Closing the Gap on the Sample Complexity of 1-Identification

Li Zitian
Department of Industrial Systems Engineering and Management
National University of Singapore
Engineering Drive 2 Block E1A #06-25 Singapore 117576
lizitian@u.nus.edu

Cheung Wang Chi
Department of Industrial Systems Engineering and Management
National University of Singapore
Engineering Drive 2 Block E1A #06-25 Singapore 117576
isecwc@nus.edu.sg

Abstract

The 1-identification problem is a fundamental pure-exploration problem in multi-armed bandits. An agent aims to determine whether there exists an arm whose mean reward exceeds a known threshold μ_0, or to output None otherwise. The agent must guarantee correctness with probability at least $1-\delta$, while minimizing the expected number of arm pulls $\mathbb{E}[\tau]$. We study the 1-identification problem and make two main contributions. First, for instances with at least one qualified arm, we derive a new lower bound on $\mathbb{E}[\tau]$ via a novel optimization formulation. Second, we propose a new algorithm and establish upper bounds that match the lower bounds up to polynomial logarithmic factors uniformly over all instances. Our result complements the analysis of $\mathbb{E}\tau$ when there are multiple qualified arms, which is an open problem in the literature.

1 Introduction

1.1 Model & Problem Formulation

We study the 1-identification problem[1], a variant of the best arm identification problem. An instance of 1-identification is specified by the tuple $\nu = ([K], \rho = \{\rho_a\}_{a\in[K]}, \mu_0, \delta)$, where $[K] \coloneqq \{1, 2, \ldots, K\}$ denotes the set of arms with $K \geq 2$, and ρ_a is the unknown reward distribution associated with arm a. Upon pulling arm a, the agent observes a random reward $R \sim \rho_a$. We denote by $r_a^\nu \coloneqq \mathbb{E}_{R\sim\rho_a}[R]$ the mean reward of arm a under instance ν, and assume that the noise $R - r_a^\nu$ is 1-sub-Gaussian. The threshold $\mu_0 \in \mathbb{R}$ is known to the agent. The confidence parameter $\delta \in (0, 1)$ controls the allowable probability of an incorrect decision. The objective of the agent is to determine whether there exists an arm $a \in [K]$ such that $r_a^\nu \geq \mu_0$. In the affirmative case, the agent must identify (at least) one such arm; otherwise, it must output the symbol None.

The 1-identification problem serves as a framework for many decision-making scenarios. For instance, a clinical trial screening phase where researchers must decide whether any of several new drug candidates meet a specific efficacy threshold to justify the cost of further development. Similarly, considering anomaly detection in power grids, an agent may need to determine if if any substation

[1]The terminology originates from Katz-Samuels and Jamieson (2020); Li and Cheung (2025). Degenne and Koolen (2019) call it Any Low Problem.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

is experiencing a voltage fluctuation that exceeds safety limits. In both cases, the agent does not necessarily need to rank every option; rather, it must quickly identify at least one "qualified" arm or confidently conclude that none exist.

Online model and algorithms. An algorithm for the 1-identification problem is defined by a triple $(\pi, \tau, \hat{a})$, where: (i) $\pi = \{\pi_t\}_{t\geq 1}$ is a (possibly randomized) sampling rule, (ii) τ is a stopping time, and (iii) $\hat{a} \in [K] \cup \{\mathsf{None}\}$ is a recommendation rule. When executed on an instance ν, the algorithm generates a random history $\{U_0\} \cup \{(A_t, X_t, U_t)\}_{t=1}^{\tau}$, where $A_t \in [K]$ is the arm selected at time t, $X_t \sim \rho_{A_t}$ is the corresponding reward, and $U_t \sim \mathrm{Unif}(0,1)$ is for internal randomization of the algorithm. The initial seed U_0 allows for randomized initialization. At each time t, the arm A_t is chosen according to $A_t = \pi_t(H(t-1))$, where $H(t-1) \coloneqq \{U_0\} \cup \{(A_s, X_s, U_s)\}_{s=1}^{t-1}$. The stopping time τ is defined with respect to the filtration $\{\sigma(H(t))\}_{t\geq 1}$. Upon stopping, the algorithm outputs a recommendation $\hat{a}$ that is measurable with respect to $\sigma(H(\tau))$. Outputting $\hat{a} \in [K]$ indicates the conclusion that $r_{\hat{a}}^{\nu} \geq \mu_0$, whereas outputting $\hat{a} = \mathsf{None}$ indicates the conclusion $\max_{a\in[K]} r_a^{\nu} < \mu_0$.

Instance classes. Following Li and Cheung (2025), an instance ν is called a *positive instance* if $\max_{a\in[K]} r_a^{\nu} > \mu_0$, and a *negative instance* if $\max_{a\in[K]} r_a^{\nu} < \mu_0$. For $\Delta > 0$, we define the margin-separated classes $\mathcal{S}_{\Delta}^{\mathrm{pos}} \coloneqq \{\nu : \max_a r_a^{\nu} - \mu_0 \geq \Delta\}$, $\mathcal{S}_{\Delta}^{\mathrm{neg}} \coloneqq \{\nu : \mu_0 - \max_a r_a^{\nu} \geq \Delta\}$. Let $\mathcal{S}^{\mathrm{pos}} \coloneqq \bigcup_{\Delta>0} \mathcal{S}_{\Delta}^{\mathrm{pos}}$, $\mathcal{S}^{\mathrm{neg}} \coloneqq \bigcup_{\Delta>0} \mathcal{S}_{\Delta}^{\mathrm{neg}}$. We define the set of correct answers $i^*(\nu)$ by $i^*(\nu) \coloneqq \begin{cases} \{a \in [K] : r_a^{\nu} \geq \mu_0\}, & \text{if } \nu \in \mathcal{S}^{\mathrm{pos}}, \\ \{\mathsf{None}\}, & \text{if } \nu \in \mathcal{S}^{\mathrm{neg}}. \end{cases}$. In this work, We ignore instances for which $\max_a r_a^{\nu} = \mu_0$.

PAC guarantees. We adopt a PAC framework and focus on algorithms with correctness guarantees.

