leaves. Number the leaf nodes on each of the star graphs from 1 to k arbitrarily (thus, G is a graph with $n = k(k+1)$ vertices). We want to solve $\mathsf{Transfer}_k$ on all of the star subgraphs *in parallel*, meaning that the center of each star should know the messages of all the leaves in its star. Clearly, when there is no noise, this problem can still be solved in k rounds, by having the i^{th} nodes in each star broadcasting in round i.

[3]High here means $1-n^{-c}$, for some constant $c > 1$, to allow union bound over the leaves.

a) Lower bound under a restriction: Suppose that, for all i, all the k leaf nodes numbered i were restricted to only broadcast together. We now describe why, under this restriction, any noise resilient protocol for our problem would require $\Omega(k \log k)$ rounds. Since all the k nodes with the same number i are restricted to broadcast together, the number of rounds needed for this group, say group i, to successfully broadcast is equal to the number of rounds needed for the node that was corrupted the most. This number is easily seen to be $\Omega(\log k)$. Since this holds for all the k groups of nodes, it implies an $\Omega(k \log k)$ lower bound.

A crucial subtlety that we omit in the analysis above is that information about the nodes' inputs may be transmitted by the order in which the groups choose to broadcast, *e.g.*, there may be a 'fancy' protocol where the nodes in group 1 would broadcast in round 1 only if they have some particular input, *etc.*. We show how to deal with this 'leak' of information when we talk about our actual construction in subsection II-C.

We end this subsection by observing that in the graph G, all the $k = \Theta(\sqrt{n})$ nodes labeled by i can broadcast successfully at the same rounds and have their respective centers hear them, thus violating Property 2 required by [EKS18] in a strong way.

C. Our Construction

The discussion in the foregoing section suggests that there are two main issues that need to be tackled. Firstly, we need to find a way to actually implement the fairly artificial restriction described above. Secondly, we need a way to control and account for the leak of information in the system. In this subsection, inspired by the discussion above, we describe a graph G for which a weak version of the above restriction holds. Later, we describe how we account for the information leak in our analysis.

a) Our network: The graph G we construct has $n = 2k^2$ nodes divided into two sets A and B of k^2 nodes each. The nodes in the sets A and B are divided into k groups of k nodes each. We will use a_{ij} (respectively, b_{ij}) where $i, j \in [k]$, to refer to the j^{th} node in the i^{th} group of A (resp., B). We now describe the edges in G:

1) *Edges between A and B:* There is an edge between node $a_{ij} \in A$ and another node $b_{i'j'} \in B$ if either $(i \neq i')$ or $(i = i'$ and $j = j')$. In words, node a_{ij} is connected to node b_{ij} and also connected to all nodes *not* in group i of B.

2) *Edges in B:* There is an edge between nodes $b_{ij}, b_{ij'} \in B$ for every i and $j \neq j'$.

Observe that the nodes in group i of A form an independent set, while the nodes in group i of B form a clique, see Figure 1.

b) Our communication task: All the nodes in the set A have a single symbol as input. Nodes in B have no input. The communication problem we consider is where all the nodes b_{ij} in group i of B have to output the input symbol of all the nodes in group i of A. The nodes in A are not required to produce an output.

In the noiseless setting, this communication task admits a simple non-adaptive and deterministic $2k$ rounds protocol: In round $i \in [k]$, all the nodes in group i of A broadcast their input. After these k rounds, node b_{ij} knows the input of a_{ij}. In round $k + j$ for $j \in [k]$, all the nodes in $\{b_{ij} \mid i \in [k]\}$ broadcast the symbol they received. After this round, all nodes in group i of B knows the input of a_{ij}, for every i.

c) Implementing the restriction: Since the nodes in A are not required to produce an output, we assume that the nodes in B do not broadcast during the protocol. This assumption is not without loss of generality and we justify it in subsection II-D.

Suppose that two nodes a_{ij} and $a_{i'j}$, where $i \neq i'$ are broadcasting in a given round. Then, since $i \neq i'$, all the nodes in the set B are connected to at least one of these two nodes. Thus, a third node in the set A cannot successfully broadcast in this round (meaning that none of its neighbors will receive the transmission). This implies that in any round, either all the nodes broadcasting are in the same group of the set A, or the number nodes that are *successfully* broadcasting is not more than 2. This is actually just a weak version of the restriction described in subsection II-B where we said that all the nodes broadcasting in a given round have to be in the same group.

D. Our Analysis

Our network and communication task are designed to make the 'basic' lower bound analysis simple: It is roughly true that all the nodes of A broadcasting in any round belong to the same group. Since there are $k = \Theta(\sqrt{n})$ nodes in any group, if these groups broadcast in a fixed order, then a given group will need to broadcast $\Omega(\log n)$ times. Since there are k groups, the total number of rounds in any simulation will be $\Omega(k \log n)$, as required for the lower bound.

This basic analysis, however, is quite far from an actual lower bound, and we need to face several challenges to be able to complete the proof. We next describe some of these challenging cases: (1) It may happen that the nodes in A convey information about their input by the order in which the groups broadcast. (2) It may also happen that a lot of information is conveyed in rounds where two nodes from different groups in A broadcast. (3) There may be a clever protocol that somehow uses broadcasts by the players in the set B. (4) Finally, it may happen that a lot of information is conveyed in rounds where a subset of a group of nodes in A broadcast, or maybe such a subset broadcasts together with a small number (< 2) of nodes outside the group.

We now show how to handle each of these cases one by one.

(1) We handle the case where information is revealed by the order in which the groups in A broadcast, using an information theoretic argument. Observe that there are k^2 nodes in the set A and each node has one bit as input. Thus, a simulation protocol needs to send $\Omega(k^2)$ bits of information from A to B. If the simulation protocol has fewer than T rounds, then there are at most k^T possible orders in which the groups in A can broadcast. This means that if $T = \Omega(k \log k)$, then the amount of information that can be conveyed using the order in which the groups broadcast is at most $\log(k^T) \ll k^2$, and thus negligible.

(2) Next, we account for the information conveyed during the rounds where two nodes from different groups of A broadcast, using a similar information theoretic argument. Recall that if nodes from different groups of A broadcast, then the number of such nodes is at most 2. Thus, in such a round, at most 2 bits of information can be transmitted 'directly'. Of course, as before, information may also be transmitted using the order of such broadcasts. However, there are at most $|A|^2 = k^4$ ways of choosing these two nodes, and thus at most $(k+k^4)^T$ ways of ordering their broadcasts (the additive term of k comes from the rounds where all the nodes broadcasting are in the same group). As before, this can only convey $\mathcal{O}(T \log k) \ll k^2$ bits of information.

(3) We need to argue that a protocol cannot really use broadcasts from the nodes in the set B. For this part of the argument, we use the fact that, in the above, we actually upper bounded the information conveyed to *all* of the set B (instead of to a given node in B).

Thus, the argument above holds even if we 'reveal' everything known to *any* of the nodes in B to *all* the nodes in the network. In this case, since nodes in B have no inputs, their broadcasts can be computed by any node in the graph, and thus are redundant.

We formalize this by considering a new model called SHARED RADIO, where all the nodes receive the same (shared) transcript. This transcript contains all the received transcripts of all the nodes in the original model (as well as some additional information, described in item 4, tailor made for our network and task). In SHARED RADIO, we require that the nodes in B do not broadcast and that the nodes in A only broadcast their input. We are able to show that a lower bound for SHARED RADIO implies a lower bound for NOISY RADIO. We then show a lower bound for SHARED RADIO using the ideas presented in this section.

(4) Finally, an extension of the information argument in the first two items does not seem to work for this case (at least not directly), as there are 2^k subsets of any given group, thus giving too many options. Before we describe how we handle this case, we would like to elaborate on why this may be a problem. At first sight, it may seem that having only a subset of nodes in a group broadcast can only give less information than having the entire group broadcast. This is, for instance, true when the noise is adversarial in case of a collision/silence. However, we work in the NOISY RADIO model, where collisions and silences are received as erasures, and it is conceivable that a subset may give some information that the entire group did not give.

For example, consider the case when a subset S of a group i in A is broadcasting, and a node v in A, but not in group i of A, is also broadcasting. Let S' be the subset of group i in B containing the nodes 'corresponding' to S. That is, $b_{ij} \in S'$ if and only if $a_{ij} \in S$. By our graph construction, the nodes in S' will have two neighbors broadcasting, and will always receive the erasure symbol λ, whereas the nodes in group i of B that are not in S' will have only one neighbor broadcasting, namely v, and will receive a noisy copy of its broadcast. In this case, when a node in the group i in B receives λ, it 'suggests' that the corresponding node in A tried to broadcast. Thus, even rounds where nodes receive λ can reveal meaningful information.

In this case, it turns out that if we 'reveal' the input of node v to all the nodes in the graph (as part of the additional information in SHARED RADIO), then, we can assume that all the nodes in group i broadcast without loss of generality. At a high level, this is because once the input of node v is known, it doesn't need to broadcast and the nodes in group i can broadcast without worrying about a collision with node v. Observe that the assumption that only an entire group can broadcast in a given round (instead of any subset), allows us to use the information theoretic bounds in the first two items.

We believe that the idea of revealing some extra information to account for signaling is more general and may find applications beyond the current work.

III. Lower Bound: Network and Protocol Construction

In this section, we construct the network and protocol used in the proof of Theorem I.1.

A. The Network Construction

Fix $k > 0$, we define a graph G over $n = 2k^2$ vertices. We partition the vertices into two sets, A and B, of k^2 vertices each. Let i, j be variables that take values in $[k]$. We index the sets A, B by the tuple (i, j) and use the notation a_{ij} (respectively, b_{ij}) to denote the $(i, j)^{\text{th}}$ vertex of A (respectively, B).

The graph G has the following edges:

- $a_{ij} \sim b_{ij}$ for all $i, j \in [k]$.
- $b_{ij} \sim b_{ij'}$ for all $i, j, j' \in [k]$ such that $j \neq j'$.
- $a_{ij} \sim b_{i'j'}$ for all $i, i', j, j' \in [k]$ such that $i \neq i'$.

We use the shorthand $A^i = \{a_{ij} \mid j \in [k]\}$ and $B^i = \{b_{ij} \mid j \in [k]\}$. Observe that each of the sets A^i is an *independent set*, and each of the sets B^i is a *clique*. For a vertex $v \in G$, the neighborhood of v is denoted using $N(v) = \{u \in G \mid u \sim v\}$. We use the convention that $v \in N(v)$.

The graph G, for $k = 3$ is drawn in Figure 1.

The graph G constructed above was designed to satisfy the following key property:

Fact III.1. *Let $i \neq i' \in [k]$ and $j, j' \in [k]$. Then, $N(a_{ij}) \cup N(a_{i'j'}) = B$.*

B. The Communication Task

In this subsection, we define the hard communication task, called $\mathsf{TribeTransfer}_k$, for $k > 0$. Consider the graph G defined in subsection III-A with $n = 2k^2$ nodes. Suppose that all nodes in the set A have an input,

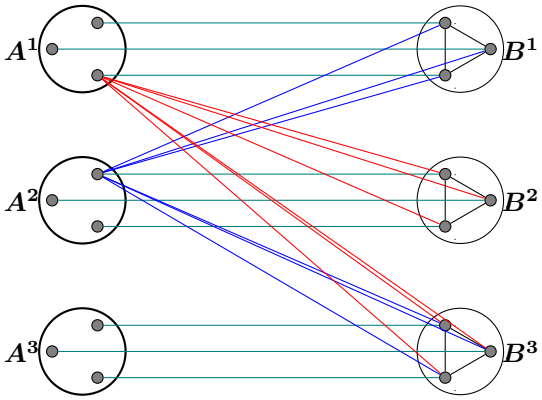

Fig. 1. Our graph construction, shown for $k = 3$. The nodes in the set A are on the left while the nodes in set B are on the right. We show all the edges between a_{ij} and b_{ij}, for all $i, j \in [k]$, in teal. Some edges in the graph are not shown to reduce clutter. Notice how any two nodes that are not in the same A^i cover all of B (Fact III.1).

i.e., node a_{ij} has input $x_{ij} \in \Gamma$ for some set Γ. The parties are required to run a broadcast protocol such that, at the end of the protocol, for all i, j, the node b_{ij} outputs X^i, which is defined as the k-length string $X^i = x_{i1}, x_{i2}, \cdots, x_{ik}$. The nodes a_{ij} are not required to produce an output.

a) Noiseless protocol for $\mathsf{TribeTransfer}_k$*:* We describe and analyze a protocol that solves $\mathsf{TribeTransfer}_k$ in the noiseless setting ($\epsilon = 0$). The protocol proceeds in $2k$ rounds. In round $i \in [k]$, all the nodes in A^i broadcast their input. Because the channel is noiseless, after these k rounds, node b_{ij} knows x_{ij} for all $i, j \in [k]$. In round $k + j$ for $j \in [k]$, all the nodes in $\{b_{ij} \mid i \in [k]\}$ broadcast the symbol they received, *i.e.*, node b_{ij} broadcasts x_{ij} for all $i \in [k]$. Observe that in round $k + j$, for every $j \neq j'$, node $b_{ij'}$ learns x_{ij}, as b_{ij} broadcasts it and none of the other neighbors of $b_{ij'}$ (other vertices in the clique B^i or a vertex from A) are broadcasting. This ensures that each b_{ij} can output X^i, and the communication task is completed. Note that this protocol is non-adaptive and deterministic.

