Electronic Colloquium on Computational Complexity, Report No. 136 (2022)

Rounds vs Communication Tradeoffs for Maximal Independent Sets

Sepehr Assadi* Gillat Kol† Zhijun Zhang‡

Abstract

We consider the problem of finding a maximal independent set (MIS) in the shared blackboard communication model with vertex-partitioned inputs. There are n players corresponding to vertices of an undirected graph, and each player sees the edges incident on its vertex – this way, each edge is known by both its endpoints and is thus *shared* by two players. The players communicate in simultaneous rounds by posting their messages on a shared blackboard visible to all players, with the goal of computing an MIS of the graph. While the MIS problem is well studied in other distributed models, and while shared blackboard is, perhaps, the simplest broadcast model, lower bounds for our problem were only known against one-round protocols.

We present a lower bound on the **round-communication tradeoff** for computing an MIS in this model. Specifically, we show that when r rounds of interaction are allowed, at least one player needs to communicate $\Omega(n^{1/20^{r+1}})$ bits. In particular, with logarithmic bandwidth, finding an MIS requires $\Omega(\log\log n)$ rounds. This lower bound can be compared with the algorithm of Ghaffari, Gouleakis, Konrad, Mitrović, and Rubinfeld [PODC 2018] that solves MIS in $O(\log\log n)$ rounds but with a logarithmic bandwidth for an *average* player. Additionally, our lower bound further extends to the closely related problem of maximal bipartite matching.

The presence of edge-sharing gives the algorithms in our model a surprising power and numerous algorithmic results exploiting this power are known. For a similar reason, proving lower bounds in this model is much more challenging, as this sharing in the players' inputs prohibits the use of standard number-in-hand communication complexity arguments. Thus, to prove our results, we devise a new round elimination framework, which we call **partial-input embedding**, that may also be useful in future work for proving *round-sensitive* lower bounds in the presence of shared inputs.

Finally, we discuss several implications of our results to multi-round (adaptive) distributed sketching algorithms, broadcast congested clique, and to the welfare maximization problem in two-sided matching markets.

*(sepehr.assadi@rutgers.edu) Department of Computer Science, Rutgers University.
†(gillat.kol@gmail.com) Department of Computer Science, Princeton University.
‡(zhijunz@princeton.edu) Department of Computer Science, Princeton University.

ISSN 1433-8092

Contents

1 Introduction

Consider the following communication model: there are n players corresponding to vertices of an undirected graph $G = (V, E)$ and each player only sees the edges incident on its vertex – this way, each edge of the graph is *shared* by the two players at its endpoints. The goal of the players is to solve some fixed problem on G, for instance, finding a spanning forest of G. To do so, the players communicate in synchronous rounds wherein all parties simultaneously write a message on a shared blackboard visible to all. The messages communicated by the players are only functions of their own inputs and the content of the blackboard. When the protocol concludes, an additional party, called the *referee*, computes the output of the protocol as a function of the blackboard content. We are interested in the tradeoff between the number of rounds of the protocol and the *per-player* communication, defined as the worst-case length of any message sent by any player in any round.

In the communication complexity terminology, this model is referred to as the multi-party communication model with *shared blackboard* and *vertex-partitioned inputs*. However, it has also been studied by different communities under different names, such as *broadcast congested clique* [DKO14, BMRT18, JLN18a, JN18], or *(adaptive) distributed sketching* [AGM12a, AGM12b, AKO20, FKN21]. At this point, there is quite a large body of algorithmic results in this model [AGM12a, AGM12a, AGM12b, AGM13, KLM+14, GMT15, MTVV15, ACK19, AKM22, LMSV11, KW14, FKN21, ACG+15] (see Section 1.2 for more details). The source of power behind these results is a crucial aspect of this model: *edge-sharing*, or in other words, the fact that each edge of the graph is seen by both its endpoints[1]. This sharing in the players' inputs makes this model an "intermediate" model lying between the *number-in-hand* model (with no input sharing) and the notorious *number-on-forehead* model (with arbitrary input sharing). As a result, lower bounds are more scarce in this model [BMN+11, DKO14, BMRT14, BMRT18, JLN18b, NY19, Yu21, AKO20].

We study the *maximal independent set (MIS)* problem in this model. While MIS is one of the most studied problems in other distributed models (see, e.g., [Lub85, Lin87, KMW16, Gha16, BBH+19]), and while shared blackboard is, perhaps, the simplest broadcast model, not much is known about MIS in this model. We do note that Luby's celebrated MIS algorithm [Lub85] implies an $O(\log n)$-round $O(1)$-per-player communication algorithm in this model. Ghaffari, Gouleakis, Konrad, Mitrovic, and Rubinfeld [GGK+18] gave an algorithm that runs in $O(\log \log n)$ rounds, but only bounds the communication of an *average* player by $O(\log n)$. I.e., the total communication by all players in a round is $O(n \log n)$, but some players may need to communicate $\omega(\log n)$ bits[2]. Moreover, Assadi, Kol, and Oshman proved that any *one*-round protocol requires almost $(n^{1/2})$ per-player communication. This state-of-affairs raises the following question:

> *What is the complexity of MIS in the shared blackboard model with vertex-partitioned inputs? In particular, what are the possible round-communication tradeoffs in this model?*

We make progress on this fundamental open question by presenting a new lower bound on the round-communication tradeoff for the MIS problem. The key contribution of our work is a new technique for proving multi-round lower bounds, even in the presence of edge-sharing. This also allows us to prove a similar lower bound for another fundamental problem, namely, the maximal bipartite matching problem.

[1] The interested reader is referred to [AGM12a] to see this in a surprising algorithm that solves graph connectivity using only a single round and $O(\log^3 n)$ communication bits per player.

[2] This algorithm is designed for the (unicast) congested clique model, but given its connection to the distributed sketching/dynamic streaming algorithm of [ACG+15]—that solves MIS as a subroutine in correlation clustering—it can be directly implemented in our model with the mentioned bounds.

Our work can be viewed as a direct continuation of two lines of work: the first line of work is on *number-in-hand* multi-round communication complexity, where we follow up on the result of Alon, Nisan, Raz, and Weinstein [ANRW15]. They give lower bounds for the bipartite maximal matching problem, where only parties on one side of the partition are allowed to communicate. The second line of work is the aforementioned lower bound of Assadi, Kol, and Oshman [AKO20], which works in our model, but only considers one-round protocols. In the following, we elaborate more on our results, techniques, and their connections to other settings.

1.1 Our Contributions

Our main result is a multi-round lower bound for computing MIS in the shared blackboard model.

Result 1. *Any r-round multi-party protocol (deterministic or randomized) in the shared blackboard model for finding a maximal independent set on n-vertex graphs requires $\Omega(n^{1/20^{r+1}})$ bits of communication per player. In particular, $\Omega(\log\log n)$ rounds are needed for protocols with $\text{polylog}(n)$ per-player communication.*

Previously, the only known lower bound for MIS in our model was the (almost) $\Omega(n^{1/2})$-communication lower bound of [AKO20] for one-round protocols. Indeed, to the best of our knowledge, there has been no prior communication lower bound in this model for any natural problem that is *sensitive* to the number of rounds (the lower bounds were either for one-round protocols, e.g., [NY19, AKO20, Yu21], or arbitrary number of rounds, e.g., [DKO14, BMRT18][3]).

The tradeoff achieved in Result 1 asymptotically matches the aforementioned $O(\log\log n)$-round algorithm of [GGK+18] for finding MIS, except that, as mentioned before, the protocol of [GGK+18] only bounds the communication of an *average* player by $O(\log n)$ bits and a few players need to communicate way more than $\text{polylog}(n)$ bits. Thus, the two results do not directly match. It remains an interesting open question to either improve the guarantee of the algorithm of [GGK+18] to per-player communication bound or improve our lower bound to average-case communication.

Our techniques in establishing Result 1 are quite general and, as a corollary to our proof, also allow us to prove a lower bound for another fundamental problem, namely, maximal matching.

Result 2. *Any r-round multi-party protocol (deterministic or randomized) in the shared blackboard model for finding a maximal matching or any constant factor approximation to maximum matching on n-vertex (bipartite) graphs requires $\Omega(n^{1/20^{r+1}})$ bits of communication per player. As such, $\Omega(\log\log n)$ rounds are needed for protocols with $\text{polylog}(n)$ per-player communication.*

As in the case of MIS, the only known lower bound prior to our work was the one-round lower bound of [AKO20]. However, for the *number-in-hand* variant of our communication model, wherein each edge of the graph is only seen by one of its endpoints, a series of papers [DNO14, ANRW15, BO17] proved a nearly-logarithmic round lower bound for the matching problem (we elaborate on this line of work later). Yet, the number-in-hand model is algorithmically much weaker than the edge-sharing model studied in our paper; for instance, the lower bound of [BO17] also holds for finding a spanning forest of the input in that model, while finding spanning forests in our model can be done with $O(\log^3 n)$ communication in just one round [AGM12a]. We refer the reader to [AKO20] for discussions on the inherit difference of number-in-hand model and our model that allows for edge-sharing and thus is "one step closer" to the notorious number-on-forehead model.

[3]Specifically, the latter ones bound the *total* communication needed to solve the problem and use this to get a lower bound on the number of rounds *times* communication per round. Such lower bounds cannot capture more nuanced round-communication tradeoffs (e.g., like the ones exhibited by [Lub85] or [GGK+18] for MIS).

Our techniques. We shall go over our techniques in detail in the streamlined overview of our approach in Section 3. For now, we only mention the high level bits of our techniques.

Our techniques unify and generalize the lower bounds of [AKO20] for one-round protocols in our model, as well as the lower bounds of [ANRW15] for multi-round protocols in the number-in-hand model. To this end, we need several substantially new ideas[4]. The main novelty of our work is in developing a new *round elimination* argument that is tailored to our edge-sharing model. Similar to standard round elimination arguments, say, the one in [ANRW15], our approach is also based on *simulating* an r-round protocol on "large" instances in only $(r-1)$ rounds for smaller "embedded" instances (with fewer players and smaller inputs). Prior work perform such a simulation by generating an *input* for the "missing" players of the large r-round instance with *low correlation* with the actual embedded $(r-1)$-round hard instance. As we argue, such an approach is doomed to fail for our model with its edge-sharing aspects. Instead, we introduce a *partial-input embedding* argument that implements this simulation via generating only the *messages* of the missing players. We then use information-theoretic tools to track the gradual increase in the *correlation* of these messages with the embedded hard instance throughout the *entire* simulation (not only the first round which is sufficient for "input-sampling" protocols of prior work).

1.2 Further Implications of Our Results to Related Models

We conclude this section by listing further implications of our results to other well-studied settings.

Broadcast congested clique. The communication model studied in our paper is equivalent to the broadcast congested clique model studied in various prior work, e.g., in [DKO14, BMRT18, JLN18a, JN18]. Specifically, our Result 1 and Result 2 imply $\Omega(\log\log n)$ round lower bounds for both MIS and maximal matching on any broadcast congested clique algorithm with $\text{polylog}(n)$ bandwidth. Incidentally, in the stronger *unicast* congested clique model, $O(\log\log n)$-round algorithms are known for both MIS [GGK+18] and maximal matching [BHH19]. We note that, as shown in [DKO14], proving lower bounds in the unicast model implies strong circuit lower bounds and thus is beyond the reach of current techniques.

Distributed sketching. Our model is also equivalent to the distributed sketching model that was initiated in the breakthrough work of [AGM12a]. Starting from the connectivity sketch of [AGM12a], there has been tremendous progress on efficient distributed sketching algorithms for various other problems in *one* round, e.g., cut sparsifiers [AGM12b], spectral sparsifiers [AGM13, KLM+14], vertex connectivity [GMT15], densest subgraph [MTVV15], $(\Delta+1)$-coloring [ACK19], Δ-coloring [AKM22], and in *multiple* rounds, e.g., minimum spanning trees [AGM12a], matchings [LMSV11, AGM12a], spanners [KW14, FKN21], and MIS and correlation clustering [ACG+15]. Given the strength of this model, proving lower bounds in this model has been a highly challenging task (see, e.g. [AKO20, FKN21]), and only a handful of lower bounds are known including $\Omega(\log^3 n)$ bits for connectivity [NY19, Yu21] and $\Omega(n^{1/2})$ bits for MIS and maximal matching [AKO20] for *one*-round sketches. Our results contribute to this line of work by providing the first *round-sensitive* lower bounds in this model, and our techniques can be of independent interest here as well.

Dynamic streaming algorithms. One key motivation of [AGM12a] in introducing graph sketching was their application to *dynamic* (semi-)streaming algorithms that can process streams of insertions and deletions of edges with $O(n\cdot\text{poly}\log(n))$ memory (*all* sketches mentioned above also imply

[4] Braverman and Oshman [BO17] gave stronger lower bounds than [ANRW15], that work for nearly logarithmic number of rounds. However, their techniques seem "too tailored" to the number-in-hand model and approximate matchings, and thus are *not* suitable for us (given the algorithm of [GGK+18] for MIS, which, even though not exactly in our model, seem quite close, it is not clear if one can get a logarithmic lower bound in our model).

dynamic streaming algorithms). Multi-round sketching protocols, similar to the ones in our model, then correspond to multi-pass streaming algorithms. Currently, the best known multi-pass dynamic semi-streaming algorithms for MIS and maximal matching require $O(\log\log n)$ passes [ACG+15] and $O(\log n)$ passes [LMSV11, AGM12a], respectively. On the lower bound front however, only single-pass lower bounds are known for either problem [AKLY16, ACK19, CDK19, DK20] (there has been recent progress on multi-pass lower bounds for computing *exact* maximum matchings [GO13, AR20, CKP+21] in logarithmic passes or even $(1+o(1))$-approximation in two passes [A22] but they do *not* apply to maximal matching in any way). While our results do *not* imply streaming lower bounds, they do rule out certain popular techniques of vertex-partitioned graph sketching for obtaining $o(\log\log n)$-pass algorithms for either problem. Thus, they can form a starting point for proving multi-pass lower bounds for all dynamic streaming algorithms as well.

Welfare maximization and interaction. A beautiful line of work initiated by [DNO14] and followed up in [ANRW15, BO17, Nis21, A17], studies the role played by the *interaction* of participating agents in the efficiency of markets. One formalization, corresponding to unit-demand agents in a matching market, is as follows: we have n agents who are interested in getting any one of their private subset of n items; the goal is to allocate these items in a way that maximizes the *welfare*, defined as the number of agents who receive an item of their liking. The market proceeds in rounds wherein the agents communicate $\text{polylog}(n)$-bit messages about their desired items. How many rounds of interaction are needed to maximize the welfare to within a constant factor?

This problem can be seen as approximating matchings on the bipartite graph consisting of agents on one side that have edges to their preferred items on the other side. The model of communication is also identical to the one in our paper with the crucial difference that only vertices on one side of the bipartition, namely, the agents, are communicating. In this model, [DNO14] gave an $O(\log n)$-round algorithm and ruled out one-round algorithms. [ANRW15] improved the lower bound to $\Omega(\log\log n)$ rounds and subsequently [BO17] obtained a nearly tight $\Omega(\frac{\log n}{\log\log n})$ lower bound (similar lower bounds are also obtained for the more general setting of combinatorial auctions in [A17]).

All these results are restricted to one-sided markets. Our Result 2 generalizes (some of) these results to *two-sided* matching markets [RS92], wherein *both* sides of the market consist of communicative agents that know in advance if they make a good match. A canonical example of two-sided matching markets is college admissions and the celebrated Gale-Shapley algorithm for stable marriage [GS62]. Another example, perhaps more closely related to the setting of our paper, is assigning users to servers in a large distributed Internet service [MS15]. Our Result 2 suggests that even when both sides of the market are able to communicate with a limited bandwidth, at least a modest amount of interaction is necessary for maximizing welfare (approximately).

2 Preliminaries

Notation. For an integer $t \in \mathbb{N}$, we write $[t]$ as a shorthand for the set $\{1, \ldots, t\}$. Let $h : A \to B$ be an arbitrary function for two sets A, B. For any subset $Z \subseteq A$, we use $h(Z) = \{h(z) \mid z \in Z\}$. For a tuple $X = (X_1, \ldots, X_t)$ and integer $i \in [t]$, we define $X_{<i} = (X_1, \ldots, X_{i-1})$ (we also define X_{-i} and $X_{\leqslant i}$ analogously). For a graph $G = (V, E)$ and a permutation σ over V, we denote by $\sigma(G)$ the graph on the same vertex set in which $\sigma(u)$ and $\sigma(v)$ are connected if and only if $(u, v) \in E$.

When there is room for confusion, we use sans-serif letters for random variables (e.g. A) and the same normal letters for their realizations (e.g. A). For random variables A, B, we use $\text{supp}(\mathsf{A})$ as the support of A, $\mathbb{H}(\mathsf{A})$ as the *Shannon entropy*, $\mathbb{I}(\mathsf{A}\,;\mathsf{B})$ as the *mutual information*, $\mathbb{D}(\mathsf{A} \mid\mid \mathsf{B})$ as the *KL-divergence*, and $\|\mathsf{A} - \mathsf{B}\|_{\text{tvd}}$ as the *total variation distance*. Necessary background on information theory, including the definitions and basic tools, is provided in Appendix A.

2.1 Multi-Party Shared Blackboard Model with Vertex-Partitioned Inputs

We work in the multi-party shared blackboard model with vertex-partitioned inputs, also known as the broadcast congested clique model in the literature. The communication model is defined formally as follows. Consider a simple graph $G = (V, E)$ with one player assigned to each of the $n = |V|$ vertices. For convenience, we identify a vertex with its associated player in the rest of this paper and use the two terms interchangeably. There is a *shared blackboard*, initially empty, that is readable and writable by all players. The player associated to a vertex $v \in V$ is presented as input with n, a unique ID of v in the range $[n]$, and IDs of all of v's neighbors $N_G(v) = \{u \in V \mid (v, u) \in E\}$. Thus, each edge $(u, v) \in E$ is *shared* by both players u and v.

Communication proceeds in $r \in \mathbb{N}$ *synchronous* rounds. For each round $t \in [r]$, the players compute their messages based on their initial input as well as the current content of the blackboard, and post them to the blackboard *simultaneously*. In a randomized protocol, the players may also use both public and private randomness. After the last round, the final content of the blackboard constitutes the *transcript*, denoted by Π, of the protocol. Then, a *referee* computes the output of the protocol depending on Π (and possibly public randomness of all players and its own private randomness). The *bandwidth* of a protocol is defined to be the *maximum* number of bits ever communicated by any player in any round.

We are interested in round-communication tradeoff of the following problems:

Maximal Independent Set. We say a protocol computes a maximal independent set (MIS) with error probability $\delta \in [0, 1]$ if the output of the referee is a valid MIS of G with probability at least $1 - \delta$ over the randomness of the protocol. The protocol may err by outputting a subset of vertices which is not independent or not maximal.

Approximate Matching. We say a protocol computes an *α-approximate matching* ($\alpha \geqslant 1$) if the output $\Gamma(\Pi)$ of the referee: (1) is always a set of *disjoint* pairs of vertices; and (2) satisfies $\mathbb{E}\,|\Gamma(\Pi) \cap E| \geqslant \mu(G)/\alpha$, where $\mu(G)$ is the size of the maximum matching of G and expectation is taken over the randomness of the protocol. This definition allows the referee to output *non-existing* edges as long as they are disjoint but only the correct ones in E are counted. This is a less restrictive error-model than requiring the algorithm to output a valid matching with certain probability and our lower bound holds even in this less restrictive setting; see also [ANRW15].