Definition 1 (δ-PAC)**.** *An algorithm is* δ-PAC*, if for any* $\nu \in \mathcal{S}^{\mathrm{pos}} \cup \mathcal{S}^{\mathrm{neg}}$, $\Pr_{\nu}(\tau < \infty,\ \hat{a} \in i^*(\nu)) \geq 1 - \delta$.

As shown in Appendix A.3 in Li and Cheung (2025), a δ-PAC algorithm may still satisfy $\mathbb{E}_{\nu}[\tau] = \infty$ for some $\nu \in \mathcal{S}^{\mathrm{pos}}$. To rule out such pathological behavior, we introduce a stronger notion.

Definition 2 (Strong δ-PAC)**.** *An algorithm is Strong* δ*-PAC if it is* δ*-PAC and, for all* $\Delta, \delta > 0$*, we have* $\sup_{\nu \in \mathcal{S}_{\Delta}^{\mathrm{pos}} \cup \mathcal{S}_{\Delta}^{\mathrm{neg}}} \mathbb{E}_{\nu}[\tau] < \infty$.

Objective. The goal of the agent is to design a Strong δ-PAC algorithm $(\pi, \tau, \hat{a})$ that minimizes the expected sampling complexity $\mathbb{E}_{\nu}[\tau]$.

1.2 Literature Review

Before illustrating our main contribution, we review the existing literature on the 1-identification problem and highlight its current limitations. We begin by unifying notation across prior works in order to facilitate comparison. Throughout this paper, we consider a reward vector $(\mu_a)_{a=1}^K \in [0,1]^K$ satisfying that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$, together with a known threshold $\mu_0 \in [0,1]$. For any $i, j \in [K] \cup \{0\}$, we define the gap $\Delta_{i,j} \coloneqq |\mu_i - \mu_j|$. In the analysis of upper bound(section 3), we only consider instance ν with $r_a^{\nu} = \mu_a$.[2] In the analysis of lower bound(section 2), we further consider instances ν' whose reward vector $r_a^{\nu'}$ is not ordered. We next recall several instance-dependent complexity measures that appear in the literature:

$$
\begin{aligned}
H_1^{\mathrm{neg}} &= \sum_{a=1}^{K} \frac{2}{\Delta_{0,a}^2}, H_1^{\mathrm{low}} = \sum_{a:\mu_a<\mu_0} \frac{2}{\Delta_{1,a}^2}, H = \frac{2}{\Delta_{0,1}^2} \\
H_1 &= \sum_{a=2}^{K} \frac{2}{\Delta_{1,a}^2}, H_0 = \sum_{a:\mu_a\geq\mu_0} \frac{2}{\Delta_{0,a}^2} \\
H^{\mathrm{u}} &= \frac{(\frac{K}{m} - 1)}{\Delta^2}, \Delta \coloneqq \min_{a:\mu_a>\mu_0} \Delta_{a,0}, m \coloneqq \arg\max\{a : \mu_a > \mu_0\} \\
H(j) &= \frac{1}{j} \sum_{a=1}^{K} \frac{1}{\max\{\Delta_{j,a}^2, \Delta_{a,0}^2\}}
\end{aligned}
\tag{1}
$$

[2]The agent does not know the monotonic ordering of $\{\mu_a\}_{a=1}^K$, and this default setting is for convenience in discussions.

The study of sampling complexity naturally separates into two regimes, depending on the instance ν is negative or positive. For negative instances, the problem is closely related to the classical Best-Arm Identification problem, and the sample complexity is well studied. In particular, Theorem 5.4 of Li and Cheung (2025) establishes that, for any δ-PAC algorithm and any negative instance ν, $\mathbb{E}_\nu[\tau] \geq \Omega\big(H_1^{\text{neg}} \log \frac{1}{\delta}\big)$. On the algorithmic side, HDoC Kano et al. (2017), APGAI Jourdan and Réda (2023), and SEE Li and Cheung (2025) guarantee $\mathbb{E}_\nu[\tau] \leq O(H_1^{\text{neg}}(\log(1/\delta) + \log H_1^{\text{neg}}))$, up to logarithmic factors. Moreover, for Gaussian rewards with unit variance, the Sticky Track-and-Stop (S-TaS) algorithm Degenne and Koolen (2019) achieves the asymptotically optimal guarantee $\limsup_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \leq T^*(\nu)$, where $T^*(\nu) = H_1^{\text{neg}}$ for negative instances.

In contrast, the sample complexity of positive instances remains unclear. Existing upper bounds suffer from various looseness and fail to match known lower bounds except in restricted instances. To illustrate with simplicity, we consider a positive instance ν satisfying $\mu_1 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$, and restrict attention to Gaussian rewards with variance 1. Table 1 summarizes representative non-asymptotic upper and lower bounds from the literature. [3] Detailed description is as follows.

Table 1: Comparison of non-asymptotic upper bounds on $\mathbb{E}\tau$ for a positive instance ν. "LB" denotes Lower Bound. $\surd_{\text{cond}}$ means the upper or lower bounds are optimal under some specific instances. The definitions of different H's are at the equation (1). We determine whether a sampling complexity upper bound is nearly optimal by comparing with the lower bound. An upper bound is nearly optimal iff the gap is up to a polynomial of $\log K, \{\log \frac{1}{\Delta_{a,0}^2}\}_{a=1}^K$.