IV. Radio Models

A. Notation

All logarithms in this paper are to the base 2. We denote random variables using capital letters and their realizations using the corresponding lower case letters. For example, a random variable X may take the value x. We often write $\Pr(\cdot \mid X = x)$, $\Pr(X = x)$, $\mathbb{E}(\cdot \mid X = x)$, *etc.*, as $\Pr(\cdot \mid x)$, $\Pr(x)$, and $\mathbb{E}(\cdot \mid x)$. The support of a random variable X, denoted $\mathrm{supp}(X)$ is the set $\{x \mid \Pr(x) > 0\}$.

Throughout this paper, we assume that the nodes of a graph with n nodes are numbered 1 to n. We also work with an alphabet set Γ that does not include a special symbol λ. We will use Γ_+ to denote the set $\Gamma \cup \{\lambda\}$.

B. The RADIO and NOISY RADIO Models

In the section, we formally define the $\mathtt{NOISY\ RADIO}_\epsilon$ model. The model RADIO is defined to be $\mathtt{NOISY\ RADIO}_0$.

a) The noise function: Let G be a graph with n nodes. For $v \in G$, define the function $\mathrm{single}^v : \Gamma_+^n \to \Gamma_+$ as follows: Let $\sigma \in \Gamma_+^n$ be a string. Let $N(v)$ be the set of neighbors of v in G. If it is not the case that for exactly one $u \in N(v)$, it holds that $\sigma_u \neq \lambda$, then we define $\mathrm{single}^v(\sigma) = \lambda$. Otherwise, if $N(v) = \{u\}$, we define $\mathrm{single}^v(\sigma) = \sigma_u$.

Finally, we define the randomized function $\text{N-single}^v(\sigma)$ to be $\mathrm{single}^v(\sigma)$ with probability $1-\epsilon$ and λ with probability ϵ. When we refer to $\text{N-single}^v(\sigma)$ for all nodes v in a set, say S, then we implicitly assume that the randomness in $\text{N-single}^v(\cdot)$ for all $v \in S$ is independent.

b) $\mathtt{NOISY\ RADIO}_\epsilon$ *protocol:* Fix $\epsilon, n \geq 0$, and let G be a graph with n nodes and let Γ be as above. A *protocol* Π in $\mathtt{NOISY\ RADIO}_\epsilon$ over G with message set Γ, is defined by a length parameter $T \in \mathbb{N}$, nT transmission functions $\{f_m^v\}_{m \in [T], v \in G}$, and n output functions $\{g^v\}_{v \in G}$. The function f_m^v has type $f_m^v : X^v \times \Gamma_+^{m-1} \to \Gamma_+$, and the function g^v has the type $g^v : X^v \times \Gamma_+^T \to Y^v$. Here, we use X^v to denote the input space of node v, Y^v to denote the output space of node v, and λ to denote the erasure symbol. We use the convention that a node that chooses to receive transmits the erasure symbol λ.

A protocol is executed as follows. At the beginning, all the nodes $v \in G$ have an input $x^v \in X^v$ and start with the empty transcript $\pi_{<1}^v$. Before round m, for $m \in [T]$, the node v has received a transcript $\pi_{<m}^v \in \Gamma_+^{m-1}$. In round m, the node v transmits the symbol $b_m^v = f_m^v(x^v, \pi_{<m}^v)$.

Let $b \in \Gamma_+^n$ be the string whose v^{th} coordinate is b_m^v. In round m, node $v \in G$ receives

$$\pi_m^v = \text{N-single}^v(b).$$

We note that the randomness used in π_m^v is fresh and independent of any randomness used anywhere else in the protocol. Observe that if $\mathrm{single}^v(b) = \lambda$ (*i.e.*, when

it is not the case that exactly one neighbor of v is broadcasting), then $\pi^v_m = \lambda$ with probability 1.

Node v appends π^v_m to the transcript $\pi^v_{<m}$ to get $\pi^v_{\leq m}$, and continues the execution of the protocol. After round T, all the nodes v output $g^v(x^v, \pi^v_{\leq T})$.

c) Non-adaptive protocols: We say that a NOISY RADIO$_\epsilon$ protocol Π is *non-adaptive* if the function $h^v_m(x^v, \pi^v_{<m}) = \mathbb{1}(f^v_m(x^v, \pi^v_{<m}) = \lambda)$ is constant for all v, m. Here, $\mathbb{1}(E)$ is the boolean indicator function for E.

C. The Lower Bound's SHARED RADIO Model

We define a noisy radio model called SHARED RADIO$_\epsilon$ and show that protocols over NOISY RADIO$_\epsilon$ can be simulated by protocols over SHARED RADIO$_\epsilon$ of the same length. While SHARED RADIO$_\epsilon$ is somewhat unrealistic, it is useful for our lower bound purposes. Note that we only define SHARED RADIO$_\epsilon$ for our graph and our communication task (subsection III-A and subsection III-B).

In the model SHARED RADIO$_\epsilon$, all the nodes have the same transcript. This transcript should be seen as the concatenation of all the n transcripts received by different nodes in NOISY RADIO$_\epsilon$. However, we also give the nodes some additional information. Firstly, if there is an A^i such that exactly one node in A^i is broadcasting, we reveal the input of this node in the transcript. Secondly, if there is an A^i such that (strictly) more than two nodes in A^i are broadcasting, then, our model behaves as if *all* the nodes in A^i are broadcasting. Finally, we also give *some* information about which subset of nodes broadcast.

In the model SHARED RADIO$_\epsilon$, the nodes in A are restricted to broadcasting only their input, while the nodes in B never broadcast.

We now formally define SHARED RADIO$_\epsilon$ and show why it can simulate protocols over NOISY RADIO$_\epsilon$ (for our graph and communication task), see Theorem IV.1. Proving a lower bound for this stronger model only makes our result stronger.

a) SHARED RADIO$_\epsilon$ protocol: Define the set $\mathsf{Cases} = \{\mathsf{many}, \mathsf{few}\} \cup \{(i, \mathsf{standard}) \mid i \in [k]\}$. We proceed to formally define the notion of a protocol in SHARED RADIO$_\epsilon$. A protocol Π has a predetermined length $T = |\Pi|$. It is defined by a collection of nT transmission functions $\{f^v_m\}_{v \in G, m \in [T]}$ and n output functions $\{g^v\}_{v \in G}$. The function f^v_m has the type $f^v_m : X^v \times \left(\Gamma^n_+\right)^{m-1} \times \mathsf{Cases}^{m-1} \to \{0, 1\}$[4], while the function g^v has the type $g^v : X^v \times \left(\Gamma^n_+\right)^T \times \mathsf{Cases}^T \to Y^v$. As in subsection IV-B, we use X^v to denote the input space of node v, and Y^v to denote the output space of node v.

The execution of this protocol proceeds as follows. At the beginning, all the nodes v have an input $x^v \in X^v$. Starting from $m = 1$, in round m, all nodes v have a transcript $t_{<m} = (\pi_{<m}, \psi_{<m}) \in \left(\Gamma^n_+\right)^{m-1} \times \mathsf{Cases}^{m-1}$. We ensure that the transcript $t_{<m}$ is the same for all the nodes v. Define the set:

$$S_m = \{v \in A \mid f^v_m(x^v, t_{<m}) = 1\}.$$

We now define $t_m = (\pi_m = \pi_{m,1} \cdots \pi_{m,n}, \psi_m) \in \Gamma^n_+ \times \mathsf{Cases}$:

- If there are at least two values of i such that $|S_m \cap A^i| > 1$, or at least three values of i such that $|S_m \cap A^i| > 0$, then we define $\psi_m = \mathsf{many}$. We also define $\pi_{m,v} = \lambda$ for all v.
- If there are at most two values of i, we have $|S_m \cap A^i| = 1$, and for all other values of i, we have $|S_m \cap A^i| = 0$, then we define $\psi_m = \mathsf{few}$. We also define $\pi_{m,v} = \lambda$ if $v \notin S_m$. Otherwise, set $\pi_{m,v} = x^v$.
- Finally, we consider the case when there exists an $i \in [k]$ such that $|S_m \cap A^i| > 1$ and $|S_m \cap \left(A \setminus A^i\right)| \leq 1$. In this case, we define $\psi_m = (i, \mathsf{standard})$. Also, for $v \in S_m \cap \left(A \setminus A^i\right)$, we set $\pi_{m,v} = x^v$. We also set, independently, for all $j \in [k]$,

$$\pi_{m,a_{ij}} = \begin{cases} x^{a_{ij}} & , \text{ with probability } 1 - \epsilon \\ \lambda & , \text{ with probability } \epsilon \end{cases}.$$

All other coordinates of π_m are set to λ.

All the players receive t_m and define $t_{\leq m}$ as $t_{<m}$ concatenated with t_m, and the execution goes on. After T rounds, node v outputs $g^v(x^v, t_{\leq T})$, finishing the protocol.

b) Lower bound for SHARED RADIO$_\epsilon$ implies a bound for NOISY RADIO$_\epsilon$: We finish this section with the following theorem, showing that every protocol for TribeTransfer$_k$ over the NOISY RADIO$_\epsilon$ model can be simulated by a protocol over the SHARED RADIO$_\epsilon$ model of the same length. Theorem IV.1 shows that in order to prove Theorem I.1, it suffices to prove an $\Omega(\log n)$ lower bound on the overhead of a noise resilient simulation in the SHARED RADIO$_\epsilon$ model.

Theorem IV.1. *Fix $k > 0$ and let G be the graph from subsection III-A with $n = 2k^2$ nodes $A \cup B$. Fix $\epsilon \geq 0$ and let Π be a protocol over NOISY RADIO$_\epsilon$. Assume*

[4]We use the convention that a node v chooses to broadcast (their input) in round m if and only if f^v_m evaluates to 1.

that when Π commences, each node $v \in A$ has an input $x^v \in \Gamma$, while nodes in B have no input.

Then, there exists a protocol Π', $|\Pi'| = |\Pi|$, in SHARED RADIO$_\epsilon$*, and a randomized function $h^v : \left(\Gamma_+^n\right)^{|\Pi|} \times \mathsf{Cases}^{|\Pi|} \to \Gamma_+^{|\Pi|}$ for all $v \in G$, such that for every input $\{x^v\}_{v\in A}$, when both Π' and Π are executed with the input x^v to node $v \in A$, it holds that*

$$\{h^v(\pi', \psi')\}_{v\in G} \equiv \{\pi^v\}_{v\in G}.$$

Here, (π', ψ') is the transcript of Π', π^v is the transcript of node v of Π, and '$\equiv$' denotes equality as joint distributions.

Proof. In this proof, we use π^G, $h^G(\pi', \psi')$, *etc.*, to denote $\{\pi^v\}_{v\in G}$, $\{h^v(\pi', \psi')\}_{v\in G}$, *etc.*[5]. The proof is by induction on $|\Pi|$. The result is straightforward for $|\Pi| = 0$. We assume that the theorem is true for $|\Pi| = T-1$ and prove it for $|\Pi| = T$. Fix a protocol Π with $|\Pi| = T$ and inputs as described by the theorem. We will construct a protocol Π' in SHARED RADIO$_\epsilon$ and a randomized function h^v for all $v \in G$ such that for every set of inputs x^A, when Π and Π' are run with input x^v to node v, then, for all $\tau^G \in \Gamma_+^{nT}$:

$$\begin{aligned}&\Pr_{\pi^G\sim\Pi(x^A)}(\pi^G = \tau^G)\\ &\quad= \Pr_{h^G,(\pi',\psi')\sim\Pi'(x^A)}(h^G(\pi', \psi') = \tau^G).\end{aligned} \tag{1}$$

Here, $\tau^G \in \Gamma_+^{nT}$ is a vector of transcripts of length T, one for each node $v \in G$.

If we use $\tau^G_{<T}$ (respectively, τ^G_T) to denote the vector of the first $T-1$ (respectively, the last) coordinate of all the transcripts in τ^G, the last equation is the same as:

$$\begin{aligned}&\Pr_{\pi^G\sim\Pi(x^A)}(\pi^G_{<T} = \tau^G_{<T} \wedge \pi^G_T = \tau^G_T)\\ &\quad= \Pr_{\substack{h^G\\(\pi',\psi')\sim\Pi'(x^A)}}(h^G_{<T}(\pi', \psi') = \tau^G_{<T} \wedge h^G_T(\pi', \psi') = \tau^G_T),\end{aligned} \tag{2}$$

where as usual $h^v_{<T}(\cdot)$ is the first $T-1$ coordinates of $h^v(\cdot)$ and $h^v_T(\cdot)$ is the last coordinate of $h^v(\cdot)$.

