Parallel Best Arm Identification in Heterogeneous Environments*

Nikolai Karpov
Indiana University
Bloomington, IN, USA
nkarpov@iu.edu

Qin Zhang†
Indiana University
Bloomington, IN, USA
qzhangcs@iu.edu

ABSTRACT

In this paper, we study the tradeoffs between the *time* and the *number of communication rounds* of the best arm identification problem in the heterogeneous collaborative learning model, where multiple agents interact with possibly different environments and they want to learn in parallel an objective function in the aggregated environment. By proving almost tight upper and lower bounds, we show that collaborative learning in the heterogeneous setting is inherently more difficult than that in the homogeneous setting in terms of the time-round tradeoff.

CCS CONCEPTS

• **Theory of computation → Communication complexity**; • **Computing methodologies → Multi-agent reinforcement learning**.

KEYWORDS

parallel learning; communication complexity; best arm identification; heterogeneous environments

ACM Reference Format:
Nikolai Karpov and Qin Zhang. 2024. Parallel Best Arm Identification in Heterogeneous Environments. In *Proceedings of the 36th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA '24), June 17–21, 2024, Nantes, France.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3626183.3659957

1 INTRODUCTION

As data continue to grow, multi-agent learning has emerged as an important direction in scalable machine learning and has attracted much attention under the name of *federated learning* [15, 16, 18], where multiple agents try to learn an objective function in parallel via communication. While the majority of work in federated learning focuses on the distributed training of neural networks, a few papers [14, 24, 25] studied parallel reinforcement learning problems in a very similar model named the *collaborative learning* (CL) model. However, most work in the literature of collaborative learning only considered the *homogeneous* environment (or, IID data), in which agents interact with the same data distribution. Real world applications of multi-agent learning often involve *heterogeneous* environments (or, non-IID data), in which agents interact with possibly different data distributions.[1] Indeed, heterogeneous environments have been identified as a key feature of the federated learning model [10].

In this paper, we investigate heterogeneous collaborative learning. We will use a basic problem in bandit theory named *best arm identification in multi-armed bandits* (BAI) as a vehicle to deliver the following message: *Collaborative learning in the heterogeneous environment is provably more difficult than that in the homogeneous environment w.r.t. communication rounds.*

In the following, we first introduce the BAI problem and the CL model, and then summarize our results and contributions. We conclude the section by discussing related work.

Best Arm Identification in Multi-Armed Bandits. In BAI, we have n arms, each of which is associated with an unknown distribution $\mathcal{D}_i$ ($i \in [n]$) with support $[0, 1]$.[2] We aim at identifying the arm whose associated distribution has the largest mean by a sequence of T pulls. In each arm pull, we choose an arm based on the previous pulls and outcomes, and obtain a sample from the arm's associated distribution. Assuming that each pull takes unit time, we call T the *time horizon*. The goal of BAI is to identify the arm with the highest mean with the smallest error probability under time horizon T.[3]

BAI is a basic problem in bandit theory and reinforcement learning, and has been studied extensively in the literature since 1950s (e.g., [2, 3, 5–7, 9, 11, 20]). The problem has numerous real-world applications, including clinical trials, article/ad/channel selection, computer game play, financial portfolio design, adaptive routing, crowd-sourced ranking, hyperparameter optimization, etc.

Let $I = \{1, 2, \ldots, n\}$ be an input instance of n arms. W.l.o.g., we assume that there is a unique best arm, which is denoted by i_*. Let μ_* be the mean of $\mathcal{D}_{i_*}$, and for any $i \in [n]$, let μ_i be the mean of $\mathcal{D}_i$. Let $\Delta_i = \mu_* - \mu_i$ be the *mean gap* of the best arm and the i-th arm. The *instance complexity* of BAI on input I is defined as:

$$H(I) \triangleq \sum_{i \in [n], i \neq i_*} 1/\Delta_i^2. \tag{1}$$

*Authors are supported in part by NSF CCF-1844234 and CCF-2006591.
†Corresponding author.

SPAA '24, June 17–21, 2024, Nantes, France

ACM ISBN 979-8-4007-0416-1/24/06…$15.00
https://doi.org/10.1145/3626183.3659957

[1] We felt that the words "IID/non-IID", which are widely used in the literature of federated learning, are somewhat confusing. In the rest of this paper, we will use the words "homogeneous" and "heterogeneous" to denote the scenarios where agents interact with identical and different data distributions, respectively.
[2] We use $[n]$ to denote $\{1, 2, \ldots, n\}$.
[3] For readers who are familiar with the bandit literature, we are considering the *fixed-time/budget* best arm identification. Another version of this problem is called *fixed-confidence*, where we want to solve BAI with a fixed error probability δ with the smallest number of pulls.

Intuitively, the term $1/\Delta_i^2$ is the number of pulls needed to separate the best arm and the i-th arm with a good probability. We will sometimes write $H \triangleq H(I)$ for convenience. It is known that under time horizon $\tilde{O}(H)$, there exists a centralized algorithm that solves BAI with probability 0.99 [2].[4] On the other hand, no centralized algorithm can solve the BAI problem with probability 0.99 under time horizon H [4].

The Collaborative Learning Model. Most study for BAI has been done in the centralized model, in which just one agent pulls the set of arms sequentially. [14, 24] studied BAI in the collaborative learning (CL) model, where there are K agents, who try to learn the best arm in parallel via communication. The learning proceeds in rounds. In each round, each agent takes a sequence of pulls (one at each time step) and observes the outcomes. At the end of each round there is a communication phase; the agents communicate with each other to exchange newly observed information and determine the number of time steps for the next round (the length of the first round is determined at the beginning of the first round). At the end of the last round, all agents have to output the same answer without any further communication. The goal of BAI in the CL model is for all agents to output the correct answer with the smallest error probability under time horizon T (i.e., the number of time steps over all rounds) and the number of rounds R. Note that the number of communication phases is $(R-1)$, since we do not allow any communication at the end of the last round.

Depending on whether the agents have real-time computing and policy-updating ability, the CL algorithms are divided into two categories: *adaptive* and *non-adaptive*. In the adaptive case, agents can change their pull policies at each time step based on new observations. While in the non-adaptive case, policy updates can only happen at the beginning of each round. In this paper, we focus on the adaptive case; the lower bound proof for the adaptive case is more challenging than the non-adaptive case due to agents' local adaptivity within a round.

Minimizing communication in the CL model is critical due to network bandwidth constraints and latency, energy consumption (think of deep-sea/outer-space exploration), and data usage (e.g., if messages are sent by mobile devices). In this paper, we mainly focus on the round complexity. Like parallel/distributed computation models such as MapReduce, initiating a new round of learning process can be very expensive due to various communication overheads. The communication cost (i.e., the total number of bits exchanged between agents) of our algorithm is optimal up to a logarithmic factor based on a recent lower bound result in [12] (see Remark 21).

Heterogeneous Environments. In the CL model studied by [24] and [14], each agent interacts with the same environment; for the BAI problem in particular, by pulling the same arm, the agents sample from the same data distribution. However, as mentioned earlier, heterogeneous environments are inherent in many real-world collaborative learning applications.

For example, in the setting of channel selection in cognitive radio networks, a base station utilizes a number of mobile devices (e.g., cell phones) to select the best channel for data transformation in a particular area. Here each mobile device represents an agent and each channel represents an arm. At each time step, an agent selects a channel and attempts to transmit a message. If the message is successfully delivered, the agent receives a reward of 1; otherwise, the reward is 0. This corresponds to the bandit setting. Since mobile devices sit at different geographic locations, the channel availability distributions they observe may be very different. The base station needs to identify the best arm with respect to the *aggregation* of local channel availability distributions. Another example is the task of item-selection in recommendation systems, where a group of servers work together to learn the globally most popular item via communication, while each server can only interact with users in a certain region (and thus get samples from a distinct data distribution).

In BAI with heterogeneous environments, by pulling the same arm, the agents sample from possibly different distributions. Let $\pi_{i,k}$ be the distribution associated with the i-th arm that the k-th agent samples from, and let $\mu_{i,k}$ be the mean of $\pi_{i,k}$. Define the global mean of the i-th arm as

$$\mu_i \triangleq \frac{1}{K} \sum_{k \in [K]} \mu_{i,k}. \tag{2}$$

Our task is to identify the arm i_* with the largest global mean, while each agent $k \in [K]$ can only pull each arm $i \in [n]$ under its local distribution $\pi_{i,k}$. In the heterogeneous setting, we define mean gaps based on the global means, that is, $\Delta_i = \mu_* - \mu_i$, and the instance complexity H again as $\sum_{i \in [n], i \neq i_*} 1/\Delta_i^2$.

Our Results. The main result of this paper is the following impossibility result.

Theorem 1 (Main Theorem). *For any $1 \le R \le \frac{\log n}{24 \log\log n}$ and any $T < Hn^{\Omega(\frac{1}{R})}/K$, any R-round T-time K-agent algorithm that solves n-arm BAI in the heterogeneous CL model has a success probability less than* 0.99.

We complement the impossibility result by the following algorithmic result.

Theorem 2. *For any $R \ge 1$ and any $T \ge c_T H n^{\frac{1}{R}}/K$ for a universal constant c_T, there exists a R-round T-time K-agent algorithm that solves n-arm BAI in the heterogeneous CL model with probability* 0.99.

We note that for a fixed time budget, the number of rounds R in the lower and upper bounds in Theorem 1 and Theorem 2 match up to a constant factor.

We would like to highlight a couple of points regarding Theorem 1. First, this is the first lower bound result that addresses the *local agent adaptivity* in the CL models. In particular, it shows that the capacity of each agent to utilize newly observed information within each round does *not* contribute to reducing the round complexity in the heterogeneous CL model. This is in stark contrast with the homogeneous CL model in which local agent adaptivity can significantly reduce the round complexity. Second, our hard input distribution for proving Theorem 1 is the first one that uses *asymmetric* arm means constructions. It exploits the heterogeneous property, enabling us to establish a higher lower bound for BAI

[4] For the convenience of presentation, we sometimes use '$\tilde{}$' on O, Ω, Θ to hide non-critical logarithmic factors. All logarithmic factors will be spelled out in our theorems explicitly.

than the one presented in the homogeneous setting [24]. We will give a more detailed technical overview in Section 2.