Algorithm	Non-asymptotic Bound	Optimality
S-TaS, proposed by Degenne and Koolen (2019) analyzed by Poiani et al. (2026)	$\mathbb{E}\tau \leq O(\frac{\log(1/\delta)}{\Delta_{0,1}^2} + \frac{K^2}{\Delta_{0,1}^4})$	$\times$
HDoC, Kano et al. (2017)	$\mathbb{E}\tau \leq O(\frac{\log \frac{K}{\delta}}{\Delta_{0,1}^2} + H_1 \log\log \frac{1}{\delta} + \frac{K}{\epsilon^2})$	$\times$
lilHDoC, Tsai et al. (2024)	$\mathbb{E}\tau \leq O(\frac{\log \frac{K}{\delta}}{\Delta_{0,1}^2} + H_1 \log\log \frac{1}{\delta} + \frac{K}{\epsilon^2})$	$\times$
APGAI, Jourdan and Réda (2023)	$\mathbb{E}\tau \leq O(H_0(\log \frac{K}{\delta}))$	$\times$
FWER-FWPD Katz-Samuels and Jamieson (2020)	$\mathbb{E}\tau \leq O\Big(H^{\text{u}} \log \frac{1}{\delta} \log\Big(H^{\text{u}} \log \frac{1}{\delta}\Big)\Big)$	$\times$
SEE, Li and Cheung (2025)	$\mathbb{E}\tau \leq O(\frac{\log \frac{1}{\delta}}{\Delta_{0,1}^2} + H(1) \log \frac{K}{\Delta_{0,1}^2})$	$\surd_{\text{cond}}$
PSEEB, This work	$\mathbb{E}\tau \leq O\Big(\log K \min_{1\leq j\leq m} \Big\{ \frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2} + H(j) \log^4 K \log \frac{1}{\Delta_{j,0}^2} \Big\}\Big)$	$\surd$
LB, Katz-Samuels and Jamieson (2020)	$\mathbb{E}\tau \geq \Omega(\frac{1}{m} H_1^{\text{low}} - \frac{1}{\Delta_{1,m+1}^2})$	$\times$
LB, Li and Cheung (2025)	$\mathbb{E}\tau \geq \Omega(H \log \frac{1}{\delta} + \frac{1}{m} H_1^{\text{low}} - \frac{1}{\Delta_{1,m+1}^2})$	$\surd_{\text{cond}}$
LB, This work	$\mathbb{E}\tau \geq \Omega\left(\min_{1\leq j\leq m} \left\{ \frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{H(j)}{\log^2(m+1)} \right\}\right)$	$\surd$

Degenne and Koolen (2019) propose the S-TaS algorithm, which attains asymptotically optimal guarantees for both negative and positive instances. In particular, they show $\limsup_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \leq \frac{2}{\Delta_{0,1}^2}$, together with a matching lower bound $\liminf_{\delta\to 0} \frac{\mathbb{E}\tau}{\log(1/\delta)} \geq \frac{2}{\Delta_{0,1}^2}$, establishing asymptotic optimality. A non-asymptotic analysis is left as a future direction in Degenne and Koolen (2019) and later addressed by Poiani et al. (2026). The resulting non-asymptotic bound, however, contains a δ-independent term $\frac{K^2}{\Delta_{0,1}^4}$, which has no counterpart in known lower bounds and leads to looseness.

[3] Due to space limit, the bounds in the Table 1 are simplified; the original bounds are larger than those reported here. Even under these simplifications, the looseness remains apparent. Following Appendix A.2 of Li and Cheung (2025), we provide a detailed derivation of the simplified bounds in Appendix A.

The HDoC algorithm of Kano et al. (2017) yields an upper bound $C_{\text{pos}}^{\text{HDoC}} := \frac{\log(K/\delta)}{\Delta_{0,1}^2} + H_1 \log\log\frac{1}{\delta} + \frac{K}{\epsilon^2}$, where $\epsilon = \min\{\min_a \Delta_{0,a},\, \min_a \Delta_{a,a+1}/2\}$. Its suboptimality arises from two sources. The first source is $H_1 \log\log\frac{1}{\delta}$, which is explicitly acknowledged in Kano et al. (2017): "Therefore, in the case that $\dots$ $K = \Omega(\frac{\log\frac{1}{\delta}}{\log\log\frac{1}{\delta}})\dots$ **there still exists a gap** between the lower bound in $\dots$ and the upper bound in $\dots$". The second source is $\frac{K}{\epsilon^2}$, which is infinite when $\epsilon = 0$. The lilHDoC algorithm Tsai et al. (2024), despite incorporating LIL-based confidence bounds and an additional warm-up phase, inherits the same fundamental limitations.

The APGAI algorithm Jourdan and Réda (2023) is originally designed for Good Arm Identification, aiming to output all the arms whose reward is above a given threshold μ_0. But the analysis in Jourdan and Réda (2023) also implies an upper bound $O(H_0 \log(K/\delta))$ for 1-identification. This bound is loose primarily due to the dependence on H_0, which might scale linearly with K, leading to a suboptimal coefficient of $\log(1/\delta)$. Moreover, if there exists an arm with $\mu_a = \mu_0$, then $H_0 = \infty$. In this case, the bound is vacuous.

Katz-Samuels and Jamieson (2020) study ϵ-good identification and 1-identification via the notion of Unverifiable Sample Complexity. Although this notion differs from the sampling complexity considered here, their results can be adapted as discussed in Appendix A.1 of Li and Cheung (2025). The resulting upper bound is suboptimal for two reasons: it scales as $\log(1/\delta)\log\log(1/\delta)$ rather than $\log(1/\delta)$, and it depends on the complexity term H^{u}, which is governed by the smallest gap above μ_0 and may be much smaller than $\Delta_{0,1}$. When $m \ll K$ and $\Delta \ll \Delta_{0,1}$, this leads to a significant gap compared to the lower bound.

Li and Cheung (2025) propose the SEE algorithm and derive explicit upper and lower bounds, decomposing $\mathbb{E}_\nu[\tau]$ into δ-dependent and δ-independent components. They show that when $m = 1$, i.e. there is only one arm whose reward is greater than μ_0, the gap between upper and lower bounds is at most polynomial in $\log K$ and $\log(1/\Delta_{0,1}^2)$, implying near-optimality. For $m > 1$, however, they comment the δ-independent term $H(1)\log\big(K/\Delta_{0,1}^2\big)$ may be improvable, which is explicitly left as an open problem.

In summary, every prior upper bound suffers from at least one of the following weaknesses. (1) A δ-independent term still scaling linearly in K even when $K/m = \Theta(1)$ (Degenne and Koolen (2019); Kano et al. (2017); Tsai et al. (2024); Jourdan and Réda (2023); Li and Cheung (2025)). (2) The coefficient of $\log 1/\delta$ is dependent on rewards of multiple arms Katz-Samuels and Jamieson (2020), the algorithm construct conservative and δ-dependent confidence intervals on too many arms. (3) A bound diverging to infinity even on positive instances Kano et al. (2017); Jourdan and Réda (2023), reflecting an ill-suited stopping/elimination rule.