Using $\tau^G_{<T}$, define the events $E(\tau^G_{<T})$ and $E'(\tau^G_{<T})$ over the probability space defined by $\Pi(x^A)$ and $(\{h^v\}_{v\in G}, \Pi'(x^A))$ respectively as follows:

$$\begin{aligned}E(\tau^G_{<T}) &\equiv \pi^G_{<T} = \tau^G_{<T},\\ E'(\tau^G_{<T}) &\equiv h^G_{<T}(\pi', \psi') = \tau^G_{<T}.\end{aligned}$$

[5]As the nodes in B do not have any input in our communication task, we use x^G and x^A interchangeably to denote $\{x^v\}_{v\in G}$.

Equation 2 simplifies to:

$$\begin{aligned}&\Pr_{\pi^G\sim\Pi(x^A)}(E(\tau^G_{<T}))\Pr_{\pi^G\sim\Pi(x^A)}(\pi^G_T = \tau^G_T \mid E(\tau^G_{<T}))\\ &\quad= \Pr_{h^G,(\pi',\psi')\sim\Pi'(x^A)}(E'(\tau^G_{<T}))\\ &\quad\times \Pr_{h^G,(\pi',\psi')\sim\Pi'(x^A)}(h^G_T(\pi', \psi') = \tau^G_T \mid E'(\tau^G_{<T})).\end{aligned} \tag{3}$$

We will now construct a protocol Π' and a randomized function h^v for all $v \in G$ such that for every set of inputs x^A, when Π and Π' are run with input x^v to node v, then Equation 3 holds. To this end, let $\hat\Pi$ denote the length $T-1$ protocol consisting of the first $T-1$ rounds of Π. We apply the induction hypothesis on $\hat\Pi$ to get $\hat\Pi'$ and $\hat h^G$.

The protocol Π' behaves identically to $\hat\Pi'$ for the first $T-1$ rounds. We will define the behavior of Π' in round T later. For node $v \in G$, define the output of the randomized function $h^v_{<T}(\pi', \psi')$ to be distributed identically as $\hat h^v(\pi'_{<T}, \psi'_{<T})$. We will define the last coordinate, $h^v_T(\pi', \psi')$, later. For now, observe that when Π and Π' are run with the same inputs x^A, we have, for all $\tau^G_{<T}$,

$$\begin{aligned}\Pr_{\pi^G\sim\Pi(x^A)}(E(\tau^G_{<T})) &= \Pr_{\pi^G\sim\Pi(x^A)}(\pi^G_{<T} = \tau^G_{<T})\\ &= \Pr_{\hat\pi^G\sim\hat\Pi(x^A)}(\hat\pi^G = \tau^G_{<T}) \quad \text{(Definition of } \hat\Pi)\\ &= \Pr_{\hat h^G,(\hat\pi',\hat\psi')\sim\hat\Pi'(x^A)}(\hat h^G(\hat\pi', \hat\psi') = \tau^G_{<T})\\ &\qquad\qquad \text{(Induction Hypothesis)}\\ &= \Pr_{\hat h^G,(\pi',\psi')\sim\Pi'(x^A)}(\hat h^G(\pi'_{<T}, \psi'_{<T}) = \tau^G_{<T})\\ &\qquad\qquad \text{(Definition of } \Pi')\\ &= \Pr_{h^G,(\pi',\psi')\sim\Pi'(x^A)}(h^G_{<T}(\pi', \psi') = \tau^G_{<T})\\ &= \Pr_{h^G,(\pi',\psi')\sim\Pi'(x^A)}(E'(\tau^G_{<T})).\\ &\qquad\qquad \text{(Definition of } h^v)\end{aligned}$$

It is sufficient now to define the behavior of the protocol Π' in round T, *i.e.*, the functions $f'^v_T : X^v \times \left(\Gamma_+^n \times \mathsf{Cases}\right)^{(T-1)} \to \{0,1\}$[6] and $h^G_T(\pi', \psi')$ such that, for all τ^G, when Π and Π' are run with the same set of inputs x^A, then

$$\begin{aligned}&\Pr_{\pi^G\sim\Pi(x^A)}(\pi^G_T = \tau^G_T \mid E(\tau^G_{<T}))\\ &= \Pr_{h^G,(\pi',\psi')\sim\Pi'(x^A)}(h^G_T(\pi', \psi') = \tau^G_T \mid E'(\tau^G_{<T})).\end{aligned} \tag{4}$$

To this end, we define the function

$$\begin{aligned}&f'^v_T(x^v, (\pi'_{<T}, \psi'_{<T}))\\ &\qquad= \mathbb{1}(f^v_T(x^v, h^v_{<T}(\pi', \psi')) \neq \lambda).\end{aligned} \tag{5}$$

[6]The output functions $g'^G(\cdot)$ are not relevant to this theorem.

Here, $\mathbb{1}(E)$ denotes the indicator function for the event E, and f_T^v is the transmission function of node v in round T of Π. Recall that $h^v_{<T}(\pi', \psi') = \hat{h}^v(\pi'_{<T}, \psi'_{<T})$ and thus, the function $f'^v_T(x^v, (\pi'_{<T}, \psi'_{<T}))$ is well defined.

Instead of defining the randomized function $h^G_T(\pi', \psi')$ *en bloc*, we observe that the claim in Equation 4 for τ^G depends only on the behavior of the function $h^G_T(\pi', \psi')$ conditioned on the event $E'(\tau^G_{<T})$. We will thus fix an arbitrary τ^G, and define $h^G_T(\pi', \psi')$ conditioned on $E'(\tau^G)$ such that Equation 4 holds. Combining the definition of $h^G_T(\pi', \psi')$ conditioned on $E'(\tau^G_{<T})$ for all such τ^G will give us the definition of $h^G_T(\pi', \psi')$.

From now on, let τ_G be fixed. We have that:

- If $E(\tau^G_{<T})$ happens, then node v broadcasts in round T of $\Pi(x^A)$ if and only if $f^v_T(x^v, \tau^v_{<T}) \neq \lambda$.
- If $E'(\tau^G_{<T})$ happens, then node v broadcasts in round T of $\Pi'(x^A)$ if and only if $f'^v_T(x^v, (\pi'_{<T}, \psi'_{<T})) = 1 \iff f^v_T(x^v, \tau^v_{<T}) \neq \lambda$.

Together, we have that, the set of players broadcasting in round T of $\Pi(x^A)$ conditioned on the event $E(\tau^G_{<T})$ is the same as the set of players broadcasting in round T of $\Pi'(x^A)$ conditioned on $E'(\tau^G_{<T})$. We denote this set by $S^*(\tau^G_{<T})$. Define ψ^*_T analogously to ψ_m in subsection IV-C, with S_m replaced by $S^*(\tau^G_{<T})$. We do a case analysis based on all the values ψ^*_T can take to define $h^v_T(\pi', \psi')$.[7]

- **Case ψ^*_T = many:** We start by defining $\sigma' \in \Gamma^n_+$ to be the string indexed by node $v \in G$ such that

$$\sigma'_v = \begin{cases} \lambda & , v \in A \\ f^v_T(\tau^v_{<T}) & , v \in B \end{cases}.$$

Observe that since nodes $v \in B$ have no input, $f^v_T(\tau^v_{<T})$ is well defined.

As ψ^*_T = many in this case, we have that, conditioned on $E(\tau^G_{<T})$, either there are three different values of i such that there exists $v \in A^i$ satisfying $f^v_T(x^v, \tau^v_{<T}) \neq \lambda$ or there are at least two different values of i such that there are at least two different $v \in A^i$ satisfying $f^v_T(x^v, \tau^v_{<T}) \neq \lambda$. In either case, by the construction of our graph (Figure 1), all the nodes in B have at least two neighbors broadcasting and receive λ with probability 1. This implies that the LHS of Equation 4 is

$$\Pr_{\pi^G \sim \Pi(x^A)}(\pi^G_T = \tau^G_T \mid E(\tau^G_{<T})) \tag{6}$$

$$= \mathbb{1}(\tau^B_T = \lambda^{|B|}) \Pr_{\pi^G \sim \Pi(x^A)}(\pi^A_T = \tau^A_T \mid E(\tau^G_{<T}))$$

$$= \mathbb{1}(\tau^B_T = \lambda^{|B|}) \Pr(\text{N-single}^A(\sigma') = \tau^A_T), \tag{7}$$

where in the last step, we use the fact that all the neighbors of a node in A are in B, and for all $v \in B$, as we are conditioning on $E(\tau^G_{<T})$, we have $\sigma'_v = f^v_T(\tau^v_{<T}) = f^v_T(\pi^v_{<T})$.

We next define $h^G_T(\pi', \psi')$ conditioned on $E'(\tau^G_{<T})$ and analyze the RHS of Equation 4. If $E'(\tau^G_{<T})$ happens, for $v \in G$, define

$$h^v_T(\pi', \psi') = \begin{cases} \text{N-single}^v(\sigma') & , v \in A \\ \lambda & , v \in B \end{cases}.$$

Recall that the randomness used in $h^v_T(\pi', \psi')$ is fresh and independent for all $v \in A$. We now analyze the RHS of Equation 4 in this case.

$$\Pr_{h^G, (\pi', \psi') \sim \Pi'(x^A)}(h^G_T(\pi', \psi') = \tau^G_T \mid E'(\tau^G_{<T}))$$

$$= \mathbb{1}(\tau^B_T = \lambda^{|B|}) \tag{8}$$

$$\times \Pr_{h^G, (\pi', \psi') \sim \Pi'(x^A)}(h^A_T(\pi', \psi') = \tau^A_T \mid E'(\tau^G_{<T}))$$

$$= \mathbb{1}(\tau^B_T = \lambda^{|B|}) \Pr(\text{N-single}^A(\sigma') = \tau^A_T), \tag{9}$$

where the probability is over the noise in N-single$(\cdot)$.

Equation 8 and Equation 6 complete the proof of Equation 4 in this case.

- **Case ψ^*_T = few:** Let $\sigma' : \Gamma^n_+ \to \Gamma^n_+$ be the function whose values are strings indexed by node $v \in G$ such that, for $y^G \in \Gamma^n_+$, we have

$$\sigma'_v(y^G) = \begin{cases} \lambda & , v \in A, y^v = \lambda \\ f^v_T(y^v, \tau^v_{<T}) & , v \in A, y^v \neq \lambda \\ f^v_T(\tau^v_{<T}) & , v \in B \end{cases}.$$

As in the previous case, we first analyze the LHS of Equation 4:

$$\Pr_{\pi^G \sim \Pi(x^A)}(\pi^G_T = \tau^G_T \mid E(\tau^G_{<T})) = \Pr(\text{N-single}^G(\sigma'(x^G)) = \tau^G_T), \tag{10}$$

where since we are conditioning on $E(\tau^G_{<T})$, we have $\sigma'_v(x^G) = f^v_T(x^v, \tau^v_{<T}) = f^v_T(x^v, \pi^v_{<T})$ for $v \in A$, and $\sigma'_v(x^G) = f^v_T(\tau^v_{<T}) = f^v_T(\pi^v_{<T})$ for $v \in B$.

We next define $h^G_T(\pi', \psi')$ conditioned on $E'(\tau^G_{<T})$ and analyze the RHS of Equation 4. If $E'(\tau^G_{<T})$ happens, define

$$h^v_T(\pi', \psi') = \text{N-single}^v(\sigma'(\pi'_T)).$$

Recall that the randomness used in $h^v_T(\pi', \psi')$ is

[7] We note that since we are sampling $(\pi', \psi') \sim \Pi'(x^A)$, it is always the case that $\psi^*_T = \psi'_T$.

fresh and independent for all $v \in G$.

As $\psi_T^* = \mathsf{few}$, we have $\pi'_{T,v} = \lambda$ if $f'^v_T(x^v, (\pi'_{<T}, \psi'_{<T})) = 0 \iff f^v_T(x^v, h^v_{<T}(\pi', \psi')) = \lambda$ and $\pi'_{T,v} = x^v$ otherwise. Thus, conditioned on $E'(\tau^G_{<T})$, we have $\sigma'(\pi'_T) = \sigma'(x^G)$. We now analyze the RHS of Equation 4. We have:

$$\begin{aligned}
&\Pr_{h^G, (\pi', \psi') \sim \Pi'(x^A)} (h^G_T(\pi', \psi') = \tau^G_T \mid E'(\tau^G_{<T})) \\
&= \Pr_{\substack{h^G_{<T} \\ (\pi', \psi') \sim \Pi'(x^A)}} (\text{N-single}^G(\sigma'(\pi'_T)) = \tau^G_T \mid E'(\tau^G_{<T})) \\
&= \Pr(\text{N-single}^G(\sigma'(x^G)) = \tau^G_T).
\end{aligned} \tag{11}$$

Equation 4 follows from Equation 10 and Equation 11 in this case.