1.1 Related Work

We summarize previous work that is closely related to this paper in the CL model, and refer readers to the book by [17] for an overview on BAI in the centralized model.

The (homogeneous) CL model was first used in the work [8] for studying multi-agent BAI, but the model was not formally defined there. The results for fixed-time BAI in [8] only consider the special case where there is only one communication phase (i.e., $R = 2$). The CL model was rigorously formulated in [24], where the authors obtained almost tight tradeoffs between the learning time and the round cost for BAI. The followup work [14] extended this line of research to the top-m arm identifications problem. [25] studied regret minimization in multi-armed bandits in essentially the same model, but it focused on the total bits of communication exchanged between the agents (or, the communication cost) instead of the number of rounds. Recently, [12] studied the tradeoff between the learning time and the communication cost in the CL model for BAI, and [1] studied linear bandits in a similar setting.

The authors of [21] studied BAI and regret minimization in multi-armed bandits in a model similar to the CL model, but mainly in the fixed-confidence setting. That is, their algorithm takes a confidence parameter δ (instead of a time horizon T) as an input, and try to use the smallest possible number of time steps to identify the best arm with probability $(1 - \delta)$. Their lower bound results are proved for the setting that agents can communicate at each time step.

In the heterogeneous CL model, [22, 23] studied regret minimization in multi-armed bandits. The authors considered the communication cost of the CL algorithms, but the cost has been embedded into the regret formulation. [19] studied BAI in the CL model where arms are partitioned into groups, and each agent can only pull arms from one particular group. This model can be thought as a special case of the heterogeneous CL model studied in this paper, where for any arm $i \in [n]$, there exists a unique agent $k \in [K]$ such that $\mu_{i,k} > 0$, while $\mu_{i,k'} = 0$ for all $k' \in [K] \backslash \{k\}$. This special case does *not* capture the inherent difficulty of the heterogeneous CL model where the information about a particular arm can spread over multiple agents, and their results cannot be generalized to the heterogeneous CL model.

2 TECHNICAL OVERVIEW OF THE MAIN RESULT

Before delving into the full proof of our main result (Theorem 1), which is very technical, we would like to provide an overview.

We note that all the parameters used in this technical overview are merely for the illustration purpose. They may *not* correspond to the actual, typically more complex, parameters used in the actual proof. We will also frequently *ignore lower-order logarithmic terms* for the sake of readability.

Generalized Round Elimination and Challenges. Let us start by briefly illustrating the generalized round elimination technique introduced in [24], and then explain the challenges in applying it in the heterogeneous setting.

Generalized round elimination can be thought as an induction on a sequence of hard distribution classes $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_R$, where $\mathcal{D}_0 = \{\phi\}$ consists of the original hard input distribution ϕ.[5] At the i-th induction step, we show that for any input distribution in $\sigma \in \mathcal{D}_{i-1}$, if the agents do not conduct enough *non-adaptive* pulls (due to the time budget constraint) in a round, then after some "input massage" which will only make the problem easier, the posterior distribution σ' belongs to $\mathcal{D}_i$. For the base case, we show that *no* 0-round CL algorithm can solve the problem for any distribution $\sigma \in \mathcal{D}_R$ with a non-trivial success probability. We can thus prove that *no* R-round algorithm for solving the problem on the original input distribution ϕ with a non-trivial success probability.

The lower bound proof using generalized round elimination in [24] was carried out on non-adaptive algorithms in the homogeneous CL setting. For adaptive algorithms, only in the case when $n \le K$ (that is, the number of arms is no more than the number of agents), we can show that adaptive pulls do *not* have much advantage against non-adaptive pulls via a coupling argument. This is why in [24], only a $\Omega\left(\frac{\log \min\{K,n\}}{\log\log \min\{K,n\}}\right)$ round lower bound can be proved for adaptive algorithms. As mentioned, for the heterogeneous CL setting, our goal is to prove an $\Omega\left(\frac{\log n}{\log\log n}\right)$ round lower bound for adaptive algorithms for any value n. To this end, we must design a new, harder input distribution that leverages the heterogeneous property of the data distributions.

An Interleaved Local Mean Construction. Our new input distribution for heterogeneous data is easier to visualize with two agents, Alice and Bob, but it can easily be extended to multiple agents.

We will focus on the $\Omega(\log n)$ round case (ignoring a $\log\log$ factor), while our lower bound result covers the entire time-round tradeoff. The formal definition of our hard input distribution and its properties can be found in Section 3.1.

Our hard input distribution has $L = \Theta(\log n)$ terms, with odd terms held by Alice and even terms held by Bob. The global mean of each arm can be written as

$$\mu = \frac{1}{2} + \sum_{\ell=1}^{L} \frac{X_\ell}{4^\ell},$$

where $X_1, \ldots, X_L \in \{0, 1\}$ are i.i.d. Bernoulli random variables with mean $\frac{1}{2}$. When $X_1 = \cdots = X_L = 1$, μ achieves its maximum possible value. The local mean of each arm at Alice's side is

$$\mu^A = \frac{1}{2} + \sum_{\ell : 1 \le 2\ell+1 \le L} \frac{2X_{2\ell+1}}{4^{2\ell+1}},$$

and that at Bob's side is

$$\mu^B = \frac{1}{2} + \sum_{\ell : 1 \le 2\ell \le L} \frac{2X_{2\ell}}{4^{2\ell}}.$$

Note that $\mu = (\mu^A + \mu^B)/2$. Let π, π^A and π^B be the underlying distributions of μ, μ^A, and μ^B.

[5] We need to use input *distribution* instead of a single hard input instance because we are proving a lower bound for randomized algorithms. By Yao's minimax lemma [26], we can instead prove a lower bound for deterministic algorithms on a hard input distribution.

Proof Intuition and New Challenges. We say an arm is at level ℓ if $X_1 = \ldots = X_\ell = 1$ and $X_{\ell+1} = 0$ (if $\ell < L$). The high-level intuition of proving an $\Omega(L)(= \Omega(\log n))$ lower bound is that Alice and Bob must learn the set of n arms level by level under a time budget $\tilde{O}(H)$, where H is the input instance complexity. That is, at the end of the ℓ-th round, they can only identify and eliminate those arms that are in the first ℓ levels, while for the remaining arms the uncertainty is still large. As a result, they need $L = \Omega(\log n)$ rounds to identify the best arm. Ideally, we hope to show that at each odd round ℓ, Alice is able to identify and eliminate those arms who are in level ℓ but not higher, while Bob is *not* able to do much as he lacks information about X_ℓ of each arm. And a similar situation holds at each even round ℓ with Alice and Bob's positions swapped.

The difficulty in formalizing the above intuition is that it is actually *possible* for each party to learn information about the bits (i.e., the X_i's) at *all levels* using their local samples and messages received from the other party. What we need to show is that this information is *not* enough to allow parties to "jump" $\omega(1)$ levels after each round given the total sample budget.

We try to formalize this intuition using generalized round elimination. There are two challenges in proving a $\Omega(\log n)$ round lower bound for BAI in the heterogeneous CL model.

(1) Explicit forms of distribution classes like those used in [24] in the homogeneous setting are difficult to obtain in the heterogeneous setting due to the intricate structures of the hard input distributions μ^A and μ^B.
(2) Since the coupling argument which reduces adaptive CL algorithms to non-adaptive CL algorithms is inapplicable when $n > K$, we have to prove the lower bound for adaptive CL algorithms directly.

In the following, we briefly illustrate how we address these two challenges.

Implicit Forms of Distribution Classes. Our first technical innovation is that we implicitly define the classes of distributions for the generalized round elimination by quantifying the relationship between each distribution in the class and the original hard input distribution. The discussion below is again a simplified version of the actual construction, whose details can be found in Section 3.2.

The distribution classes for Alice and Bob are defined in a similar way. Here we use Alice for example, and define the distribution classes $\mathcal{D}^A_\ell$ ($\ell = 0, 1, \ldots$) for Alice. The combined distribution class will be denoted by $\mathcal{D}_\ell = (\mathcal{D}^A_\ell, \mathcal{D}^B_\ell)$, where $\mathcal{D}^B_\ell$ is the one defined for Bob.

Let S^A_ℓ be the set of all possible local means at Alice's side for arms in levels $\ell, \ldots, L$, and let $\varsigma = n^{1/L}$. For each level $\ell = 0, 1, \ldots, L$, define input class $\mathcal{D}^A_\ell$ to be the set of distributions σ^A with support S^A_ℓ such that

$$\forall x, y \in S^A_\ell : \frac{\Pr_{\mu^A \sim \sigma^A}[\mu^A = x]}{\Pr_{\mu^A \sim \sigma^A}[\mu^A = y]} = \frac{\Pr_{\mu^A \sim \pi^A}[\mu^A = x]}{\Pr_{\mu^A \sim \pi^A}[\mu^A = y]} \cdot e^{\pm \frac{\ell}{\varsigma}}. \quad (3)$$

Note that $\mathcal{D}^A_0 = \{\pi^A\}$ where π^A is the original input distribution at Alice's side. Intuitively, Equation (3) states that for any distribution $\sigma^A \in \mathcal{D}^A_\ell$, the ratio between the probability mass on any two possible mean values in σ^A is close to that in the original input distribution π^A. Consequently, if the original input distribution π^A is quite "uncertain", then any distribution $\sigma^A \in \mathcal{D}^A_\ell$ is also quite uncertain. The extra $e^{\pm \frac{\ell}{\varsigma}}$ is a relaxation term that counts the influence of the pull outcomes in the first ℓ rounds on the posterior distribution of π^A.

We have the following lemma. Its formal statement can be found in Lemma 11 in Section 3.2. We slightly rewrite and simplify the statement here for the illustration purpose.

Lemma 3. *For any $\ell \in \{0, 1, \ldots, L-1\}$, any distribution $\sigma^A \in \mathcal{D}^A_\ell$, and any good sequence of pull outcomes $\theta = (\theta_1, \ldots, \theta_q)$ in the current round, the posterior distribution of σ^A after observing a sequence of pull outcomes being θ and conditioning on the mean of the arm $\mu^A \in S^A_{\ell+1}$, denoted by $(\sigma^A \mid \theta, \mu^A \in S^A_{\ell+1})$, belongs to the distribution class $\mathcal{D}^A_{\ell+1}$.*

On the other hand, we can also show that the pull sequence θ is good with high probability if its length is not too large, which holds if there is a time budget constraint.