3 Technical Overview

As our proof is quite dense and technical and involves various information theoretic maneuvers that are daunting to parse, we use this section to unpack our main ideas and give a streamlined overview of our approach. We emphasize that this section oversimplifies many details and the discussions will be informal for the sake of intuition.

The starting point of our approach is a lower bound of [ANRW15] for approximate matchings in the *number-in-hand* multi-party communication model. We first give a detailed discussion of this result as our techniques need to inevitably subsume this work (since our result implies theirs as well). We then discuss the challenges of extending this result to our model that allows for edge-sharing and present a technical overview of our work. We stick with approximate matchings in this overview as it is easier to work with and to compare with [ANRW15].

3.1 A Detailed Overview of [ANRW15]

[ANRW15] considers the same communication setting as ours on bipartite graphs $G = (L \sqcup R, E)$ with the key difference that the players are only associated with vertices in L, and thus each edge is seen by only a single player. They prove that any protocol that uses $\text{polylog}(n)$ communication per player and computes an $O(1)$-approximate matching requires $\Omega(\log \log n)$ rounds in this model.

The proof in [ANRW15] is via **round elimination**: to lower bound polylog (n)-communication r-round protocols π_r, they start with $p_r \approx n^{4/5}$ *independent* $(r-1)$-round "hard" instances $I_1, \ldots, I_{p_r}$, called *principal* instances. These instances are supported on disjoint sets of $\approx n^{1/5}$ vertices each, and are then "embedded" in a single graph G to form an r-round instance I. This instance is such that the *first* message of π_r cannot reveal much information about principal instances and thus π_r cannot solve them in its remaining $r-1$ rounds given their (inductive) hardness.

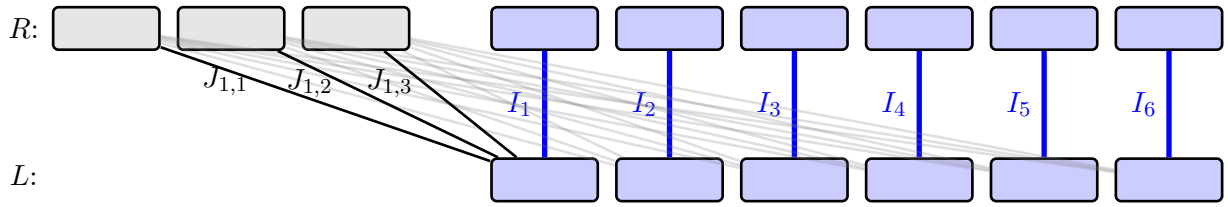

Figure 1: An illustration of the lower bound instances of [ANRW15] with parameters $f_r = 3$ and $p_r = 6$. The top right vertices (blue) are used in principal instances, while top left vertices (gray) are fooling instances. The heavy (blue) edges are from principal instances and the light (gray) edges are from fooling instances – to avoid clutter, only the edges in fooling instances of the first principal instance are drawn (solid black edges). To find a large matching in this graph, one needs to find sufficiently large matchings in many of the principal instances.

To limit the information revealed by π_r about principal instances, [ANRW15] further "packs" the graph, for every principal instance $i \in [p_r]$, with $f_r \approx n^{2/5}$ *fooling* instances $J_{i,*} := J_{i,1}, \ldots, J_{i,f_r}$. This packing ensures that: (1) these fooling instances are supported on a small set of vertices on the R-side of the bipartition and so π_r still has to solve *most of* the underlying principal instances in order to solve I; and (2) each player in I "plays" in $f_r + 1$ instances, consisting of only one principal instance, while being *oblivious* to which instance is the principal one. An ingenious idea in [ANRW15] is that these fooling instances need *not* actually be hard $(r-1)$-round instances! Instead, they form a product distribution where for each vertex $v \in L$, only the marginal distribution of v is the same under fooling and principal instances. This ensures that in the first round (and only in this round), v cannot distinguish between principal and fooling instances.

Round elimination embedding. We can now discuss how [ANRW15] *eliminates* the first round of π_r and obtains an $(r-1)$-round protocol σ for solving an $(r-1)$-round hard instance $I^\star$.

Embedding argument of [ANRW15]:

(i) The players in σ sample the first message $M^{(1)}$ of π_r using *public* randomness.

(ii) Then, they will sample an index $i \in [p_r]$ uniformly and let $I_i = I^\star$ in the instance I.

(iii) Next, they sample $J_{i,1}, \ldots, J_{i,f_r}$ conditioned on $M^{(1)}$ and $I_i = I^\star$ using *private* randomness. This is a non-trivial sampling process which, on a high level, is doable only because fooling instances are product distributions (with only the marginals matching principal ones).

More specifically, each player v *independently* sample its own input $J_{i,*}(v)$ in all the fooling instances, conditioned on only its actual input $I_i(v)$ in its principal instance I_i, and $M^{(1)}$.

(iv) Finally, the players of σ sample the remaining $p_r - 1$ principal instances I_{-i} and $(p_r-1)\cdot f_r$ fooling instances $J_{-i,*}$ conditioned on $M^{(1)}$ to have a complete instance I.

At this point, the players in σ already have the first message $M^{(1)}$ of π_r as well as inputs of all underlying instances without any communication. So, they can continue running π_r from its second round, by each player of σ on $I^\star$ communicating the messages of corresponding player of π_r in I_i, and simulating messages of π_r for players outside I_i with no communication. As π_r will also need to solve I_i for a random $i \in [p_r]$, this gives a $(r-1)$-round protocol σ for $I^\star = I_i$.

At a high level, the correctness of this approach can be argued as follows:

- The *right* distribution of all underlying variables for π_r can be expressed as (by chain rule):

$$\mathsf{M}^{(1)} \times (\mathsf{I}_i \mid \mathsf{M}^{(1)}) \times (\mathsf{J}_{i,*} \mid \mathsf{I}_i, \mathsf{M}^{(1)}) \times (\mathsf{I}_{-i}, \mathsf{J}_{-i,*} \mid \mathsf{J}_{i,*}, \mathsf{I}_i, \mathsf{M}^{(1)}). \tag{1}$$

- The distribution sampled from in the protocol σ on the other hand is:

$$\underbrace{\mathsf{M}^{(1)}}_{\text{publicly}} \times \underbrace{\mathsf{I}_i}_{\text{input}} \times \underbrace{(\bigtimes_v \mathsf{J}_{i,*}(v) \mid \mathsf{I}_i(v), \mathsf{M}^{(1)})}_{\text{privately}} \times \underbrace{(\mathsf{I}_{-i}, \mathsf{J}_{-i,*} \mid \mathsf{M}^{(1)})}_{\text{publicly}}. \tag{2}$$

Let us show that these distributions are $o(1)$-close in total variation distance, which implies that π_r also works (almost) as good on sampled instances (see Fact A.5), giving us the desired $(r-1)$-round protocol σ for $I^\star$. Here, the first terms are the same. For the second terms,

$$\|\mathsf{I}_i - (\mathsf{I}_i \mid \mathsf{M}^{(1)})\|_{\text{tvd}}^2 \leqslant \mathbb{I}(\mathsf{I}_i\,;\mathsf{M}^{(1)}) \leqslant \frac{1}{f_r+1} \cdot \mathbb{I}(\mathsf{J}_{i,*}, \mathsf{I}_i\,;\mathsf{M}_i^{(1)}) \leqslant o(1). \tag{3}$$

In Eq (3), the first inequality is standard (see Fact A.4 and Fact A.8). The second inequality uses the fact that the players in I_i in π_r are oblivious to origins of their edges in I_i vs. $J_{i,*} = J_{i,1}, \ldots, J_{i,f_r}$ (by the marginal indistinguishability of these instances); thus, the information revealed by their messages $M_i^{(1)}$ is "spread" over these instances; also, other players of π_r cannot reveal any information about these instances as they do not see them. The final inequality holds because the messages communicated by $\approx n^{1/5}$ players in I_i have collective size much smaller than $f_r \approx n^{2/5}$.

Finally, the third and fourth terms in Eq (1) and (2) also have the same distributions in both cases, which at a high level, follows from the rectangle property of communication protocols: for instance, since $J_{i,*}(u)$ and $J_{i,*}(v)$ were independent originally, they remain independent even after conditioning on $M^{(1)}$ – this is sufficient to show the equivalence of corresponding distributions. This concludes the closeness of these distributions and our overview of the work of [ANRW15].

3.2 Our Approach and New Ideas

The very first obvious challenge in using construction of [ANRW15] in our model is that it can be easily solved in just a single round once *both* sides of the bipartite graph can speak (the maximum matching of instances created is incident on vertices with degree one in R who can just communicate their edge directly on the blackboard). This brings us to the first and most obvious of our ideas.

3.2.1 Idea One: Symmetrizing the Input Distribution

The first step is to symmetrize the input distribution in [ANRW15]. Basically, to create a hard r-round instance, we again start with $(r-1)$-round hard principal instances $I_1, \ldots, I_{p_r}$. We then also add f_r sets of vertices $\mathcal{F}_1, \ldots, \mathcal{F}_{f_r}$ called the *fooling blocks* and use vertices on *both* sides of each principal instance I_i, called *principal block* $\mathcal{P}_i$, and the fooling blocks to form fooling instances $J_{i,1}, \ldots, J_{i,f_r}$ – as before, these fooling instances are *not* hard $(r-1)$-round distributions, but only that the input of principal blocks match the "right" distribution *marginally*.

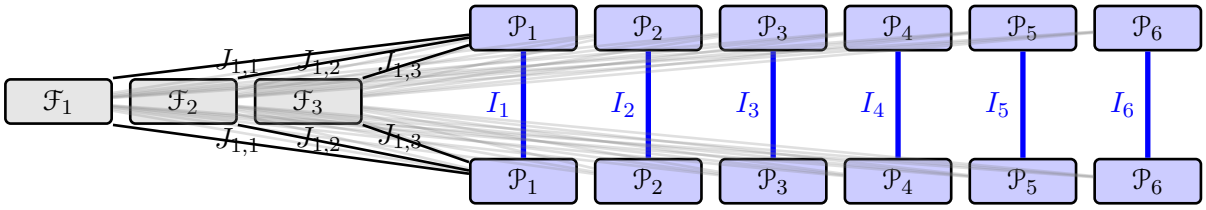

Figure 2: An illustration of our lower bound instances with parameters $f_r = 3$ and $p_r = 6$. The top and bottom vertices (blue) are principal blocks, while middle left vertices (gray) are fooling blocks. The heavy (blue) edges are from principal instances and the light (gray) edges are from fooling instances – to avoid clutter, only the edges in fooling instances of the first principal instance are drawn (solid black edges). Note that fooling blocks participate only in fooling instances while principal blocks participate both in principal and fooling instances.

This step of symmetrizing the input distribution is a straightforward extension of [ANRW15], and we claim no novelty in this part. The interesting part is how to analyze this distribution in our model in light of the following key differences from [ANRW15]: (1) in addition to principal blocks, vertices in $\mathcal{F}_1, \ldots, \mathcal{F}_{f_r}$ can now also communicate; and (2) there is an edge-sharing aspect in our model; in particular, sharing of edges between fooling blocks and principal blocks allows fooling blocks to communicate even about edges directly inside principal instances (!), and yet fooling blocks themselves are not even fooled anymore in the distribution. We discuss our approach for handling these parts in the following three subsections.

3.2.2 Idea Two: Bounding Revealed Information on Average

Our goal as before is to do a round elimination argument and embed an $(r-1)$-round instance inside an r-round one. Our embedding argument in the *first* round is going to be the same as that of [ANRW15], except that we also sample the first message $M_F^{(1)}$ of fooling blocks using public randomness (there are *no* such players in [ANRW15]). We will then have all the messages of round one, namely, $M^{(1)} = (M_P^{(1)}, M_F^{(1)})$, as well as edges incident on the principal block $\mathcal{P}_i$, namely, $I_i, J_{i,*}$, inside I without having done any communication.

Specifically, we design a protocol σ that given an $(r-1)$-round instance $I^\star$, creates an r-round instance I and uses a polylog(n)-communication r-round protocol π_r on I to solve $I^\star$ as follows.

Our embedding argument – first round:

(*i*) Players in σ sample the first message $M_P^{(1)}, M_F^{(1)}$ of principal and fooling blocks *publicly*.

(*ii*) Then, they will sample an index $i \in [p_r]$ uniformly and let $I_i = I^\star$ in the instance I; thus, players in σ will play the role of principal block $\mathcal{P}_i$ in π_r from now on.

(*iii*) Next, they sample $J_{i,*}$ conditioned only on $M_P^{(1)}$ and $I_i = I^\star$ using *private* randomness by each vertex v of σ independently sampling $J_{i,*}(v)$ only conditioned on $I_i(v), M_P^{(1)}$.

Let us argue that the joint distribution of obtained random variables at this point is close to that of the actual distribution induced by π_r (similar to Eq (1) and (2) for [ANRW15]):

- The *right* distribution of the underlying variables for π_r can be expressed as:

$$(\mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)}) \times (\mathsf{I}_i \mid \mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)}) \times (\mathsf{J}_{i,*} \mid \mathsf{I}_i, \mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)}). \tag{4}$$

- The distribution sampled from in the protocol σ is:

$$\underbrace{(\mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)})}_{\text{publicly}} \times \underbrace{\mathsf{I}_i}_{\text{input}} \times (\underbrace{\bigtimes_v \mathsf{J}_{i,*}(v) \mid \mathsf{I}_i(v), \mathsf{M}_P^{(1)}}_{\text{privately}}). \tag{5}$$

The first terms are the same. For the second terms, similar to Eq (3), we have,

$$\|\mathsf{I}_i - (\mathsf{I}_i \mid \mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)})\|_{\text{tvd}}^2 \leqslant \mathbb{I}(\mathsf{I}_i ; \mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)}) = \mathbb{I}(\mathsf{I}_i ; \mathsf{M}_P^{(1)}) + \mathbb{I}(\mathsf{I}_i ; \mathsf{M}_F^{(1)} \mid \mathsf{M}_P^{(1)}), \tag{6}$$

using the chain rule of mutual information (Fact A.1-(6)) in the equality. The first term in RHS above can still be bounded by $o(1)$ by the same logic that principal blocks are oblivious to identity of principal instance edges in their input. But such a statement is not true about fooling blocks in the second term, as those vertices themselves are not fooled. Consider the following 1-bit protocol.

Example. Suppose we direct each edge of the graph randomly to one of its endpoints using public randomness. Principal blocks[a] send the XOR of their *outgoing* edges and fooling blocks send the XOR of their *incoming* edges incident on $J_{i,*}$ for some $i \in [p_r]$. Taking the XOR of messages sent by $\mathcal{P}_i$, $M_{P,i}^{(1)}$, and fooling blocks, $M_F^{(1)}$, reveals XOR of all edges inside I_i as each such edge will be outgoing for exactly one endpoint and edges in $J_{-i,*}$ cancel out in this XOR. This reveals one bit of information about I_i, making $\mathbb{I}(\mathsf{I}_i ; \mathsf{M}_F^{(1)} \mid \mathsf{M}_P^{(1)}) \geqslant 1$. (Ideas like this are used in actual distributed sketching protocols, e.g., in [AGM12a, KLM+14].)

[a] A player can know whether it is principal or fooling simply based on its degree.

Instead, we show that fooling blocks cannot reveal much about I_i for an *average* $i \in [p_r]$:

$$\mathop{\mathbb{E}}_i[\mathbb{I}(\mathsf{I}_i ; \mathsf{M}_F^{(1)} \mid \mathsf{M}_P^{(1)})] \leqslant \frac{1}{p_r} \cdot \mathbb{I}(\mathsf{I}_1, \ldots, \mathsf{I}_{p_r} ; \mathsf{M}_F^{(1)} \mid \mathsf{M}_P^{(1)}) \leqslant o(1), \tag{7}$$

where in the second inequality we used the fact that the polylog(n)-bit messages of *all* $f_r \approx n^{2/5}$ fooling blocks of size $\approx n^{1/5}$ cannot reveal more than $o(p_r)$ information as $p_r \approx n^{4/5}$ (this idea is similar to the "public-vs-private" vertices of [AKO20] for one-round lower bounds in the distributed sketching model). This allows us to bound the LHS of Eq (6) on average for $i \in [p_r]$. A similar

type of argument can be applied to the third terms also to "drop" the conditioning on $M_F^{(1)}$, while changing the distribution only by $o(1)$ in total variation distance. This implies that

$$\mathop{\mathbb{E}}_{i} \|(\mathsf{J}_{i,*} \mid \mathsf{I}_i, \mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)}) - (\mathsf{J}_{i,*} \mid \mathsf{I}_i, \mathsf{M}_P^{(1)})\|_{\text{tvd}} \leqslant o(1).$$

By [ANRW15], the second distribution here matches the product distribution sampled privately by the players (the third term of Eq (5)). This is now sufficient for simulating the *first round* of π_r (almost) faithfully with no communication as $i \in [p_r]$ is also chosen randomly in the embedding[5].

It is tempting to consider our job done as we successfully simulated the first round of π_r with no communication, and thus we *eliminated* a round. But in fact, this is just the start of **the unique challenges of our model**. Unlike [ANRW15], it is not clear how we can continue running π_r in the subsequent rounds: in σ, we have only decided on the input of principal block $\mathcal{P}_i$ in I – the input to other principal blocks and all fooling blocks are still undecided, and so π_r is not well defined for the subsequent rounds. We now need to deviate entirely from [ANRW15] to handle this.

3.2.3 Idea Three: Partial-Input Embedding and Non-Simultaneous Simulation

To continue running π_r from its second round onwards, we should be able to simulate *all* players in I, not only the principal block $\mathcal{P}_i$ responsible for $I_i = I^\star$. Let us consider a standard approach.

Standard approach for handling remaining instances. The standard approach is to sample input of remaining players in π_r using public randomness and let the "actual" players of σ simulate them "in their head" with no communication (this corresponds to step (iv) of embedding of [ANRW15]). This approach fails completely for us. Consider the fooling blocks first: at this point in the protocol σ, the players have sampled $J_{i,*}$ *privately* which was necessary in the first round (given the correlation of $J_{i,*}(v)$ with $I_i(v)$ via $M_P^{(1)}$ and that $I_i(v)$ was only known to v). But given that the other endpoints of these edges are in fooling blocks, this means that *no* single player of σ can even know the edges incident on a single vertex in fooling blocks, leaving no player to simulate players of π_r in fooling blocks (or sampling rest of their inputs).

A more subtle issue happens when it comes to the rest of principal blocks, which on the surface, should be fine given they share no edges with principal block $\mathcal{P}_i$. To be able to sample instances $I_{-i}, J_{-i,*}$ publicly in the last step of embedding, we need the following two distributions to be close:

$$\underbrace{(\mathsf{I}_{-i}, \mathsf{J}_{-i,*} \mid \mathsf{J}_{i,*}, \mathsf{I}_i, \mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)})}_{\text{right distribution}} \quad \textit{vs.} \quad \underbrace{(\mathsf{I}_{-i}, \mathsf{J}_{-i,*} \mid \mathsf{M}_P^{(1)}, \mathsf{M}_F^{(1)})}_{\text{"input-sampling"-protocol distribution}}.$$

Yet, even a 1-bit communication protocol can turn these two distributions far from each other:

> **Example.** Suppose principal blocks remain silent and each fooling block sends the XOR of their incident edges. Then conditioned on the messages $M_F^{(1)}$, once we additionally know $J_{i,*}$, we learn the parity of edges in $J_{-i,*}$ which changes the distribution of $J_{-i,*}$ by $\Omega(1)$.