On the lower-bound side, Li and Cheung (2025); Katz-Samuels and Jamieson (2020) are both unable to achieve joint optimality on both δ-dependent and δ-independent terms, because their arguments bound the δ-dependent and δ-independent parts *separately* rather than jointly. The takeaway from Table 1 is that PSEEB is the first algorithm whose upper bound matches a *jointly* optimal lower bound (Theorem 1) up to $\text{poly}(\log K,\ \log(1/\Delta_{j,0}^2))$, uniformly over all positive instances.

There are other research topics which are less related to the 1-identification. We put the literature review in the Appendix A.2.

1.3 Main Contribution

In this paper, we develop novel upper and lower sample complexity bounds for the 1-identification problem. Our results characterize the sample complexity in both positive and negative instances and close the gap for instances with multiple qualified arms. We first establish a new lower bound for positive instances.

Theorem 1. *For any δ-PAC alg and any reward vector $\{\mu_a\}_{a=1}^K \in [0,1]^K$ satisfying that $\mu_1 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$, any tolerance level δ such that $\delta < \min\{10^{-8}, \frac{1}{64m^2}\}$, we can find a permutation $\sigma : [K] \to [K]$ and define a positive instance ν with $r_{\sigma(a)}^\nu = \mu_a, a \in [K]$ with unit-variance Gaussian reward, such that $\mathbb{E}_{alg,\nu}\tau \geq \Omega\left(\min_{1\leq j\leq m}\left\{\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{H(j)}{\log^2(m+1)}\right\}\right)$.*

Theorem 1 is our main contribution, achieving nearly joint optimality on both δ-dependent and δ-independent term. To the best of our knowledge, the "Simulator" framework Simchowitz et al. (2017); Katz-Samuels and Jamieson (2020) was the **only known method** for deriving δ-independent lower bounds. However, results in Katz-Samuels and Jamieson (2020) are only tight under restricted settings ($\mu_1 = \cdots = \mu_m > \mu_0$), without joint optimality on both parts. The bottleneck is constructing a "Simulator" that depends on the sampling path and the reward vector. Our proposed method reduces this complex construction to finding a feasible solution for a convex optimization problem, which not only significantly simplifies the proof burden but also yields a tighter, jointly optimal result. We provide sketch proof in Section 2 and full proof in Appendix B.

On the algorithmic side, we combine and refine ideas from Katz-Samuels and Jamieson (2020); Li and Cheung (2025) to propose a new algorithm, *Parallel Sequential Exploration–Exploitation on Brackets* (Algorithm 1, abbreviated as PSEEB). We establish the following guarantees.

Theorem 2. *Algorithm 1 is δ-PAC.*

Theorem 3. *Apply Algorithm 1 to a 1-identification instance ν, we have*

$$\mathbb{E}_\nu \tau \leq \begin{cases} O\left(\log K \min\limits_{1\leq j\leq m}\left\{\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \log^3 K(\log(4K) + \log_2 \frac{1}{\Delta_{j,0}^2})H(j)\right\}\right) & \nu \text{ is positive} \\ O\left((\log K)H_1^{neg}\log\frac{H_1^{neg}}{\delta}\right) & \nu \text{ is negative} \end{cases}$$

Combining Theorems 2 and 3, we conclude that Algorithm 1 is Strong δ-PAC. Moreover, the gap between the lower bound in Theorem 1 and the upper bound in Theorem 3 is at most polynomial in $\log K$ and $\{\log(1/\Delta_{0,a}^2)\}_{a=1}^m$. This shows that PSEEB is nearly optimal for all 1-identification instances and identifies $\min_{1\leq j\leq m}\left\{\frac{\log(1/\delta)}{\Delta_{j,0}^2} + H(j)\right\}$ as the appropriate instance-dependent complexity measure on positive instances.

Compared to prior work, our results resolve an open problem posed in Li and Cheung (2025) by providing the first nearly optimal upper bound for instances with multiple qualified arms.

1.4 Intuition of Sample Complexity in the Positive Case

For a positive instance ν, the intuition of sample complexity $O\Big(\log K \min_{1\leq j\leq m}\big\{\log(1/\delta)/\Delta_{j,0}^2 + H(j)\log^4 K \log 1/\Delta_{j,0}^2\big\}\Big)$ in Theorem 3 is as follows. A δ-PAC algorithm suffice to confirm one arm $j \in [m]$ satisfying $\mu_j > \mu_0$ with high confidence level, which means the coefficient of $\log 1/\delta$ should be dependent on only one of $\{\frac{1}{\Delta_{j,0}^2}\}_{j=1}^m$. Then, determining which j is appropriate remains an issue. If j is large, $\Delta_{j,0}$ could be small, leading to a large coefficient of $\log 1/\delta$. At the same time, a larger j guarantees a higher probability of "capturing" an arm in $[j]$, if we sample a fixed-size subset of $[K]$, which helps to avoid spending samples on all the arms.

Consider a "lucky" agent with prior knowledge $j^* = \arg\min_{1\leq j\leq m}\{(\log(1/\delta))/\Delta_{j,0}^2 + H(j)\}$. This agent samples K/j^* arms in $[K]$ uniformly at random, successfully capturing an arm j^* with reward μ_{j^*}. However, the arm's unknown position requires the agent $H(j^*)$ pulls to identify j^* at a low confidence level. Subsequently, it must pull arm j^* another $O(\log(1/\delta)/\Delta_{j^*,0}^2)$ times to confirm j^* is qualified with confidence $1-\delta$. Algorithm 1 maintains multiple parallel copies, wishing j^* is captured in an early bracket. Since the agent cannot pre-identify the correct bracket and low-confidence identification may be erroneous, the expected sample complexity includes an additional factor of $\log^4 K \log\frac{1}{\Delta_{j,0}^2}$ for the rough identification phase.