- **Case $\psi_T^* = (i, \mathsf{standard})$:** Let $\sigma' : \Gamma_+^n \to \Gamma_+^n$ be the function whose values are strings indexed by node $v \in G$ such that, for $y^G \in \Gamma_+^n$, we have

$$\sigma'_v(y^G) = \begin{cases} \lambda & , v \in A, y^v = \lambda \\ f^v_T(y^v, \tau^v_{<T}) & , v \in A, y^v \neq \lambda \\ f^v_T(\tau^v_{<T}) & , v \in B \end{cases}.$$

As in the previous case, we first analyze the LHS of Equation 4. In this case, conditioned on $E(\tau^G_{<T})$, there are at least two different $v \in A^i$ satisfying $f^v_T(x^v, \tau^v_{<T}) \neq \lambda$. Thus, by the construction of our graph (Figure 1), all the nodes in $B \setminus B^i$ have at least two neighbors broadcasting and receive λ with probability 1. This implies that the LHS of Equation 4 is

$$\begin{aligned}
&\Pr_{\pi^G \sim \Pi(x^A)} (\pi^G_T = \tau^G_T \mid E(\tau^G_{<T})) \\
&= \mathbb{1}(\tau_T^{B \setminus B^i} = \lambda^{|B \setminus B^i|}) \\
&\quad \times \Pr_{\pi^{A \cup B^i} \sim \Pi(x^A)} (\pi_T^{A \cup B^i} = \tau_T^{A \cup B^i} \mid E(\tau^G_{<T})) \\
&= \mathbb{1}(\tau_T^{B \setminus B^i} = \lambda^{|B \setminus B^i|}) \\
&\quad \times \Pr(\text{N-single}^{A \cup B^i}(\sigma'(x^G)) = \tau_T^{A \cup B^i}).
\end{aligned} \tag{12}$$

where in the last step, we use the fact that if $E(\tau^G_{<T})$ happens, then $\sigma'_v(x^G) = f^v_T(x^v, \tau^v_{<T}) = f^v_T(x^v, \pi^v_{<T})$, as in the previous case.

We next define $h^G_T(\pi', \psi')$ conditioned on $E'(\tau^G_{<T})$ and analyze the RHS of Equation 4. If $E'(\tau^G_{<T})$ happens, we define:

$$h^v_T(\pi', \psi') = \begin{cases} \text{N-single}^v(\sigma'(\pi'_T)) & , v \in A \\ \lambda & , v \in B \setminus B^i \\ \lambda & , v = b_{ij} \in B^i \\ & \quad \pi'_{T,a_{ij}} = \lambda \\ \text{single}^v(\sigma'(\pi'_T)) & , v = b_{ij} \in B^i \\ & \quad \pi'_{T,a_{ij}} \neq \lambda \end{cases}.$$

We note that the randomness used in $h^v_T(\pi', \psi')$ is fresh and independent, for all $v \in G$. As $\psi_T^* = (i, \mathsf{standard})$, we have for all $v \in A \setminus A^i$, $\pi'_{T,v} = \lambda \iff f'^v_T(x^v, (\pi'_{<T}, \psi'_{<T})) = 0 \iff f^v_T(x^v, h^v_{<T}(\pi', \psi')) = \lambda$ and $\pi'_{T,v} = x^v$ otherwise. Also, we have for all v that $\pi'_{T,v} \neq \lambda \implies \pi'_{T,v} = x^v$. Thus, conditioned on $E'(\tau^G_{<T})$, we have

$$h^v_T(\pi', \psi') = \begin{cases} \text{N-single}^v(\sigma'(x^G)) & , v \in A \\ \lambda & , v \in B \setminus B^i \\ \lambda & , v = b_{ij} \in B^i \\ & \quad \pi'_{T,a_{ij}} = \lambda \\ \text{single}^v(\sigma'(x^G)) & , v = b_{ij} \in B^i \\ & \quad \pi'_{T,a_{ij}} \neq \lambda \end{cases}.$$

We now analyze the RHS of Equation 4. We have:

$$\begin{aligned}
&\Pr_{h^G, (\pi', \psi') \sim \Pi'(x^A)} (h^G_T(\pi', \psi') = \tau^G_T \mid E'(\tau^G_{<T})) \\
&= \mathbb{1}(\tau_T^{B \setminus B^i} = \lambda^{|B \setminus B^i|}) \\
&\quad \times \Pr_{\substack{h^G \\ (\pi', \psi') \sim \Pi'(x^A)}} (h_T^{A \cup B^i}(\pi', \psi') = \tau_T^{A \cup B^i} \mid E'(\tau^G_{<T})).
\end{aligned}$$

Again, since $\psi_T^* = (i, \mathsf{standard})$, we have that $\pi'_{T,u}$, for all $u \in A^i$, is λ with probability ϵ, independently (and also independent of the randomness in h^G). Thus, we get

$$\begin{aligned}
&\Pr_{h^G, (\pi', \psi') \sim \Pi'(x^A)} (h^G_T(\pi', \psi') = \tau^G_T \mid E'(\tau^G_{<T})) \\
&= \mathbb{1}(\tau_T^{B \setminus B^i} = \lambda^{|B \setminus B^i|}) \\
&\quad \times \Pr(\text{N-single}^{A \cup B^i}(\sigma'(x^G)) = \tau_T^{A \cup B^i}).
\end{aligned} \tag{13}$$

Equation 4 follows from Equation 12 and Equation 13 in this case.

Together, the three cases above prove Equation 4 for all (τ^G, x^A).

□

V. Lower Bound: Analysis

In this section, we prove Theorem I.1. For the rest of this lower bound proof, we fix the graph G constructed in subsection III-A. For simplicity of exposition, we fix

a set Γ, set $\epsilon = 1/3$, and denote by SHARED RADIO the SHARED RADIO$_\epsilon$ model with error rate ϵ over G and message set Γ. The proof of Theorem I.1 follows directly from the following theorem in combination with Theorem IV.1.

Theorem V.1. *Fix $k \in \mathbb{N}$ such that $k > 2^{100}$, and let Π be a protocol for* TribeTransfer$_k$ *in the* SHARED RADIO *model such that $|\Pi| < \frac{k \log k}{10000}$. There exists a set $I \subseteq [k]$ such that $|I| \geq 4k/5$ and for all $i \in I$ and all $j \in [k]$, we have*

$$\Pr(b_{ij} \text{ outputs } X^i) \leq \frac{1}{4}.$$

A. Notation

A general protocol will be denoted using Π, and T will be reserved for $|\Pi|$. The notation X denotes the random vector of all the n inputs and x is one realization of X. The input for node v will be denoted using x^v (or X^v). The inputs for the nodes in A^i will be denoted using X^i.

The random variable $\mathcal{T}_m$ takes values in $\Gamma_+^n \times \mathsf{Cases}$ and denotes the transcript received by the players in round m. A particular realization of $\mathcal{T}_m = (\Pi_m, \Psi_m)$ will be denoted as $t_m = (\pi_m, \psi_m)$. We define $\mathcal{T}_{\leq m}$ to be the concatenation of the transcripts received in the first m rounds. Analogously, define $\mathcal{T}_{<m}$ to be the concatenation of the transcripts received in the first $m-1$ rounds. We slightly abuse notation and shorten $\mathcal{T}_{\leq T}$ to $\mathcal{T}$, $\Pi_{\leq T}$ to Π, *etc.*. The distinction between the random variable Π and the protocol Π should be clear from context.

a) The functions $\mathrm{succ}_i(\cdot)$, $\mathrm{ct}_i(\cdot)$, *and* $\mathrm{use}_i(\cdot)$*:* For $i \in [k]$ and a transcript $(\pi_{\leq m}, \psi_{\leq m})$, we define

$$\mathrm{succ}_i(\pi_{\leq m}) = \{v \in A^i \mid \exists l \in [m] : \pi_{l,v} \neq \lambda\},$$
$$\mathrm{succ}(\pi_{\leq m}) = \cup_{i \in [k]} \mathrm{succ}_i(\pi_{\leq m}).$$

Thus, the value of $\mathrm{succ}_i(\pi_{\leq m})$ is a subset of A^i. Intuitively, it is the subset of nodes in A^i that have "successfully" broadcast in the first m rounds. Also, define

$$\mathrm{ct}_i(\pi_{\leq m}) = |\mathrm{succ}_i(\pi_{\leq m})|,$$
$$\mathrm{ct}(\pi_{\leq m}) = \sum_{i \in [k]} \mathrm{ct}_i(\pi_{\leq m}).$$

Sometimes, we need these functions for a transcript of a single round. In this case, we have

$$\mathrm{succ}_i(\pi_m) = \{v \in A^i \mid \pi_{m,v} \neq \lambda\}.$$

The functions $\mathrm{succ}(\cdot)$, $\mathrm{ct}_i(\cdot)$, and $\mathrm{ct}(\cdot)$, for the transcript of a single round are defined similarly.

Define, for all $i \in [k]$, the set valued function

$$\mathrm{use}_i(\psi_{\leq m}) = \{m' \leq m \mid \psi_{m'} = (i, \mathsf{standard})\}.$$

Observe that, for any realization ψ of Ψ,

$$\sum_{i \in [k]} |\mathrm{use}_i(\psi)| \leq T. \tag{14}$$

B. The Proof of Theorem V.1

For the rest of the text, we fix k and a protocol Π for TribeTransfer$_k$ in SHARED RADIO.

Recall that $\mathcal{T}_{\leq m} = (\Pi_{\leq m}, \Psi_{\leq m})$ is a random variable that denotes the transcript of Π in the first m rounds. Unless stated otherwise, the probabilities and expectations below are over the players' inputs, and the randomness used by the players and by the channel. We prove Theorem V.1 in 3 steps:

Theorem V.2. *If $|\Pi| = T \leq \frac{k \log k}{10000}$, there is a set $I' \subseteq [k]$ such that $|I'| \geq 9k/10$, and for all $i \in I'$, we have*

$$\Pr\left(\mathrm{ct}_i(\Pi) \geq k - \sqrt{k}\right) \leq \frac{1}{10}.$$

Theorem V.2 shows that for all $i \in I'$, a significant number of nodes in A^i were 'erased' out during the execution of Π. One may conclude that nodes in B^i would not be able to output X^i. This is fallacious as it may be the case that information about X^i was revealed in some other ways, *e.g.*, the rounds in which the nodes in A^i chose to broadcast. In what follows, we upper bound this information leak.

Recall that the alphabet in our protocols is Γ. Define $\gamma = \log(|\Gamma|)$.

Theorem V.3. *We have*

$$\mathbb{I}(X : \mathcal{T}) \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi)] + 20\gamma T \log k.$$

Theorem V.4. *Suppose that, for some $i \in [k]$, we have $\mathbb{I}(X^i : \mathcal{T}) \leq \gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)] + f(k)$ for some function $f : \mathbb{N} \to \mathbb{R}^+$. Then,*

$$\Pr_t\left(2^{-\mathbb{H}_\infty(X^i|t)} \geq \frac{10f(k)+2}{\gamma k + 1 - \gamma\,\mathrm{ct}_i(\pi)}\right) \leq \frac{1}{10}.$$

Here, $t = (\pi, \psi)$ is sampled according to $\mathcal{T}$.

We first present the proof of Theorem V.1 assuming the three results above.

Proof of Theorem V.1. Using Lemma A.10 and linearity of expectation, we can rewrite Theorem V.3 as:

$$\sum_{i \in [k]} \mathbb{I}(X^i : \mathcal{T}) \leq \sum_{i \in [k]} \gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)] + 20\gamma T \log k. \tag{15}$$

Define the set:

$$I = \left\{ i \in [k] \mid \mathbb{I}(X^i : \mathcal{T}) \leq \gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)] + \frac{200\gamma T \log k}{k} \right\}.$$

We claim that $|I| \geq 9k/10$. Otherwise, using the fact that $\mathbb{I}(X^i : \mathcal{T}) = \gamma k - \mathbb{H}(X^i \mid \mathcal{T}) \geq \gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)]$, we have the following:

$$\begin{aligned}
&\sum_{i\in[k]} \mathbb{I}(X^i : \mathcal{T}) \\
&= \sum_{i\in I} \mathbb{I}(X^i : \mathcal{T}) + \sum_{i\in[k]\setminus I} \mathbb{I}(X^i : \mathcal{T}) \\
&\geq \sum_{i\in I} \gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)] \\
&\quad + \sum_{i\in[k]\setminus I} \left(\gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)] + \frac{200\gamma T \log k}{k}\right) \\
&> \sum_{i\in[k]} \gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)] + 20\gamma T \log k,
\end{aligned}$$

a contradiction to Equation 15.

Let I' be the set promised by Theorem V.2. Observe that $|I \cap I'| \geq 4k/5$, and for all $i \in I \cap I'$, we apply Theorem V.4 to get that

$$\Pr_t\left(2^{-\mathbb{H}_\infty(X^i|t)} \geq \frac{2000\gamma T \log k + 2k}{\gamma k^{3/2}}\right) \leq \frac{1}{5}.$$

The probability that b_{ij} outputs X_i is at most the probability of the most likely X^i given the observed transcript t, given by $2^{-\mathbb{H}_\infty(X^i|t)}$. Thus,

$$\Pr(b_{ij} \text{ outputs } X^i) \leq \mathop{\mathbb{E}}_t\left[2^{-\mathbb{H}_\infty(X^i|t)}\right] \leq \frac{1}{5} + k^{-0.1} < \frac{1}{4}.$$

□

C. Proof of Theorem V.2

We present our proof of Theorem V.2.