Lemma 3 helps in establishing the foundation of the induction in the round elimination without having to go through the spelling of the posterior distributions after a round of arm pulls.

A Lower Bound for Adaptive CL Algorithms. Our second technical contribution is to prove the lower bound for adaptive CL algorithms directly, instead of via a reduction from a lower bound for non-adaptive CL algorithms. The details can be found in Section 3.3.

Let us first recall the proof for non-adaptive algorithms in [24]. After the first round of pulls, we set a threshold η and *publish* those arms who have been pulled more than η times in the first round; we call these arms the *heavy arms*. By publishing an arm we mean revealing its mean to all agents; note that this will only make the problem easier, and consequently make the lower bound proof stronger. This arm publishing procedure is what we formerly referred to as the " input massage".

We use the arm publishing procedure to ensure that the means of remaining arms belong to the next class (i.e., $\mu^A \in S^A_{\ell+1}$). For the set of distribution classes $\mathcal{D}^A_\ell$ ($\ell = 0, 1, \ldots$) used in this paper, we can use Lemma 3 to show that the posterior distribution of some $\sigma^A \in \mathcal{D}^A_\ell$, after the publishing procedure, belongs to the next distribution class $\mathcal{D}^A_{\ell+1}$.

In order for the induction to proceed, we need to make sure that if we publish all heavy arms, the probability of the best arm being published is small, since otherwise the problem would already be solved and the round elimination process *cannot* continue. This is easy to do with non-adaptive algorithms, because the whole pull sequence and consequently the set of heavy arms are determined at the beginning of each round. If the time budget is small, then the number of heavy arms must be small. Consequently, the probability that the set of heavy arms contain the best arm is also small, because all arms are almost equally uncertain at the beginning of each round. In other words, the set of heavy arms would be an almost *random subset* of all arms.

Adaptive algorithms, however, can utilize their adaptivity to look for arms with high means and make more pulls on those arms.

To handle this challenge, we choose to explicitly analyze for each heavy arm its probability of being the best arm after the first round of pulls, and then show that the sum of these probabilities is small. This analysis is much more complicated than that for the non-adaptive algorithms. We try to illustrate the main ideas below.

The key to the analysis for each individual arm is that, because of the interleaved mean structure, Alice misses most information of half of the terms held by Bob. Without this information, her adaptivity cannot help much in the task of identifying which arm is more likely to be the best arm. On the other hand, the time budget constraint also prevents Alice from extracting and revealing to Bob too much information about her local means of arms which are *not* published in the next round (see the algorithm *Arm Publishing and Additional Pulls* in Section 3.3 for details on how we publish arms). A similar argument holds for Bob. Despite appearing natural, it is highly non-trivial to put this intuition into a formal proof since we need to carefully bound the "help" of the historical information exchange. The adaptivity of the algorithm further complicates the description of the posterior distribution of the arms after one round of pulls. Fortunately, our implicit representation of the distribution classes is flexible enough to handle this additional complexity.

Finally, we would like to mention that due to technical needs, in each step of our induction we have to "consume" multiple, but still $O(1)$, levels out of the L levels of arms, but this will *not* change the asymptotic round bound.

Generalizing to K Parties. Finally, we comment that we can easily generalize the lower bound for 2 agents to K agents via a reduction. See Section 3.4 for details.

3 THE IMPOSSIBILITY RESULT

In this section, we give the proof to Theorem 1.

We start with the case when there are two agents (i.e., $K = 2$), and then generalize the results to all K. Below are a few notations that we will be using in this section.

- R: The number of rounds used by the algorithm. We will focus on the range $1 \le R \le \frac{\log n}{24 \log\log n}$.
- $L \triangleq 6R$: The number of terms in the means of arms in the hard input distribution.
- $\eta \triangleq n^{\frac{1}{2L}} = n^{\frac{1}{12R}}$: Intuitively, it is the ratio between the maximum contributions of consecutive terms in the mean construction. For $1 \le R \le \frac{\log n}{24 \log\log n}$, we always have $\eta \ge \log^2 n$.
- $\zeta \triangleq \frac{\sqrt{\eta}}{2^7} = \frac{n^{\frac{1}{24R}}}{2^7}$: A parameter related to the time of the CL algorithm.
- $\gamma \triangleq \frac{\eta}{2^7} = \frac{n^{\frac{1}{12R}}}{2^7} = \Theta(\zeta^2)$: A parameter for the convenience of the presentation.
- $\mathrm{Ber}(\mu)$ denotes the Bernoulli distribution with mean μ.

For convenience, when we write $c = a \pm b$ (or $c \pm d = a \pm b$), we mean $c \in [a - b, a + b]$ (or $[c - d, c + d] \subseteq [a - b, a + b]$). Without this simplification, some formulas may be difficult to read.

We will use the following standard concentration bound.

Lemma 4 (Chernoff-Hoeffding Inequality). *Let $X_1, \ldots, X_n \in [a_i, b_i]$ be independent random variables. Let $X = \sum_{i=1}^n X_i$. For any $t \ge 0$, it holds that*

$$\Pr[X \ge \mathbf{E}[X] + t] \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right), \quad \text{and}$$
$$\Pr[X \le \mathbf{E}[X] - t] \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

In the rest of this section, we first introduce the hard input distribution that we use to prove the lower bound and discuss its properties. We then introduce the classes of distributions on which we will perform the generalized round elimination. After these preparation steps, we present our main lower bound proof for $K = 2$, and then extend it to the general case.

3.1 The Hard Input Distribution (When $K = 2$) and Its Properties.

Define random variable

$$\mu = \mu(X_1, \ldots, X_L) = \frac{1}{2} + \sum_{\ell=1}^{L} \frac{X_\ell}{\eta^\ell}, \tag{4}$$

where for each $\ell \in [L]$, $X_\ell \sim \mathrm{Ber}\left(\eta^{-2}\right)$ are drawn independently. Let π be the distribution of random variable μ.

Let $(\mu_1, \ldots, \mu_n) \sim \pi^{\otimes n}$, where μ_i is the global mean of arm i. We divide each μ_i into two local means μ_i^A and μ_i^B for Alice and Bob respectively, where

$$\mu^A = \frac{1}{2} + \sum_{\ell : 1 \le 2\ell+1 \le L} \frac{2X_{2\ell+1}}{\eta^{2\ell+1}}, \quad \text{and} \quad \mu^B = \frac{1}{2} + \sum_{\ell : 1 \le 2\ell \le L} \frac{2X_{2\ell}}{\eta^{2\ell}}.$$

That is, Alice takes all odd terms in the summation of (4), and Bob takes all even terms in the summation of (4); the factor 2 is just to make sure that $\mu = (\mu^A + \mu^B)/2$. It is clear that μ^A and μ^B are independent, because they depend on disjoint subsets of $\{X_1, \ldots, X_L\}$. Let π^A and π^B be the underlying distributions of random variables μ^A and μ^B, respectively. We can write $\pi = (\pi^A, \pi^B)$.

Key Properties of the Support of Distribution $\pi = (\pi^A, \pi^B)$. For each $\ell \in \{0, 1, \ldots, L\}$, we define the following two sets:

$$S_\ell^A \triangleq \left\{ \frac{1}{2} + \sum_{k : 1 \le 2k+1 \le \ell} \frac{2}{\eta^{2k+1}} + \sum_{k : \ell < 2k+1 \le L} \frac{2X_{2k+1}}{\eta^{2k+1}} \;\middle|\; X_{2k+1} \in \{0, 1\} \right\}, \tag{5}$$

and

$$S_\ell^B \triangleq \left\{ \frac{1}{2} + \sum_{k : 1 \le 2k \le \ell} \frac{2}{\eta^{2k}} + \sum_{k : \ell < 2k \le L} \frac{2X_{2k}}{\eta^{2k}} \;\middle|\; X_{2k} \in \{0, 1\} \right\}. \tag{6}$$

Intuitively, the set S_ℓ^A consists of values in $\mathrm{supp}(\pi^A)$ with $X_1 = X_3 = \ldots = X_{\ell'} = 1$, where ℓ' is the largest odd integer no more than ℓ. And the set S_ℓ^B consists of values in $\mathrm{supp}(\pi^B)$ with $X_2 = X_4 = \ldots = X_{\ell'} = 1$, where ℓ' is the largest even integer no more than ℓ. It is easy to see that

$$\mathrm{supp}(\pi^A) = S_0^A \supset S_1^A = S_2^A \supset S_3^A = S_4^A \supset \ldots,$$

and

$$\mathrm{supp}(\pi^B) = S_0^B = S_1^B \supset S_2^B = S_3^B \supset S_4^B = \ldots.$$

Let $\theta = (\theta_1, \ldots, \theta_q) \in \{0,1\}^q$ be a sequence of q pull outcomes on an arm with mean x. For convenience, we write

$$p(\theta \mid x) \triangleq \Pr_{\Theta \sim \mathrm{Ber}(x)^{\otimes q}}[\Theta = \theta]. \tag{7}$$

We have

$$p(\theta \mid x) = \prod_{j=1}^{q} x^{\theta_j}(1-x)^{1-\theta_j}. \tag{8}$$

The following two lemmas give key properties of the sets S_ℓ^A and S_ℓ^B. Intuitively, it says that if we can only pull the arm whose mean is $x \in S_\ell^A$ (or $x \in S_\ell^B$) for a small number of times, then it is hard to differentiate its true mean x from other values in S_ℓ^A (or S_ℓ^B) based on the pull outcomes. Due to the space constraints, we leave the proof of this technical lemma to the full version of this paper [13].