All in all, when it comes to our edge-sharing model, the standard approach of sampling the remaining instances inherently fails: (i) fooling blocks are directly incident on edges in $J_{i,*}$ which are part of the input to players in $\mathcal{P}_i$ in π_r; (ii) worse yet, the messages of fooling blocks even correlate inputs of the rest of principal vertices with those of $\mathcal{P}_i$, meaning that *all* principal players can reveal information about I_i not only the ones in $\mathcal{P}_i$ that are directly incident on it.

[5] [ANRW15] also works with a random $i \in [p_r]$ but only to ensure that the underlying instance I_i needs to be solved by π_r as most but not all principal instances are solved in π_r – *all* information-theoretic guarantees for π_r mentioned for the embedding of [ANRW15] hold for *arbitrary* $i \in [p_r]$ unlike ours.

Our approach for handling remaining instances. A key idea we use in the rest of our protocol is what we call **partial-input embedding**: we only generate the rest of the input for players $\mathcal{P}_i$ and for all the remaining players, we will simulate them solely by *sampling their messages* without ever committing to their input. Thus, our embedding keeps going even beyond the first round as we will need to generate the messages of remaining players throughout the entire execution of π_r.

In particular, after running the embedding part of the first round, for any round $t > 1$, the players in the protocol σ will simulate the t-th round of π_r as follows:

> Our embedding argument – after first round:
>
> (i) The players in σ communicate messages of $\mathcal{P}_i$ using the current content of the blackboard $M^{(<t)}$, and their inputs $J_{i,*}, I_i$ sampled for the first round, and send the messages $M^{(t)}_{P,i}$.
>
> (ii) *After* this message is revealed, the players use *public* randomness to sample the t-th message of remaining players $M^{(t)}_{-i} := (M^{(t)}_{P,-i}, M^{(t)}_F)$ conditioned on public knowledge $M^{(<t)}, M^{(t)}_{P,i}$.

It is worth pointing out a rather strange aspect of this embedding. In π_r itself, the messages $M^{(t)}_{P,i}$ and $M^{(t)}_{-i}$ are communicated *simultaneously* with each other. Yet, in our simulation of π_r, we are crucially using messages principal block $\mathcal{P}_i$ to help us generate the remaining messages! We will discuss the necessity of this **non-simultaneous simulation** of a round in the next subsection.

As before, let us examine the underlying distributions in the first t rounds for $t > 1$:

- The right distribution of the underlying variables up until this point in π_r is:

$$\underbrace{(\mathsf{M}^{(<t)}, \mathsf{J}_{i,*}, \mathsf{I}_i)}_{\text{prior rounds}} \times (\mathsf{M}^{(t)}_{P,i} \mid \mathsf{M}^{(<t)}, \mathsf{J}_{i,*}, \mathsf{I}_i) \times (\mathsf{M}^{(t)}_{-i} \mid \mathsf{M}^{(t)}_{P,i}, \mathsf{M}^{(<t)}, \mathsf{J}_{i,*}, \mathsf{I}_i). \tag{8}$$

- The distribution sampled from in the protocol σ is:

$$\underbrace{(\mathsf{M}^{(<t)}, \mathsf{J}_{i,*}, \mathsf{I}_i)}_{\text{prior rounds}} \times \underbrace{(\bigtimes_v \mathsf{M}^{(t)}_{P,i}(v) \mid \mathsf{M}^{(<t)}, \mathsf{J}_{i,*}(v), \mathsf{I}_i(v))}_{\text{communication}} \times \underbrace{(\mathsf{M}^{(t)}_{-i} \mid \mathsf{M}^{(t)}_{P,i}, \mathsf{M}^{(<t)})}_{\text{publicly}}. \tag{9}$$

The first terms can be shown to be $o(1)$-close inductively (with base case being success of our simulation in the first round). The second terms are identical since the messages $M^{(t)}_{P,i}$ in π_r are simply generated simultaneously by each vertex $v \in \mathcal{P}_i$ looking at its own neighborhood $J_{i,*}(v), I_i(v)$ and the blackboard $M^{(<t)}$. For the last terms to be close, similar to Eq (3) and (6), we need to bound the mutual information between $M^{(t)}_{-i}$ and $J_{i,*}, I_i$ at this point of the protocol, namely:

$$\|(\mathsf{M}^{(t)}_{-i} \mid \mathsf{M}^{(t)}_{P,i}, \mathsf{M}^{(<t)}) - (\mathsf{M}^{(t)}_{-i} \mid \mathsf{M}^{(t)}_{P,i}, \mathsf{M}^{(<t)}, \mathsf{J}_{i,*}, \mathsf{I}_i)\|^2_{\text{tvd}} \leqslant \mathbb{I}(\mathsf{M}^{(t)}_{-i}; \mathsf{J}_{i,*}, \mathsf{I}_i \mid \mathsf{M}^{(t)}_{P,i}, \mathsf{M}^{(<t)}). \tag{10}$$

Yet, while the RHS of this equation may seem similar to that of Eq (7), this is a much more challenging term to bound as we shall discuss in the next subsection. For now, we only mention that our proof eventually bounds this information term *on average* for $i \in [p_r]$ with $o(1)$ which allows us to continue the simulation.

Having shown the $o(1)$-closeness of the distribution of π_r and the one used in our embedding, the proof ends as follows. The players of σ can continue running π_r by playing the role of principal block $\mathcal{P}_i$ in π_r explicitly with proper communication and keep sampling messages of remaining players as done in the embedding. At the end of the last round, they will obtain an almost faithful

simulation of the entire protocol π_r which allows them to solve $I^\star = I_i$ as π_r likely needs to solve I_i for a random $i \in [p_r]$. This will then give us an $(r-1)$-round protocol for $I^\star$ which in turn allows us to use the inductive hardness of these instances to infer the lower bound for r-round protocols.

3.2.4 Idea Four: Bounding Gradual Correlation of Players' Inputs

The main technical part of our proof is to bound the information term in the RHS of Eq (10), namely, the information other players can reveal about the input of principal block $\mathcal{P}_i$ in a single round. By the definition of $M^{(t)}_{-i} = (M^{(t)}_{P,-i}, M^{(t)}_F)$ and chain rule (Fact A.1-(6)), we have,

$$\text{RHS of Eq (10)} = \mathbb{I}(\mathsf{M}^{(t)}_{P,-i}\,; \mathsf{J}_{i,*}, \mathsf{I}_i \mid \mathsf{M}^{(t)}_{P,i}, \mathsf{M}^{(<t)}) + \mathbb{I}(\mathsf{M}^{(t)}_{F}\,; \mathsf{J}_{i,*}, \mathsf{I}_i \mid \mathsf{M}^{(t)}_{P}, \mathsf{M}^{(<t)}). \tag{11}$$

Recall that by the construction of the instance I, we have $\mathsf{J}_{i,*}, \mathsf{I}_i \perp \mathsf{J}_{-i,*}, \mathsf{I}_{-i}$. By the rectangle property of communication protocols, if the input of players are independent of each other, then even after communication, their corresponding input remains independent. *Assuming* we have this conditional independence here, one can easily prove both of the following properties:

$$\mathbb{I}(\mathsf{M}^{(t)}_{P,-i}\,; \mathsf{J}_{i,*}, \mathsf{I}_i \mid \mathsf{M}^{(t)}_{P,i}, \mathsf{M}^{(<t)}) = 0, \qquad \text{(by Fact A.1-(2))}$$

$$\mathop{\mathbb{E}}_{i}[\mathbb{I}(\mathsf{M}^{(t)}_{F}\,; \mathsf{J}_{i,*}, \mathsf{I}_i \mid \mathsf{M}^{(t)}_{P}, \mathsf{M}^{(<t)})] \leqslant \frac{1}{p_r} \cdot \mathbb{I}(\mathsf{M}^{(t)}_{F}\,; \mathsf{J}, \mathsf{I} \mid \mathsf{M}^{(t)}_{P}, \mathsf{M}^{(<t)}) \leqslant o(1). \qquad \text{(similar to Eq (7))}$$

So then what is the problem here? **Short answer: edge-sharing between the players!**

While $\mathsf{J}_{i,*}, \mathsf{I}_i \perp \mathsf{J}_{-i,*}, \mathsf{I}_{-i}$ is true initially, having fooling blocks that are able to see (subsets of) *both* these sets from the other endpoints, means that their messages can *correlate* these inputs as well. In other words, it can be that $\mathsf{J}_{i,*}, \mathsf{I}_i \not\perp \mathsf{J}_{-i,*}, \mathsf{I}_{-i} \mid \mathsf{M}^{(<t)}_F$ already from the second round. What is even more problematic is that even principal blocks in $\mathcal{P}_i$ and $\mathcal{P}_{-i}$ will see messages of these fooling blocks, so after the second round, even messages of other principal blocks correlate their originally independent inputs – more formally, this means that $\mathsf{J}_{i,*}, \mathsf{I}_i \not\perp \mathsf{J}_{-i,*}, \mathsf{I}_{-i} \mid \mathsf{M}^{(t)}_P$ (with no direct conditioning on fooling blocks' messages) can also happen after the second round!

The following example helps to motivate our approach.

Example. Consider the following two protocols:

- Protocol 1: in the second round, every principal block *except for* $\mathcal{P}_i$ sends XOR of their edges to fooling blocks[a] $J_{-i,*}$, while fooling blocks send XOR of all their edges in J.
- Protocol 2: in the second round, every principal block sends XOR of their edges in J while fooling blocks send XOR of all their edges in J.

In the first protocol, conditioned on $M^{(2)}_F$, the messages $M^{(2)}_{P,-i}$ reveal the XOR of edges in $J_{i,*}$, and thus the first mutual information term in Eq (11) is 1 bit (note that here $M^{(2)}_{P,i} = \varnothing$).

In the second protocol, while $M^{(2)}_{P,-i}, M^{(2)}_F$ still reveal the XOR of $J_{i,*}$, given that $M^{(2)}_{P,i}$ is already this XOR itself, the mutual information term in Eq (11) is 0 bit.

[a] Identity of fooling blocks can be known to everyone in the second round.

This example shows that one can have protocols that for *some* values of $i \in [p_r]$, principal blocks in $\mathcal{P}_{-i}$ can reveal non-trivial information about inputs of a principal block $\mathcal{P}_i$ also. But the given protocol (Protocol 1) is quite sensitive to the choice of index i, and for other indices $j \neq i$, this

revealing of information no longer happens in this specific protocol. On the other hand, making the protocol less sensitive to the choice of i by "symmetrizing" the actions of players breaks its information-revealing property as players in $\mathcal{P}_i$ themselves will reveal the information offered by others. We exploit this by bounding the first term of Eq (11) *on average* for $i \in [p_r]$. Note that this is precisely the step that our non-simultaneous simulation of a round, alluded to in Section 3.2.3, kicks in: the messages of $M^{(2)}_{P,-i}$ are still correlated heavily with $J_{i,*}, I_i$ even in Protocol 2; but conditioning on $M^{(2)}_{P,i}$ allows us to "break" this correlation and thus generate these messages even in the absence of public knowledge of $J_{i,*}, I_i$. We argue this is true for all protocols in the following.

To continue, by using chain rule (Fact A.1-(6)) on the first term of Eq (11), we get that,

$$\mathbb{I}(\mathsf{M}^{(t)}_{P,-i}\,;\mathsf{J}_{i,*},\mathsf{I}_i \mid \mathsf{M}^{(t)}_{P,i},\mathsf{M}^{(<t)}) = \mathbb{I}(\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P}\,;\mathsf{J}_{i,*},\mathsf{I}_i) - \mathbb{I}(\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i}\,;\mathsf{J}_{i,*},\mathsf{I}_i) \tag{12}$$

where RHS is all the information revealed by the protocol about $J_{i,*}, I_i$ *minus* the information revealed already by players $\mathcal{P}_i$ and content of the blackboard. Now, in the absence of any conditioning, one can use the fact that $\mathsf{J}_{i,*},\mathsf{I}_i \perp \mathsf{J}_{-i,*},\mathsf{I}_i$ to bound:

$$\text{First term of Eq (12) on average:}\quad \mathop{\mathbb{E}}_{i}[\mathbb{I}(\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P}\,;\mathsf{J}_{i,*},\mathsf{I}_i)] \leqslant o(1) + \frac{1}{p_r}\cdot\mathbb{I}(\mathsf{M}^{(\leqslant t)}_{P}\,;\mathsf{J},\mathsf{I} \mid \mathsf{M}^{(<t)}_{F}),$$

i.e., argue that fooling blocks can only reveal $o(1)$ bits about the input of an average principal block and the rest is the *average* information revealed by principal blocks themselves about the entire input. The second term of Eq (12) is lower bounded by (via a simple application of chain rule and non-negativity of mutual information),

$$\text{Second term of Eq (12) on average:}\quad \mathop{\mathbb{E}}_{i}[\mathbb{I}(\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i}\,;\mathsf{J}_{i,*},\mathsf{I}_i)] \geqslant \mathop{\mathbb{E}}_{i}[\mathbb{I}(\mathsf{M}^{(\leqslant t)}_{P}\,;\mathsf{J}_{i,*},\mathsf{I}_i) \mid \mathsf{M}^{(<t)}_{F}].$$

Last step of the proof is to bound the second terms of the two equations above by showing that

$$\mathbb{I}(\mathsf{M}^{(\leqslant t)}_{P}\,;\mathsf{J},\mathsf{I} \mid \mathsf{M}^{(<t)}_{F}) \leqslant \sum_{i=1}^{p_r}\mathbb{I}(\mathsf{M}^{(\leqslant t)}_{P}\,;\mathsf{J}_{i,*},\mathsf{I}_i \mid \mathsf{M}^{(<t)}_{F}).$$

In words, this means that the total information revealed by principal blocks about the entire instance is bounded by the sum of the information revealed by them about each individual principal block's input $J_{i,*}, I_i$ for $i \in [p_r]$ *after* we condition on the messages of fooling blocks. This step requires a detailed calculation that at its core boils down to the fact that once we condition on $M^{(<t)}_F$, we can "isolate" the information revealed by each message $M^{(t)}_{P,i}$ solely to $J_{i,*}, I_i$ – in other words, the principal blocks cannot generate correlation with other principal blocks' inputs on their own beyond what is already forced by fooling blocks.

Plugging in these bounds all together in Eq (12) bounds the RHS by $o(1)$. A similar exercise, allows us to bound the second term in Eq (11) by $o(1)$ also, which bounds the total information revealed about $J_{i,*}, I_i$ by players other than the ones in $\mathcal{P}_i$ by $o(1)$. This concludes the $o(1)$ bound on the mutual information term in Eq (10), and implies the correctness of our simulation.

To conclude, we managed to simulate *all* rounds of π_r almost faithfully by continuing the embedding throughout the protocol and as a result solve the underlying instance $I^\star$ in $(r-1)$ rounds using a protocol with polylog(n)-size messages. We can now repeat this argument for $(r-1)$-round protocols and since in each recursion, the size of underlying instances drops by a factor of $\approx n^{1/5}$, we will end up with a non-trivial instance for any $r = o(\log\log n)$ that needs to be solved by a 0-round protocol – a contradiction that implies our desired lower bound.

4 A Hard Distribution for Maximal Independent Set

The following is a formal restatement of Result 1.

Theorem 1 (Result 1, formal)**.** *For $r \geqslant 0$ and any r-round multi-party protocol (deterministic or randomized) in the shared blackboard model for computing a maximal independent set on n-vertex graphs with constant error probability, there must exist* some *vertex communicating at least $\Omega(n^{1/20^{r+1}})$ bits in* some *round.*

In this section, we give a recursive definition of the hard distribution for maximal independent set that we are going to use for our proofs in Section 5. The base case is the following hard distribution $\mathcal{D}^{(0)}_{\mathsf{MIS}}$ for protocols without any communication.

Distribution 1. The hard distribution $\mathcal{D}^{(0)}_{\mathsf{MIS}}$ for protocols computing a maximal independent set without any communication.
Parameters: bandwidth k, number of vertices $n_0 = 2k$.

1. Let E be an arbitrary, *fixed* perfect matching over n_0 vertices.
2. For $e \in E$, drop e with probability $1/2$ independently.
3. Return the graph G sampled above.

An immediate observation about $\mathcal{D}^{(0)}_{\mathsf{MIS}}$ is that any valid maximal independent set uniquely determines the set of matching edges that is dropped from E: for $e = (u, v) \in E$, e is dropped from E if and only if both of u, v are present in the maximal independent set. So for any deterministic referee, it can output a valid maximal independent set with probability at most 2^{-k} over $\mathcal{D}^{(0)}_{\mathsf{MIS}}$ if it gets no information from the vertices. Note that this distributional bound naturally generalizes to randomized referees by an averaging argument, which is summarized in the following lemma.

Lemma 4.1 (Base Case)**.** *Any 0-round protocol for computing a maximal independent set can only succeed with probability 2^{-k} over $\mathcal{D}^{(0)}_{\mathsf{MIS}}$.*

Building upon $\mathcal{D}^{(0)}_{\mathsf{MIS}}$, we construct the r-round hard distribution $\mathcal{D}^{(r)}_{\mathsf{MIS}}$ recursively. Assume we are given the $(r-1)$-round hard distribution $\mathcal{D}^{(r-1)}_{\mathsf{MIS}}$ over n_{r-1} vertices. The construction consists of two steps: first defining an auxiliary "half distribution" $\mathcal{H}^{(r)}_{\mathsf{MIS}}$ and then using $\mathcal{H}^{(r)}_{\mathsf{MIS}}$ to get the desired $\mathcal{D}^{(r)}_{\mathsf{MIS}}$, as shown below. The "half instances" roughly correspond to the hard instances we talk about in Section 3. See Figure 3 for an illustration.

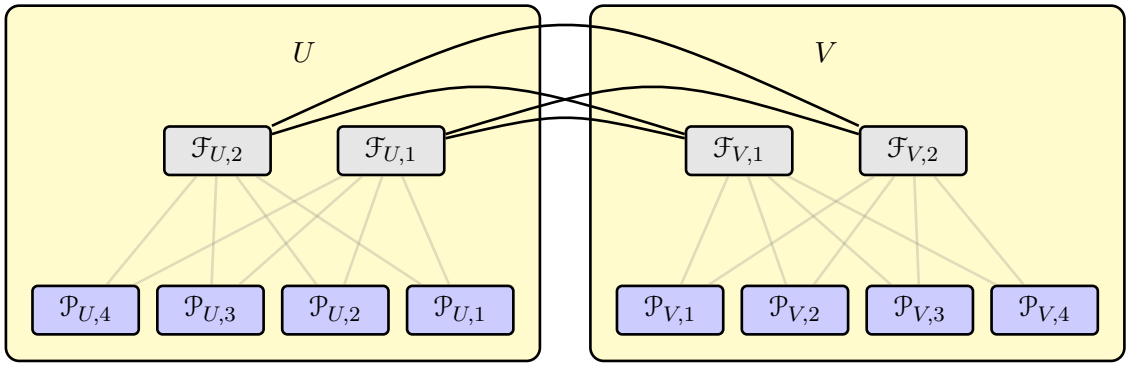

Figure 3: An illustration of our lower bound instances for maximal independent set with parameters $\hat{f}_r = 2$ and $\hat{p}_r = 4$. The bottom vertices (blue) are principal blocks, while top vertices (gray) are fooling blocks. The heavy (solid black) edges fully connect fooling vertices from two "half instances" (yellow boxes). Note that these are the only edges across two "half instances". To find a maximal independent set in this graph, one needs to find maximal independent sets in all principal instances of at least one of "half instances".