Proof of Theorem V.2. Recall (Π, Ψ) is a random variable that denotes the transcript of Π. We begin by defining the following events that depend on (Π, Ψ):

$$E_1^i \equiv \sum_{m\in \mathrm{use}_i(\psi)} \mathrm{ct}_i(\Pi_{\leq m}) - \mathrm{ct}_i(\Pi_{<m}) \geq k - k^{\frac{2}{3}}.$$

$$E_2^i \equiv |\mathrm{use}_i(\Psi)| \leq \frac{\log k}{25}.$$

$$E_3 \equiv \exists I \subseteq [k] : |I| > \frac{k}{200} \wedge \forall i \in I : \mathrm{ct}_i(\Pi) \geq k - \sqrt{k}.$$

$$E_4 \equiv \exists I \subseteq [k] : |I| > \frac{k}{400} \wedge \forall i \in I : E_i^1.$$

$$E_5 \equiv \exists i \in [k] : E_1^i \wedge E_2^i.$$

We will show that

$$\Pr(E_3) \overset{(a)}{\leq} \Pr(E_4) \overset{(b)}{\leq} \Pr(E_5) \overset{(c)}{\leq} \sum_{i\in[k]} \Pr(E_1^i \mid E_2^i) \overset{(d)}{\leq} \frac{1}{200}. \quad (16)$$

Before proving Equation 16, we show that $\Pr(E_3) \leq \frac{1}{200}$ indeed implies the theorem. Consider the quantity $\zeta = \sum_{i\in[k]} \Pr(\mathrm{ct}_i(\Pi) \geq k - \sqrt{k})$. Define the random variable $Y^i = \mathbb{1}(\mathrm{ct}_i(\Pi) \geq k - \sqrt{k})$. It holds that $\zeta = \sum_{i\in[k]} \mathbb{E}[Y^i] = \mathbb{E}\left[\sum_{i\in[k]} Y^i\right]$. Assuming $\Pr(E_3) \leq \frac{1}{200}$, we have:

$$\zeta \leq k \cdot \Pr(E_3) + \frac{k}{200}\Pr(\overline{E_3}) \leq \frac{k}{100}.$$

This implies the theorem as there can be at most $\frac{k}{10}$ values of i such that $\Pr(\mathrm{ct}_i(\pi) \geq k - \sqrt{k}) \geq \frac{1}{10}$.

We now prove Equation 16.

To show inequality (a), we assume $E_3 \wedge \overline{E_4}$ and derive a contradiction. Consider a realization (π, ψ) of (Π, Ψ) such that $E_3 \wedge \overline{E_4}$ holds. Then, if $(\Pi, \Psi) = (\pi, \psi)$, we have a set $I \subseteq [k] : |I| > \frac{k}{400}$ such that for all $i \in I$,

$$\mathrm{ct}_i(\pi) - \left(\sum_{m\in \mathrm{use}_i(\psi)} \mathrm{ct}_i(\pi_{\leq m}) - \mathrm{ct}_i(\pi_{<m})\right) \geq k^{\frac{2}{3}} - \sqrt{k}.$$

This implies $\sum_{i\in I}\sum_{m\notin \mathrm{use}_i(\psi)} \mathrm{ct}_i(\pi_{\leq m}) - \mathrm{ct}_i(\pi_{<m}) \geq \frac{k}{400}(k^{\frac{2}{3}} - \sqrt{k}) \geq k^{\frac{8}{5}}$. Since the summand $\mathrm{ct}_i(\pi_{\leq m}) - \mathrm{ct}_i(\pi_{<m})$ is always positive, we can take the sum over all $i \in [k]$ instead of $i \in I$. With this change, we can rearrange to

$$\sum_{m\in[T]} \sum_{i:m\notin \mathrm{use}_i(\psi)} \mathrm{ct}_i(\pi_{\leq m}) - \mathrm{ct}_i(\pi_{<m}) \geq k^{\frac{8}{5}}.$$

Let I_m be the set of all i's such that $m \notin \mathrm{use}_i(\psi)$. Observe that at most 2 positions of π_m that are in A^i, for some $i \in I_m$, are different from λ (this can be seen by going over the possible values for ψ_m). Thus, we get

$$\sum_{m\in[T]} \sum_{i:m\notin \mathrm{use}_i(\psi)} \mathrm{ct}_i(\pi_{\leq m}) - \mathrm{ct}_i(\pi_{<m}) \leq 2T < k^{\frac{8}{5}},$$

a contradiction.

To show inequality (b), we show that $E_4 \implies E_5$. Consider a realization (π, ψ) of (Π, Ψ) such that E_4 happens. Then, there are more than $\frac{k}{400}$ values of i such that E_1^i happens. Since the sets $\mathrm{use}_i(\psi) \subseteq [T]$ are disjoint for all i and $T \leq \frac{k \log k}{10000}$, there is at least one value of i satisfying E_1^i, such that E_2^i also happens, thus E_5 happens.

Inequality (c) follows from a simple union bound and the fact that $\Pr(A \wedge B) \leq \Pr(A \mid B)$ for any two events A, B.

Finally, we show inequality (d) by showing that for any i and any ψ such that E_2^i happens when $\Psi = \psi$, we have $\Pr(E_1^i \mid \psi) \leq \frac{1}{200k}$. Fix any such ψ. Then, E_1^i can happen only if there is a set $J \subseteq [k], |J| \geq k - k^{\frac{2}{3}}$ such that for all $j \in J$, there exists $m \in \mathrm{use}_i(\psi)$ such that $\pi_{m,a_{ij}} \neq \lambda$. Since for all $j \in [k]$ and $m \in \mathrm{use}_i(\psi)$, we have $\pi_{m,a_{ij}} \neq \lambda$ independently with probability $\frac{2}{3}$ and we have $|\mathrm{use}_i(\psi)| \leq \frac{\log k}{25}$, there exists

$m \in \mathrm{use}_i(\psi)$ such that $\pi_{m,a_{ij}} \neq \lambda$ with probability $1-(1-\frac{2}{3})^{|\mathrm{use}_i(\psi)|} \leq 1-k^{-\frac{1}{10}}$. We union bound over all possible values of J to get

$$\Pr(E_1^i \mid \psi) \leq \underbrace{k^{k^{\frac{2}{3}}}}_{\text{Values of } J} \cdot \left(1-k^{-\frac{1}{10}}\right)^{k-k^{\frac{2}{3}}} \leq k^{k^{\frac{2}{3}}} \cdot \mathrm{e}^{-k^{-\frac{1}{10}}(k-k^{\frac{2}{3}})} \ll \frac{1}{200k},$$

as desired. □

D. Proof of Theorem V.3

The following theorem is the main ingredient in the proof of Theorem V.3. Recall that $\mathcal{T} = (\Pi, \Psi)$.

Theorem V.5. *It holds that*

$$\mathbb{I}(X : \mathcal{T} \mid \Psi) \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi)] + 15\gamma T \log k.$$

Proof of Theorem V.3 assuming Theorem V.5. For any round $m \in [T]$, Ψ_m takes at most k^5 different values. Thus, we have

$$\mathbb{H}(\Psi) \leq 5T \log k.$$

We have

$$\begin{aligned}\mathbb{I}(X : \mathcal{T}) &\leq \mathbb{I}(X : \mathcal{T} \mid \Psi) + \mathbb{H}(\Psi) \quad \text{(by Lemma A.11)}\\ &\leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi)] + 15\gamma T \log k + 5T \log k \quad \text{(by Theorem V.5)}\\ &< \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi)] + 20\gamma T \log k.\end{aligned}$$

□

It remains to prove Theorem V.5. Theorem V.5 follows from Lemma V.7 below (by setting $m = T$) that shows that if one cannot signal using the types of the rounds, then the amount of new information that the transcript of a single round can give about the inputs is (roughly) at most the number of new input symbols revealed in this round.

Lemma V.6. *Let $m \in [T]$. In the event that $\Psi_m = (i, \mathsf{standard})$ for some i, it holds that*

$$\begin{aligned}&\mathbb{I}(X : \mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m})\\ &\quad \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{\leq m}) - \mathrm{ct}(\Pi_{<m})] + 2\log k + \gamma.\end{aligned}$$

Proof. It holds that

$$\begin{aligned}&\mathbb{I}(X : \mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m}) = \mathbb{I}(X : \Pi_m \mid \mathcal{T}_{<m}\Psi_{\leq m})\\ &= \mathbb{I}(X : \Pi_m, \mathrm{succ}(\Pi_m) \mid \mathcal{T}_{<m}\Psi_{\leq m})\\ &\leq \mathbb{I}(X : \mathrm{succ}(\Pi_m) \mid \mathcal{T}_{<m}\Psi_{\leq m})\\ &\quad + \mathbb{H}(\Pi_m \mid \mathrm{succ}(\Pi_m)\mathcal{T}_{<m}\Psi_{\leq m}).\\ &\qquad\qquad \text{(by Fact A.8, Fact A.9)}\end{aligned}$$

Observe that, in the event that $\Psi_m = (i, \mathsf{standard})$, we have $\mathbb{I}(X : \mathrm{succ}(\Pi_m) \mid \mathcal{T}_{<m}\Psi_{\leq m}) \leq \gamma + 2\log k$. This is because $\mathrm{succ}_i(\Pi_m)$ is determined by noise independently from the input X, and the rest of $\mathrm{succ}(\Pi_m)$ has entropy at most $\gamma + 2\log k$, as at most one $\pi_{m,v}$ may be different from λ for $v \notin A^i$.

We next claim that

$$\begin{aligned}&\mathbb{H}(\Pi_m \mid \mathrm{succ}(\Pi_m)\mathcal{T}_{<m}\Psi_{\leq m})\\ &\quad = \mathbb{H}(\Pi_m \mid \mathrm{succ}(\Pi_m)\,\mathrm{ct}(\Pi_{\leq m})\mathcal{T}_{<m}\Psi_{\leq m})\\ &\quad \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{\leq m}) - \mathrm{ct}(\Pi_{<m})],\end{aligned}$$

and the assertion follows. The first step is because $\mathrm{ct}(\Pi_{\leq m})$ is determined by $\mathrm{succ}(\Pi_m)\mathcal{T}_{<m}\Psi_{\leq m}$. The second step is because, given any realization $(S, c, \pi_{<m}, \psi_{\leq m})$ of $\mathrm{succ}(\Pi_m)\,\mathrm{ct}(\Pi_{\leq m})\mathcal{T}_{<m}\Psi_{\leq m}$, the value of Π_m is determined by its value in coordinates in the subset S (as the rest of the coordinates are λ). Even amongst these coordinates, some are already determined by $\pi_{<m}$. The number of coordinates that are not determined is exactly $c - \mathrm{ct}(\pi_{<m})$. □

Lemma V.7. *Let $m \in [T]$. It holds that*

$$\mathbb{I}(X : \mathcal{T}_{\leq m} \mid \Psi_{\leq m}) \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{\leq m})] + 15\gamma m \log k.$$

Proof. Proof by induction on m. The statement is trivially true for $m = 0$. We assume it holds for $m-1$ and prove it for m. We have

$$\begin{aligned}&\mathbb{I}(X : \mathcal{T}_{\leq m} \mid \Psi_{\leq m}) \qquad (17)\\ &\quad = \mathbb{I}(X : \mathcal{T}_{<m} \mid \Psi_{\leq m}) + \mathbb{I}(X : \mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m})\\ &\qquad\qquad \text{(by Fact A.9)}\\ &\quad \leq \mathbb{I}(X : \mathcal{T}_{<m} \mid \Psi_{<m}) \qquad (18)\\ &\qquad + \mathbb{I}(X : \mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m}) + \mathbb{H}(\Psi_m)\\ &\qquad\qquad \text{(by Lemma A.11)}\\ &\quad \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{<m})] + 15\gamma(m-1)\log k \qquad (19)\\ &\qquad + \mathbb{I}(X : \mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m}) + 5\log k.\\ &\qquad\qquad \text{(by the induction hypothesis)}\end{aligned}$$

We now proceed by a case analysis:

- $\psi_m = (i, \mathsf{standard})$**:** In this case, we use Lemma V.6 to get

$$\begin{aligned}&\mathbb{I}(X : \mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m}) \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{\leq m}) - \mathrm{ct}(\Pi_{<m})]\\ &\qquad\qquad + 2\log k + \gamma.\end{aligned}$$

  Plugging it in Equation 17 gives

$$\begin{aligned}&\mathbb{I}(X : \mathcal{T}_{\leq m} \mid \Psi_{\leq m})\\ &\quad \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{<m})] + 15\gamma m \log k\\ &\qquad + \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{\leq m}) - \mathrm{ct}(\Pi_{<m})]\\ &\quad \leq \gamma \cdot \mathbb{E}[\mathrm{ct}(\Pi_{\leq m})] + 15\gamma m \log k,\end{aligned}$$

  and the assertion follows.