Lemma 5. *For any $x \in S_\ell^A$, let $\Theta = (\Theta_1, \ldots, \Theta_q)$ be a sequence of $q \in \left[\eta^3, \frac{\eta^{2\ell-1}}{2^7}\right]$ pull outcomes on an arm with mean x. For any $y \in S_\ell^A$ $(y \neq x)$, we have*

$$\Pr_{\Theta \sim \mathrm{Ber}(x)^{\otimes q}}\left[\frac{p(\Theta \mid y)}{p(\Theta \mid x)} < e^{-\frac{2}{\eta}}\right] \leq e^{-\frac{\eta}{2^{10}}},$$

and

$$\Pr_{\Theta \sim \mathrm{Ber}(x)^{\otimes q}}\left[\frac{p(\Theta \mid y)}{p(\Theta \mid x)} > e^{\frac{2}{\eta}}\right] \leq e^{-\frac{\eta}{2^{10}}}.$$

The following lemma is symmetric to Lemma 5, and can be proved using a similar line of arguments.

Lemma 5′. *For any $x \in S_\ell^B$, let $\Theta = (\Theta_1, \ldots, \Theta_q)$ be a sequence of $q \in [\eta^3, \frac{\eta^{2\ell-1}}{2^7}]$ pull outcomes on an arm with mean x. For any $y \in S_\ell^B$ $(y \neq x)$, we have*

$$\Pr_{\Theta \sim \mathrm{Ber}(x)^{\otimes q}}\left[\frac{p(\Theta \mid y)}{p(\Theta \mid x)} < e^{-\frac{2}{\eta}}\right] \leq e^{-\frac{\eta}{2^{10}}}, \tag{9}$$

and

$$\Pr_{\Theta \sim \mathrm{Ber}(x)^{\otimes q}}\left[\frac{p(\Theta \mid y)}{p(\Theta \mid x)} > e^{\frac{2}{\eta}}\right] \leq e^{-\frac{\eta}{2^{10}}}. \tag{10}$$

Instance Complexity under Distribution $\pi^{\otimes n}$. We now try to bound the instance complexity of an input sampled from distribution $\pi^{\otimes n}$. Let $\mu_* = \frac{1}{2} + \sum_{\ell=1}^{L} \frac{1}{\eta^\ell}$. The following event stands for the case when there is only one best arm with mean μ_*.

$$\mathcal{E}_0 : \exists \text{ unique } i^* \in [n] \text{ s.t. } \mu_{i^*} = \mu_*. \tag{11}$$

The following lemma shows that $\mathcal{E}_0$ holds with at least a constant probability.

Lemma 6. $\Pr_{(\mu_1,\ldots,\mu_n)\sim\pi^{\otimes n}}[\mathcal{E}_0] \geq 1/e$.

Proof. We have

$$\begin{aligned}
\Pr_{(\mu_1,\ldots,\mu_n)\sim\pi^{\otimes n}}[\mathcal{E}_0] &= \sum_{i=1}^{n}\left(\Pr[\mu_i = \mu_*]\prod_{j\in[n], j\neq i}\Pr[\mu_j \neq \mu_*]\right) \\
&= n\cdot\frac{1}{\eta^{2L}}\cdot\left(1-\frac{1}{\eta^{2L}}\right)^{n-1} \\
&= \left(1-\frac{1}{n}\right)^{n-1} \geq \frac{1}{e}.
\end{aligned}$$

□

We now try to upper bound the instance complexity of inputs sampled from distribution $\pi^{\otimes n}$, conditioned on event $\mathcal{E}_0$.

Lemma 7. $\mathbf{E}_{(\mu_1,\ldots,\mu_n)\sim\pi^{\otimes n}}[H \mid \mathcal{E}_0] \leq \eta^{2+2L}L$.

Proof. Conditioned on $\mathcal{E}_0$, let i^* be the unique best arm with mean μ_*. We can write

$$\begin{aligned}
&\mathbf{E}_{(\mu_1,\ldots,\mu_n)\sim\pi^{\otimes n}}[H \mid \mathcal{E}_0] \\
=\ &\mathbf{E}_{(\mu_1,\ldots,\mu_n)\sim\pi^{\otimes n}}\left[\sum_{i\in[n], i\neq i^*}(\mu_* - \mu_i)^{-2}\,\middle|\,\max_{i\neq i^*}\{\mu_i\} < \mu_*\right] \\
=\ &(n-1)\,\mathbf{E}_{\mu\sim\pi}\left[(\mu_* - \mu)^{-2} \mid \mu < \mu_*\right].
\end{aligned} \tag{12}$$

To upper bound (12), we partition the values in $\mathrm{supp}(\pi)$ into L disjoint sets. For each $\ell \in [L]$, we define

$$P_\ell \triangleq \left\{\frac{1}{2} + \sum_{k=1}^{\ell-1}\frac{1}{\eta^k} + \frac{0}{\eta^\ell} + \sum_{k=\ell+1}^{L}\frac{X_k}{\eta^k}\,\middle|\,(X_\ell, \ldots, X_L) \in \{0,1\}^{L-\ell+1}\right\}. \tag{13}$$

Clearly, we have $\bigcup_{\ell=1}^{L} P_\ell = \mathrm{supp}(\pi) \setminus \{\mu_*\}$, and for any $\ell \in [L]$ and any $\mu \in P_\ell$,

$$\mu_* - \mu \geq \frac{1}{\eta^\ell}. \tag{14}$$

Plugging (14) to (12), we have

$$\begin{aligned}
(12) &\leq (n-1)\cdot\sum_{\ell=1}^{L}\left(\Pr_{\mu\sim\pi}[\mu \in P_\ell \mid \mu < \mu_*]\cdot\eta^{2\ell}\right) \\
&= (n-1)\cdot\sum_{\ell=1}^{L}\left(\frac{\Pr_{\mu\sim\pi}[\mu \in P_\ell, \mu < \mu_*]}{\Pr_{\mu\sim\pi}[\mu < \mu_*]}\cdot\eta^{2\ell}\right) \\
&= (n-1)\cdot\sum_{\ell=1}^{L}\left(\frac{\left(\frac{1}{\eta^2}\right)^{\ell-1}\left(1-\frac{1}{\eta^2}\right)}{1-\frac{1}{n}}\cdot\eta^{2\ell}\right) \\
&= n\cdot\left(1-\frac{1}{\eta^2}\right)\cdot\sum_{\ell=1}^{L}\eta^2 \\
&\leq \eta^2 Ln = \eta^{2+2L}L.
\end{aligned}$$

□

Define event

$$\mathcal{E}_1 : \mathcal{E}_0 \text{ holds } \wedge (H < 2\eta^{2+2L}L). \tag{15}$$

By Markov's inequality and Lemma 7, we have

$$\Pr[H \geq 2\eta^{2+2L}L \mid \mathcal{E}_0] \leq 1/2,$$

which, combined with Lemma 6, gives the following lemma.

Lemma 8. $\Pr[\mathcal{E}_1] \geq 1/(2e)$.

Hard Input Distribution. The hard input distribution we use for proving the lower bound is $(\pi^{\otimes n} \mid \mathcal{E}_1)$. That is, the probability mass is uniformly distributed among the support of $\pi^{\otimes n}$ *except* those instances in which there is 0 or multiple arms with means μ_* (i.e., when $\mathcal{E}_0$ does *not* hold) and those of which the instance complexity is more than $2\eta^{2+2L}L$.

In our lower bound proof in Section 3.3, we will spend most of our time working on the input distribution $\pi^{\otimes n}$. We will switch to $(\pi^{\otimes n} \mid \mathcal{E}_1)$ at the end of the proof.

3.2 Classes of Hard Distributions

In this section, we define the classes of hard distributions that we use for the generalized round elimination. We start by introducing a concept called *good pull outcome sequences*.

Good Pull Outcome Sequences. We say a pull outcome sequence $\theta = (\theta_1, \ldots, \theta_q)$ *good* w.r.t. S_ℓ^A if for any $x, y \in S_\ell^A$, $p(\theta|x)$ and $p(\theta|y)$ are close. More precisely, we define the set of good pull outcome sequences w.r.t. S_ℓ^A as

$$G_\ell^A \triangleq \left\{ \theta \;\middle|\; \forall x, y \in S_\ell^A : \frac{p(\theta \mid x)}{p(\theta \mid y)} \in \left[e^{-\frac{4}{\eta}}, e^{\frac{4}{\eta}} \right] \right\}. \tag{16}$$

Similarly, we define the set of good pull outcome sequences w.r.t. S_ℓ^B as

$$G_\ell^B \triangleq \left\{ \theta \;\middle|\; \forall x, y \in S_\ell^B : \frac{p(\theta \mid x)}{p(\theta \mid y)} \in \left[e^{-\frac{4}{\eta}}, e^{\frac{4}{\eta}} \right] \right\}. \tag{17}$$

The following lemma says that when the length of sequence q is not large, then with high probability, the pull outcome sequence is good.

Lemma 9. *Let $\Theta = (\Theta_1, \ldots, \Theta_q)$ be a sequence of $q \in [\eta^3, \frac{\eta^{2\ell-1}}{2^7}]$ pull outcomes on an arm with mean μ^A. For any distribution σ^A with support S_ℓ^A, we have $\Pr_{\mu^A \sim \sigma^A, \Theta \sim \mathrm{Ber}(\mu^A)^{\otimes q}} \left[\Theta \notin G_\ell^A \right] \le n^{-10}$.*

Proof. By the law of total probability, we write

$$\begin{aligned} &\Pr_{\mu^A \sim \sigma^A, \Theta \sim \mathrm{Ber}(\mu^A)^{\otimes q}} \left[\Theta \notin G_\ell^A \right] \\ = &\sum_{z \in S_\ell^A} \left(\Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\Theta \notin G_\ell^A \right] \Pr_{\mu^A \sim \sigma^A} [\mu^A = z] \right). \end{aligned} \tag{18}$$

By definition, $\theta \notin G_\ell^A$ if and only if there exists a pair of $x, y \in S_\ell^A$ such that

$$\frac{p(\theta \mid x)}{p(\theta \mid y)} > e^{\frac{4}{\eta}} \quad \text{or} \quad \frac{p(\theta \mid x)}{p(\theta \mid y)} < e^{-\frac{4}{\eta}}. \tag{19}$$

We first consider the case $\frac{p(\theta|x)}{p(\theta|y)} > e^{\frac{4}{\eta}}$. In this case, for any $z \in S_\ell^A$, we have

$$\frac{p(\theta \mid x)}{p(\theta \mid z)} > e^{\frac{2}{\eta}} \quad \text{or} \quad \frac{p(\theta \mid y)}{p(\theta \mid z)} < e^{-\frac{2}{\eta}}.$$

Consequently,

$$\begin{aligned} &\Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\frac{p(\Theta \mid x)}{p(\Theta \mid y)} > e^{\frac{4}{\eta}} \right] \\ \le &\Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\frac{p(\Theta \mid x)}{p(\Theta \mid z)} > e^{\frac{2}{\eta}} \right] + \Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\frac{p(\Theta \mid y)}{p(\Theta \mid z)} < e^{-\frac{2}{\eta}} \right] \\ \le &\; 2e^{-\frac{\eta}{2^{10}}}, \end{aligned} \tag{20}$$

where in the last inequality we have used Lemma 5.