Distribution 2. The "half distribution" $\mathcal{H}^{(r)}_{\mathsf{MIS}}$ over graphs with vertex set V ($r \geqslant 1$).
Parameters: bandwidth k, number of **fooling blocks** $\hat{f}_r = k^6 \cdot n_{r-1}^3$, number of **principal blocks** $\hat{p}_r = k^6 \cdot n_{r-1}^3 \cdot \hat{f}_r$, number of vertices $\hat{n}_r = (n_{r-1} - 1) \cdot \hat{f}_r + n_{r-1} \cdot \hat{p}_r$, and vertex set V with $|V| = \hat{n}_r$.

1. Partition V into disjoint sets of vertices $\mathcal{P}_1, \ldots, \mathcal{P}_{\hat{p}_r}, \mathcal{F}_1, \ldots, \mathcal{F}_{\hat{f}_r}$ such that $\forall i \in [\hat{p}_r] : \ |\mathcal{P}_i| = n_{r-1}$ and $\forall j \in [\hat{f}_r] : \ |\mathcal{F}_j| = n_{r-1} - 1$. Define $\mathcal{P}(V) := \bigcup_{i \in [\hat{p}_r]} \mathcal{P}_i$ and $\mathcal{F}(V) := \bigcup_{j \in [\hat{f}_r]} \mathcal{F}_j$.
2. For $i \in [\hat{p}_r]$, sample an independent instance of $\mathcal{D}^{(r-1)}_{\mathsf{MIS}}$ on $\mathcal{P}_i$.
3. For $u \in \mathcal{P}(V)$ and $j \in [\hat{f}_r]$, sample an independent instance of $\mathcal{D}^{(r-1)}_{\mathsf{MIS}}$ on $\mathcal{F}_j \cup \{u\}$ and only keep the edges adjacent to u (dropping all the edges between vertices in $\mathcal{F}_j$).
4. Return the graph G sampled above.

Distribution 3. The hard distribution $\mathcal{D}^{(r)}_{\mathsf{MIS}}$ for r-round protocols computing a maximal independent set ($r \geqslant 1$).
Parameters: bandwidth k, number of **fooling blocks** $f_r = 2\hat{f}_r$, number of **principal blocks** $p_r = 2\hat{p}_r$, number of vertices $n_r = 2\hat{n}_r$.

1. Let U and V be two disjoint sets of vertices, each of size $\hat{n}_r$. Sample two independent instances of $\mathcal{H}^{(r)}_{\mathsf{MIS}}$ on U and V.
2. For $u \in \mathcal{F}(U)$ and $v \in \mathcal{F}(V)$, add an edge (u, v).
3. Let G' be the graph sampled above. Sample a uniformly random permutation σ over $U \cup V$ and return $G = \sigma(G')$.

Remark 4.2. *A few remarks are in order.*

1. *In the construction of the "half distribution" $\mathcal{H}^{(r)}_{\mathsf{MIS}}$, we call the sets of vertices $\mathcal{P}_1, \ldots, \mathcal{P}_{\hat{p}_r}$ the* principal blocks, *and the sets of vertices $\mathcal{F}_1, \ldots, \mathcal{F}_{\hat{f}_r}$ the* fooling blocks. *All vertices in $\mathcal{P}(V)$ and $\mathcal{F}(V)$ are the* principal vertices *and the* fooling vertices, *respectively.*

2. *With a slight abuse of notation, we write $\sigma(\mathcal{P}_1), \ldots, \sigma(\mathcal{P}_{p_r})$ to denote all p_r principal blocks of $\sigma(U \cup V)$, and similarly $\sigma(\mathcal{F}_1), \ldots, \sigma(\mathcal{F}_{f_r})$ for all fooling blocks, in the construction of the hard distribution $\mathcal{D}^{(r)}_{\mathsf{MIS}}$.*

3. *It is not hard to see that $n_r \leqslant k^{20^{r+1}}$ for $r \geqslant 0$. Indeed, $n_0 = 2k \leqslant k^{20}$ and by induction, the number of fooling blocks is $\hat{f}_r \leqslant k^6 \cdot k^{3 \cdot 20^r} \leqslant k^{9 \cdot 20^r}$, the number of principal blocks is $\hat{p}_r \leqslant k^6 \cdot k^{3 \cdot 20^r} \cdot \hat{f}_r \leqslant k^{18 \cdot 20^r}$, and thus $n_r \leqslant 2 \cdot 2 \cdot k^{20^r} \cdot \hat{p}_r \leqslant k^{20^{r+1}}$ for $r \geqslant 1$. Throughout the paper we assume the bandwidth parameter k is at least some sufficiently large constant.*

One important property about $\mathcal{D}^{(r)}_{\mathsf{MIS}}$, which justifies our use of two "half instances", is that any valid maximal independent set for G must also be maximal for the induced subgraph on either $\sigma(\mathcal{P}(U))$ or $\sigma(\mathcal{P}(V))$. The implication is that solving a hard instance drawn from $\mathcal{D}^{(r)}_{\mathsf{MIS}}$ requires to solve at least one of the "half instances" drawn from $\mathcal{H}^{(r)}_{\mathsf{MIS}}$. Formally, we have the following claim.

Claim 4.3. *Let Γ be any valid maximal independent set for a graph G drawn from $\mathcal{D}^{(r)}_{\mathsf{MIS}}$. Then at least one of the following must hold:*

1. *$\Gamma \cap \sigma(\mathcal{P}(U))$ is a valid maximal independent set for the induced subgraph on $\sigma(\mathcal{P}(U))$.*

2. *$\Gamma \cap \sigma(\mathcal{P}(V))$ is a valid maximal independent set for the induced subgraph on $\sigma(\mathcal{P}(V))$.*

Proof. Without loss of generality we assume σ is simply the identity permutation throughout the proof. Suppose for now that Γ contains one fooling vertex $f \in \mathcal{F}(U)$. Note that our construction in Distribution 3 fully connects $\mathcal{F}(U)$ to $\mathcal{F}(V)$ so none of $\mathcal{F}(V)$ is contained in Γ. Furthermore, those are the only edges between the two "half instances" on U and V. Altogether, it shows $\mathcal{P}(V)$ has no neighbor chosen by Γ. Since Γ is a valid maximal independent set for G, its restriction to $\mathcal{P}(V)$, i.e. $\Gamma \cap \mathcal{P}(V)$, must be a valid maximal independent set for the induced subgraph on $\mathcal{P}(V)$.

The case is symmetric when Γ contains one fooling vertex $f \in \mathcal{F}(V)$. It is also not hard to see that both statements in the claim must hold if none of the fooling vertices is contained in Γ. This concludes the proof. ∎

Note that our construction in Distribution 3 has no edge between principal blocks, so Claim 4.3 further implies that solving an r-round instance requires to solve at least half of the principal $(r-1)$-round instances.

5 The Lower Bound for Maximal Independent Set

We prove the following theorem in this section. Theorem 1 is a straightforward corollary by an averaging argument, namely the easy direction of Yao's minimax principle [Yao77]. Note that by the third statement of Remark 4.2, $n_r \leqslant k^{20^{r+1}}$ so we know $k \geqslant n_r^{1/20^{r+1}}$.

Theorem 2. *For $r = o(\log k)$, any r-round protocol for computing a maximal independent set that communicates at most k bits per vertex in every round can only succeed with probability less than 0.1 over $\mathcal{D}^{(r)}_{\mathsf{MIS}}$.*

Our proof to Theorem 2 for r-round protocols in general is by repeatedly applying the following round elimination lemma.

Lemma 5.1 (Round Elimination)**.** *For $r = o(\log k)$ and $\delta \in [0,1]$, if there exists an r-round protocol for computing a maximal independent set that communicates at most k bits per vertex in every round and succeeds with probability δ over $\mathcal{D}^{(r)}_{\mathsf{MIS}}$, then there also exists an $(r-1)$-round protocol for computing a maximal independent set that communicates at most k bits per vertex in every round and succeeds with probability $\delta/2 - 1/n_{r-1}$ over $\mathcal{D}^{(r-1)}_{\mathsf{MIS}}$.*

Before proving Lemma 5.1, which is the main part of this section, we first show it easily implies Theorem 2.

Proof of Theorem 2. Suppose for the purpose of contradiction that there exists an r-round protocol that communicates at most k bits per vertex in every round and that has success probability 0.1 over $\mathcal{D}^{(r)}_{\mathsf{MIS}}$. Applying Lemma 5.1 for r times, we obtain a 0-round protocol having success probability

$$\begin{aligned}
\frac{0.1}{2^r} - \sum_{t\in[r]} \frac{1}{2^{t-1}\cdot n_{t-1}} &\geqslant \frac{0.1}{2^r} - \frac{1}{n_0}\cdot\sum_{t\in[r]}\frac{1}{2^{t-1}} && \text{(as } n_{t-1} \text{ is increasing)}\\
&\geqslant \frac{0.1}{2^r} - \frac{2}{n_0}\\
&= \frac{1}{k^{o(1)}},
\end{aligned}$$

over $\mathcal{D}^{(0)}_{\mathsf{MIS}}$, where the last step follows from the assumption $r = o(\log k)$. Recall that $n_0 = 2k$ so the second term above is $\Theta(1/k)$ and can be ignored. However, the existence of such a 0-round protocol contradicts the lower bound of Lemma 4.1. This concludes the proof of the theorem. ∎

We prove Lemma 5.1 in the rest of this section. To this end, fix any r-round protocol π on n_r vertices that communicates at most k bits per vertex in every round and succeeds with probability δ over $\mathcal{D}^{(r)}_{\mathsf{MIS}}$. By an averaging argument, we may assume without loss of generality π is deterministic. Before proceeding to the actual proof, let us first define the following random variables with respect to π when its input is drawn from $\mathcal{D}^{(r)}_{\mathsf{MIS}}$.

- Σ: the random permutation σ over n_r vertices;
- B_i: the edges within the i-th principal block $\Sigma(\mathcal{P}_i)$ for $i \in [p_r]$;
- T_i: the edges between the i-th principal block $\Sigma(\mathcal{P}_i)$ and all fooling vertices $\Sigma(\mathcal{F}(U \cup V))$ for $i \in [p_r]$;

- $\mathsf{G}_i := (\mathsf{B}_i, \mathsf{T}_i)$: all edges incident to the i-th principal block $\Sigma(\mathcal{P}_i)$ for $i \in [p_r]$ (there is no edge between principal blocks by our construction in Distribution 3);
- $\mathsf{G} := (\mathsf{G}_1, \ldots, \mathsf{G}_{p_r})$: the set of all sampled edges (the edges between $\Sigma(\mathcal{F}(U))$ and $\Sigma(\mathcal{F}(V))$ are always present and thus not included here; there is no other edge between fooling blocks by our construction in Distribution 3);
- $\mathsf{M}_{P,i}^{(t)}$: the messages sent by the i-th principal block $\Sigma(\mathcal{P}_i)$ in the t-th round for $i \in [p_r]$ and $t \in [r]$;
- $\mathsf{M}_{P}^{(t)} := (\mathsf{M}_{P,1}^{(t)}, \ldots, \mathsf{M}_{P,p_r}^{(t)})$: the messages sent by all principal blocks in the t-th round for $t \in [r]$;
- $\mathsf{M}_{F}^{(t)}$: the messages sent by all fooling blocks in the t-th round for $t \in [r]$;
- $\mathsf{M}^{(t)} := (\mathsf{M}_{P}^{(t)}, \mathsf{M}_{F}^{(t)})$: all messages sent in the t-th round for $t \in [r]$.

Note that $\mathsf{M}^{(<t)}$ is exactly the content of the blackboard at the beginning of the t-th round. For any vertex $u \in \Sigma(\mathcal{P}_i)$, we further define $\mathsf{B}_i(u)$ as the subset of B_i representing only edges incident to u. $\mathsf{T}_i(u), \mathsf{G}_i(u)$ are similarly defined. Let $\mathsf{M}_{P,i}^{(t)}(u)$ be the message sent by $u \in \Sigma(\mathcal{P}_i)$ in the t-th round. Fix any Σ, $\mathsf{M}_{P,i}^{(t)}$ is a function of $\mathsf{M}^{(<t)}$ and G_i while $\mathsf{M}_{P,i}^{(t)}(u)$ is only a function of $\mathsf{M}^{(<t)}$ and $\mathsf{G}_i(u)$. After all r rounds of communication, the referee has to output the solution based solely on $\mathsf{M}^{(\leqslant r)}$ since we have assumed π to be deterministic.

Algorithm 1 presents the complete simulation protocol for round elimination, formalizing our discussion in Section 3. At a high level, we construct the following $(r-1)$-round (randomized) protocols $\tau_1, \ldots, \tau_{p_r}$ on n_{r-1} vertices that are essentially simulating π on n_r vertices. At the end of the proof, we will show there exists some index $i^* \in [p_r]$ such that τ_{i*} simulates π sufficiently well and is able to solve instances of $\mathcal{D}_{\mathsf{MIS}}^{(r-1)}$ with the desired probability.

Algorithm 1. The $(r-1)$-round protocol τ_i, for any *fixed* $i \in [p_r]$, simulating π for computing a maximal independent set.

1. Sample Σ uniformly at random using public randomness. Identify the vertices of τ_i with $\Sigma(\mathcal{P}_i)$ in π, and with a slight abuse of notation, any vertex u of τ_i is used interchangeably with its counterpart in $\Sigma(\mathcal{P}_i)$[a]. In addition, each vertex u of τ_i identifies its input given in τ_i with $\mathsf{B}_i(u)$ in π.
2. Do the following *without* any communication:
   (a) Sample $\mathsf{M}_{P,i}^{(1)}$, conditioned on Σ, using public randomness.
   (b) For each vertex u of τ_i, independently sample $\mathsf{T}_i(u)$, conditioned on $\mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}, \Sigma$, using *private* randomness.
   (c) Sample $\mathsf{M}_{P,-i}^{(1)}$, conditioned on $\mathsf{M}_{P,i}^{(1)}, \Sigma$, using public randomness.
   (d) Sample $\mathsf{M}_{F}^{(1)}$, conditioned on $\mathsf{M}_{P}^{(1)}, \Sigma$, using public randomness.
3. For every $t \in [2, r]$, do the following with one round of communication:
   (a) For each vertex u of τ_i, generate and broadcast $\mathsf{M}_{P,i}^{(t)}(u)$ as in π, based on $\mathsf{G}_i(u), \mathsf{M}^{(<t)}, \Sigma$.
   (b) Sample $\mathsf{M}_{P,-i}^{(t)}$, conditioned on $\mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma$, using public randomness.
   (c) Sample $\mathsf{M}_{F}^{(t)}$, conditioned on $\mathsf{M}^{(<t)}, \mathsf{M}_{P}^{(t)}, \Sigma$, using public randomness.
4. Let Γ be the output of the referee of π when given $\mathsf{M}^{(\leqslant r)}$. The referee of τ_i finally outputs $\Gamma \cap \Sigma(\mathcal{P}_i)$.

[a] At a high level, the vertices of τ_i are going to play the role of the i-th principal block in π and jointly simulate all other vertices of π using public randomness. That is, they proceed with π as if they were $\Sigma(\mathcal{P}_i)$.

As discussed in Section 3, to prove Lemma 5.1, our goal is to find an index $i^* \in [p_r]$ such that τ_{i*} simulates π almost perfectly. Concretely, it is sufficient to have the distribution of the final blackboard $\mathsf{M}^{(\leqslant r)}$ sampled by τ_{i*} be close to the true distribution generated by π. These two distributions would be identical if τ_{i*} *were* able to do the sampling process in Algorithm 1 such that each random variable newly sampled in any step is drawn conditioned on all previously sampled random variables. Unfortunately, this is impossible because B_{i*} is the input to τ_{i*}, which is not publicly known by all vertices: each vertex is only given the edges incident to it, essentially its "local view". What τ_{i*} can actually do is to sample new random variables conditioned on all random variables previously sampled using *public* randomness. The hope is that the joint distribution of all sampled random variables is not affected by much as τ_{i*} drops conditioning on B_{i*} as well as all random variables sampled using private randomness, namely T_{i*} in Algorithm 1. In fact, we will show this is true *on average* over all possible $i \in [p_r]$, and thus it is sufficient to pick the best index as i^*.

Table 1 makes a detailed comparison between the sampled distribution by τ_i and the true distribution in π. Note that B_i is given as the input to τ_i and by our construction in Distribution 3, it has exactly the same distribution as any principal block in π.

Conditioning r.v. / Sampled r.v.	Sampled distribution by τ_i	True distribution in π
B_i		
$\mathsf{M}_{P,i}^{(1)}$	Σ	B_i, Σ
$\mathsf{T}_i(u)$	$\mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}, \Sigma$	
T_i		$\mathsf{B}_i, \mathsf{M}_{P,i}^{(1)}, \Sigma$
$\mathsf{M}_{P,-i}^{(t)}$	$\mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma$	$\mathsf{G}_i, \mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma$
$\mathsf{M}_{F}^{(t)}$	$\mathsf{M}^{(<t)}, \mathsf{M}_{P}^{(t)}, \Sigma$	$\mathsf{G}_i, \mathsf{M}^{(<t)}, \mathsf{M}_{P}^{(t)}, \Sigma$

Table 1: Sampled distribution by τ_i v.s. True distribution in π

Remark 5.2. *A couple of remarks about Algorithm 1 and Table 1.*

1. *In Algorithm 1, $\mathsf{T}_i(u)$ is sampled independently by each vertex u using private randomness. This means the sampled T_i in fact follows a product distribution, conditioned on $\mathsf{B}_i, \mathsf{M}_{P,i}^{(1)}, \Sigma$. We will prove in Lemma 5.3 that this generates precisely the true distribution of T_i.*
2. *Recall that π is assumed to be deterministic so $\mathsf{M}_{P,i}^{(t)}$ is a function of $\mathsf{G}_i, \mathsf{M}^{(<t)}, \Sigma$. More specifically, $\mathsf{M}_{P,i}^{(t)}(u)$ is a function of $\mathsf{G}_i(u), \mathsf{M}^{(<t)}, \Sigma$ for each vertex u. τ_i indeed generates them using this approach as shown in Algorithm 1.*

Our next step is to prove every pair of the conditional distributions are close. At the end, we will put them together to show the final blackboards $\mathsf{M}^{(\leqslant r)}$ are also close. The comparison between the conditional distributions is split into three parts. Lemma 5.3 proves we can indeed sample the first round message $\mathsf{M}_{P,i}^{(1)}$ publicly and thus eliminate the first round of communication. The first statement of Remark 5.2 is made precise by Lemma 5.4. A similar conditional decomposition lemma is established in [ANRW15]. Lemma 5.5 formalizes the intuition of directly sampling the messages of all other blocks.

With a slight abuse of notation, we may also use $< u$ to denote all vertices $v < u$ in $\Sigma(\mathcal{P}_i)$ for some $i \in [p_r]$ that can be inferred from context. Similarly, $-u$ is used as a shorthand for $\Sigma(\mathcal{P}_i) \setminus \{u\}$.