- **Otherwise:** In these cases, because at most two coordinates of Π_m may be different from λ,

$$\mathbb{I}(X : \mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m}) \leq \mathbb{H}(\mathcal{T}_m \mid \mathcal{T}_{<m}\Psi_{\leq m}) < 2\gamma + 4\log k.$$

Here, $4\log k$ comes from the number of ways of choosing the (at most) two coordinates that are different from λ and 2γ comes from the number of possible values of these coordinates. Plugging it in Equation 17 gives the assertion, as $\mathrm{ct}(\Pi_{<m}) \leq \mathrm{ct}(\Pi_{\leq m})$.

□

E. Proof of Theorem V.4

We will need the following lemma connecting $\mathbb{H}(X \mid t)$ and $\mathbb{H}_\infty(X \mid t)$. The argument in this lemma is similar to that in Lemma 2.6 in [BEGH16] (where it was said to be a special case of Fano's inequality). See [KR13] also.

Lemma V.8. *Let $(X, \mathcal{T})$ be a pair of discrete random variables. Let t be a realization of $\mathcal{T}$. Define the set $\Omega_t = \{x \mid \Pr(x, t) \neq 0\}$. If $|\Omega_t| > 0$, it holds that*

$$2^{-\mathbb{H}_\infty(X|t)} \leq 1 - \frac{\mathbb{H}(X \mid t) - 1}{\log(|\Omega_t|) + 1}.$$

Proof. Assume that $|\Omega_t| > 1$ as otherwise the claim is trivial. Let x^* maximize $\Pr(x, t)$ and define $p^* = \Pr(x^* \mid t)$. We have

$$\begin{aligned}
&\mathbb{H}(X \mid t) \\
&= -\sum_{x \in \Omega_t} \Pr(x \mid t) \log(\Pr(x \mid t)) \\
&= -p^* \log(p^*) - \sum_{x \in \Omega_t \setminus \{x^*\}} \Pr(x \mid t) \log(\Pr(x \mid t)) \\
&= -p^* \log(p^*) - (1 - p^*) \log(1 - p^*) \\
&\quad - \sum_{x \in \Omega_t \setminus \{x^*\}} \Pr(x \mid t) \log\left(\frac{\Pr(x \mid t)}{1 - p^*}\right) \\
&= h(p^*) + (1 - p^*)\, \mathbb{H}(X \mid \mathcal{T} = t, X \neq x^*) \\
&\leq h(p^*) + (1 - p^*) \log(|\Omega_t|) \\
&\leq 1 + (1 - p^*) \log(|\Omega_t|),
\end{aligned}$$

where $h(x) = -x \log x - (1 - x) \log(1 - x)$ is defined on $(0, 1)$. The proof is done as $p^* = 2^{-\mathbb{H}_\infty(X|t)}$ by definition. □

Proof of Theorem V.4. Recall that

$$\mathbb{I}(X^i : \mathcal{T}) = \mathbb{H}(X^i) - \mathbb{H}(X^i \mid \mathcal{T}) = \gamma k - \mathop{\mathbb{E}}_t[\mathbb{H}(X^i \mid t)].$$

Let Z be the random variable that whenever $\mathcal{T} = t = (\pi, \psi)$ receives the value $\gamma k - \mathbb{H}(X^i \mid t) - \gamma\, \mathrm{ct}_i(\pi)$. By our assumption in the theorem statement

$$\begin{aligned}
\mathbb{E}[Z] &= \gamma k - \mathop{\mathbb{E}}_t[\mathbb{H}(X^i \mid t)] - \gamma \cdot \mathop{\mathbb{E}}_\pi[\mathrm{ct}_i(\pi)] \\
&= \mathbb{I}(X^i : \mathcal{T}) - \gamma \cdot \mathbb{E}[\mathrm{ct}_i(\Pi)] \leq f(k).
\end{aligned}$$

Let t be a realization of $\mathcal{T}$. Denote

$$\Omega_t = \{x \in \mathrm{supp}(X^i) \mid \Pr(x, t) \neq 0\}.$$

Note that

$$\log(|\Omega_t|) \leq \gamma k - \gamma\, \mathrm{ct}_i(\pi). \tag{20}$$

In addition,

$$\mathbb{H}(X^i \mid t) \leq \log(|\Omega_t|) = \gamma k - \gamma\, \mathrm{ct}_i(\pi),$$

implying that $Z \geq 0$.

Since Z is a non-negative random variable whose expectation is at most $f(k)$, by Markov's inequality, we have

$$\Pr(Z \geq 10 f(k)) \leq \frac{1}{10}.$$

Let t be a realization of $\mathcal{T}$, such that $Z < 10 f(k)$. It holds that

$$\begin{aligned}
2^{-\mathbb{H}_\infty(X^i|t)} &\leq 1 - \frac{\mathbb{H}(X^i \mid t) - 1}{\log(|\Omega_t|) + 1} \quad \text{(by Lemma V.8)} \\
&\leq 1 - \frac{\gamma k - \gamma\, \mathrm{ct}_i(\pi) - 10 f(k) - 1}{\gamma k + 1 - \gamma\, \mathrm{ct}_i(\pi)} \\
&\qquad \text{(by Equation 20 and } Z < 10 f(k)) \\
&\leq \frac{10 f(k) + 2}{\gamma k + 1 - \gamma\, \mathrm{ct}_i(\pi)},
\end{aligned}$$

and the assertion follows. □

References

[AGS16] Shweta Agrawal, Ran Gelles, and Amit Sahai. Adaptive protocols for interactive communication. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 595–599. IEEE, 2016. 3, 4

[BE14] Mark Braverman and Klim Efremenko. List and unique coding for interactive communication in the presence of adversarial noise. In *Foundations of Computer Science (FOCS)*, pages 236–245, 2014. 3

[BEGH16] Mark Braverman, Klim Efremenko, Ran Gelles, and Bernhard Haeupler. Constant-rate coding for multiparty interactive communication is impossible. In *Symposium on Theory of Computing (STOC)*, pages 999–1010. ACM, 2016. 3, 5, 16

[BGMO17] Mark Braverman, Ran Gelles, Jieming Mao, and Rafail Ostrovsky. Coding for interactive communication correcting insertions and deletions. *IEEE Transactions on Information Theory*, 63(10):6256–6270, 2017. 3

[BKN14] Zvika Brakerski, Yael Tauman Kalai, and Moni Naor. Fast interactive coding against adversarial noise. *Journal of the ACM (JACM)*, 61(6):35, 2014. 3

[BR11] Mark Braverman and Anup Rao. Towards coding for maximum errors in interactive communication. In

Symposium on Theory of computing (STOC), pages 159–166. ACM, 2011. 3

[Bra12] Mark Braverman. Towards deterministic tree code constructions. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 161–167. ACM, 2012. 3

[CGK+14] Keren Censor-Hillel, Seth Gilbert, Fabian Kuhn, Nancy A. Lynch, and Calvin C. Newport. Structuring unreliable radio networks. *Distributed Computing*, 27(1):1–19, 2014. 4

[CHHZ17] Keren Censor-Hillel, Bernhard Haeupler, D. Ellis Hershkowitz, and Goran Zuzic. Broadcasting in noisy radio networks. In *Symposium on Principles of Distributed Computing (PODC)*, pages 33–42, 2017. 4

[CHHZ18] Keren Censor-Hillel, Bernhard Haeupler, D. Ellis Hershkowitz, and Goran Zuzic. Erasure correction for noisy radio networks. *CoRR*, abs/1805.04165, 2018. 1, 4

[CK85] Imrich Chlamtac and Shay Kutten. On broadcasting in radio networks-problem analysis and protocol design. *IEEE Trans. Communications*, 33(12):1240–1246, 1985. 1

[EGH16] Klim Efremenko, Ran Gelles, and Bernhard Haeupler. Maximal noise in interactive communication over erasure channels and channels with feedback. *IEEE Transactions on Information Theory*, 62(8):4575–4588, 2016. 3

[EKS18] Klim Efremenko, Gillat Kol, and Raghuvansh Saxena. Interactive coding over the noisy broadcast channel. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 507–520. ACM, 2018. 1, 3, 4, 5, 19

[FK00] Uriel Feige and Joe Kilian. Finding OR in a noisy broadcast network. *Inf. Process. Lett.*, 73(1-2):69–75, 2000. 4

[Gal88] Robert G. Gallager. Finding parity in a simple broadcast network. *IEEE Transactions on Information Theory*, 34(2):176–180, 1988. 1, 4

[Gam87] Abbas El Gamal. Open problems presented at the 1984 workshop on specific problems in communication and computation sponsored by bell communication research. *Appeared in "Open Problems in Communication and Computation", by Thomas M. Cover and B. Gopinath (editors). Springer-Verlag*, 1987. 1, 4

[GH14] Ran Gelles and Bernhard Haeupler. Capacity of interactive communication over erasure channels and channels with feedback. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1296–1311. SIAM, 2014. 3

[GHK+16] Ran Gelles, Bernhard Haeupler, Gillat Kol, Noga Ron-Zewi, and Avi Wigderson. Towards optimal deterministic coding for interactive communication. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1922–1936. Society for Industrial and Applied Mathematics, 2016. 3

[GHLN12] Mohsen Ghaffari, Bernhard Haeupler, Nancy A. Lynch, and Calvin C. Newport. Bounds on contention management in radio networks. In *International Symposium on Distributed Computing (DISC)*, pages 223–237, 2012. 4

[GHM18] Ofer Grossman, Bernhard Haeupler, and Sidhanth Mohanty. Algorithms for noisy broadcast with erasures. In *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018. 4

[GHS14] Mohsen Ghaffari, Bernhard Haeupler, and Madhu Sudan. Optimal error rates for interactive coding i: Adaptivity and other settings. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 794–803. ACM, 2014. 2, 3

[GK17] Ran Gelles and Yael Tauman Kalai. Constant-rate interactive coding is impossible, even in constant-degree networks. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, pages 21:1–21:13, 2017. 3

[GKS08] Navin Goyal, Guy Kindler, and Michael Saks. Lower bounds for the noisy broadcast problem. *SIAM Journal on Computing*, 37(6):1806–1841, 2008. 1, 3, 4, 5

[GLN13] Mohsen Ghaffari, Nancy A. Lynch, and Calvin C. Newport. The cost of radio network broadcast for different models of unreliable links. In *Principles of Distributed Computing (PODC)*, pages 345–354, 2013. 4

[GMS11] Ran Gelles, Ankur Moitra, and Amit Sahai. Efficient and explicit coding for interactive communication. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 768–777. IEEE, 2011. 3

[GMS14] Ran Gelles, Ankur Moitra, and Amit Sahai. Efficient coding for interactive communication. *IEEE Transactions on Information Theory*, 60(3):1899–1913, 2014. 3

[Hae14] Bernhard Haeupler. Interactive channel capacity revisited. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 226–235. IEEE, 2014. 3

[KKP01] Evangelos Kranakis, Danny Krizanc, and Andrzej Pelc. Fault-tolerant broadcasting in radio networks. *Journal of Algorithms*, 39(1):47–67, 2001. 4

[KLN+10] Fabian Kuhn, Nancy A. Lynch, Calvin C. Newport, Rotem Oshman, and Andréa W. Richa. Broadcasting in unreliable radio networks. In *Symposium on Principles of Distributed Computing (PODC)*, pages 336–345, 2010. 4

[KM05] Eyal Kushilevitz and Yishay Mansour. Computation in noisy radio networks. *SIAM J. Discrete Math.*, 19(1):96–108, 2005. 4

[KPP10] Evangelos Kranakis, Michel Paquette, and Andrzej Pelc. Communication in random geometric radio networks with positively correlated random faults. *Ad Hoc & Sensor Wireless Networks*, 9(1-2):23–52, 2010. 4

[KR13] Gillat Kol and Ran Raz. Interactive channel capacity. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 715–724. ACM, 2013. 3, 16

[MS14] Cristopher Moore and Leonard J Schulman. Tree codes and a conjecture on exponential sums. In *Innovations in theoretical computer science (ITCS)*, pages 145–154. ACM, 2014. 3

[New04] Ilan Newman. Computing in fault tolerance broadcast networks. In *Computational Complexity Conference (CCC)*, pages 113–122, 2004. 4

[Pel07] David Peleg. Time-efficient broadcasting in radio networks: A review. In *Distributed Computing and Internet Technology ICDCIT*, pages 1–18, 2007. 2

[RS94] Sridhar Rajagopalan and Leonard J. Schulman. A coding theorem for distributed computation. In *Symposium on the Theory of Computing (STOC)*, pages 790–799, 1994. 3

[Sch92] Leonard J Schulman. Communication on noisy channels: A coding theorem for computation. In *Foundations of Computer Science (FOCS)*, pages 724–733. IEEE, 1992. 3, 19

[Sch93] Leonard J Schulman. Deterministic coding for interactive

communication. In *Symposium on Theory of computing (STOC)*, pages 747–756. ACM, 1993. 3

[Sch96] Leonard J. Schulman. Coding for interactive communication. *IEEE Transactions on Information Theory*, 42(6):1745–1756, 1996. 3

APPENDIX A
INFORMATION THEORY PRELIMINARIES

A. Entropy

Definition A.1 (Entropy)**.** *The (binary) entropy of a discrete random variable X is defined as*

$$\mathbb{H}(X) = \sum_{x \in \mathrm{supp}(X)} \Pr(x) \log \frac{1}{\Pr(x)} = \mathop{\mathbb{E}}_{x \sim X}\left[\log \frac{1}{\Pr(x)}\right].$$