By a similar argument, we can show

$$\Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\frac{p(\Theta \mid x)}{p(\Theta \mid y)} < e^{-\frac{4}{\eta}} \right] \le 2e^{-\frac{\eta}{2^{10}}}. \tag{21}$$

By (20), (21), and the definition of G_ℓ^A in (16), we have

$$\begin{aligned} &\Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\Theta \notin G_\ell^A \right] \\ \le &\sum_{x, y \in S_\ell^A} \left(\Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\frac{p(\Theta \mid x)}{p(\Theta \mid y)} > e^{\frac{4}{\eta}} \right] + \right. \\ &\left. \Pr_{\Theta \sim \mathrm{Ber}(z)^{\otimes q}} \left[\frac{p(\Theta \mid x)}{p(\Theta \mid y)} < e^{-\frac{4}{\eta}} \right] \right) \\ \le &\; 4 \left| S_\ell^A \right|^2 e^{-\frac{\eta}{2^{10}}}, \end{aligned} \tag{22}$$

where in the last inequality we have taken a union bound on all pairs $(x, y) \in S_\ell^A \times S_\ell^A$.

Plugging (22) to (18), we have

$$\begin{aligned} &\Pr_{\mu^A \sim \sigma^A, \Theta \sim \mathrm{Ber}(\mu^A)^{\otimes q}} \left[\Theta \notin G_\ell^A \right] \\ \le &\sum_{z \in S_\ell^A} \left(4 \left| S_\ell^A \right|^2 e^{-\frac{\eta}{2^{10}}} \Pr_{\mu^A \sim \sigma^A} [\mu^A = z] \right) \\ = &\; 4 \left| S_\ell^A \right|^2 e^{-\frac{\eta}{2^{10}}} \le n^{-10}, \end{aligned}$$

where the last inequality is due to $\eta \ge \log^2 n$. □

The following lemma is symmetric to Lemma 9, and can be proved using a similar line of arguments.

Lemma 9′. *Let $\Theta = (\Theta_1, \ldots, \Theta_q)$ be a sequence of $q \in [\eta^3, \frac{\eta^{2\ell-1}}{2^7}]$ pull outcomes on an arm with mean μ^B. For any distribution σ^B with support S_ℓ^B, we have $\Pr_{\mu^B \sim \sigma^B, \Theta \sim \mathrm{Ber}(\mu^B)^{\otimes q}} \left[\Theta \notin G_\ell^B \right] \le n^{-10}$.*

Classes of Distributions $\mathcal{D}_\ell^A$, $\mathcal{D}_\ell^B$, and $\mathcal{D}_\ell$ ($\ell = 0, 1, \ldots, L$). We are now ready to define classes of input distributions on which we will perform the induction.

For $\ell \in \{0, 1, \ldots, L\}$, we define $\mathcal{D}_\ell^A$ to be the class of distributions σ^A with support S_ℓ^A such that

$$\forall x, y \in S_\ell^A : \frac{\Pr_{\mu^A \sim \sigma^A}[\mu^A = x]}{\Pr_{\mu^A \sim \sigma^A}[\mu^A = y]} = \frac{\Pr_{\mu^A \sim \pi^A}\left[\mu^A = x\right]}{\Pr_{\mu^A \sim \pi^A}\left[\mu^A = y\right]} \cdot e^{\pm \frac{4\ell}{\eta}}. \tag{23}$$

Similarly, we define $\mathcal{D}_\ell^B$ to be the class of distributions σ^B with support S_ℓ^B such that

$$\forall x, y \in S_\ell^B : \frac{\Pr_{\mu^B \sim \sigma^B}[\mu^B = x]}{\Pr_{\mu^B \sim \sigma^B}[\mu^B = y]} = \frac{\Pr_{\mu^B \sim \pi^B}\left[\mu^B = x\right]}{\Pr_{\mu^B \sim \pi^B}\left[\mu^B = y\right]} \cdot e^{\pm \frac{4\ell}{\eta}}. \tag{24}$$

Let $\mathcal{D}_\ell = (\mathcal{D}_\ell^A, \mathcal{D}_\ell^B)$. We say a distribution $\sigma = (\sigma^A, \sigma^B) \in \mathcal{D}_\ell$ iff $\sigma^A \in \mathcal{D}_\ell^A$ and $\sigma^B \in \mathcal{D}_\ell^B$.

We have the following simple fact.

Fact 10. *$\mathcal{D}_0^A = \{\pi^A\}$, $\mathcal{D}_0^B = \{\pi^B\}$, and $\mathcal{D}_0 = \{\pi\}$.*

The following lemma shows a key property of distribution classes $\mathcal{D}_\ell^A$. Intuitively, if the mean of an arm follows a distribution $\sigma^A \in \mathcal{D}_\ell^A$, then after observing a good sequence of pulls that belongs to G_k^A for a $k \ge \ell + 1$, the posterior distribution of the arm belongs to distribution class $\mathcal{D}_k^A$.

Lemma 11. *For any $\ell \in \{0, 1, \ldots, L-1\}$, any $k \in \{\ell+1, \ldots, L\}$, any distribution $\sigma^A \in \mathcal{D}^A_\ell$, and any good sequence of pull outcomes $\theta = (\theta_1, \ldots, \theta_q) \in G^A_k$, the posterior distribution of σ^A after observing a sequence of pull outcomes being θ and conditioning on the mean of the arm $\mu^A \in S^A_k$, denoted by $(\sigma^A \mid \theta, \mu^A \in S^A_k)$, belongs to the distribution class $\mathcal{D}^A_k$.*

Proof. Fix two arbitrary fixed values $x, y \in S^A_k$. By Bayes' theorem, we have

$$\begin{aligned}
&\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\mu^A = x \mid \Theta = \theta, \mu^A \in S^A_k] \\
=\ & \frac{\Pr_{\substack{\mu^A\sim\sigma^A,\\ \Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}}[\Theta = \theta, \mu^A \in S^A_k \mid \mu^A = x] \Pr_{\mu^A\sim\sigma^A}[\mu^A = x]}{\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\Theta = \theta, \mu^A \in S^A_k]} \\
=\ & \frac{\Pr_{\Theta\sim\mathrm{Ber}(x)^{\otimes q}}[\Theta = \theta] \Pr_{\mu^A\sim\sigma^A}[\mu^A = x]}{\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\Theta = \theta, \mu^A \in S^A_k]} \\
=\ & \frac{p(\theta \mid x) \cdot \Pr_{\mu^A\sim\sigma^A}[\mu^A = x]}{\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\Theta = \theta, \mu^A \in S^A_k]}, \qquad (25)
\end{aligned}$$

where in the second equality we have used the fact $\mu^A = x \in S^A_k$, and in the third equality we have used the definition of $p(\theta|x)$ in (7).

Similarly, we have

$$\begin{aligned}
&\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\mu^A = y \mid \Theta = \theta, \mu^A \in S^A_k] \\
=\ & \frac{p(\theta \mid y) \cdot \Pr_{\mu^A\sim\sigma^A}[\mu^A = y]}{\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\Theta = \theta, \mu^A \in S^A_k]}.
\end{aligned}$$

We next have

$$\begin{aligned}
&\frac{\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\mu^A = x \mid \Theta = \theta, \mu^A \in S^A_k]}{\Pr_{\mu^A\sim\sigma^A,\Theta\sim\mathrm{Ber}(\mu^A)^{\otimes q}}[\mu^A = y \mid \Theta = \theta, \mu^A \in S^A_k]} \\
\overset{(25),(26)}{=}\ & \frac{\Pr_{\mu^A\sim\sigma^A}[\mu^A = x]}{\Pr_{\mu^A\sim\sigma^A}[\mu^A = y]} \cdot \frac{p(\theta \mid x)}{p(\theta \mid y)} \qquad (26) \\
=\ & \frac{\Pr_{\mu^A\sim\pi^A}[\mu^A = x]}{\Pr_{\mu^A\sim\pi^A}[\mu^A = y]} \cdot e^{\pm\frac{4\ell}{\eta}} \cdot e^{\pm\frac{4}{\eta}} \qquad (27) \\
=\ & \frac{\Pr_{\mu^A\sim\pi^A}[\mu^A = x]}{\Pr_{\mu^A\sim\pi^A}[\mu^A = y]} \cdot e^{\pm\frac{4k}{\eta}}, \qquad (28)
\end{aligned}$$

where from (26) to (27) we have used the definition of distribution class $\mathcal{D}^A_\ell$ in (23) and the fact $S^A_k \subseteq S^A_\ell$, as well as the definition of G^A_ℓ and the fact $\theta \in G^A_k$. From (27) to (28) we have used the fact $k \geq \ell + 1$.

By (28) and the definition of $\mathcal{D}^A_k$ in (23), we have

$$(\sigma^A \mid \theta, \mu^A \in S^A_k) \in \mathcal{D}^A_k.$$

□

The following lemma is symmetric to Lemma 11, and can be proved using a similar line of arguments.

Lemma 11′. *For any $\ell \in \{0, 1, \ldots, L-1\}$, any $k \in \{\ell+1, \ldots, L\}$, any distribution $\sigma^B \in \mathcal{D}^B_\ell$, and any good sequence of pull outcomes $\theta = (\theta_1, \ldots, \theta_q) \in G^B_k$, the posterior distribution of σ^B after observing a sequence of pull outcomes being θ and conditioning on the mean of the arm $\mu^B \in S^B_k$, denoted by $(\sigma^B \mid \theta, \mu^B \in S^B_k)$, belongs to the distribution class $\mathcal{D}^B_k$.*

Let

$$\mu^A_* = \frac{1}{2} + 2 \sum_{\ell:1\leq 2\ell+1\leq L} \frac{1}{\eta^{2\ell+1}}$$

be the mean of local best arm at Alice's side, and let

$$\mu^B_* = \frac{1}{2} + 2 \sum_{\ell:1\leq 2\ell\leq L} \frac{1}{\eta^{2\ell}}$$

be the mean of local best arm at Bob's side. The following lemma shows that an arm whose mean is distributed according to $\sigma \in \mathcal{D}^A_\ell$ has a small probability being a local best arm.