2 Lower Bound of Sample Complexity, for Positive Instances

In this section, we continue to work with a reward vector $\{\mu_a\}_{a=1}^K$ satisfying $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$ and consider the instance ν defined by $r_a^\nu = \mu_a$. Throughout the lower-bound analysis, we assume unit-variance Gaussian rewards. For any instance ν' and any permutation $\sigma : [K] \to [K]$ (a bijection), we define the permuted instance $\sigma(\nu')$ by $r_{\sigma(a)}^{\sigma(\nu')} \coloneqq r_a^{\nu'}, \forall a \in [K]$. We begin by introducing a symmetry notion that allows us to restrict attention to algorithms that do not prefer specific arm labels.

Definition 3 (Symmetric algorithm)**.** *An algorithm* $\widetilde{\text{alg}}$ *is* symmetric *if for any instance* ν *with reward vector* $\{\mu_a\}_{a=1}^K$*, any sequence* $\{a_t\}_{t=1}^T \in [K]^T$*, any integer* $T \geq 1$*, and any two permutations* $\sigma', \sigma'' : [K] \to [K]$*, it holds that*

$$\Pr_{\sigma'(\nu),\, \widetilde{\text{alg}}}\big(\forall t \in [T],\, A_t = \sigma'(a_t),\, \tau > T\big) = \Pr_{\sigma''(\nu),\, \widetilde{\text{alg}}}\big(\forall t \in [T],\, A_t = \sigma''(a_t),\, \tau > T\big).$$

Intuitively, under a symmetric algorithm $\widetilde{\text{alg}}$, the reward distribution of the pulled arm A_t at time t is invariant under any permutation on the arms' indexes. In particular, we have $\mathbb{E}_{\sigma'(\nu),\, \widetilde{\text{alg}}}[\tau] = \mathbb{E}_{\sigma''(\nu),\, \widetilde{\text{alg}}}[\tau]$, for permutations σ', σ''. A similar definition also appears in (Simchowitz et al., 2017, Definition 12), which further requires the invariance of output distribution; this extra requirement is not needed here. The following lemma shows that it suffices to consider symmetric algorithms for lower bounds.

Lemma 4. *Let* S *denote the set of all permutations of* $[K]$*. For any* δ*-PAC algorithm* alg*, there exists a* δ*-PAC symmetric algorithm* $\widetilde{\text{alg}}$ *such that for every instance* ν'*,* $\max_{\sigma \in S} \mathbb{E}_{\text{alg},\, \sigma(\nu')}[\tau] \geq \mathbb{E}_{\widetilde{\text{alg}},\, \nu'}[\tau]$.

The proof of Lemma 4 is in Appendix B.1. It is obtained by randomizing a permutation of arm labels prior to executing alg. Given Lemma 4, to prove Theorem 1, it suffices to establish the following lower bound for symmetric algorithms.

Theorem 5. *For any symmetric and* δ*-PAC alg and any reward vector* $\{\mu_a\}_{a=1}^K \in [0,1]^K$ *satisfying that* $\mu_1 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$*, any tolerance level* δ *such that* $\delta < \min\{10^{-8}, \frac{1}{64m^2}\}$*, define a positive instance* ν *with* $r_a^\nu = \mu_a, a \in [K]$ *with unit-variance Gaussian reward. We have* $\mathbb{E}_{alg,\nu}\tau \geq \Omega\left(\min_{1 \leq j \leq m}\left\{\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{H(j)}{\log^2(m+1)}\right\}\right)$.

The assumption $\delta < \min\{10^{-8}, 1/(64m^2)\}$ is to ensure $\delta \cdot m^2$ is upper bounded by an absolute constant. The upper bound eliminates the possibility of being δ-PAC by outputting a uniformly chosen arm without any arm pulls.

Sketch Proof of Theorem 5. We outline the main steps; the full proof is in Appendix B.2. The main idea is to establish Lemma 6,

Lemma 6. *Assume* $\delta < \min\{10^{-8}, 1/(64m^2)\}$*, for any symmetric and* δ*-PAC alg,* $\mathbb{E}_{\nu,alg}\tau$ *is lower bounded by the* $\frac{1}{3200}$ *times optimal value of the following optimization problem*

$$\min_{v, \{p_j\}_{j=1}^m} \quad v \tag{2}$$

$$\text{s.t.} \quad 0 \leq p_j \leq 1 \quad \forall j \in [m] \tag{C1}$$

$$v \geq 0 \tag{C2}$$

$$v \geq \sum_{j=1}^m \frac{p_j \log\frac{1}{\delta}}{\Delta_{j,0}^2} \tag{C3}$$

$$\sum_{j=1}^m p_j \geq \frac{1}{2} \tag{C4}$$

$$v \geq \sum_{a=1}^K \frac{p_j}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}\left(1 + \log\frac{1}{p_j}\right)} \quad \forall j \in [m]. \tag{C5}$$

Given Lemma 6, we proceed to prove the Theorem 1 as follows. We first prove the above optimization problem (2) is convex. Then, we turn to construct an explicit feasible dual solution, and prove the objective value of this feasible solution is lower bounded by $\Omega\left(\min_{1 \leq j \leq m}\left\{\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{H(j)}{\log^2(m+1)}\right\}\right)$. By the strong duality, we complete the proof of Theorem 5.

Denote event $\mathcal{E}_j = \left\{N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a} = j\right\}$. To prove Lemma 6, we show taking $v = 3200\mathbb{E}_\nu\tau$, $p_j = \Pr_\nu(\mathcal{E}_j)$ for $j \in [m]$ is a feasible solution to (2). The argument repeatedly applies the

transportation (change-of-measure) inequality (Kaufmann et al., 2016, Lemma 18) to relate $\Pr_\nu(\mathcal{E}_j)$ to probabilities under suitable alternative instances. We prove the feasibility of $v = 3200\mathbb{E}_\nu\tau$, $p_j = \Pr_\nu(\mathcal{E}_j)$ by going through the constraint (C1), (C2), (C3), (C4) (C5).

For (C1), (C2), they are evident. As the expected pulling times must be non-negative and the probability must be in the interval $[0, 1]$.