Definition A.2 (Conditional Entropy)**.** *The entropy of a discrete random variable X given another random variable Y is defined as*

$$\mathbb{H}(X \mid Y) = \mathop{\mathbb{E}}_{y \sim Y}\left[\mathbb{H}(X \mid Y = y)\right].$$

Fact A.3. *We have $\mathbb{H}(XY) = \mathbb{H}(X) + \mathbb{H}(Y \mid X) \leq \mathbb{H}(X) + \mathbb{H}(Y)$. Equality holds if X and Y are independent.*

Fact A.4. *If the random variable X takes values in the set Ω, it holds that*

$$0 \leq \mathbb{H}(X) \leq \log|\Omega|.$$

B. Min-Entropy

Definition A.5 (Min-Entropy)**.** *The min-entropy of a discrete random variable X is*

$$\mathbb{H}_\infty(X) = \min_{x:\Pr(x)\neq 0} \log \frac{1}{\Pr(x)}.$$

Fact A.6. *If the random variable X takes values in the set Ω, it holds that*

$$0 \leq \mathbb{H}_\infty(X) \leq \mathbb{H}(X) \leq \log|\Omega|.$$

C. Mutual Information

Definition A.7 (Mutual Information)**.** *The mutual information between two discrete random variables X and Y is defined as*

$$\mathbb{I}(X : Y) = \mathbb{H}(X) - \mathbb{H}(X \mid Y) = \mathbb{H}(Y) - \mathbb{H}(Y \mid X).$$

We can also define conditional mutual information as

$$\begin{aligned}\mathbb{I}(X : Y \mid Z) &= \mathbb{H}(X \mid Z) - \mathbb{H}(X \mid YZ) \\ &= \mathbb{H}(Y \mid Z) - \mathbb{H}(Y \mid XZ).\end{aligned}$$

Fact A.8. *We have $0 \leq \mathbb{I}(X : Y \mid Z) \leq \mathbb{H}(X)$.*

Fact A.9 (Chain Rule)**.** *If A, B, C, D are random variables, then*

$$\mathbb{I}(AB : C \mid D) = \mathbb{I}(A : C \mid D) + \mathbb{I}(B : C \mid AD).$$

Lemma A.10 (Super-additivity of information)**.** *Let $n > 0$ and $X_1, X_2, \cdots, X_n$ be mutually independent random variables. Then, for any random variable Y,*

$$\sum_{i \in [n]} \mathbb{I}(X_i : Y) \leq \mathbb{I}(X_1 X_2 \cdots X_n : Y).$$

Proof. We have

$$\begin{aligned}&\mathbb{I}(X_1 X_2 \cdots X_n : Y) \\ &= \sum_{i \in [n]} \mathbb{I}(X_i : Y \mid X_1 \cdots X_{i-1}) && \text{((Chain rule))} \\ &= \sum_{i \in [n]} \mathbb{H}(X_i \mid X_1 \cdots X_{i-1}) - \mathbb{H}(X_i \mid Y X_1 \cdots X_{i-1}) \\ &= \sum_{i \in [n]} \mathbb{H}(X_i) - \mathbb{H}(X_i \mid Y X_1 \cdots X_{i-1}) \\ & && \text{((Independence of } X_i\text{))} \\ &\geq \sum_{i \in [n]} \mathbb{H}(X_i) - \mathbb{H}(X_i \mid Y) \\ &= \sum_{i \in [n]} \mathbb{I}(X_i : Y).\end{aligned}$$

□

Lemma A.11. *Let X, Y, Z be random variables. We have:*

$$|\mathbb{I}(X : Y) - \mathbb{I}(X : Y \mid Z)| \leq \mathbb{H}(Z).$$

Proof. We prove both directions separately. Firstly,

$$\begin{aligned}&\mathbb{I}(X : Y \mid Z) + \mathbb{H}(Z) \\ &\geq \mathbb{I}(X : Y \mid Z) + \mathbb{I}(X : Z) \\ &= \mathbb{I}(X : YZ) = \mathbb{I}(X : Y) + \mathbb{I}(X : Z \mid Y) \\ &\geq \mathbb{I}(X : Y).\end{aligned}$$

For the other direction, note that:

$$\begin{aligned}&\mathbb{I}(X : Y) + \mathbb{H}(Z) \\ &\geq \mathbb{I}(X : Y) + \mathbb{I}(X : Z \mid Y) \\ &= \mathbb{I}(X : YZ) = \mathbb{I}(X : Z) + \mathbb{I}(X : Y \mid Z) \\ &\geq \mathbb{I}(X : Y \mid Z).\end{aligned}$$

□

APPENDIX B
CODING WITH LOGARITHMIC OVERHEAD

In this section we prove Theorem I.2. That is, we describe an $\mathcal{O}(\log n)$ overhead scheme to make any non-adaptive noiseless protocol noise resilient. We can restrict attention to deterministic protocols as randomized protocols are distributions over deterministic ones.

For the rest of the text, fix a graph G. Fix a deterministic, non-adaptive protocol Π that works over

the noiseless RADIO model over G. Fix inputs x^v for every $v \in G$. Let T be the number of rounds in Π. Recall from subsection IV-B, that the protocol Π is defined by transmission functions, $f_m^v : X^v \times \Gamma_+^{m-1} \to \Gamma_+$, and output functions, $g^v : X^v \times \Gamma_+^T \to Y^v$, for every round $m \in [T]$ and node $v \in G$. We next show how to transform Π to a noise resilient protocol.

Observe that if $T < n^{100}$, the protocol Π can be simulated by repeating each broadcast $c_0 \cdot \log n$ times, for a sufficiently large constant c_0. The reason is that the repetitions allow each node to retrieve the value broadcast in a given round with probability greater than $1 - n^{-1000}$. We can apply the union bound over all rounds and nodes and claim that the probability of the noise effecting the output is negligible. Thus, we assume $T > n^{100}$ throughout.

a) Consistent transcripts: Since we assume that the protocol Π is non-adaptive, the following set is well defined, and known to all nodes in advance:

$$S_{u\to v} = \{m \in [T] \mid u \text{ is the unique node in } N(v) \text{ that is broadcasting in round } m\}.$$

For the rest of the text, the notation $\pi[S]$ means the string obtained from π by deleting the all the coordinates not in S. We also use the notation $s\|s'$ to denote the concatenation of two strings, s and s'.

Consider a set of transcripts $\{\pi^v \in \Gamma_+^*\}_{v\in G}$, one for each node $v \in G$. For a vertex $u \in G$, define the *transmitted transcript* $\rho^u \in \Gamma_+^*$ of u with respect to π^u to be the string whose t^{th} symbol is $f_t^u(x^u, \pi^u_{<t})$, for all $t \in [|\pi^u|]$. Let $m \in \{0, \cdots, T\}$. We say that the a set $\{\pi_v\}$ is m-consistent if $|\pi^v| = m$ for every $v \in G$ and

$$\forall v \in G\ \forall u \neq v \in G : \rho^u[S_{u\to v}] = \pi^v[S_{u\to v}]. \quad (21)$$

We define m-consistent for $m > T$ to be the same as T-consistent. Given a T-consistent transcript, one can recover the noiseless transcripts for every v. Thus, it suffices for our simulation protocol to generate a T-consistent transcripts.

A. The Simulation Protocol Π'

At a high level, our simulation protocol Π' simulates the noiseless protocol Π round by round, repeating each round $c_0 \log n$ times. The constant c_0 will be fixed in the analysis. Periodically, our protocol Π' has check stages to verify if any errors were introduced, and re-simulate if that is the case.

In order to control the overhead of our simulation, we break the noiseless protocol Π into *chunks* of n^5 rounds each. A *level* 0 *simulation* in our protocol simulates one chunk in Π followed by a 'check' stage. For $l \geq 0$, a level l simulation has two level $l-1$ simulations followed by a check stage. Finally, the protocol Π' has 2^ℓ successive executions of level ℓ simulations. Here, ℓ is the unique integer such that $10T \leq 2^{2\ell} n^5 < 40T$. Similar recursive structures can be found, for example, in [Sch92], [EKS18].

Let $m \in [T]$. When we say a node v simulates a *virtual round* m of Π with transcript π^v, we mean that the node v broadcasts $f_m^v(x^v, \pi^v)$, where x^v is the input of v. As usual, if $f_m^v(x^v, \pi^v) = \lambda$, the node v receives. We also ensure that $|\pi^v| = m-1$. If $m > T$, simulating a virtual round means that v receives.

Our protocol Π' for simulating Π is described in Algorithm 3. A level l simulation, for $l \geq 1$, is described in Algorithm 2. It simulates (roughly) the first $n^5 2^l$ rounds of Π. Algorithm 1 has a level 0 simulation. It simulates the first n^5 rounds of Π. The protocol in the check stages, $\texttt{CHECK}_k$, is described in Algorithm 4. This protocol assumes that each node v has a pair of transcripts π_1^v and π_2^v. At the end of an execution each node v outputs an integer m^v. The protocol has the property that if $\{\pi_1^v\}$ is m_1-consistent, then all the outputs m^v are the same value m_2, and m_2 is the maximum such that $\{\pi_1^v \| \pi_2^v\}$ is $(m_1 + m_2)$-consistent. $\texttt{CHECK}_k$ uses the *'rumor spreading'* protocol described in the following proposition:

Proposition B.1. *Let H be a connected graph with n nodes, and assume that each node $v \in H$ has an input $m^v \in [t]$. There exists a deterministic, non-adaptive, noiseless broadcast protocol $\texttt{RS}_H$ with $\mathcal{O}(n^2 \log t)$ rounds such that when the protocol ends all the nodes output $\min_{v\in H} m^v$.*

The above proposition can be proved by having the nodes broadcast their input over a fixed spanning tree.

a) Simulation length: The for loop of $\texttt{CHECK}_k$ is has $\mathcal{O}(n^2)$ iterations and each iteration requires $\mathcal{O}(k \log n)$ rounds (we use the assumption that the transcript π_2^v has at most $2^{k-1}n^5$ rounds for all v). By Proposition B.1, the execution of $\texttt{RS}_G$ uses another $\mathcal{O}(n^2 k \log(n))$ rounds, as it is run with inputs in $[2^{k-1}n^5]$. Since every message is repeated $100k \log n$ times, $\texttt{CHECK}_k$ has $n^2 \operatorname{poly}(k) \operatorname{polylog}(n)$ rounds. Thus, assuming that $T > n^{100}$, the total number of rounds in

Π' is at most

$$2^\ell \left(\underbrace{2^\ell n^5 \cdot c_0 \log(n)}_{\text{Simulation rounds}} + \cdot \underbrace{\sum_{j=0}^{\ell} \frac{2^\ell n^5}{2^j n^5} \cdot n^2 \operatorname{poly}(j+1) \operatorname{polylog}(n)}_{\text{Check rounds}} \right)$$
$$= \mathcal{O}\left(2^{2\ell} n^5 \log n\right)$$
$$= \mathcal{O}\left(T \log n\right).$$

Algorithm 1 Level 0 simulation

Input: A transcript π^v for node v.
Output: A string s^v for node v such that $|s^v| \leq n^5$.

1: $\forall v \in G : s^v \leftarrow \varepsilon$, the empty string.
2: **for** $i \in [n^5]$ **do**
3: All nodes v simulate virtual round $|\pi^v| + i$ with transcript $\pi^v \| s^v$. This is repeated $c_0 \log n$ times and the majority symbol received is appended to s^v.
4: **end for**
5: Execute $\texttt{CHECK}_1$ with inputs π^v and s^v. Let the output of node v be m^v.
6: $\forall v \in G : s^v \leftarrow s^v_{\leq m^v}$.

Algorithm 2 Level l simulation, $l \geq 1$

Input: A transcript π^v for node v.
Output: A string s^v for node v such that $|s^v| \leq n^5 2^l$.

1: Execute a level $l-1$ simulation with inputs π^v to get outputs s_1^v.
2: Execute a level $l-1$ simulation with inputs $\pi^v \| s_1^v$ to get outputs s_2^v.
3: $\forall v \in G : s^v \leftarrow s_1^v \| s_2^v$.
4: Execute $\texttt{CHECK}_{l+1}$ with inputs π^v and s^v. Let the output of node v be m^v.
5: $\forall v \in G : s^v \leftarrow s^v_{\leq m^v}$.

Algorithm 3 The Simulation Protocol Π'

Input: An input x^v for node v. Note that the rest of the protocols are also using x^v.
Output: A string s^v for node v.

1: $\forall v \in G : \pi^v \leftarrow \varepsilon$, the empty string.
2: **for** $i \in [2^\ell]$ **do**
3: Execute a level ℓ simulation with inputs π^v to get outputs s^v.
4: $\forall v \in G : \pi^v \leftarrow \pi^v \| s^v$.
5: **end for**
6: Output $s^v \leftarrow \pi^v$.