Lemma 12. *For any $\ell \in \{0, 1, \ldots, L\}$, and any $\sigma^A \in \mathcal{D}^A_\ell$, we have $\Pr_{\mu^A\sim\sigma^A}[\mu^A = \mu^A_*] \leq e^{\frac{4\ell}{\eta}} \eta^{-2d_1}$, where $d_1 = |\{k \in \mathbb{Z} \mid \ell < 2k+1 \leq L\}|$ is the number of odd integers in the set $\{\ell+1, \ldots, L\}$.*

Proof. We first define a few quantities. Let

$$\rho_{\max} = \max_{x\in S^A_\ell} \frac{\Pr_{\mu^A\sim\sigma^A}[\mu^A = x]}{\Pr_{\mu^A\sim\pi^A}[\mu^A = x \mid \mu^A \in S^A_\ell]}, \qquad (29)$$

and slightly abusing the notation, let $x \in S^A_\ell$ be the value that achieves $\rho_{\max}$. Let

$$\rho_{\min} = \min_{y\in S^A_\ell} \frac{\Pr_{\mu^A\sim\sigma^A}[\mu^A = y]}{\Pr_{\mu^A\sim\pi^A}[\mu^A = y \mid \mu^A \in S^A_\ell]}, \qquad (30)$$

and let $y \in S^A_\ell$ be the value that achieves $\rho_{\min}$. It is clear that $\rho_{\min} \leq 1 \leq \rho_{\max}$. We also have

$$\begin{aligned}
\frac{\rho_{\max}}{\rho_{\min}} &= \frac{\Pr_{\mu^A\sim\sigma^A}[\mu^A = x]}{\Pr_{\mu^A\sim\sigma^A}[\mu^A = y]} \cdot \frac{\Pr_{\mu^A\sim\pi^A}[\mu^A = y \mid \mu^A \in S^A_\ell]}{\Pr_{\mu^A\sim\pi^A}[\mu^A = x \mid \mu^A \in S^A_\ell]} \\
&= \frac{\Pr_{\mu^A\sim\sigma^A}[\mu^A = x]}{\Pr_{\mu^A\sim\sigma^A}[\mu^A = y]} \cdot \frac{\Pr_{\mu^A\sim\pi^A}[\mu^A = y]}{\Pr_{\mu^A\sim\pi^A}[\mu^A = x]} \\
&= e^{\pm\frac{4\ell}{\eta}}. \qquad (31)
\end{aligned}$$

where in the second equation we have used the fact that $x, y \in S^A_\ell$, and in the last equation we have used the fact that $\sigma^A \in \mathcal{D}^A_\ell$.

We thus have $e^{-\frac{4\ell}{\eta}} \leq \rho_{\min} \leq 1 \leq \rho_{\max} \leq e^{\frac{4\ell}{\eta}}$. The last inequality $\rho_{\max} \leq e^{\frac{4\ell}{\eta}}$ implies

$$\begin{aligned}
\Pr_{\mu^A\in\sigma^A}[\mu^A = \mu^A_*] &\leq \Pr_{\mu^A\sim\pi^A}\left[\mu^A = \mu^A_* \,\middle|\, \mu^A \in S^A_\ell\right] \cdot e^{\frac{4\ell}{\eta}} \\
&= \left(\frac{1}{\eta^2}\right)^{d_1} \cdot e^{\frac{4\ell}{\eta}}, \qquad (32)
\end{aligned}$$

where $d_1 = \{k \in \mathbb{Z} \mid \ell < 2k+1 \leq L\}$. □

The following lemma is similar to Lemma 12, and can be proved using a similar line of arguments.

Lemma 12′. *For any $\ell \in \{0,1,\dots,L\}$, and any $\sigma^B \in \mathcal{D}^B_\ell$, we have $\Pr_{\mu^B\sim\sigma^B}[\mu^B = \mu^B_*] \le e^{\frac{4\ell}{\eta}}\eta^{-2d_0}$, where $d_0 = |\{k \in \mathbb{Z} \mid \ell < 2k \le L\}|$ is the number of even integers in the set $\{\ell+1,\dots,L\}$.*

3.3 The Lower Bound for $K = 2$

In this section, we show the following lower bound result for the case of two agents.

Theorem 13. *For any $1 \le R \le \frac{\log n}{24\log\log n}$, any R-round 2-agent algorithm that solves n-arm BAI in the heterogeneous CL model with probability 0.99 needs to use at least $Hn^{\frac{1}{25R}}$ time.*

By Yao's Minimax Lemma, we can just prove for any deterministic algorithm over the hard input distribution $(\pi^{\otimes n} \mid \mathcal{E}_1)$.

We will first analyze the success probability of any deterministic algorithm $\mathcal{A}$ on input distribution $\pi^{\otimes n}$. We say $\mathcal{A}$ succeeds on an input instance I if $\mathcal{A}$ outputs an index i such that $\mu_i = \mu_*$. Note that there could be multiple $i \in [n]$ such that $\mu_i = \mu_*$ and $\mathcal{A}$ can output any index in this set.

The Induction Step. Let the quantity λ_r be the largest success probability of a $(R-r)$-round $2\zeta\eta^{2+2L}L$-time algorithm on some input distribution in $\mathcal{D}^{\otimes\kappa}_{6r}$ for some $\kappa \in [n]$. That is,

$$\lambda_r \triangleq \max_{\kappa\in[n]} \max_{\nu\in\mathcal{D}^{\otimes\kappa}_{6r}} \max_{\mathcal{A}} \Pr_{I\sim\nu}[\mathcal{A} \text{ succeeds on } I], \tag{33}$$

where $\max_{\mathcal{A}}$ runs over all algorithms $\mathcal{A}$ that use $(R-r)$ rounds and $2\zeta\eta^{2+2L}L$ time.

The following lemma connects the error probabilities λ_r and λ_{r+1}, and is the key for the induction.

Lemma 14. *For any $r = 1,\dots,R-1$, it holds that*

$$\lambda_r \le \lambda_{r+1} + 4e^{\frac{10L}{\eta}}L^2\eta^{-\frac{5}{2}} + n^{-5}.$$

The rest of Section 3.3 devotes to the proof of Lemma 14. Slightly abusing the notation, let κ be the value that maximizes the error in the definition of λ_r (the first max in (33)). We write $\nu = (\sigma^A_1,\dots,\sigma^A_\kappa,\sigma^B_1,\dots,\sigma^B_\kappa)$. Since $\nu \sim \mathcal{D}^{\otimes\kappa}_{6r} = \left((\mathcal{D}^A_{6r})^{\otimes\kappa},(\mathcal{D}^B_{6r})^{\otimes\kappa}\right)$, we have for any $i \in [\kappa]$, $\sigma^A_i \in \mathcal{D}^A_{6r}$ and $\sigma^B_i \in \mathcal{D}^B_{6r}$.

Consider the *first round* of the collaborative learning process. Let random variables $\mathcal{H}^A$ and $\mathcal{H}^B$ be the pull history (i.e., the sequence of ⟨arm index, pull outcome⟩ pairs) of Alice and Bob, respectively. Let random variables Θ^A_i and Θ^B_i be the sequence of pull outcomes in the pull history $\mathcal{H}^A$ and $\mathcal{H}^B$ projecting on arm i, respectively. Let t^A_i be the number of pulls Alice makes on arm i, and let t^B_i be the number of pulls Bob makes on arm i.

For $\ell = \{0,1,\dots,L\}$, we introduce the following sets of arms.

$$E^A_\ell = \{i \mid \gamma\eta^{2(\ell-1)} < t^A_i \le \gamma\eta^{2\ell}\}, \tag{34}$$

$$E^B_\ell = \{i \mid \gamma\eta^{2(\ell-1)} < t^B_i \le \gamma\eta^{2\ell}\}. \tag{35}$$

To facilitate the analysis, we augment the algorithm after the first round of pulls by publishing a set of arms, as well as making some additional pulls on the remaining arms so as to massage the posterior mean distribution. By publishing arm i we mean revealing its local means μ^A_i and μ^B_i (and thus also its global mean $\mu_i = (\mu^A_i + \mu^B_i)/2$) to both Alice and Bob. We remove arm i from the set of arms if $\mu_i \ne \mu_*$, otherwise we just output arm i and be done. Note that such an augmentation only leads to a stronger lower bound, since the success probability of the augmented algorithm can only increase compared with the algorithm before the augmentation. We also include all additional pulls to $\mathcal{H}^A$ and $\mathcal{H}^B$.

Arm Publishing and Additional Pulls

(1) Publish all arms in the following set:
$$E = E^A \cup E^B$$
$$\text{where} \quad E^A = \bigcup_{\ell=6(r+1)}^{\infty} E^A_\ell \quad \text{and} \quad E^B = \bigcup_{\ell=6(r+1)}^{\infty} E^B_\ell.$$
(2) For each arm $i \in [\kappa]\backslash E$, Alice makes additional pulls on it until her number of pulls on arm i reaches $\gamma\eta^{2(6(r+1)-1)}$, and Bob makes additional pulls on it until his number of pulls on arm i reaches $\gamma\eta^{2(6(r+1)-1)}$.
(3) Let $P^A = \left\{i \;\middle|\; \mu^A_i \in S^A_{6(r+1)}\right\}$, and $P^B = \left\{i \;\middle|\; \mu^B_i \in S^B_{6(r+1)}\right\}$. Publish all arms in $[\kappa]\backslash(P^A \cap P^B)$.