Lemma 5.3. *Let $\epsilon_r = r/(k^4 \cdot n_{r-1}^2)$. For each $i \in [p_r]$,*

$$\mathbb{I}(\mathsf{M}_{P,i}^{(1)}; \mathsf{B}_i \mid \Sigma) \leqslant \epsilon_r.$$

Proof. Assume without loss of generality that $i \in [\hat{p}_r]$. That is, we only consider the principal blocks on the side of U in Distribution 3. For any $u \in \Sigma(\mathcal{P}_i)$, $\mathsf{T}_i(u)$ is independent of $\mathsf{T}_i(< u)$ given B_i, Σ by our construction in Distribution 3. This implies $\mathsf{G}_i(u)$ and $\mathsf{G}_i(< u)$ are independent conditioned on B_i, Σ. Using the second statement of Remark 5.2, we know $\mathsf{M}_{P,i}^{(1)}(u)$ and $\mathsf{M}_{P,i}^{(1)}(< u)$ are independent conditioned on B_i, Σ as well by the data processing inequality (Fact A.1-(7)). Then we can get

$$\begin{aligned}
&\mathbb{I}(\mathsf{M}_{P,i}^{(1)}; \mathsf{B}_i \mid \Sigma) \\
=& \sum_{u \in \Sigma(\mathcal{P}_i)} \mathbb{I}(\mathsf{M}_{P,i}^{(1)}(u); \mathsf{B}_i \mid \mathsf{M}_{P,i}^{(1)}(< u), \Sigma) \quad \text{(by the chain rule of mutual information (Fact A.1-(6)))}
\end{aligned}$$

$$\leqslant \sum_{u\in\Sigma(\mathcal{P}_i)} \mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{B}_i \mid \Sigma). \qquad \text{(as } \mathsf{M}^{(1)}_{P,i}(u) \perp \mathsf{M}^{(1)}_{P,i}(<u) \mid \mathsf{B}_i,\Sigma \text{ and by Proposition A.3)}$$

Since the vertices are symmetric, it suffices to show an individual term above is upper bounded by ϵ_r/n_{r-1} as $\Sigma(\mathcal{P}_i)$ contains n_{r-1} vertices. So we fix the vertex u in the following.

One crucial observation is that u is simultaneously participating in $\hat{f}_r+1$ independent instances drawn from $\mathcal{D}^{(r-1)}_{\mathsf{MIS}}$: the principal one with $\Sigma(\mathcal{P}_i)\backslash\{u\}$ and $\hat{f}_r$ fooling ones with each of the fooling blocks $\Sigma(\mathcal{F}_j)$ for $j\in[\hat{f}_r]$. Collectively these $\hat{f}_r+1$ instances constitute $\mathsf{G}_i(u)$. Fix an ordering Λ for subsets of vertices with size $n_{r-1}-1$. Let $\mathsf{S}_1,\ldots,\mathsf{S}_{\hat{f}_r+1}$ denote these $\hat{f}_r+1$ instances in the order consistent with Λ and $\mathsf{S}_{<j}=(\mathsf{S}_1,\ldots,\mathsf{S}_{j-1})$ for $j\in[\hat{f}_r+1]$. Note that $\mathsf{G}_i(u)=(\mathsf{S}_1,\ldots,\mathsf{S}_{\hat{f}_r+1})$. Define Z to be the set of all these $\hat{f}_r+1$ *blocks* of vertices, i.e. $\mathsf{Z}:=\{\Sigma(\mathcal{P}_i)\backslash\{u\}\}\cup\left\{\Sigma(\mathcal{F}_j)\mid j\in[\hat{f}_r]\right\}$. We emphasize that Z records the *partition* of all u's possible neighbors into $\hat{f}_r+1$ blocks, but not which one corresponds to the principal instance. This is important because S_j are mutually independent conditioned on Z whereas they are not necessarily independent conditioned only on the set of all u's possible neighbors. Let W be the rank of the principal block among Z according to the order defined by Λ, so $\mathsf{S}_{\mathsf{W}}=\mathsf{B}_i$. Given Z, W is uniformly distributed over $[\hat{f}_r+1]$ because Σ is a uniformly random permutation. Intuitively, u cannot distinguish between all $\hat{f}_r+1$ instances by itself, implying that $\mathsf{M}^{(1)}_{P,i}(u)$ should only reveal little information about the principal instance B_i. Formally, we have

$$\begin{aligned}
&\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{B}_i\mid\Sigma)\\
&=\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{S}_{\mathsf{W}}\mid\Sigma,\mathsf{Z},\mathsf{W}) && \text{(as } \mathsf{Z},\mathsf{W} \text{ are completely determined by } \Sigma \text{ for any fixed } i,u)\\
&\leqslant\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{S}_{\mathsf{W}}\mid\mathsf{Z},\mathsf{W}) && \text{(as } \mathsf{M}^{(1)}_{P,i}(u)\perp\Sigma\mid\mathsf{S}_{\mathsf{W}},\mathsf{Z},\mathsf{W} \text{ and by Proposition A.3)}\\
&=\sum_{j\in[\hat{f}_r+1]}\Pr(\mathsf{W}=j)\cdot\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{S}_j\mid\mathsf{Z},\mathsf{W}=j)\\
&=\frac{1}{\hat{f}_r+1}\cdot\sum_{j\in[\hat{f}_r+1]}\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{S}_j\mid\mathsf{Z}),
\end{aligned}$$

as the joint distribution of $(\mathsf{M}^{(1)}_{P,i}(u),\mathsf{S}_j,\mathsf{Z})$ is independent of the event $\mathsf{W}=j$. Continuing,

$$\begin{aligned}
&\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{B}_i\mid\Sigma)\\
&\leqslant\frac{1}{\hat{f}_r+1}\cdot\sum_{j\in[\hat{f}_r+1]}\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{S}_j\mid\mathsf{Z})\\
&\leqslant\frac{1}{\hat{f}_r+1}\cdot\sum_{j\in[\hat{f}_r+1]}\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{S}_j\mid\mathsf{S}_{<j},\mathsf{Z}) && \text{(as } \mathsf{S}_j\perp\mathsf{S}_{<j}\mid\mathsf{Z} \text{ and by Proposition A.2)}\\
&=\frac{1}{\hat{f}_r+1}\cdot\mathbb{I}(\mathsf{M}^{(1)}_{P,i}(u)\,;\mathsf{G}_i(u)\mid\mathsf{Z}) && \text{(by the chain rule of mutual information (Fact A.1-(6)))}\\
&\leqslant\frac{1}{\hat{f}_r}\cdot\mathbb{H}(\mathsf{M}^{(1)}_{P,i}(u)\mid\mathsf{Z})\\
&\qquad\text{(by the definition of mutual information and non-negativity of entropy (Fact A.1-(1)))}\\
&\leqslant\frac{1}{\hat{f}_r}\cdot\mathbb{H}(\mathsf{M}^{(1)}_{P,i}(u)) && \text{(as conditioning can only reduce entropy (Fact A.1-(3)))}
\end{aligned}$$

$$\leqslant \frac{k}{\hat{f}_r}. \qquad \text{(by the assumption on } \pi\text{'s communication and Fact A.1-(1))}$$

Plugging in $\hat{f}_r$ as defined in Distribution 3, we finally get the desired upper bound $k/\hat{f}_r = 1/(k^5 \cdot n_{r-1}^3) \leqslant \epsilon_r/n_{r-1}$. This concludes the proof by our argument at the beginning. ∎

Lemma 5.4. *For each $i \in [p_r]$, and fixed $\mathsf{B}_i, \mathsf{M}_{P,i}^{(1)}, \Sigma$,*

$$\operatorname{dist}(\mathsf{T}_i \mid \mathsf{B}_i, \mathsf{M}_{P,i}^{(1)}, \Sigma) \sim \bigtimes_{u \in \Sigma(\mathcal{P}_i)} \operatorname{dist}(\mathsf{T}_i(u) \mid \mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}, \Sigma).$$

Proof. Let $\overline{\mathsf{B}}_i(u)$ denote the subset of B_i representing edges not incident to u[6] for $u \in \Sigma(\mathcal{P}_i)$. So $\mathsf{B}_i = (\mathsf{B}_i(u), \overline{\mathsf{B}}_i(u))$. It suffices to show $\mathbb{I}(\mathsf{T}_i(u)\,; \mathsf{T}_i(-u), \overline{\mathsf{B}}_i(u) \mid \mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}, \Sigma) = 0$. Using the second statement of Remark 5.2, we have

$$\mathbb{I}(\mathsf{M}_{P,i}^{(1)}(u)\,; \mathsf{T}_i(-u), \overline{\mathsf{B}}_i(u) \mid \mathsf{T}_i(u), \mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}(-u), \Sigma) = 0, \tag{13}$$

because $\mathsf{M}_{P,i}^{(1)}(u)$ is completely determined by $\mathsf{G}_i(u) = (\mathsf{B}_i(u), \mathsf{T}_i(u)), \Sigma$. Similarly, we also have

$$\mathbb{I}(\mathsf{M}_{P,i}^{(1)}(-u)\,; \mathsf{T}_i(u) \mid \mathsf{T}_i(-u), \overline{\mathsf{B}}_i(u), \mathsf{B}_i(u), \Sigma) = 0, \tag{14}$$

because $\mathsf{M}_{P,i}^{(1)}(-u)$ is completely determined by $\mathsf{B}_i = (\mathsf{B}_i(u), \overline{\mathsf{B}}_i(u)), \mathsf{T}_i(-u), \Sigma$. Combining Eq (13) and (14), we then get

$$\begin{aligned}
&\mathbb{I}(\mathsf{T}_i(u)\,; \mathsf{T}_i(-u), \overline{\mathsf{B}}_i(u) \mid \mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}, \Sigma) \\
&= \mathbb{I}(\mathsf{T}_i(u)\,; \mathsf{T}_i(-u), \overline{\mathsf{B}}_i(u) \mid \mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}(u), \mathsf{M}_{P,i}^{(1)}(-u), \Sigma) \\
&\leqslant \mathbb{I}(\mathsf{T}_i(u)\,; \mathsf{T}_i(-u), \overline{\mathsf{B}}_i(u) \mid \mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}(-u), \Sigma) && \text{(by Eq (13) and Proposition A.3)} \\
&\leqslant \mathbb{I}(\mathsf{T}_i(u)\,; \mathsf{T}_i(-u), \overline{\mathsf{B}}_i(u) \mid \mathsf{B}_i(u), \Sigma) && \text{(by Eq (14) and Proposition A.3)} \\
&= 0,
\end{aligned}$$

by our construction in Distribution 3. ∎

Lemmas 5.3 and 5.4 together ensure Algorithm 1 simulates the input and the first round of communication with little bias. Building upon this, Lemma 5.5 takes care of all remaining rounds. This is accomplished using the novel idea of non-simultaneous simulation as discussed in Section 3.2.3.

Lemma 5.5. *Let $\epsilon_r = 1/(k^4 \cdot n_{r-1}^2)$. For each $t \in [r]$,*

1. $\mathbb{E}_{i \in [p_r]}\, \mathbb{I}(\mathsf{M}_{P,-i}^{(t)}\,; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma) \leqslant \epsilon_r$.
2. $\mathbb{E}_{i \in [p_r]}\, \mathbb{I}(\mathsf{M}_{F}^{(t)}\,; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_{P}^{(t)}, \Sigma) \leqslant \epsilon_r$.

Before going into the actual proof of Lemma 5.5, we first present the following technical claim. Roughly, it shows what is revealed about G as a whole is no more than the sum of the information revealed about individual G_i by each principal block itself, justifying Section 3.2.4.

[6] Note that $\overline{\mathsf{B}}_i(u) \neq \mathsf{B}_i(-u)$ since each edge (u, v) appears in both $\mathsf{B}_i(u)$ and $\mathsf{B}_i(v)$.

Claim 5.6. *For each $t \in [r]$,*

$$\mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) \leqslant \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_P^{(<t)}, \mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_F^{(<t)}, \Sigma). \tag{15}$$

Proof. The proof is by rewriting both sides of the above inequality using the chain rule of mutual information (Fact A.1-(6)) for multiple times. For the left hand side of Eq (15), we have

$$\begin{aligned}
&\mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) \\
=& \sum_{t'\in[t]} \mathbb{I}(\mathsf{M}_P^{(t')}; \mathsf{G} \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) \\
=& \sum_{t'\in[t]} \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G} \mid \mathsf{M}_{P,<i}^{(t')}, \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) \\
\leqslant& \sum_{t'\in[t]} \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G} \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma), && \text{(by Proposition A.3)}
\end{aligned}$$

where we use the observation that $\mathsf{M}_{P,i}^{(t')}$ is fully determined by G, Σ, as π is deterministic, and thus conditionally independent of $\mathsf{M}_{P,<i}^{(t')}$. Continuing,

$$\begin{aligned}
&\mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) \\
\leqslant& \sum_{t'\in[t]} \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G} \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) \\
=& \sum_{t'\in[t]} \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G}_i, \mathsf{G}_{-i} \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) \\
=& \sum_{t'\in[t]} \sum_{i\in[p_r]} \left[\mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) + \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G}_{-i} \mid \mathsf{G}_i, \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma)\right] \\
=& \sum_{t'\in[t]} \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma),
\end{aligned}$$

as $\mathsf{M}_{P,i}^{(t')}$ is fully determined by $\mathsf{G}_i, \mathsf{M}^{(<t')}, \Sigma$ using the second statement of Remark 5.2. The right hand side of Eq (15) can be bounded as follows.

$$\begin{aligned}
&\sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_P^{(<t)}, \mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_F^{(<t)}, \Sigma) \\
=& \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t)}, \mathsf{M}_F^{(<t)}, \Sigma) + \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_P^{(<t)}; \mathsf{G}_i \mid \mathsf{M}_F^{(<t)}, \Sigma) \\
=& \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t)}, \mathsf{M}_F^{(<t)}, \Sigma) + \sum_{i\in[p_r]} \sum_{t'\in[t-1]} \mathbb{I}(\mathsf{M}_P^{(t')}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) \\
=& \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t)}, \mathsf{M}_F^{(<t)}, \Sigma) + \sum_{i\in[p_r]} \sum_{t'\in[t-1]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}, \mathsf{M}_{P,-i}^{(t')}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) \\
\geqslant& \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t)}, \mathsf{M}_F^{(<t)}, \Sigma) + \sum_{i\in[p_r]} \sum_{t'\in[t-1]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma) \\
& \qquad \text{(by the non-negativity and chain rule of mutual information (Fact A.1-(6)))}
\end{aligned}$$

$$= \sum_{i\in[p_r]} \sum_{t'\in[t]} \mathbb{I}(\mathsf{M}_{P,i}^{(t')}; \mathsf{G}_i \mid \mathsf{M}_P^{(<t')}, \mathsf{M}_F^{(<t)}, \Sigma).$$

We finally reach the desired conclusion by putting the two sides together. ∎

Now we are ready to prove Lemma 5.5.

Proof of Lemma 5.5.

Proof of the first statement: Instead of bounding the expectation directly, for convenience, we are going to work with the following summation:

$$\begin{aligned}
&\sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,-i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma) \\
&= \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}; \mathsf{G}_i \mid \Sigma) - \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \Sigma), \qquad (16)
\end{aligned}$$

as $\mathsf{M}_P^{(t)} = (\mathsf{M}_{P,i}^{(t)}, \mathsf{M}_{P,-i}^{(t)})$ and by the chain rule of mutual information (Fact A.1-(6)). The first term above can be upper bounded as

$$\begin{aligned}
&\sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}; \mathsf{G}_i \mid \Sigma) \\
&\leqslant \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}; \mathsf{G}_i \mid \mathsf{G}_{<i}, \Sigma) && \text{(as } \mathsf{G}_i \perp \mathsf{G}_{<i} \mid \Sigma \text{ and by Proposition A.2)} \\
&= \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}; \mathsf{G} \mid \Sigma) && \text{(by the chain rule of mutual information (Fact A.1-(6)))} \\
&= \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}, \mathsf{M}_F^{(<t)}; \mathsf{G} \mid \Sigma) \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma). && \text{(by the chain-rule of mutual information (Fact A.1-(6)))}
\end{aligned}$$

Plugging into Eq (16), we have

$$\begin{aligned}
&\sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,-i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma) \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) - \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \Sigma) \\
&= \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) - \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_P^{(<t)}, \mathsf{M}_F^{(<t)}, \mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \Sigma) \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) - \sum_{i\in[p_r]} \mathbb{I}(\mathsf{M}_P^{(<t)}, \mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_F^{(<t)}, \Sigma) \\
&\qquad \text{(by the non-negativity and chain rule of mutual information (Fact A.1-(6)))} \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(<t)}) && \text{(by Claim 5.6)} \\
&\leqslant k \cdot (n_{r-1} - 1) \cdot f_r \cdot (t-1), && \text{(by the subadditivity of entropy (Fact A.1-(4)))}
\end{aligned}$$

since there are f_r fooling blocks of $n_{r-1} - 1$ fooling vertices each, and every fooling vertex communicates at most k bits in each of the first $t-1$ rounds. Going back to the expectation, we finally get

$$\mathop{\mathbb{E}}_{i\in[p_r]} \mathbb{I}(\mathsf{M}_{P,-i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma)$$

$$
\begin{aligned}
&= \frac{1}{p_r} \cdot \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}_{P,-i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma) \\
&\leqslant \frac{k \cdot n_{r-1} \cdot f_r \cdot r}{p_r} \\
&\leqslant \epsilon_r,
\end{aligned}
$$

by the assumption $r = o(\log k)$.

Proof of the second statement: The proof is quite similar to the first one. Our goal is still to the bound the following summation:

$$
\begin{aligned}
&\sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}_F^{(t)}; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}, \Sigma) \\
&= \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}^{(\leqslant t)}; \mathsf{G}_i \mid \Sigma) - \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}; \mathsf{G}_i \mid \Sigma), \qquad (17)
\end{aligned}
$$

as $\mathsf{M}^{(t)} = (\mathsf{M}_P^{(t)}, \mathsf{M}_F^{(t)})$ and by the chain rule of mutual information (Fact A.1-(6)). Again we bound the first term above as follows.

$$
\begin{aligned}
&\sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}^{(\leqslant t)}; \mathsf{G}_i \mid \Sigma) \\
&\leqslant \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}^{(\leqslant t)}; \mathsf{G}_i \mid \mathsf{G}_{<i}, \Sigma) && \text{(as } \mathsf{G}_i \perp \mathsf{G}_{<i} \mid \Sigma \text{ and by Proposition A.2)} \\
&= \mathbb{I}(\mathsf{M}^{(\leqslant t)}; \mathsf{G} \mid \Sigma) && \text{(by the chain rule of mutual information (Fact A.1-(6)))} \\
&= \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}, \mathsf{M}_F^{(<t)}, \mathsf{M}_F^{(t)}; \mathsf{G} \mid \Sigma) \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}, \mathsf{M}_F^{(<t)}; \mathsf{G} \mid \Sigma) && \text{(by the chain-rule of mutual information (Fact A.1-(6)))} \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(t)}) + \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma). \\
& && \text{(by the chain-rule of mutual information (Fact A.1-(6)))}
\end{aligned}
$$

Plugging into Eq (17), we have

$$
\begin{aligned}
&\sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}_F^{(t)}; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}, \Sigma) \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(t)}) + \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) - \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}; \mathsf{G}_i \mid \Sigma) \\
&= \mathbb{H}(\mathsf{M}_F^{(t)}) + \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) - \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}_P^{(<t)}, \mathsf{M}_F^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \mathsf{M}_{P,-i}^{(t)}; \mathsf{G}_i \mid \Sigma) \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(t)}) + \mathbb{H}(\mathsf{M}_F^{(<t)}) + \mathbb{I}(\mathsf{M}_P^{(\leqslant t)}; \mathsf{G} \mid \mathsf{M}_F^{(<t)}, \Sigma) - \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}_P^{(<t)}, \mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}_F^{(<t)}, \Sigma). \\
& && \text{(by the non-negativity and chain rule of mutual information (Fact A.1-(6)))} \\
&\leqslant \mathbb{H}(\mathsf{M}_F^{(t)}) + \mathbb{H}(\mathsf{M}_F^{(<t)}) && \text{(by Claim 5.6)} \\
&\leqslant k \cdot (n_{r-1} - 1) \cdot f_r \cdot t, && \text{(by the subadditivity of entropy (Fact A.1-(4)))}
\end{aligned}
$$

by counting the total communication of all fooling vertices. The desired upper bound on the expectation is derived similarly to the first statement. ∎

Technically, it is actually possible to prove the following similar to Claim 5.6:

$$\mathbb{I}(\mathsf{M}_P^{(t)}; \mathsf{G} \mid \mathsf{M}^{(<t)}, \Sigma) \leqslant \sum_{i \in [p_r]} \mathbb{I}(\mathsf{M}_{P,i}^{(t)}; \mathsf{G}_i \mid \mathsf{M}^{(<t)}, \Sigma),$$

which would justify the intuition we provided for Claim 5.6 even better. However, the proof of Lemma 5.5 is complicated by the fact that neither $\mathsf{G}_i \perp \mathsf{G}_{<i} \mid \mathsf{M}_P^{(\leqslant t)}, \Sigma$ nor $\mathsf{G}_i \perp \mathsf{G}_{<i} \mid \mathsf{M}^{(\leqslant t)}, \Sigma$ is true. In general $\mathsf{M}_P^{(t)}$ depends on $\mathsf{M}_F^{(<t)}$, which in turn is able to correlate G_i and G_{-i}. At the core of the proof of Lemma 5.5 is applying the chain rule of mutual information (Fact A.1-(6)) over all G_i. To have the chain rule go through in the correct direction, what we need is the conditional independence between all G_i. As a result, we are forced to rewrite the summation as in Eq (16) and (17) such that the conditional independence between all G_i hold, and then conduct a more careful analysis to bound the amount of correlation caused by the messages of fooling vertices. The current form of Claim 5.6 turns out to be more appropriate for this purpose.