For (C3), notice that $\mathbb{1}(\tau = +\infty)+\mathbb{1}(\tau < +\infty, \hat{a} \in [K]\cup\{\mathsf{None}\}) = 1$, we have $\sum_{j=1}^m \mathbb{1}(\hat{a} = j) \leq 1$, further $\sum_{j=1}^m \mathbb{1}(\mathcal{E}_j) \leq 1$. What's more, the event $\mathcal{E}_j$ guarantees that $\tau\mathbb{1}(\mathcal{E}_j) \geq N_j(\tau)\mathbb{1}(\mathcal{E}_j) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\mathbb{1}(\mathcal{E}_j)$. These observations concludes

$$3200\mathbb{E}\tau \geq \mathbb{E}\tau \geq \sum_{j=1}^m \mathbb{E}\tau\mathbb{1}(\mathcal{E}_j) \geq \sum_{j=1}^m \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\Pr_\nu(\mathcal{E}_j).$$

For (C4), we start by proving

$$\Pr_\nu(\tau < +\infty, \hat{a} \in [m]) \geq 1 - 5\delta, \tag{3}$$

which must hold regardless of whether $\{a : \mu_a = \mu_0\}$ is empty. The main idea is to utilize the assumption that the algorithm is δ-PAC.

Then, we are ready to prove (C4) by contradiction. Denote $\bar{\mathcal{E}}_j = \{N_j(\tau) < \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a} = j\}$. If $\sum_{j=1}^m \Pr_\nu(\mathcal{E}_j) < \frac{1}{2}$, together with (3), we can find a $j_0 \in [m]$ such that

$$\Pr_\nu\left(\bar{\mathcal{E}}_{j_0}\right) \geq \frac{1}{8m} \geq \sqrt{\delta}. \tag{4}$$

We consider an alternative instance $\nu'_{j_0}(\epsilon)$ with $r_{a'}^{\nu'_{j_0}} = \begin{cases}\mu_0 - \epsilon & a' = j_0\\ r_{a'}^\nu & else\end{cases}$. The δ-PAC requirement of algorithm means that

$$\Pr_{\nu'_{j_0}(\epsilon)}(\bar{\mathcal{E}}_{j_0}) < \delta \tag{5}$$

Combining (5),(4), we have $\log\frac{\Pr_\nu(\bar{\mathcal{E}}_{j_0})}{\Pr_{\nu'_{j_0}(\epsilon)}(\bar{\mathcal{E}}_{j_0})} \geq \frac{1}{2}\log\frac{1}{\delta}$ holds for all $\epsilon > 0$. Meanwhile, Lemma 9, which utilizes the Transportation Equality (Kaufmann et al., 2016, Lemma 18), suggests that

$$\begin{aligned}&\log\frac{\Pr_\nu\left(\bar{\mathcal{E}}_{j_0}\right)}{\Pr_{\nu'_{j_0}(\epsilon)}(\bar{\mathcal{E}}_{j_0})}\\ \leq&\sqrt{2\frac{\log\frac{1}{\delta}}{100\Delta_{j_0,0}^2}(\Delta_{j_0,0}+\epsilon)^2\log\frac{1}{\sqrt{\delta}}} + \sqrt{2\pi\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}(\Delta_{j_0,0}+\epsilon)^2} + \frac{1}{2}\frac{\log\frac{1}{\delta}}{100\Delta_{j_0,0}^2}(\Delta_{j_0,0}+\epsilon)^2\end{aligned} \tag{6}$$

Taking $\epsilon \to 0$, we can show the right hand side of (6) is less than $\frac{1}{2}\log\frac{1}{\delta}$ for $\delta < 10^{-8}$, which leads to a contradiction. Lemma 10 merges all the above ideas and conclude $\sum_{j=1}^m \Pr_\nu(\mathcal{E}_j) \geq \frac{1}{2}$.

For (C5), we first prove Lemma 12, asserting that

$$\mathbb{E}_\nu N_a(\tau) \geq \frac{1}{4}T_j(a)\Pr_\nu\left(N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a} = j\right). \tag{7}$$

for each $j \in [m], a \in [K], j \neq a$, where $T_j(a) = \lceil\frac{1}{200\max\{\Delta_{j,0}^2,\Delta_{a,j}^2\}\left(1+\log\frac{1}{\Pr_\nu(\mathcal{E}_j)}\right)}\rceil$. Given Lemma 12 holds, sum up (7) for each $a \in [K]$ and use the fact that $\frac{1}{\max\{\Delta_{j,0}^2,\Delta_{a,j}^2\}} \geq \frac{1}{4\max\{\Delta_{a,0}^2,\Delta_{a,j}^2\}}$, we prove (C5). Then, the remaining work is to prove (7). In the following, we present the sketch idea, which highly relies on the definition of $T_j(a)$.

For $n \in \mathbb{N}$, we define $\tau_n^j(a)$ by looking for an time index t, such that $N_j(t) + N_a(t) = n, A_{t+1} \in \{j, a\}$. Such a time index t is unique and $t < \tau$, if it exists. We take $\tau_n^j(a) = t$ if t exists, or

$\tau_n^j(a) = +\infty$ if t doesn't exist. Since A_{t+1} is fully determined by the history up to time index t, for $t \in \mathbb{N}$. $\tau_n^j(a)$ is a stopping time regrading the filtration $\{\sigma(H(t))\}_{t\geq 1}$.

Denote $\tilde{\mathcal{E}}^{aj}(n) = \{\tau_n^j(a) < \tau\}$, $\tilde{\mathcal{E}}_j^{aj}(n) = \{A_{\tau_n^j(a)+1} = j, \tau_n^j(a) < \tau\}$, $\tilde{\mathcal{E}}_a^{aj}(n) = \{A_{\tau_n^j(a)+1} = a, \tau_n^j(a) < \tau\}$. We prove two inequalities.

$$\Pr_\nu(\tilde{\mathcal{E}}^{aj}(n)) \geq \Pr_\nu(\mathcal{E}_j), \forall n \leq T_j(a) - 1. \tag{8}$$

$$\Pr_\nu(\tilde{\mathcal{E}}_a^{aj}(n)) \geq \frac{1}{4}\Pr_\nu(\tilde{\mathcal{E}}^{aj}(n)), \forall n \leq T_j(a) - 1. \tag{9}$$

Given (8) and (9), from the fact that we only pull arm a at $\tau_n^j(a) + 1$, we can conclude

$$\begin{aligned}\mathbb{E}_\nu N_a(\tau) = \sum_{n=0}^{+\infty}\Pr_\nu(\tilde{\mathcal{E}}_a^{aj}(n)) \geq \sum_{n=0}^{T_j(a)-1}\Pr_\nu(\tilde{\mathcal{E}}_a^{aj}(n)) &\overset{(9)}{\geq} \sum_{n=0}^{T_j(a)-1}\frac{1}{4}\Pr_\nu(\tilde{\mathcal{E}}^{aj}(n)) \\ &\overset{(8)}{\geq}\frac{1}{4}\sum_{n=0}^{T_j(a)-1}\Pr_\nu(\mathcal{E}_j) = \frac{1}{4}T_j(a)\Pr_\nu(\mathcal{E}_j).\end{aligned}$$

Thus, to prove (7), we suffice to show (8) (9) are both correct.