Let $T = \{i \in [\kappa] \mid \mu_i = \mu_*\}$ be the set of best arms. We try to analyze the probability that the augmented algorithm correctly outputs an arm in T, which is upper bounded by the sum of the probabilities of the following three events:

(1) $T \cap E^A \ne \emptyset$.
(2) $T \cap E^B \ne \emptyset$.
(3) $\tilde{\mathcal{A}}$ succeeds on $\left(P^A \cap P^B\right)\backslash E$, where $\tilde{\mathcal{A}}$ is the $(R-(r+1))$-round algorithm obtained from $\mathcal{A}$ conditioned on the pull history of the first round being $\mathcal{H}^A$ and $\mathcal{H}^B$.

The following lemma upper bounds the first probability. Its proof is quite technical and lengthy; due to space constraints, we leave it to the full version of this paper [13].

Lemma 15. $\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}\left[T \cap E^A \ne \emptyset\right] \le 2e^{\frac{10L}{\eta}}L^2\eta^{-\frac{5}{2}} + n^{-6}$.

The following lemma upper bounds the second probability. It is symmetric to Lemma 15, and can be proved using a similar line of arguments.

Lemma 15′. $\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}\left[T \cap E^B \ne \emptyset\right] \le 2e^{\frac{10L}{\eta}}L^2\eta^{-\frac{5}{2}} + n^{-6}$.

The next lemma upper bounds the third probability.

Lemma 16. *Let $\tilde{\mathcal{A}}$ be the $(R-(r+1))$-round algorithm obtained from $\mathcal{A}$, conditioned on the pull history of the first round being $\mathcal{H}^A$ and $\mathcal{H}^B$. We have*

$$\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}\left[\tilde{A} \text{ succeeds on } \left(P^A \cap P^B\right)\backslash E\right] \le \lambda_{r+1} + 2n^{-9}.$$

Before proving Lemma 16, we begin with some preparation. Define two events

$$\chi^A : \exists i \in P^A\backslash E \quad \text{s.t.} \quad \Theta^A_i \notin G^A_{6(r+1)}, \tag{36}$$

$$\chi^B : \exists i \in P^B\backslash E \quad \text{s.t.} \quad \Theta^B_i \notin G^B_{6(r+1)}. \tag{37}$$

In the next two lemmas, we show that χ^A and χ^B do *not* happen with high probability.

Lemma 17. $\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\chi^A] \le n^{-9}$.

Proof. Recall that each arm in $P^A\backslash E$ has been pulled for $q = \gamma\eta^{2(6(r+1)-1)} \in [\eta^3, \frac{\eta^{2(6(r+1))-1}}{2^7}]$ times.

$$\begin{aligned}
&\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\chi^A] \\
\le\ & \sum_{i=1}^{\kappa} \Pr_{\mu_i^A\sim\sigma_i^A,\Theta_i^A\sim\mathrm{Ber}(\mu_i^A)^{\otimes q}} \left[\Theta_i^A \notin G^A_{6(r+1)} \;\middle|\; \mu_i^A \in S^A_{6(r+1)}\right] && (38)\\
\le\ & n\cdot n^{-10} = n^{-9}, && (39)
\end{aligned}$$

where from (38) to (39) we have used Lemma 9. □

The following lemma is symmetric to Lemma 17, and can be proved using a similar line of arguments.

Lemma 17′. $\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\chi^B] \le n^{-9}$.

Proof of Lemma 16. For the convenience of writing, we further introduce the following event.

$$\psi : \tilde{\mathcal{A}} \text{ succeeds on } \left(P^A \cap P^B\right)\backslash E. \tag{40}$$

We write

$$\begin{aligned}
&\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\psi] \\
\le\ & \Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\psi,\neg\chi^A,\neg\chi^B] + \Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\chi^A] \\
&+ \Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\chi^B] && (41)\\
\le\ & \Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\psi,\neg\chi^A,\neg\chi^B] + 2n^{-9} && (42)\\
=\ & \sum_{(h^A,h^B)} \left(\Pr_{I\sim\nu}[\psi,\neg\chi^A,\neg\chi^B \mid (\mathcal{H}^A,\mathcal{H}^B) = (h^A,h^B)]\right. \\
&\left. \Pr_{\mathcal{H}^A,\mathcal{H}^B}[(\mathcal{H}^A,\mathcal{H}^B) = (h^A,h^B)]\right) + 2n^{-9}, && (43)
\end{aligned}$$

where from (41) to (42) we have used Lemma 17 and Lemma 17′.

Consider a fixed pull history (h^A,h^B). For any $i \in (P^A\cap P^B)\backslash E$, its sequence of pull outcomes (θ_i^A,θ_i^B) in the first round is fully determined by (h^A,h^B). We consider two cases.

Case I: χ^A or χ^B holds. In this case, we have

$$\Pr_{I\sim\nu}[\psi,\neg\chi^A,\neg\chi^B \mid (\mathcal{H}^A,\mathcal{H}^B) = (h^A,h^B)] = 0. \tag{44}$$

Case II: $\neg\chi^A$ and $\neg\chi^B$ holds. In this case, by the definition of χ^A in (36) and χ^B in (37), we have for any $i \in (P^A\cap P^B)\backslash E$, $\theta_i^A \in G^A_{6(r+1)}$ and $\theta_i^B \in G^B_{6(r+1)}$. The posterior distribution of the local mean of arm i at Alice's side can be written as

$$\begin{aligned}
\tilde{\sigma}_i^A &= \left(\sigma_i^A \;\middle|\; \mu_i^A \in S^A_{6(r+1)}, (\mathcal{H}^A,\mathcal{H}^B) = (h^A,h^B)\right)\\
&= \left(\sigma_i^A \mid \mu_i^A \in S^A_{6(r+1)}, \Theta_i^A = \theta_i^A \in G^A_{6(r+1)}\right) \in \mathcal{D}^A_{6(r+1)}.
\end{aligned}$$

Similarly, the posterior distribution of the local mean of arm i at Bob's side can be written as

$$\tilde{\sigma}_i^B = \left(\sigma_i^B \;\middle|\; \mu_i^B \in S^B_{6(r+1)}, (\mathcal{H}^A,\mathcal{H}^B) = (h^A,h^B)\right) \in \mathcal{D}^B_{6(r+1)}.$$

Thus, for any $i \in (P^A\cap P^B)\backslash E$, we have $\tilde{\sigma}_i = (\tilde{\sigma}_i^A,\tilde{\sigma}_i^B) \in \mathcal{D}_{6(r+1)}$. Recall that $\tilde{\mathcal{A}}$ is a $(R-(r+1))$-round algorithm working on a set of arms $(P^A\cap P^B)\backslash E$ with $\tilde{\kappa} = \left|(P^A\cap P^B)\backslash E\right| \le n$, conditioned on the first round pull history being $(\mathcal{H}^A,\mathcal{H}^B)$. By the definition of λ_{r+1} in (33) and the fact that conditioned on the pull history $(\mathcal{H}^A,\mathcal{H}^B)$, the distribution of the $\tilde{\kappa}$ arms belongs to $\tilde{\sigma}^{\otimes\tilde{\kappa}} \in \mathcal{D}^{\otimes\tilde{\kappa}}_{6(r+1)}$,

$$\Pr_{I\sim\nu}[\psi,\neg\chi^A,\neg\chi^B \mid (\mathcal{H}^A,\mathcal{H}^B) = (h^A,h^B)] \le \lambda_{r+1}. \tag{45}$$

Combining (43), (44) and (45), we have

$$\begin{aligned}
&\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\psi] \\
\le\ & \sum_{(h^A,h^B)} \left(\lambda_{r+1} \Pr_{\mathcal{H}^A,\mathcal{H}^B}[(\mathcal{H}^A,\mathcal{H}^B) = (h^A,h^B)]\right) + 2n^{-9}\\
\le\ & \lambda_{r+1} + 2n^{-9}.
\end{aligned}$$

□

Summing Up. Combining Lemma 15, Lemma 15′, and Lemma 16,

$$\begin{aligned}
&\Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[\mathcal{A} \text{ succeeds on } I] \\
\le\ & \Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[T\cap E^A \ne \emptyset] + \Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}[T\cap E^B \ne \emptyset]\\
&+ \Pr_{I\sim\nu,\mathcal{H}^A,\mathcal{H}^B}\left[\tilde{A} \text{ succeeds on } \left(P^A\cap P^B\right)\backslash E\right] && (46)\\
\le\ & 4\left(e^{\frac{10L}{\eta}} L^2\eta^{-\frac{5}{2}} + n^{-6}\right) + (\lambda_{r+1} + 2n^{-9}) && (47)\\
\le\ & \lambda_{r+1} + 4e^{\frac{10L}{\eta}} L^2\eta^{-\frac{5}{2}} + n^{-5}. && (48)
\end{aligned}$$

Since (48) holds for any algorithm $\mathcal{A}$, distribution ν, and $\kappa \in [n]$, we have $\lambda_r \le \lambda_{r+1} + 4e^{\frac{10L}{\eta}} L^2\eta^{-\frac{5}{2}} + n^{-5}$.

The Base Case. In the base case we consider 0-round algorithm (i.e., when $r = R$). We have the following lemma.

Lemma 18. *For $R = \frac{L}{6}$, $\lambda_R \le e^{\frac{48R}{\eta}}\eta^{-2}$.*

Proof. Any 0-round algorithm needs to output an arm i as the best arm without making any pulls. For any i with mean $\mu_i \sim \sigma_i \in \mathcal{D}_{6R}$, by Lemma 12 and Lemma 12′, we have

$$\begin{aligned}
\Pr_{\mu_i\sim\sigma_i}[\mu_i = \mu_*] &= \Pr_{\mu_i^A\sim\sigma_i^A}[\mu_i^A = \mu_*^A] \Pr_{\mu_i^B\sim\sigma_i^B}[\mu_i^B = \mu_*^B] && (49)\\
&\le e^{\frac{4\cdot 6R}{\eta}}\eta^{-2d_1}\cdot e^{\frac{4\cdot 6R}{\eta}}\eta^{-2d_0} && (50)\\
&= e^{\frac{48R}{\eta}}\eta^{-2(d_0+d_1)}, && (51)
\end{aligned}$$

where $d_0+d_1 = |\{k \mid 6R < k \le L\}|$. For $R = \frac{L}{6}$, we have $\Pr_{\mu_i\sim\sigma_i}[\mu_i = \mu_*] \le e^{\frac{48R}{\eta}}\eta^{-2}$. □

Putting Things Together (Proof for Theorem 13). By Lemma 18 and Lemma 14, we have

$$\begin{aligned}
\lambda_0 &\le \lambda_R + R\cdot\left(4e^{\frac{10L}{\eta}} L^2\eta^{-\frac{5}{2}} + n^{-5}\right)\\
&\le e^{\frac{48R}{\eta}}\eta^{-2} + (L/6)\cdot\left(4e^{\frac{10L}{\eta}} L^2\eta^{-\frac{5}{2}} + n^{-5}\right) \le \eta^{-1}.
\end{aligned}$$

Therefore, any R-round collaborative algorithm that uses $2\zeta\eta^{2+2L}L$ time (i.e., each agent can make at most $2\zeta\eta^{2+2L}L$ pulls in total) can succeed with probability at most η^{-1}.