Combining Lemmas 5.3 to 5.5, the following corollary follows directly from Pinsker's inequality (Fact A.8). It essentially captures our initial intuition that the final blackboard $\mathsf{M}^{(\leqslant r)}$ sampled by τ_i is close to the true distribution on average over all possible $i \in [p_r]$.

Corollary 5.7. *Let μ be the true distribution for $(\mathsf{G}, \mathsf{M}^{(\leqslant r)}, \Sigma)$ in π and for $i \in [p_r]$, μ_i be the marginal distribution of μ for $(\mathsf{G}_i, \mathsf{M}^{(\leqslant r)}, \Sigma)$. For $i \in [p_r]$, let ν_i be the distribution of $(\mathsf{G}_i, \mathsf{M}^{(\leqslant r)}, \Sigma)$ defined by*

$$\begin{aligned}\nu_i(\mathsf{G}_i, \mathsf{M}^{(\leqslant r)}, \Sigma) := \mu(\mathsf{B}_i, \Sigma) \cdot \mu(\mathsf{M}_{P,i}^{(1)} \mid \Sigma) \cdot \prod_{u \in \Sigma(\mathcal{P}_i)} \mu(\mathsf{T}_i(u) \mid \mathsf{B}_i(u), \mathsf{M}_{P,i}^{(1)}, \Sigma) \\ \cdot \prod_{t \in [r]} \mu(\mathsf{M}_{P,-i}^{(t)} \mid \mathsf{M}^{(<t)}, \mathsf{M}_{P,i}^{(t)}, \Sigma) \cdot \prod_{t \in [r]} \mu(\mathsf{M}_F^{(t)} \mid \mathsf{M}^{(<t)}, \mathsf{M}_P^{(t)}, \Sigma) \\ \cdot \prod_{t \in [2,r]} \prod_{u \in \Sigma(\mathcal{P}_i)} \mu(\mathsf{M}_{P,i}^{(t)}(u) \mid \mathsf{G}_i(u), \mathsf{M}^{(<t)}, \Sigma).\end{aligned}$$

It holds that

$$\mathop{\mathbb{E}}_{i \in [p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i \sim \mu} \|\mu_i(\mathsf{M}^{(\leqslant r)}, \Sigma \mid \mathsf{B}_i) - \nu_i(\mathsf{M}^{(\leqslant r)}, \Sigma \mid \mathsf{B}_i)\|_{\mathrm{tvd}} \leqslant \frac{1}{k \cdot n_{r-1}}.$$

Proof. Firstly, we convert the statements of Lemmas 5.3 and 5.5 to the language of total variation distance. For each $i \in [p_r]$, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathsf{B}_i, \Sigma) \sim \mu} \|\mu_i(\mathsf{M}_{P,i}^{(1)} \mid \mathsf{B}_i, \Sigma) - \nu_i(\mathsf{M}_{P,i}^{(1)} \mid \mathsf{B}_i, \Sigma)\|_{\mathrm{tvd}} \\ =\ &\mathop{\mathbb{E}}_{(\mathsf{B}_i, \Sigma) \sim \mu} \|\mu(\mathsf{M}_{P,i}^{(1)} \mid \mathsf{B}_i, \Sigma) - \mu(\mathsf{M}_{P,i}^{(1)} \mid \Sigma)\|_{\mathrm{tvd}} \\ \leqslant\ &\mathop{\mathbb{E}}_{(\mathsf{B}_i, \Sigma) \sim \mu} \sqrt{\mathbb{D}(\mu(\mathsf{M}_{P,i}^{(1)} \mid \mathsf{B}_i, \Sigma) \mid\mid \mu(\mathsf{M}_{P,i}^{(1)} \mid \Sigma))} && \text{(by Pinsker's inequality (Fact A.8))} \\ \leqslant\ &\sqrt{\mathop{\mathbb{E}}_{(\mathsf{B}_i, \Sigma) \sim \mu} \mathbb{D}(\mu(\mathsf{M}_{P,i}^{(1)} \mid \mathsf{B}_i, \Sigma) \mid\mid \mu(\mathsf{M}_{P,i}^{(1)} \mid \Sigma))} && \text{(by the concavity of } \sqrt{\cdot}) \\ =\ &\sqrt{\mathbb{I}(\mathsf{M}_{P,i}^{(1)}; \mathsf{B}_i \mid \Sigma)} && \text{(by Fact A.4)} \\ \leqslant\ &\epsilon_r^{1/2}. && \text{(by Lemma 5.3)}\end{aligned}$$

Meanwhile, for each $t \in [r]$, we can get

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{(\mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma)\sim\mu} \|\mu_i(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma) - \nu_i(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma)\|_{\mathrm{tvd}} \\
=\ &\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{(\mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma)\sim\mu} \|\mu(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma) - \mu(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma)\|_{\mathrm{tvd}} \\
\leqslant\ &\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{(\mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma)\sim\mu} \sqrt{\mathbb{D}(\mu(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma) \mid\mid \mu(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma))} \\
&\qquad\qquad \text{(by Pinsker's inequality (Fact A.8))} \\
\leqslant\ &\sqrt{\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{(\mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma)\sim\mu} \mathbb{D}(\mu(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma) \mid\mid \mu(\mathsf{M}^{(t)}_{P,-i} \mid \mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma))} \\
&\qquad\qquad \text{(by the concavity of } \sqrt{\cdot}) \\
=\ &\sqrt{\mathop{\mathbb{E}}_{i\in[p_r]} \mathbb{I}(\mathsf{M}^{(t)}_{P,-i}\,;\mathsf{G}_i \mid \mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P,i},\Sigma)} \qquad \text{(by Fact A.4)} \\
\leqslant\ &\epsilon_r^{1/2}, \qquad \text{(by Lemma 5.5)}
\end{aligned}$$

and similarly

$$\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{(\mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P},\Sigma)\sim\mu} \|\mu_i(\mathsf{M}^{(t)}_{F} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P},\Sigma) - \nu_i(\mathsf{M}^{(t)}_{F} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\mathsf{M}^{(t)}_{P},\Sigma)\|_{\mathrm{tvd}} \leqslant \epsilon_r^{1/2}.$$

We also trivially have

$$\mathop{\mathbb{E}}_{(\mathsf{B}_i,\mathsf{M}^{(1)}_{P,i},\Sigma)\sim\mu} \|\mu_i(\mathsf{T}_i \mid \mathsf{B}_i,\mathsf{M}^{(1)}_{P,i},\Sigma) - \nu_i(\mathsf{T}_i \mid \mathsf{B}_i,\mathsf{M}^{(1)}_{P,i},\Sigma)\|_{\mathrm{tvd}} = 0,$$

by Lemma 5.4 for each $i \in [p_r]$, and

$$\mathop{\mathbb{E}}_{(\mathsf{G}_i,\mathsf{M}^{(<t)},\Sigma)\sim\mu} \|\mu_i(\mathsf{M}^{(t)}_{P,i} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\Sigma) - \nu_i(\mathsf{M}^{(t)}_{P,i} \mid \mathsf{G}_i,\mathsf{M}^{(<t)},\Sigma)\|_{\mathrm{tvd}} = 0,$$

by the second statement of Remark 5.2 for each $t \in [2, r]$. Additionally observe that

$$\mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu} \|\mu_i(\Sigma \mid \mathsf{B}_i) - \nu_i(\Sigma \mid \mathsf{B}_i)\|_{\mathrm{tvd}} = 0,$$

since Σ is a uniformly random permutation drawn independent of B_i. Combining all these conditional distributions using the chain rule of total variation distance (Fact A.6), it holds that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu} \|\mu_i(\mathsf{M}^{(\leqslant r)},\mathsf{T}_i,\Sigma \mid \mathsf{B}_i) - \nu_i(\mathsf{M}^{(\leqslant r)},\mathsf{T}_i,\Sigma \mid \mathsf{B}_i)\|_{\mathrm{tvd}} \\
\leqslant\ &\epsilon_r^{1/2} \cdot (2r+1) \\
=\ &\frac{(2r+1)}{k^2 \cdot n_{r-1}} \\
\leqslant\ &\frac{1}{k \cdot n_{r-1}},
\end{aligned}$$

by the linearity of expectation and the assumption $r = o(\log k)$. This concludes the proof as marginalization can never increase total variation distance (Fact A.7). ∎

In Corollary 5.7, note that μ and μ_i are the true distributions in π while ν_i is the distribution sampled by τ_i. We conclude this section by finishing the proof of Lemma 5.1.

Proof of Lemma 5.1. For each $i \in [p_r]$, define $\mathsf{O}_i^{\pi} \in \{0,1\}$ to be 1 if and only if the referee of π outputs a valid maximal independent set Γ for an r-round instance such that $\Gamma \cap \Sigma(\mathcal{P}_i)$ is also a valid maximal independent set for the induced subgraph on $\Sigma(\mathcal{P}_i)$. Also define $\mathsf{O}_i^{\tau} \in \{0,1\}$ to be 1 if and only if the referee of τ_i outputs a valid maximal independent set for an $(r-1)$-round instance. Recall that the referee of π is a deterministic function of $\mathsf{M}^{(\leqslant r)}$, so for each $i \in [p_r]$, the referee of τ_i is a deterministic function of $\mathsf{M}^{(\leqslant r)}$ and Σ, by Algorithm 1.

Firstly imagine the idealized situation where τ_i *were* able to sample $\mathsf{M}^{(\leqslant r)}, \Sigma \mid \mathsf{B}_i$ precisely following μ_i. Since the marginal distribution for each principal instance in $\mathcal{D}_{\mathsf{MIS}}^{(r)}$ is the same as $\mathcal{D}_{\mathsf{MIS}}^{(r-1)}$, by the linearity of expectation we get

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \Pr_{(\mathsf{M}^{(\leqslant r)},\Sigma|\mathsf{B}_i)\sim\mu_i} (\mathsf{O}_i^{\tau} = 1) \\
=\ &\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu} \Pr_{(\mathsf{M}^{(\leqslant r)},\Sigma|\mathsf{B}_i)\sim\mu} (\mathsf{O}_i^{\tau} = 1) && (\text{as } \mu_i(\mathsf{M}^{(\leqslant r)}, \mathsf{B}_i, \Sigma) = \mu(\mathsf{M}^{(\leqslant r)}, \mathsf{B}_i, \Sigma)) \\
=\ &\mathop{\mathbb{E}}_{i\in[p_r]} \Pr_{(\mathsf{M}^{(\leqslant r)},\mathsf{G},\Sigma)\sim\mu} (\mathsf{O}_i^{\pi} = 1) && (\text{by Algorithm 1}) \\
=\ &\mathop{\mathbb{E}}_{(\mathsf{G},\Sigma)\sim\mu} \Pr_{i\in[p_r]} (\mathsf{O}_i^{\pi} = 1) && (\text{as } \mathsf{M}^{(\leqslant r)} \text{ is fully determined by } \mathsf{G}, \Sigma) \\
\geqslant\ &\delta/2, && (18)
\end{aligned}$$

because π succeeds with probability δ by assumption, and conditioned on this event, at least half of the principal instances are solved by Claim 4.3. Now consider the real success probability of τ_i over ν_i. By Fact A.5, we have for each $i \in [p_r]$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathsf{B}_i\sim\nu_i} \Pr_{(\mathsf{M}^{(\leqslant r)},\Sigma|\mathsf{B}_i)\sim\nu_i} (\mathsf{O}_i^{\tau} = 1) \\
=\ &\mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \Pr_{(\mathsf{M}^{(\leqslant r)},\Sigma|\mathsf{B}_i)\sim\nu_i} (\mathsf{O}_i^{\tau} = 1) && (\text{as } \nu_i(\mathsf{B}_i) = \mu_i(\mathsf{B}_i)) \\
\geqslant\ &\mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \Pr_{(\mathsf{M}^{(\leqslant r)},\Sigma|\mathsf{B}_i)\sim\mu_i} (\mathsf{O}_i^{\tau} = 1) - \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \|\mu_i(\mathsf{M}^{(\leqslant r)}, \Sigma \mid \mathsf{B}_i) - \nu_i(\mathsf{M}^{(\leqslant r)}, \Sigma \mid \mathsf{B}_i)\|_{\text{tvd}}. && (19)
\end{aligned}$$

Combining Corollary 5.7 with Eq (18) and (19), we finally get

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\nu_i} \Pr_{(\mathsf{M}^{(\leqslant r)},\Sigma|\mathsf{B}_i)\sim\nu_i} (\mathsf{O}_i^{\tau} = 1) \\
\geqslant\ &\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \Pr_{(\mathsf{M}^{(\leqslant r)},\Sigma|\mathsf{B}_i)\sim\mu_i} (\mathsf{O}_i^{\tau} = 1) - \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \|\mu_i(\mathsf{M}^{(\leqslant r)}, \Sigma \mid \mathsf{B}_i) - \nu_i(\mathsf{M}^{(\leqslant r)}, \Sigma \mid \mathsf{B}_i)\|_{\text{tvd}} \\
\geqslant\ &\delta/2 - 1/n_{r-1},
\end{aligned}$$

as desired. Picking the index $i^* \in [p_r]$ that maximizes the success probability of τ_{i^*} concludes the proof. ∎

6 The Lower Bound for Approximate Bipartite Matching

In this section we adapt the techniques for maximal independent set to prove the following formal version of Result 2.

Theorem 3 (Result 2, formal)**.** *For $r \geqslant 0$ and any r-round multi-party protocol (deterministic or randomized) in the shared blackboard model for computing a maximal matching or any constant factor approximation to maximum matching on n-vertex (bipartite) graphs, there must exist* some *vertex communicating at least $\Omega(n^{1/20^{r+1}})$ bits in* some *round.*

Intuitively, Distribution 3 and Algorithm 1 make little use of any property specific to independent sets so most of the lemmas hold for matchings as well. For convenience, we first make minor adjustment to the hard distributions in Section 6.1 to better fit the need of approximation, and then present the lower bound proof for approximate matching for *general* graphs in Section 6.2. It is worth noticing that our constructed instances may not be bipartite in general. Fortunately, Section 6.3 gives a simple reduction to the bipartite case, concluding the proof of Theorem 3.

6.1 A Hard Distribution for Approximate Matching

We use the following base case for approximate matching. The idea is to have maximum matchings of a fixed size that remain hard to approximate. This will help simplify the calculation in later proofs a lot.

Distribution 4. The hard distribution $\mathcal{D}^{(0)}_{\mathsf{MM}}$ for protocols computing an approximate matching without any communication.
Parameters: bandwidth k, number of vertices $n_0 = 2k$.

1. Let U and V be two disjoint sets of vertices, each of size k. Sample two vertices $u \in U, v \in V$ uniformly at random and independently.
2. Add an edge (u, v).
3. Return the graph G sampled above.

It is easy to see any graph G drawn from $\mathcal{D}^{(0)}_{\mathsf{MM}}$ always has a maximum matching of size 1. Recall that protocols for approximate matching are required to output a valid matching (though potentially containing non-existing edges), which is of size at most k. Since the chosen edge (u, v) is sampled uniformly at random from k^2 possibilities, no protocols can achieve an approximation ratio better than $k^2/k = k$ if no information is revealed by the vertices. This is summarized in the following lemma.

Lemma 6.1 (Base Case)**.** *Any 0-round protocol for computing an approximate matching has an approximation ratio no better than k over $\mathcal{D}^{(0)}_{\mathsf{MM}}$.*

The construction for r-round hard distributions $\mathcal{D}^{(r)}_{\mathsf{MM}}$ is almost the same as for $\mathcal{D}^{(r)}_{\mathsf{MIS}}$. In fact, it can be even simplified in the sense that the "half instances" are sufficient for the purpose of constructing a hard distribution. Concretely, we construct $\mathcal{D}^{(r)}_{\mathsf{MM}}$ recursively as follows.

Distribution 5. The hard distribution $\mathcal{D}^{(r)}_{\mathsf{MM}}$ for r-round protocols computing an approximate matching ($r \geqslant 1$).
Parameters: bandwidth k, number of fooling blocks $f_r = k^6 \cdot n_{r-1}^3$, number of principal blocks $p_r = k^6 \cdot n_{r-1}^3 \cdot f_r$, number of vertices $n_r = (n_{r-1} - 1) \cdot f_r + n_{r-1} \cdot p_r$, and vertex set V with $|V| = n_r$.

1. Partition V into disjoint sets of vertices $\mathcal{P}_1, \ldots, \mathcal{P}_{p_r}, \mathcal{F}_1, \ldots, \mathcal{F}_{f_r}$ such that $\forall i \in [p_r] : \ |\mathcal{P}_i| = n_{r-1}$ and $\forall j \in [f_r] : \ |\mathcal{F}_j| = n_{r-1} - 1$. Define $\mathcal{P}(V) := \bigcup_{i \in [p_r]} \mathcal{P}_i$ and $\mathcal{F}(V) := \bigcup_{j \in [f_r]} \mathcal{F}_j$.
2. For $i \in [p_r]$, sample an independent instance of $\mathcal{D}^{(r-1)}_{\mathsf{MM}}$ on $\mathcal{P}_i$.
3. For $u \in \mathcal{P}(V)$ and $j \in [f_r]$, sample an independent instance of $\mathcal{D}^{(r-1)}_{\mathsf{MM}}$ on $\mathcal{F}_j \cup \{u\}$ and only keep the edges adjacent to u (dropping all the edges between vertices in $\mathcal{F}_j$).
4. Let G' be the graph sampled above. Sample a uniformly random permutation σ over V and return $G = \sigma(G')$.

It is not hard to verify that $n_r \leqslant k^{20^{r+1}}$ still holds for $r \geqslant 0$. At a high level, the number of fooling vertices is rather small as $f_r \ll p_r$, so their contribution to the size of maximum matchings is limited. On the other hand, a vast majority of matching edges should come from within the principal blocks so a good approximation ratio for r-round instances implies good approximation ratios over all principal $(r-1)$-round instances on average. Claim 6.2 provides a useful lower bound on the size of maximum matchings for graphs drawn from $\mathcal{D}^{(r)}_{\mathsf{MM}}$. It will be used in the proof at the very end of this section.