For (8), we notice that $\frac{1}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}} \leq \frac{1}{\Delta_{j,0}^2}$, and that $T_j(a) < \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}$. That means for $n \leq T_j(a)$, we have $\mathcal{E}_j \Rightarrow \{N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\} \Rightarrow \{\tau_n^j(a) < \tau\}$, which suggests (8).

For (9), we introduce an alternative instance $\nu_{j,a}$, with $r_{a'}^{\nu_{j,a}} = \begin{cases} r_a^\nu & a' = j \\ r_j^\nu & a' = a \\ r_{a'}^\nu & else \end{cases}$. The instance distribution $\nu_{j,a}$ swaps the position of arm a and j compared to the instance ν. Notice that

$$\frac{1}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}} \leq \frac{1}{\Delta_{j,a}^2}, \quad T_j(a) \leq \frac{O(1)}{\Delta_{j,a}^2(1 + \log\frac{1}{\Pr_\nu(\mathcal{E}_j)})} \tag{10}$$

Lemma 8 and 11 combine (10) and the Transportation Equality(Lemma 18 in Kaufmann et al. (2016)), to conclude

$$\mathrm{kl}\Big(\Pr_\nu(\tilde{\mathcal{E}}_j^{aj}(n)), \Pr_{\nu_{ja}}(\tilde{\mathcal{E}}_j^{aj}(n))\Big) \leq 0.4\Pr_\nu(\tilde{\mathcal{E}}^{aj}(n)), \forall n \leq T_j(a) - 1 \tag{11}$$

where $\mathrm{kl}(x, y) = x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y}$. $1 + \log\frac{1}{\Pr_\nu(\mathcal{E}_j)}$ in the denominator of $T_j(a)$ is crucial to show (11). (11) suggests that the algorithm is unable to distinguish ν and ν_{ja} at the round t such that $N_a(t) + N_j(t) \leq T_j(a)$. For a symmetric algorithm, we have $\Pr_{\nu_{ja}}(\tilde{\mathcal{E}}_j^{aj}(n)) = \Pr_\nu(\tilde{\mathcal{E}}_a^{aj}(n))$, which means

$$\mathrm{kl}\Big(\Pr_\nu(\tilde{\mathcal{E}}_j^{aj}(n)), \Pr_\nu(\tilde{\mathcal{E}}_a^{aj}(n))\Big) \leq 0.4\Pr_\nu(\tilde{\mathcal{E}}^{aj}(n)), \forall n \leq T_j(a) - 1 \tag{12}$$

By the Pinsker's Inequality $\mathrm{kl}(x, y) \geq 2(x-y)^2$ and $\Pr_\nu(\tilde{\mathcal{E}}_j^{aj}(n)) + \Pr_\nu(\tilde{\mathcal{E}}_a^{aj}(n)) = \Pr_\nu(\tilde{\mathcal{E}}^{aj}(n))$, we can derive

$$\Pr_\nu(\tilde{\mathcal{E}}_a^{aj}(n)) \geq \frac{1}{4}\Pr_\nu(\tilde{\mathcal{E}}^{aj}(n)), \forall n \leq T_j(a) - 1.$$

The above inequality suggests that we have proved (9), hence (C5) is justified. □

3 Algorithm Design for the Upper Bound of the Sample Complexity

In this section, we present our proposed algorithm, *Parallel Sequential Exploration Exploitation on Brackets* (PSEEB), together with a high-level discussion of its design and analysis.

Algorithm 1 merges and adapts the bracketing idea of Katz-Samuels and Jamieson (2020) with the Sequential Exploration–Exploitation (SEE) framework of Li and Cheung (2025). In Katz-Samuels and Jamieson (2020), each bracket is generated by an independent random shuffle. In contrast, PSEEB

Algorithm 1: Parallel Sequential Exploration Exploitation on Brackets(PSEEB)

Input: Threshold μ_0, Error tolerance level δ, Arm number K;
Randomly generate a permutation of $\sigma : [K] \to [K]$, take $B_b = \{\sigma^{-1}(i)\}_{i=1}^{\min\{2^{b-1},K\}}$ for $b = 1, 2, \cdots, \lceil \log_2 K \rceil + 1$; // $\sigma^{-1}(i)$ `denotes the arm index in position` i
Create algorithm copies $\{\text{alg}_b\}_{b=1}^{\lceil \log_2 K \rceil + 1}$ by specifying alg_b through calling Algorithm 3 with $B \leftarrow B_b$, $\mu_0 \leftarrow \mu_0$, $\delta \leftarrow \frac{\delta}{\lceil \log_2 K \rceil + 1}$, $C \leftarrow 1.01$, $K \leftarrow K$. Samples across brackets are not shared;
for *round index* $t = 1, 2, \cdots$ **do**
  for *Bracket Index* $b = 1, 2, \cdots, \lceil \log_2 K \rceil + 1$ **do**
    Resume alg_b, execute one arm pull requested by alg_b, return the realized reward to alg_b;
    if *alg_b terminates with output* $\hat{a} \in [K] \cup \{\textit{None}\}$ **then**
      Terminate the whole algorithm and return $\hat{a}$.
    end
  end
end