Recall the definition of event $\mathcal{E}_1$ in (15): $\exists$ a unique $i^* \in [n]$ such that $\mu_{i^*} = \mu_*$ *and* the instance complexity $H = H(I) \le 2\eta^{2+2L}L$ where $I \sim (\pi^{\otimes n} \mid \mathcal{E}_1)$.

By Lemma 8, $\Pr[\mathcal{E}_1] \ge 1/(2e)$. We thus have

$$\begin{aligned}\Pr_{I\sim(\pi^{\otimes n}|\mathcal{E}_1)}[\mathcal{A} \text{ succeeds on } I] &\le \frac{\Pr_{I\sim\pi^{\otimes n}}[\mathcal{A} \text{ succeeds on } I]}{\Pr_{I\sim\pi^{\otimes n}}[\mathcal{E}_1]} \\ &\le \lambda_0 \cdot (2e) \\ &\overset{(52)}{\le} \frac{2e}{\eta} < 0.9\,.\end{aligned}$$

Therefore, any R-round ($1 \le R \le \frac{\log n}{24\log\log n}$) collaborative algorithm that succeeds on input distribution $(\pi^{\otimes n} \mid \mathcal{E}_1)$ with probability at least 0.9 needs time at least $2\zeta\eta^{2+2L}L \ge H \cdot \zeta \ge H \cdot n^{\frac{1}{25R}}$.

3.4 General K

We now consider the general case where there are K agents. The following theorem is a restatement of Theorem 1.

Theorem 19. *For any $1 \le R \le \frac{\log n}{24\log\log n}$, any R-round K-agent algorithm that solves n-arm BAI in the heterogeneous CL mode with probability 0.99 uses time at least $Hn^{\Omega(\frac{1}{R})}/K$.*

Proof. We prove the general K case by a reduction from the $K = 2$ case. Suppose there exists a R-round algorithm for BAI in the heterogeneous CL model with n arms using K agents and uses time smaller than $Hn^{\frac{1}{26R}}/K$, we show that there also exists a R-round algorithm for the same problem using 2 agents and uses time smaller than $Hn^{\frac{1}{25R}}$, contradicting Theorem 13.

The reduction works as follows. Given any algorithm $\mathcal{A}$ for the K-agent case, we construct an algorithm $\mathcal{A}'$ for the 2-agent case: We divide the K agents to two groups each having $K/2$ agents. Let Alice simulate the first group, and Bob simulate the second group. In each round, the sequence of arm pulls Alice makes is simply the *concatenation* of arm pulls made by the $K/2$ agents that she simulates, and the sequence of arm pulls Bob makes is the concatenation of arm pulls made by the $K/2$ agents that he simulates. The messages sent by Alice in each communication step is a concatenation of the messages sent by agents in the group she simulates in the corresponding communication step in $\mathcal{A}$; similar for Bob. Now if $\mathcal{A}$ uses time at most $Hn^{\frac{1}{26R}}/K$, then $\mathcal{A}'$ uses time at most $Hn^{\frac{1}{26R}}/K \cdot (K/2) < Hn^{\frac{1}{25R}}$, contradicting to Theorem 13. □

4 THE ALGORITHM

In this section, we present a CL algorithm that gives Theorem 2. Our algorithm is *non-adaptive*. It follows the successive elimination approach, and can be seen as a generalization of the algorithm for the heterogeneous CL setting in [12] to the entire time-round tradeoff curve.

Intuitively, we partition the learning process into R rounds with predefined lengths $t_1, \dots, t_R$. In each round r, each of the K agents simply pulls each remaining arm for t_r times. At the end of each round, the K agents communicate and compute the global empirical means of each arm, and then select the n_r arms with the highest global empirical means and proceed to the next round, where $n_1, \dots, n_R$ are also predefined. We set n_R to be 1 so that at the end of the R-round, there will be just one arm left, which can be proven to be the best arm with high probability.

Algorithm 1: CL-Heterogeneous(I, R, T)

Input: a set of n arms I, round parameter R, number of agents K, and time horizon T.
Output: the arm with the largest global mean.
Initialize $I_0 = I$;
set $T_0 \leftarrow 0$, $T_r \leftarrow \left\lfloor \frac{n^{r/R}T}{n^{1+1/R}R} \right\rfloor$ for $r = 1, \dots, R$;
set $n_r \leftarrow \left\lfloor \frac{n}{n^{r/R}} \right\rfloor$ for $r = 0, \dots, R-1$, and $n_R \leftarrow 1$;
for $r = 0, 1, \dots, R-1$ **do**
  each agent pulls each arm in I_r for $(T_{r+1} - T_r)$ times;
  the k-th agent computes the local empirical mean $\hat{\mu}_{i,k}^{(r)}$ for $i \in I_r$;
  let $\hat{\mu}_i^{(r)} \leftarrow \frac{1}{K}\sum_{k\in[K]} \hat{\mu}_{i,k}^{(r)}$;
  let I_{r+1} be the set of n_{r+1} arms in I_r with the highest global empirical means $\hat{\mu}_i^{(r)}$;
return the single element in I_R.

The algorithm is described in Algorithm 1. It gives the following guarantees. Due to the space constraints, we leave its proof to the full version of this paper [13].

Theorem 20. *For any $R \ge 1$, Algorithm 1 solves BAI in the heterogeneous CL model with K agents and n arms using T time steps and R rounds, with a success probability at least*

$$1 - 2nR \cdot \exp\left(-KT/(2n^{\frac{1}{R}}RH)\right). \tag{52}$$

Note that Theorem 2 (in the introduction) is an immediate corollary of Theorem 20.

Remark 21. *We note that the total messages exchanged between the agents in Algorithm 1 is $O(nK)$ words, which is optimal (up to a logarithmic factor) based on a lower bound result in [12].*

REFERENCES

[1] Sanae Amani, Tor Lattimore, András György, and Lin F. Yang. Distributed contextual linear bandits with minimax optimal communication cost. *CoRR*, abs/2205.13170, 2022.
[2] Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *COLT*, pages 41–53, 2010.
[3] Robert Bechhofer. A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. *Biometrics*, 14(3):408–429, 1958.
[4] Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *COLT*, pages 590–604, 2016.
[5] Lijie Chen, Jian Li, and Mingda Qiao. Towards instance optimal bounds for best arm identification. In *COLT*, pages 535–592, 2017.
[6] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for multi-armed bandit and markov decision processes. In *COLT*, pages 255–270, 2002.
[7] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(Jun):1079–1105, 2006.
[8] Eshcar Hillel, Zohar Shay Karnin, Tomer Koren, Ronny Lempel, and Oren Somekh. Distributed exploration in multi-armed bandits. In *NIPS*, pages 854–862, 2013.
[9] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil' UCB : An optimal exploration algorithm for multi-armed bandits. In *COLT*, pages 423–439, 2014.

[10] Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista A. Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D'Oliveira, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaïd Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrède Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Mariana Raykova, Hang Qi, Daniel Ramage, Ramesh Raskar, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao. Advances and open problems in federated learning. *CoRR*, abs/1912.04977, 2019.

[11] Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *ICML*, pages 1238–1246, 2013.

[12] Nikolai Karpov and Qin Zhang. Communication-efficient collaborative best arm identification. In *AAAI*, 2023.

[13] Nikolai Karpov and Qin Zhang. Collaborative best arm identification with limited communication on non-iid data. *CoRR*, abs/2207.08015, 2024.

[14] Nikolai Karpov, Qin Zhang, and Yuan Zhou. Collaborative top distribution identifications with limited interaction (extended abstract). In *FOCS*, pages 160–171. IEEE, 2020.

[15] Jakub Konečný, Brendan McMahan, and Daniel Ramage. Federated optimization: Distributed optimization beyond the datacenter. *CoRR*, abs/1511.03575, 2015.

[16] Jakub Konečný, H. Brendan McMahan, Daniel Ramage, and Peter Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. *CoRR*, abs/1610.02527, 2016.

[17] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[18] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Aarti Singh and Xiaojin (Jerry) Zhu, editors, *AISTATS*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 2017.

[19] Aritra Mitra, Hamed Hassani, and George J. Pappas. Exploiting heterogeneity in robust federated best-arm identification. *CoRR*, abs/2109.05700, 2021.

[20] Edward Paulson. A sequential procedure for selecting the population with the largest mean from k normal populations. *The Annals of Mathematical Statistics*, pages 174–180, 1964.

[21] Clémence Réda, Sattar Vakili, and Emilie Kaufmann. Near-optimal collaborative learning in bandits. *CoRR*, abs/2206.00121, 2022.

[22] Chengshuai Shi and Cong Shen. Federated multi-armed bandits. In *AAAI*, pages 9603–9611. AAAI Press, 2021.

[23] Chengshuai Shi, Cong Shen, and Jing Yang. Federated multi-armed bandits with personalization. In Arindam Banerjee and Kenji Fukumizu, editors, *AISTATS*, volume 130 of *Proceedings of Machine Learning Research*, pages 2917–2925. PMLR, 2021.

[24] Chao Tao, Qin Zhang, and Yuan Zhou. Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits. In David Zuckerman, editor, *FOCS*, pages 126–146. IEEE Computer Society, 2019.

[25] Yuanhao Wang, Jiachen Hu, Xiaoyu Chen, and Liwei Wang. Distributed bandit learning: Near-optimal regret with efficient communication. In *ICLR*. OpenReview.net, 2020.

[26] Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity (extended abstract). In *FOCS*, pages 222–227, 1977.