Claim 6.2. *Let Γ be any valid maximum matching for a graph G drawn from $\mathcal{D}^{(r)}_{\mathsf{MM}}$. Then,*

$$|\Gamma| \geqslant \frac{n_r}{2k} \cdot \left(1 - \sum_{t \in [r]} \frac{f_t}{p_t}\right) \geqslant \frac{n_r}{4k}. \tag{20}$$

Proof. The base case of $r = 0$ holds trivially. For $r \geqslant 1$, we know by induction that any principal $(r-1)$-round instance has a maximum matching of size at least $\frac{n_{r-1}}{2k} \cdot (1 - \sum_{t \in [r-1]} f_t/p_t)$. Since all p_r principal blocks are disjoint by our construction in Distribution 5, we get

$$\begin{aligned}
|\Gamma| &\geqslant p_r \cdot \frac{n_{r-1}}{2k} \cdot \left(1 - \sum_{t \in [r-1]} \frac{f_t}{p_t}\right) \\
&\geqslant \frac{p_r}{f_r + p_r} \cdot \frac{n_r}{2k} \cdot \left(1 - \sum_{t \in [r-1]} \frac{f_t}{p_t}\right) && (\text{as } n_r \leqslant n_{r-1} \cdot (f_r + p_r)) \\
&\geqslant \left(1 - \frac{f_r}{p_r}\right) \cdot \frac{n_r}{2k} \cdot \left(1 - \sum_{t \in [r-1]} \frac{f_t}{p_t}\right) \\
&\geqslant \frac{n_r}{2k} \cdot \left(1 - \sum_{t \in [r]} \frac{f_t}{p_t}\right).
\end{aligned}$$

The last inequality of Eq (20) follows from the simple fact that

$$\sum_{t\in[r]} \frac{f_t}{p_t} = \sum_{t\in[r]} \frac{1}{k^6 \cdot n_{t-1}^3} \leqslant \frac{1}{2},$$

since $r = o(\log k)$ by assumption. This concludes the proof. ∎

6.2 Proof of the Lower Bound for Approximate Matching

The version of Theorem 3 for general graphs is a straightforward corollary of the following distributional lower bound by Yao's minimax principal [Yao77]. We point out that since any maximal matching is also a 2-approximate matching, it is sufficient to prove the hardness of approximate matching.

Theorem 4. *For $r = o(\log k)$, any r-round protocol for computing an approximate matching for general graphs that communicates at most k bits per vertex in every round has an approximation ratio no better than $\Omega(k)$ over $\mathcal{D}^{(r)}_{\mathsf{MM}}$.*

The proof of Theorem 4 is again via a round elimination lemma as shown in Lemma 6.3.

Lemma 6.3 (Round Elimination)**.** *For $r = o(\log k)$ and $\alpha = \omega(1/n_{r-1})$, if there exists an r-round protocol for computing an approximate matching that communicates at most k bits per vertex in every round and has an approximation ratio of α^{-1} over $\mathcal{D}^{(r)}_{\mathsf{MM}}$, then there also exists an $(r-1)$-round protocol for computing an approximate matching that communicates at most k bits per vertex in every round and has an approximation ratio of $(\alpha - C/n_{r-1})^{-1}$ over $\mathcal{D}^{(r-1)}_{\mathsf{MM}}$, for some universal constant $C > 0$.*

Before proving Lemma 6.3, which is the main part of this section, we first show it easily implies Theorem 4.

Proof of Theorem 4. Suppose for the purpose of contradiction that there exists an r-round protocol that communicates at most k bits per vertex and that has an approximation ratio of $\alpha^{-1} = o(k)$ over $\mathcal{D}^{(r)}_{\mathsf{MM}}$. Applying Lemma 6.3 for r times, we obtain a 0-round protocol having an approximation ratio of

$$\left(\alpha - C \cdot \sum_{t\in[r]} \frac{1}{n_{t-1}}\right)^{-1} \leqslant \left(\alpha - \frac{2C}{n_0}\right)^{-1} = o(k),$$

over $\mathcal{D}^{(0)}_{\mathsf{MM}}$, as n_{t-1} is doubly exponentially increasing, and $\alpha = \omega(1/k), n_0 = 2k$. However, the existence of such a 0-round protocol contradicts the lower bound of Lemma 6.1, concluding the proof. ∎

To prove Lemma 6.3, we use the same approach for simulation as in Algorithm 1. Fix a deterministic r-round protocol π on n_r vertices that communicates at most k bits per vertex in every round and has an approximation ratio of α^{-1} over $\mathcal{D}^{(r)}_{\mathsf{MM}}$. We define exactly the same set of random variables as in Section 5 and construct the $(r-1)$-round (randomized) protocols $\tau_1, \ldots, \tau_{p_r}$ on n_{r-1} vertices, which are identical to Algorithm 1 except for the processing of the final output. Specifically, let Γ be the output of the referee of π when given $\mathsf{M}^{(\leqslant r)}$. The referee of τ_i finally outputs $\Gamma \cap (\Sigma(\mathcal{P}_i) \times \Sigma(\mathcal{P}_i))$, namely the *edges* within $\Sigma(\mathcal{P}_i)$, for computing approximate matchings. See Algorithm 2 for a recap of the complete simulation protocol for round elimination.

Algorithm 2. The $(r-1)$-round protocol τ_i, for any *fixed* $i \in [p_r]$, simulating π for computing an approximate matching.

1. Sample $\mathsf{\Sigma}$ uniformly at random using public randomness. Identify the vertices of τ_i with $\mathsf{\Sigma}(\mathcal{P}_i)$ in π, and with a slight abuse of notation, any vertex u of τ_i is used interchangeably with its counterpart in $\mathsf{\Sigma}(\mathcal{P}_i)$. In addition, each vertex u of τ_i identifies its input given in τ_i with $\mathsf{B}_i(u)$ in π.
2. Do the following *without* any communication:
   (a) Sample $\mathsf{M}^{(1)}_{P,i}$, conditioned on $\mathsf{\Sigma}$, using public randomness.
   (b) For each vertex u of τ_i, independently sample $\mathsf{T}_i(u)$, conditioned on $\mathsf{B}_i(u), \mathsf{M}^{(1)}_{P,i}, \mathsf{\Sigma}$, using *private* randomness.
   (c) Sample $\mathsf{M}^{(1)}_{P,-i}$, conditioned on $\mathsf{M}^{(1)}_{P,i}, \mathsf{\Sigma}$, using public randomness.
   (d) Sample $\mathsf{M}^{(1)}_{F}$, conditioned on $\mathsf{M}^{(1)}_{P}, \mathsf{\Sigma}$, using public randomness.
3. For every $t \in [2, r]$, do the following with one round of communication:
   (a) For each vertex u of τ_i, generate and broadcast $\mathsf{M}^{(t)}_{P,i}(u)$ as in π, based on $\mathsf{G}_i(u), \mathsf{M}^{(<t)}, \mathsf{\Sigma}$.
   (b) Sample $\mathsf{M}^{(t)}_{P,-i}$, conditioned on $\mathsf{M}^{(<t)}, \mathsf{M}^{(t)}_{P,i}, \mathsf{\Sigma}$, using public randomness.
   (c) Sample $\mathsf{M}^{(t)}_{F}$, conditioned on $\mathsf{M}^{(<t)}, \mathsf{M}^{(t)}_{P}, \mathsf{\Sigma}$, using public randomness.
4. Let Γ be the output of the referee of π when given $\mathsf{M}^{(\leqslant r)}$. The referee of τ_i finally outputs $\Gamma \cap (\mathsf{\Sigma}(\mathcal{P}_i) \times \mathsf{\Sigma}(\mathcal{P}_i))$.

It is not hard to verify that Lemmas 5.3 to 5.5 and Corollary 5.7 also hold for approximate matching. In fact, each of them follows verbatim as the proofs work in a black-box way. Using all these results, we conclude this section with the proof of Lemma 6.3.

Proof of Lemma 6.3. For any r-round instance, let O be the size of its maximum matching and for each $i \in [p_r]$, O_i be the size of the maximum matching for the induced subgraph on $\mathsf{\Sigma}(\mathcal{P}_i)$. It always holds that

$$\mathsf{O} \geqslant \sum_{i \in [p_r]} \mathsf{O}_i, \tag{21}$$

since the union of maximum matchings for all principal $(r-1)$-round instances is always a valid matching for the r-round instance. Define O^{π} to be the number of valid edges in $\Gamma \cap E$, where E is the set of input edges of the r-round instance, and for each $i \in [p_r]$, O^{π}_i to be the number of valid edges in $\Gamma \cap E \cap (\mathsf{\Sigma}(\mathcal{P}_i) \times \mathsf{\Sigma}(\mathcal{P}_i))$. It holds that

$$\mathsf{O}^{\pi} \leqslant n_{r-1} \cdot f_r + \sum_{i \in [p_r]} \mathsf{O}^{\pi}_i, \tag{22}$$

because of the fact that the number of disjoint edges incident to fooling vertices is bounded by the total number of fooling vertices. (Recall that the output Γ of the referee of π must be a set of

disjoint edges.) Also define O_i^τ to be the number of valid edges (i.e. excluding non-existing edges) output by the referee of τ_i for an $(r-1)$-round instance. Note that the referee of π is a deterministic function of $\mathsf{M}^{(\leqslant r)}$ by assumption, so for each $i \in [p_r]$, the referee of τ_i is a deterministic function of $\mathsf{M}^{(\leqslant r)}$ and $\mathsf{\Sigma}$, by Algorithm 2.

Again imagine the idealized situation where τ_i *were* able to sample $\mathsf{M}^{(\leqslant r)}, \mathsf{\Sigma} \mid \mathsf{B}_i$ precisely following μ_i. By the linearity of expectation, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{\Sigma}\mid\mathsf{B}_i)\sim\mu_i} [\mathsf{O}_i^\tau] \\
&= \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{\Sigma}\mid\mathsf{B}_i)\sim\mu} [\mathsf{O}_i^\tau] && (\text{as } \mu_i(\mathsf{M}^{(\leqslant r)}, \mathsf{B}_i, \mathsf{\Sigma}) = \mu(\mathsf{M}^{(\leqslant r)}, \mathsf{B}_i, \mathsf{\Sigma})) \\
&= \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{G},\mathsf{\Sigma})\sim\mu} [\mathsf{O}_i^\pi] && (\text{by Algorithm 2}) \\
&= \mathop{\mathbb{E}}_{(\mathsf{G},\mathsf{\Sigma})\sim\mu} \mathop{\mathbb{E}}_{i\in[p_r]} [\mathsf{O}_i^\pi] && (\text{as } \mathsf{M}^{(\leqslant r)} \text{ is fully determined by } \mathsf{G}, \mathsf{\Sigma}) \\
&\geqslant \mathop{\mathbb{E}}_{(\mathsf{G},\mathsf{\Sigma})\sim\mu} \left[\frac{\mathsf{O}^\pi - n_{r-1}\cdot f_r}{p_r}\right] && (\text{by Eq (22)}) \\
&= \mathop{\mathbb{E}}_{(\mathsf{G},\mathsf{\Sigma})\sim\mu} \left[\frac{\mathsf{O}^\pi}{p_r}\right] - \frac{n_{r-1}\cdot f_r}{p_r} \\
&\geqslant \alpha \cdot \mathop{\mathbb{E}}_{(\mathsf{G},\mathsf{\Sigma})\sim\mu} \left[\frac{\mathsf{O}}{p_r}\right] - \frac{n_{r-1}\cdot f_r}{p_r} && (\text{as } \pi \text{ has an approximation ratio of } \alpha^{-1}) \\
&\geqslant \alpha \cdot \mathop{\mathbb{E}}_{(\mathsf{G},\mathsf{\Sigma})\sim\mu} \mathop{\mathbb{E}}_{i\in[p_r]} [\mathsf{O}_i] - \frac{n_{r-1}\cdot f_r}{p_r} && (\text{by Eq (21)}) \\
&= \alpha \cdot \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\nu_i} [\mathsf{O}_i] - \frac{n_{r-1}\cdot f_r}{p_r}, && (23)
\end{aligned}$$

as $\mu(\mathsf{B}_i) = \nu_i(\mathsf{B}_i)$. Meanwhile, for each $i \in [p_r]$, Fact A.5 bounds the real expected matching size of τ_i over ν_i as

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathsf{B}_i\sim\nu_i} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{\Sigma}\mid\mathsf{B}_i)\sim\nu_i} [\mathsf{O}_i^\tau] \\
&= \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{\Sigma}\mid\mathsf{B}_i)\sim\nu_i} [\mathsf{O}_i^\tau] && (\text{as } \nu_i(\mathsf{B}_i) = \mu_i(\mathsf{B}_i)) \\
&\geqslant \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{\Sigma}\mid\mathsf{B}_i)\sim\mu_i} [\mathsf{O}_i^\tau] - \frac{n_{r-1}}{2} \cdot \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \|\mu_i(\mathsf{M}^{(\leqslant r)}, \mathsf{\Sigma} \mid \mathsf{B}_i) - \nu_i(\mathsf{M}^{(\leqslant r)}, \mathsf{\Sigma} \mid \mathsf{B}_i)\|_{\text{tvd}}, && (24)
\end{aligned}$$

since the size of any matching is at most half of the total number of vertices. Combining Corollary 5.7 with Eq (23) and (24), we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\nu_i} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{\Sigma}\mid\mathsf{B}_i)\sim\nu_i} [\mathsf{O}_i^\tau] \\
&\geqslant \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \mathop{\mathbb{E}}_{(\mathsf{M}^{(\leqslant r)},\mathsf{\Sigma}\mid\mathsf{B}_i)\sim\mu_i} [\mathsf{O}_i^\tau] - \frac{n_{r-1}}{2} \cdot \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\mu_i} \|\mu_i(\mathsf{M}^{(\leqslant r)}, \mathsf{\Sigma} \mid \mathsf{B}_i) - \nu_i(\mathsf{M}^{(\leqslant r)}, \mathsf{\Sigma} \mid \mathsf{B}_i)\|_{\text{tvd}} \\
&\geqslant \alpha \cdot \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\nu_i} [\mathsf{O}_i] - \frac{n_{r-1}\cdot f_r}{p_r} - \frac{n_{r-1}}{2} \cdot \frac{1}{k\cdot n_{r-1}} \\
&\geqslant \left(\alpha - \frac{n_{r-1}\cdot f_r}{p_r \cdot n_{r-1}/(4k)} - \frac{1/(2k)}{n_{r-1}/(4k)}\right) \cdot \mathop{\mathbb{E}}_{i\in[p_r]} \mathop{\mathbb{E}}_{\mathsf{B}_i\sim\nu_i} [\mathsf{O}_i],
\end{aligned}$$

as $\mathsf{O}_i \geqslant n_{r-1}/(4k)$ for each $i \in [p_r]$ by Claim 6.2. Therefore, picking the best index $i^* \in [p_r]$ shows

τ_{i*} has an approximation ratio of at most

$$\left(\alpha - \frac{n_{r-1} \cdot f_r}{p_r \cdot n_{r-1}/(4k)} - \frac{1/(2k)}{n_{r-1}/(4k)}\right)^{-1} \leqslant \left(\alpha - O(\frac{1}{n_{r-1}})\right)^{-1},$$

as claimed. ∎

6.3 A Reduction to Bipartite Graphs

The following reduction, together with the lower bound for approximate matching for general graphs shown in Section 6.2, concludes the proof of Theorem 3.

Lemma 6.4. *For $r, \alpha \geqslant 1$, if there exists an r-round protocol for computing an α-approximate bipartite matching for bipartite graphs, then there also exists an r-round protocol for computing a 2α-approximate matching for general graphs, with exactly the same bandwidth.*

Proof. Let π be a protocol for bipartite graphs. We construct a corresponding protocol π' for general graphs as follows. On a given input graph $G = (V, E)$ with n vertices, the vertices in π' jointly sample $z \in \{0,1\}^n$ uniformly at random using public randomness. Let $L = \{v \in V \mid z_v = 0\}$ and $R = \{v \in V \mid z_v = 1\}$. Note that all vertices agree on L, R since they can be easily computed from z. Also let $G' = (V, E')$ be the bipartite subgraph of G where only edges across the cut (L, R) are preserved, i.e. $E' = \{(u, v) \in E \mid z_u \neq z_v\}$. Then all vertices run π on G' as if the neighbors of a vertex $v \in V$ are $N_{G'}(v) = \{u \in N_G(v) \mid z_u \neq z_v\}$. The referee of π' simply outputs the answer given by the referee of π.

It is easy to see π' has the same number of rounds and exactly the same bandwidth as π. Moreover, observe that G' is essentially a random bipartition of G so half of the original edges are dropped in expectation. In particular, we have $\mathbb{E}_z[\mu(G')] \geqslant \mu(G)/2$. (Recall that $\mu(\cdot)$ denotes the size of the maximum matching.) Therefore, an α-approximate bipartite matching for G' (over the randomness of z) is also a 2α-approximate matching for G by definition. ∎

Acknowledgement

We are thankful to the anonymous reviewers of FOCS 2022 for helpful comments on the presentation of the paper.

Sepehr Assadi is supported in part by a National Science Foundation CAREER award CCF-2047061, a gift from Google Research, and a Fulcrum award from Rutgers Research Council. Gillat Kol is supported by a National Science Foundation CAREER award CCF-1750443 and by a BSF grant No. 2018325.