Algorithm 4 The level k check round $\texttt{CHECK}_k$

Input: Transcripts π_1^v, π_2^v for every player v. The transcript π_2^v has at most $2^{k-1} n^5$ rounds (for all v).
Output: A round number m^v for each node v.

1: $\forall v \in G : \pi^v \leftarrow \pi_1^v \| \pi_2^v$.
2: $\forall v \in G : r^v \leftarrow 2^{k-1} n^5$.
3: ρ^v is constructed from π^v as in section B, for all $v \in G$.
4: **for all** Ordered Edge $e = (u, v) \in G$ **do**
5: Sample a hash function $h : \Gamma_+^* \to \Gamma^{100k \log n}$.
6: Using binary search, find the largest $m \in \{|\pi_1^v|, \cdots, 2^{k-1} n^5 + |\pi_1^v|\}$ such that $h\left(\rho^u_{\leq m}[S_{u \to v} \setminus [|\pi_1^u|]]\right) = h\left(\pi^v_{\leq m}[S_{u \to v} \setminus [|\pi_1^v|]]\right)$. Each broadcast in this step is repeated $100k \log n$ times and majority is taken.
7: $r^v \leftarrow \min(r^v, m - |\pi_1^v|)$.
8: **end for**
9: Execute $\texttt{RS}_G$ with inputs r^v. Repeat each broadcast in this execution $100k \log n$ times and take majority. Let m^v be the value received by node v.
10: Output m^v.

B. Correctness Analysis for Π'

In this section we prove that Π' indeed simulates Π. We divide this section into 4 subsections, one for each one of our protocols. Throughout, for a string valued variable s^v, we use σ^v to denote $|s^v|$. If the value of σ^v is the same for all v, then, we use σ to denote the common value. Otherwise, define $\sigma = 0$. Analogously, for a variable s_1^v, we define σ_1^v and σ_1, *etc.*.

For the rest of this section, all probabilities and expectations are over the randomness used by the nodes, as well as the noise in the channel.

1) Analyzing $\texttt{CHECK}_k$:

Lemma B.2. *Consider an execution of $\texttt{CHECK}_k$ with inputs π_1^v, π_2^v for node v. As in the protocol, we define $\pi^v = \pi_1^v \| \pi_2^v$. Suppose that $\{\pi_1^v\}$ is m_1-consistent and let m_2 be the largest integer such that $\{\pi^v_{\leq m_1 + m_2}\}$ is $(m_1 + m_2)$-consistent. Let m^v be the output of node v. It holds that*

$$\Pr(\exists v \in G : m^v \neq m_2) \leq n^{-30k}.$$

Proof. Each broadcast in $\texttt{CHECK}_k$ is repeated $100k \log n$ times. Since the total number of rounds is $\operatorname{poly}(k, n)$, we can use a union bound to claim that every broadcast is received correctly by all the nodes except with probability at most n^{-40k}. We condition on this event throughout the rest of the proof.

Since the hash function h has range $\Gamma^{100k\log n}$, the probability that $h\left(\rho^u_{\le m}[S_{u\to v}\setminus[|\pi^u_1|]]\right) = h\left(\pi^v_{\le m}[S_{u\to v}\setminus[|\pi^v_1|]]\right)$ when $\rho^u_{\le m}[S_{u\to v}\setminus[|\pi^u_1|]] \neq \pi^v_{\le m}[S_{u\to v}\setminus[|\pi^v_1|]]$ is at most n^{-100k}. By a union bound, the probability that this happens for some u, v during an execution of $\texttt{CHECK}_k$ is at most n^{-40k}. We also condition on this event not happening throughout the rest of the proof. Both of our assumptions hold except with probability at most n^{-30k}.

Let $\hat{r}^v$ be the value of r^v right after the for loop. Using Proposition B.1 and the fact that $\{\pi^v_1\}$ is m_1-consistent, we have

$$\forall v \in G : m^v = \min_{v\in G} \hat{r}^v.$$

Given our conditioning, we also have

$$\begin{aligned}\hat{r}^v &= \min_{e=(u,v)\in G} \max\{m \mid h\left(\rho^u_{\le m}[S_{u\to v}\setminus[|\pi^u_1|]]\right)\\ &= h\left(\pi^v_{\le m}[S_{u\to v}\setminus[|\pi^v_1|]]\right)\} - |\pi^v_1|\\ &= \min_{e=(u,v)\in G} \max\{m \mid \rho^u_{\le m}[S_{u\to v}] = \pi^v_{\le m}[S_{u\to v}]\} - m_1\\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{(as } \{\pi^v_1\} \text{ is } m_1\text{-consistent)}\\ &= m_2.\end{aligned}$$

Combining the two equations above and using the fact that $m_2 \ge 0$ gives the result. □

Corollary B.3. *For $l, m \ge 0$, if the input $\{\pi^v\}$ to level l simulation is m-consistent, then the output s^v satisfies,*

$$\Pr\left(\{\pi^v \| s^v\} \text{ is not } (m+\sigma)\text{-consistent}\right) \le n^{-30(l+1)}.$$

Proof. A level l simulation ends with an execution of $\texttt{CHECK}_{l+1}$ with π^v as the first input. Use Lemma B.2. □

2) Analyzing a Level 0 Simulation: Recall the definitions of s^v and σ at the beginning of subsection B-B. The following lemma shows that, with high probability, an execution of a level 0 simulation simulates the next n^5 rounds of Π.

Lemma B.4. *There exists a $c_0 > 0$ such that for all $m \ge 0$, if the input $\{\pi^v\}$ to a level 0 simulation is m-consistent, then,*

$$\mathbb{E}[\sigma] \ge n^5(1 - n^{-20}).$$

Proof. We pick c_0 large enough so that the probability (union bounded over all edges in the graph and all the n^5 rounds in a level 0 simulation) of an incorrect reception is at most n^{-50}.

Since we assume that the input $\{\pi^v\}$ to the level 0 simulation is m-consistent, the transcripts π^v are a valid simulation of the first m rounds of Π. Thus, if there is no incorrect reception, the transcripts π^v, s^v that are input to $\texttt{CHECK}_1$ satisfy that $\{\pi^v \| s^v\}$ is $(m+n^5)$-consistent. By Lemma B.2, we are done. □

3) Analyzing a Level l Simulation: Recall the definitions of s^v_1, s^v_2, s^v, σ, σ_1, σ_2 at the beginning of subsection B-B. We have the following guarantee from a level l simulation.

Lemma B.5. *If the input $\{\pi^v\}$ to a level l simulation is m-consistent for some $m \ge 0$, then,*

$$\mathbb{E}[\sigma] \ge (\mathbb{E}[\sigma_1] + \mathbb{E}[\sigma_2])\left(1 - n^{-10l}\right).$$

Proof. Consider an execution of a level l simulation. Define the events:

$$E_1 \equiv \{\pi^v \| s^v_1\} \text{ is } (m+\sigma_1)\text{-consistent.}$$
$$E_2 \equiv \{\pi^v \| s^v_1 \| s^v_2\} \text{ is } (m+\sigma_1+\sigma_2)\text{-consistent.}$$

By Corollary B.3 applied to the first execution of the $l-1$ simulation, $\Pr(\bar{E}_1) \le n^{-30l}$, as we assume that $\{\pi^v\}$ is m-consistent. By Corollary B.3 applied to the second execution of the $l-1$ simulation, $\Pr(\bar{E}_2 \mid E_1) \le n^{-30l}$. Therefore,

$$\begin{aligned}\Pr(\bar{E}_2) &= \Pr(\bar{E}_1)\cdot\Pr(\bar{E}_2 \mid \bar{E}_1) + \Pr(E_1)\cdot\Pr(\bar{E}_2 \mid E_1)\\ &\le n^{-30l}\cdot 1 + 1\cdot n^{-30l} = 2\cdot n^{-30l}.\end{aligned}$$

The input to $\texttt{CHECK}_{l+1}$ is π^v, s^v. If E_2 occurs, this input satisfies that $\{\pi^v \| s^v\}$ is $(m+\sigma_1+\sigma_2)$-consistent. By Lemma B.2, we have

$$\Pr(\sigma \neq \sigma_1+\sigma_2 \mid E_2) \le n^{-30(l+1)}.$$

This means

$$\Pr(\sigma \neq \sigma_1+\sigma_2) \le n^{-30(l+1)} + \Pr(\bar{E}_2) \le n^{-10l}.$$

The result then follows as $\sigma \ge 0$. □

Lemma B.6. *For $l \ge 0$, if the input $\{\pi^v\}$ to a level l simulation is m-consistent for some $m \ge 0$, then,*

$$\mathbb{E}[\sigma] \ge 0.9\cdot 2^l n^5.$$

Proof. We prove a stronger statement by induction on l. Namely, that

$$\mathbb{E}[\sigma] \ge 2^l n^5\left(1 - n^{-20} - 2\sum_{i=1}^{l} n^{-8i}\right).$$

The base case is due to Lemma B.4. Consider an execution of a level l simulation. As above, define the event:

$$E_1 \equiv \{\pi^v \| s^v_1\} \text{ is } (m+\sigma_1)\text{-consistent.}$$

We get

$$\mathbb{E}[\sigma] \ge (\mathbb{E}[\sigma_1] + \mathbb{E}[\sigma_2])\left(1 - n^{-10l}\right) \quad \text{(by Lemma B.5)}$$

$$\geq \left(\mathbb{E}[\sigma_1] + \Pr(E_1)\,\mathbb{E}[\sigma_2 \mid E_1]\right)\left(1 - n^{-10l}\right) \quad \text{(as } \sigma_2 \geq 0\text{)}$$

$$\geq 2^{l-1}n^5\left(1 - n^{-20} - 2\sum_{i=1}^{l-1} n^{-8i}\right) \times \left(1 + \left(1 - \frac{\Pr(\bar{E}_1)}{2}\right)\right)\left(1 - n^{-10l}\right) \quad \text{(by induction hypothesis)}$$

$$\geq 2^{l}n^5\left(1 - n^{-20} - 2\sum_{i=1}^{l-1} n^{-8i}\right) \times \left(1 - \frac{\Pr(\bar{E}_1)}{2}\right)\left(1 - n^{-10l}\right)$$

$$\geq 2^{l}n^5\left(1 - n^{-20} - 2\sum_{i=1}^{l-1} n^{-8i}\right)\left(1 - n^{-8l}\right) \quad \text{(by Corollary B.3)}$$

$$\geq 2^{l}n^5\left(1 - n^{-20} - 2\sum_{i=1}^{l} n^{-8i}\right).$$

□

4) Analyzing Π': A level ℓ simulation gives a guarantee for the expected progress (Lemma B.6). In order to get concentration, the protocol Π' has several executions of level ℓ simulations. For the output of Π', we have:

Lemma B.7. $\Pr\left(\{s^v\} \textit{ is not } T\textit{-consistent}\right) \leq \exp\left(-2^{\ell-4}\right) + n^{-6\ell} < n^{-5}$.

Proof. For $i \in [2^\ell]$, define the events

$$E_i \equiv \{\pi_i^v\} \text{ is } \Big(\sum_{j\in[i]} \sigma_j\Big)\text{-consistent.}$$

$$E \equiv \bigwedge_{i=1}^{2^\ell} E_i.$$

Observe that:

$$\Pr(\bar{E}) \leq \Pr(\bar{E}_1) + \sum_{i=2}^{2^\ell} \Pr(\bar{E}_i \mid E_{i-1}) \leq \sum_{i=1}^{2^\ell} n^{-20(\ell+1)} \leq n^{-6\ell}. \quad \text{(by Corollary B.3)}$$

Condition on E for the rest of this proof. Due to E, we know that $\{s^v\}$ is $\sum_{i=1}^{2^\ell}\sigma_i$-consistent. It is enough to show that $\sum_{i=1}^{2^\ell}\sigma_i \geq T$. Also, due to E, for all the level ℓ simulations, we have by Markov's inequality and Lemma B.6

$$\Pr(\sigma < 2^{l-1}n^5) < 1/2. \tag{22}$$

Furthermore, the bound in Equation 22 holds for each simulation independently. Define Y_i to be the indicator random variable that is 1 iff $\sigma_i \geq 2^{\ell-1}n^5$. Using a Chernoff bound, we have

$$\Pr\left(\sum_{i=1}^{2^\ell} Y_i < \frac{1}{2}\cdot\left(\frac{1}{2}\cdot 2^\ell\right)\right) \leq \exp\left(-2^{\ell-4}\right).$$

Since $\Pr(\bar{E}) \leq n^{-6\ell}$, we get that except with probability at most $\exp\left(-2^{\ell-4}\right) + n^{-6\ell}$,

$$\sum_{i=1}^{2^\ell}\sigma_i \geq \sum_{i\in[2^\ell]:Y_i=1}\sigma_i \geq 2^{\ell-1}n^5\sum_{i\in[2^\ell]} Y_i \geq 2^{\ell-1}n^5\cdot 2^{\ell-2} = 2^{\ell-3}n^5 \geq T,$$

where the last inequality is by the definition of ℓ. □

Proof of Theorem I.2. The output of our simulation protocol Π' is T consistent with high probability (Lemma B.7). This implies that Π' simulates Π successfully with high probability. □