References

[A17] S. Assadi. Combinatorial auctions do need modest interaction. In *Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17, Cambridge, MA, USA, June 26-30, 2017*, pages 145–162, 2017. 4

[A22] S. Assadi. A two-pass (conditional) lower bound for semi-streaming maximum matching. In J. S. Naor and N. Buchbinder, editors, *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, pages 708–742. SIAM, 2022. 4

[ACG+15] K. J. Ahn, G. Cormode, S. Guha, A. McGregor, and A. Wirth. Correlation clustering in data streams. In F. R. Bach and D. M. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 2237–2246. JMLR.org, 2015. 1, 3, 4

[ACK19] S. Assadi, Y. Chen, and S. Khanna. Polynomial pass lower bounds for graph streaming algorithms. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019.*, pages 265–276, 2019. 1, 3, 4

[AGM12a] K. J. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 459–467. SIAM, 2012. URL http://dl.acm.org/citation.cfm?id=2095116.2095156. 1, 2, 3, 4, 9

[AGM12b] K. J. Ahn, S. Guha, and A. McGregor. Graph sketches: sparsification, spanners, and subgraphs. In *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2012, Scottsdale, AZ, USA, May 20-24, 2012*, pages 5–14, 2012. doi:10.1145/2213556.2213560. 1, 3

[AGM13] K. J. Ahn, S. Guha, and A. McGregor. Spectral sparsification in dynamic graph streams. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 16th International Workshop, APPROX 2013, and 17th International Workshop, RANDOM 2013, Berkeley, CA, USA, August 21-23, 2013. Proceedings*, pages 1–10, 2013. 1, 3

[AKLY16] S. Assadi, S. Khanna, Y. Li, and G. Yaroslavtsev. Maximum matchings in dynamic graph streams and the simultaneous communication model. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1345–1364, 2016. 4

[AKM22] S. Assadi, P. Kumar, and P. Mittal. Brooks' theorem in graph streams: A single-pass semi-streaming algorithm for Δ-coloring. *CoRR*, abs/2203.10984, 2022. 1, 3

[AKO20] S. Assadi, G. Kol, and R. Oshman. Lower bounds for distributed sketching of maximal matchings and maximal independent sets. In Y. Emek and C. Cachin, editors, *PODC '20: ACM Symposium on Principles of Distributed Computing, Virtual Event, Italy, August 3-7, 2020*, pages 79–88. ACM, 2020. 1, 2, 3, 9

[ANRW15] N. Alon, N. Nisan, R. Raz, and O. Weinstein. Welfare maximization with limited interaction. In V. Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1499–1512. IEEE Computer Society, 2015. ii, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20

[AR20] S. Assadi and R. Raz. Near-quadratic lower bounds for two-pass graph streaming algorithms. In S. Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 342–353. IEEE, 2020. 4

[BBH+19] A. Balliu, S. Brandt, J. Hirvonen, D. Olivetti, M. Rabie, and J. Suomela. Lower bounds for maximal matchings and maximal independent sets. In D. Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 481–497. IEEE Computer Society, 2019. 1

[BHH19] S. Behnezhad, M. Hajiaghayi, and D. G. Harris. Exponentially faster massively parallel maximal matching. In D. Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1637–1649. IEEE Computer Society, 2019. 3

[BMN+11] F. Becker, M. Matamala, N. Nisse, I. Rapaport, K. Suchan, and I. Todinca. Adding a referee to an interconnection network: What can(not) be computed in one round. In *25th IEEE International Symposium on Parallel and Distributed Processing, IPDPS 2011, Anchorage, Alaska, USA, 16-20 May, 2011 - Conference Proceedings*, pages 508–514. IEEE, 2011. 1

[BMRT14] F. Becker, P. Montealegre, I. Rapaport, and I. Todinca. The simultaneous number-in-hand communication model for networks: Private coins, public coins and determinism. In M. M. Halldórsson, editor, *Structural Information and Communication Complexity - 21st International Colloquium, SIROCCO 2014, Takayama, Japan, July 23-25, 2014. Proceedings*, volume 8576 of *Lecture Notes in Computer Science*, pages 83–95. Springer, 2014. 1

[BMRT18] F. Becker, P. Montealegre, I. Rapaport, and I. Todinca. The impact of locality on the detection of cycles in the broadcast congested clique model. In M. A. Bender, M. Farach-Colton, and M. A. Mosteiro, editors, *LATIN 2018: Theoretical Informatics - 13th Latin American Symposium, Buenos Aires, Argentina, April 16-19, 2018, Proceedings*, volume 10807 of *Lecture Notes in Computer Science*, pages 134–145. Springer, 2018. 1, 2, 3

[BO17] M. Braverman and R. Oshman. A rounds vs. communication tradeoff for multi-party set disjointness. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 144–155, 2017. 2, 3, 4

[CDK19] G. Cormode, J. Dark, and C. Konrad. Independent sets in vertex-arrival streams. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, pages 45:1–45:14, 2019. 4

[CKP+21] L. Chen, G. Kol, D. Paramonov, R. R. Saxena, Z. Song, and H. Yu. Almost optimal super-constant-pass streaming lower bounds for reachability. In S. Khuller and V. V.

Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 570–583. ACM, 2021. 4

[CT06] T. M. Cover and J. A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006. 40

[DK20] J. Dark and C. Konrad. Optimal lower bounds for matching and vertex cover in dynamic graph streams. In S. Saraf, editor, *35th Computational Complexity Conference, CCC 2020, July 28-31, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 169 of *LIPIcs*, pages 30:1–30:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. 4

[DKO14] A. Drucker, F. Kuhn, and R. Oshman. On the power of the congested clique model. In M. M. Halldórsson and S. Dolev, editors, *ACM Symposium on Principles of Distributed Computing, PODC '14, Paris, France, July 15-18, 2014*, pages 367–376. ACM, 2014. 1, 2, 3

[DNO14] S. Dobzinski, N. Nisan, and S. Oren. Economic efficiency requires interaction. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 233–242, 2014. 2, 4

[FKN21] A. Filtser, M. Kapralov, and N. Nouri. Graph spanners by sketching in dynamic streams and the simultaneous communication model. In D. Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1894–1913. SIAM, 2021. 1, 3

[GGK+18] M. Ghaffari, T. Gouleakis, C. Konrad, S. Mitrovic, and R. Rubinfeld. Improved massively parallel computation algorithms for mis, matching, and vertex cover. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing, PODC 2018, July 23-27, 2018*, pages 129–138, 2018. 1, 2, 3

[Gha16] M. Ghaffari. An improved distributed algorithm for maximal independent set. In R. Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 270–277. SIAM, 2016. 1

[GMT15] S. Guha, A. McGregor, and D. Tench. Vertex and hyperedge connectivity in dynamic graph streams. In *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 241–247, 2015. 1, 3

[GO13] V. Guruswami and K. Onak. Superlinear lower bounds for multipass graph processing. In *Proceedings of the 28th Conference on Computational Complexity, CCC 2013, K.lo Alto, California, USA, 5-7 June, 2013*, pages 287–298, 2013. 4

[GS62] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962. 4

[JLN18a] T. Jurdzinski, K. Lorys, and K. Nowicki. Communication complexity in vertex partition whiteboard model. In Z. Lotker and B. Patt-Shamir, editors, *Structural Information and Communication Complexity - 25th International Colloquium, SIROCCO 2018, Ma'ale HaHamisha, Israel, June 18-21, 2018, Revised Selected Papers*, volume 11085 of *Lecture Notes in Computer Science*, pages 264–279. Springer, 2018. 1, 3

[JLN18b] T. Jurdzinski, K. Lorys, and K. Nowicki. Communication complexity in vertex partition whiteboard model. In Z. Lotker and B. Patt-Shamir, editors, *Structural Information and Communication Complexity - 25th International Colloquium, SIROCCO 2018, Ma'ale HaHamisha, Israel, June 18-21, 2018, Revised Selected Papers*, volume 11085 of *Lecture Notes in Computer Science*, pages 264–279. Springer, 2018. 1

[JN18] T. Jurdzinski and K. Nowicki. Connectivity and minimum cut approximation in the broadcast congested clique. In Z. Lotker and B. Patt-Shamir, editors, *Structural Information and Communication Complexity - 25th International Colloquium, SIROCCO 2018, Ma'ale HaHamisha, Israel, June 18-21, 2018, Revised Selected Papers*, volume 11085 of *Lecture Notes in Computer Science*, pages 331–344. Springer, 2018. 1, 3

[KLM+14] M. Kapralov, Y. T. Lee, C. Musco, C. Musco, and A. Sidford. Single pass spectral sparsification in dynamic streams. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 561–570, 2014. doi:10.1109/FOCS.2014.66. 1, 3, 9

[KMW16] F. Kuhn, T. Moscibroda, and R. Wattenhofer. Local computation: Lower and upper bounds. *J. ACM*, 63(2):17:1–17:44, 2016. 1

[KW14] M. Kapralov and D. P. Woodruff. Spanners and sparsifiers in dynamic streams. In *ACM Symposium on Principles of Distributed Computing, PODC '14, Paris, France, July 15-18, 2014*, pages 272–281, 2014. 1, 3

[Lin87] N. Linial. Distributive graph algorithms-global solutions from local data. In *28th Annual Symposium on Foundations of Computer Science, Los Angeles, California, USA, 27-29 October 1987*, pages 331–335. IEEE Computer Society, 1987. 1

[LMSV11] S. Lattanzi, B. Moseley, S. Suri, and S. Vassilvitskii. Filtering: a method for solving graph problems in mapreduce. In *SPAA 2011: Proceedings of the 23rd Annual ACM Symposium on Parallelism in Algorithms and Architectures, San Jose, CA, USA, June 4-6, 2011 (Co-located with FCRC 2011)*, pages 85–94, 2011. doi:10.1145/1989493.1989505. 1, 3, 4

[Lub85] M. Luby. A simple parallel algorithm for the maximal independent set problem. In R. Sedgewick, editor, *Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 6-8, 1985, Providence, Rhode Island, USA*, pages 1–10. ACM, 1985. 1, 2

[MS15] B. M. Maggs and R. K. Sitaraman. Algorithmic nuggets in content delivery. *Comput. Commun. Rev.*, 45(3):52–66, 2015. 4

[MTVV15] A. McGregor, D. Tench, S. Vorotnikova, and H. T. Vu. Densest subgraph in dynamic graph streams. In *Mathematical Foundations of Computer Science 2015 - 40th International Symposium, MFCS 2015, Milan, Italy, August 24-28, 2015, Proceedings, Part II*, pages 472–482, 2015. 1, 3

[Nis21] N. Nisan. The demand query model for bipartite matching. In D. Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 592–599. SIAM, 2021. 4

[NY19] J. Nelson and H. Yu. Optimal lower bounds for distributed and streaming spanning forest computation. In T. M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 1844–1860. SIAM, 2019. 1, 2, 3

[RS92] A. E. Roth and M. Sotomayor. Two-sided matching. *Handbook of game theory with economic applications*, 1:485–541, 1992. 4

[Yao77] A. C. Yao. Probabilistic computations: Toward a unified measure of complexity (extended abstract). In *18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977*, pages 222–227, 1977. 17, 31

[Yu21] H. Yu. Tight distributed sketching lower bound for connectivity. In D. Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1856–1873. SIAM, 2021. 1, 2, 3

Appendix

A Basic Tools From Information Theory

We now briefly introduce some definitions from information theory that are needed in this paper. For a random variable A, we use $\mathrm{supp}(\mathsf{A})$ to denote the support of A and $\mathrm{dist}(\mathsf{A})$ to denote its distribution. When it is clear from context, we may abuse the notation and use A directly instead of $\mathrm{dist}(\mathsf{A})$, for example, write $A \sim \mathsf{A}$ to mean $A \sim \mathrm{dist}(\mathsf{A})$, i.e., A is sampled from the distribution of random variable A.

We denote the *Shannon entropy* of a random variable A by $\mathbb{H}(\mathsf{A})$, which is defined as:

$$\mathbb{H}(\mathsf{A}) = \sum_{A \in \mathrm{supp}(\mathsf{A})} \Pr(\mathsf{A} = A) \cdot \log \frac{1}{\Pr(\mathsf{A} = A)}.$$

The *conditional entropy* of A conditioned on B is denoted by $\mathbb{H}(\mathsf{A} \mid \mathsf{B})$ and defined as:

$$\mathbb{H}(\mathsf{A} \mid \mathsf{B}) = \mathop{\mathbb{E}}_{B \sim \mathsf{B}} \left[\mathbb{H}(\mathsf{A} \mid \mathsf{B} = B)\right],$$

where $\mathbb{H}(\mathsf{A} \mid \mathsf{B} = B)$ is defined in a standard way by using the distribution of A conditioned on the event $\mathsf{B} = B$ in the previous equation. The *mutual information* of two random variables A and B is denoted by $\mathbb{I}(\mathsf{A}\,;\mathsf{B})$ and defined as:

$$\mathbb{I}(\mathsf{A}\,;\mathsf{B}) = \mathbb{H}(\mathsf{A}) - \mathbb{H}(\mathsf{A} \mid \mathsf{B}) = \mathbb{H}(\mathsf{B}) - \mathbb{H}(\mathsf{B} \mid \mathsf{A}).$$

The *conditional mutual information* $\mathbb{I}(\mathsf{A}\,;\mathsf{B} \mid \mathsf{C})$ is $\mathbb{H}(\mathsf{A} \mid \mathsf{C}) - \mathbb{H}(\mathsf{A} \mid \mathsf{B}, \mathsf{C})$ and hence by linearity of expectation:

$$\mathbb{I}(\mathsf{A}\,;\mathsf{B} \mid \mathsf{C}) = \mathop{\mathbb{E}}_{C \sim \mathsf{C}} \left[\mathbb{I}(\mathsf{A}\,;\mathsf{B} \mid \mathsf{C} = C)\right].$$

We also use the following standard measures of distance (or divergence) between distributions.

KL-divergence. For two distributions μ and ν, the *Kullback-Leibler divergence* between μ and ν is denoted by $\mathbb{D}(\mu \mid\mid \nu)$ and defined as:

$$\mathbb{D}(\mu \mid\mid \nu) = \mathop{\mathbb{E}}_{a \sim \mu} \left[\log \frac{\mu(a)}{\nu(a)}\right].$$

Total variation distance. We denote the total variation distance between two distributions μ and ν on the same support Ω by $\|\mu - \nu\|_{\mathrm{tvd}}$, defined as:

$$\|\mu - \nu\|_{\mathrm{tvd}} = \max_{\Omega' \subseteq \Omega} \left(\mu(\Omega') - \nu(\Omega')\right) = \frac{1}{2} \cdot \sum_{x \in \Omega} \left|\mu(x) - \nu(x)\right|.$$

We refer the interested readers to the textbook by Cover and Thomas [CT06] for an excellent introduction to the field of information theory.

A.1 Useful Properties of Entropy and Mutual Information

We use the following basic properties of entropy and mutual information throughout.

Fact A.1 (cf. [CT06])**.** *Let* A*,* B*,* C*, and* D *be four (possibly correlated) random variables.*

1. $0 \leqslant \mathbb{H}(\mathsf{A}) \leqslant \log|\mathrm{supp}(\mathsf{A})|$. *The right equality holds iff* $\mathrm{dist}(\mathsf{A})$ *is uniform.*
2. $\mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C}) \geqslant 0$. *The equality holds iff* A *and* B *are* independent *conditioned on* C.
3. Conditioning on a random variable reduces entropy*:* $\mathbb{H}(\mathsf{A} \mid \mathsf{B},\mathsf{C}) \leqslant \mathbb{H}(\mathsf{A} \mid \mathsf{B})$. *The equality holds iff* $\mathsf{A} \perp \mathsf{C} \mid \mathsf{B}$.
4. Subadditivity of entropy*:* $\mathbb{H}(\mathsf{A},\mathsf{B} \mid \mathsf{C}) \leqslant \mathbb{H}(\mathsf{A} \mid C) + \mathbb{H}(\mathsf{B} \mid \mathsf{C})$.
5. Chain rule for entropy*:* $\mathbb{H}(\mathsf{A},\mathsf{B} \mid \mathsf{C}) = \mathbb{H}(\mathsf{A} \mid \mathsf{C}) + \mathbb{H}(\mathsf{B} \mid \mathsf{C},\mathsf{A})$.
6. Chain rule for mutual information*:* $\mathbb{I}(\mathsf{A},\mathsf{B};\mathsf{C} \mid \mathsf{D}) = \mathbb{I}(\mathsf{A};\mathsf{C} \mid \mathsf{D}) + \mathbb{I}(\mathsf{B};\mathsf{C} \mid \mathsf{A},\mathsf{D})$.
7. Data processing inequality*: for a deterministic function* $f(\mathsf{A})$*,* $\mathbb{I}(f(\mathsf{A});\mathsf{B} \mid \mathsf{C}) \leqslant \mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C})$.

We also use the following propositions, regarding the effect of conditioning on mutual information.

Proposition A.2. *For random variables* $\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D}$*, if* $\mathsf{A} \perp \mathsf{D} \mid \mathsf{C}$*, then,*

$$\mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C}) \leqslant \mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C},\mathsf{D}).$$

Proof. Since A and D are independent conditioned on C, by Fact A.1-(3), $\mathbb{H}(\mathsf{A} \mid \mathsf{C}) = \mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{D})$ and $\mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{B}) \geqslant \mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{B},\mathsf{D})$. We have,

$$\begin{aligned}\mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C}) &= \mathbb{H}(\mathsf{A} \mid \mathsf{C}) - \mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{B}) = \mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{D}) - \mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{B}) \\ &\leqslant \mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{D}) - \mathbb{H}(\mathsf{A} \mid \mathsf{C},\mathsf{B},\mathsf{D}) = \mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C},\mathsf{D}). \qquad \blacksquare\end{aligned}$$

Proposition A.3. *For random variables* $\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D}$*, if* $\mathsf{A} \perp \mathsf{D} \mid \mathsf{B},\mathsf{C}$*, then,*

$$\mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C}) \geqslant \mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C},\mathsf{D}).$$

Proof. Since $\mathsf{A} \perp \mathsf{D} \mid \mathsf{B},\mathsf{C}$, by Fact A.1-(3), $\mathbb{H}(\mathsf{A} \mid \mathsf{B},\mathsf{C}) = \mathbb{H}(\mathsf{A} \mid \mathsf{B},\mathsf{C},\mathsf{D})$. Moreover, since conditioning can only reduce the entropy (again by Fact A.1-(3)),

$$\begin{aligned}\mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C}) &= \mathbb{H}(\mathsf{A} \mid \mathsf{C}) - \mathbb{H}(\mathsf{A} \mid \mathsf{B},\mathsf{C}) \geqslant \mathbb{H}(\mathsf{A} \mid \mathsf{D},\mathsf{C}) - \mathbb{H}(\mathsf{A} \mid \mathsf{B},\mathsf{C}) \\ &= \mathbb{H}(\mathsf{A} \mid \mathsf{D},\mathsf{C}) - \mathbb{H}(\mathsf{A} \mid \mathsf{B},\mathsf{C},\mathsf{D}) = \mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C},\mathsf{D}). \qquad \blacksquare\end{aligned}$$

A.2 Measures of Distance Between Distributions

The following states the relation between mutual information and KL-divergence.

Fact A.4. *For random variables* $\mathsf{A},\mathsf{B},\mathsf{C}$*,*

$$\mathbb{I}(\mathsf{A};\mathsf{B} \mid \mathsf{C}) = \mathop{\mathbb{E}}_{(b,c)\sim(\mathsf{B},\mathsf{C})}\Big[\mathbb{D}(\mathrm{dist}(\mathsf{A} \mid \mathsf{B}=b,\mathsf{C}=c) \mid\mid \mathrm{dist}(\mathsf{A} \mid \mathsf{C}=c))\Big].$$

We use the following basic properties of total variation distance.

Fact A.5. *Suppose* μ *and* ν *are two distributions for a random variable* X*, then,*

$$\mathop{\mathbb{E}}_{\mu}[\mathsf{X}] \leqslant \mathop{\mathbb{E}}_{\nu}[\mathsf{X}] + \|\mu-\nu\|_{\mathrm{tvd}} \cdot \max_{X_0 \in \mathrm{supp}(\mathsf{X})} X_0.$$

Fact A.6. *Suppose μ and ν are two distributions for the tuple $(\mathsf{X}_1, \ldots, \mathsf{X}_t)$, then,*

$$\|\mu(\mathsf{X}_1, \ldots, \mathsf{X}_t) - \nu(\mathsf{X}_1, \ldots, \mathsf{X}_t)\|_{\mathrm{tvd}} \leqslant \sum_{i=1}^{n} \mathop{\mathbb{E}}_{(X_1,\ldots,X_{i-1})\sim\mu} \|\mu(\mathsf{X}_i \mid X_1, \ldots, X_{i-1}) - \nu(\mathsf{X}_i \mid X_1, \ldots, X_{i-1})\|_{\mathrm{tvd}}.$$

Fact A.7. *Suppose μ and ν are two distributions for the pair (X, Y), then,*

$$\|\mu(\mathsf{X}) - \nu(\mathsf{X})\|_{\mathrm{tvd}} \leqslant \|\mu(\mathsf{X}, \mathsf{Y}) - \nu(\mathsf{X}, \mathsf{Y})\|_{\mathrm{tvd}}.$$

The following Pinsker's inequality bounds the total variation distance between two distributions based on their KL-divergence.

Fact A.8 (Pinsker's inequality)**.** *For any distributions μ and ν, $\|\mu - \nu\|_{\mathrm{tvd}} \leqslant \sqrt{\frac{1}{2} \cdot \mathbb{D}(\mu \mid\mid \nu)}$.*

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il