Algorithm 2: (Informal) Sequential Exploration Exploitation

Input: Action subset B, threshold μ_0, tolerance level δ, C, K;
Specify $\delta_k = \frac{1}{3^k}$, $\alpha_k = 5^k$, $\beta_k = 2^k$, $T_k^{\text{et}} \propto \beta_k \log \frac{K}{\delta_k}$, $T_k^{\text{et}} \propto \beta_k \log \frac{K\alpha_k}{\delta}$;
for *Phase* $k = 1, 2, \cdots$ **do**
  while *True* **do**
    if $\exists a \in B$, *LCB*$_a(t, \delta_k) > \mu_0$ **then**
      Consecutively keep pulling a up to T_k^{et} times. If $\text{LCB}_a(t, \delta/\alpha_k) > \mu_0$ during the consecutive pulling, Stop and output a ▷ Exploit by keeping pulling;
    if $\forall a \in B$, *UCB*$_a(t, \delta_k) \le \mu_0$, $B = [K]$ *and* $\delta_k < \delta/3$ **then**
      Stop and output None;
    end
    if *At least 1 of 3 conditions holds for all* $a \in B$ *(1)* $N_a(t) \ge T_k^{ee}$ *(2)* *UCB*$_a(t, \delta_k) < \mu_0$ *(3)* *UCB*$_a(t, \delta_k) < \max_{a'}$ *UCB*$_{a'}(t, \delta_k)$ **then**
      Break, Go to the start of phase $k + 1$;
    end
    Pick an arm A_t^{ee} who simultaneously fulfills (1) UCB is highest among B (2) UCB is higher than μ_0 (3) Pulling times is less than T_k^{ee} ▷ Explore by maximum UCB Rule;
  end
end

defines *all* brackets from a *single* random permutation of $[K]$, yielding a nested family $\{B_b\}_{b=1}^{\lceil \log_2 K \rceil + 1}$ satisfying $B_1 \subset B_2 \subset \cdots \subset B_{\lceil \log_2 K \rceil + 1} = [K]$. For a bracket index $1 \le b \le \lceil \log_2 K \rceil$, we have $|B_b| = 2^b$, and $|B_{\lceil \log_2 K \rceil + 1}| = \min\{2^{\lceil \log_2 K \rceil}, K\}$. This monotone structure is crucial for our analysis, in particular for Lemma 7, which guarantees that with $\Omega(1)$ probability, a bracket B_b with size $|B_b| \ge \frac{K}{j}$ satisfies $[j] \cap B_b \ne \emptyset$.

Lemma 7. *Given a qualified arm index* $j \in [m]$*, denote bracket index* $b_j = \min\{b \in [\lceil \log_2 K \rceil + 1] : B_b \cap [j] \ne \emptyset\}$*. We have* $\Pr(b_j \ge \tilde{b}) \le \exp\left(-\frac{j|B_{\tilde{b}-1}|}{K}\right), \forall \tilde{b} \in [\lceil \log_2 K \rceil + 1]$*. Here we denote* $B_0 = \emptyset$ *and take* $|B_0| = 0$*.*

Then, PSEEB instantiates $\lceil \log_2 K \rceil + 1$ independent algorithm copies via Algorithm 3 (given in Appendix C). Algorithm 2 presents the sketch idea of Algorithm 3. This procedure is adapted from Algorithms 2–5 of Li and Cheung (2025), named as Sequential Exploration Exploitation(SEE). SEE splits the sampling into two parts, exploration and exploitation. The samples across two parts are not shared and independent. We adapt SEE with modified budget allocation and stopping rules to accommodate bracketing. Algorithm copy $b \in [\lceil \log_2 K \rceil + 1]$ takes bracket B_b as an input parameter and only pull arms in B_b. We wish to use a tolerance level $\delta_k > \delta$ to define upper/lower

confidence bounds $\mathrm{UCB}_a(t,\delta_k)$, $\mathrm{LCB}_a(t,\delta_k)$. Once we identify an candidate arm $a \in B_b$ such that $\mathrm{LCB}_a(t,\delta_k) \geq \mu_0$, we keep pulling it and see whether $\mathrm{LCB}_a(t,\delta) \geq \mu_0$ will hold during the consecutive pulling. Detailed discussion of Algorithm 3 is in Appendix C.

Algorithm 1 manages all the algorithm copies in a round-robin manner. In each global round (Line 4), Algorithm 1 execute one of the algorithm copies with one arm pull, and the overall algorithm terminates as soon as some copy outputs $\hat{a} \in [K] \cup \{\mathsf{None}\}$. By construction, only the copy $\mathrm{alg}_{\lceil \log_2 K \rceil + 1}$ (which operates on $B_{\lceil \log_2 K \rceil + 1} = [K]$) is permitted to output None. For any other bracket B_b with $|B_b| < K$, the corresponding copy never outputs None; even if it becomes convinced that $\max_{a \in B_b} \mu_a < \mu_0$, it has to continue sampling.

Let τ denote the stopping time of PSEEB, and let $\tau(B_b)$ denote the number of rounds executed by copy alg_b up to its own termination (possibly ∞). Since the Algorithm 1 terminates upon the first terminating copy and adopts the output, we have $\tau \leq (\lceil \log_2 K \rceil + 1)\tau(B_b)$ holds for all $b \in [\lceil \log_2 K \rceil + 1]$, and hence $\mathbb{E}\tau \leq [\lceil \log_2 K \rceil + 1]\mathbb{E}\tau(B_b)$.

Theorems 2 and 3 are derived from several technical lemmas. We provide a high level summary here, with proof sketch in Appendix C.3, full proof in Appendix C.4.

To prove Theorem 2, the first step is prove $\Pr(\tau(B_{\lceil \log_2 K \rceil + 1}) = +\infty) = 0$, whose main idea is very similar to the proof of Theorem 5.2 in Li and Cheung (2025). Denote $\kappa_b^{\mathrm{ee}}, \kappa_b^{\mathrm{et}}$ as the minimum phase indices after which the exploration and exploitation concentration events of copy b hold, we can prove $\Pr(\tau(B_{\lceil \log_2 K \rceil + 1}) = +\infty) \leq \Pr(\max\{\kappa^{\mathrm{ee}}_{\lceil \log_2 K \rceil + 1}, \kappa^{\mathrm{et}}_{\lceil \log_2 K \rceil + 1}\} = +\infty) = 0$. Given the first step, we are able to conclude $\Pr\left(\tau < +\infty\right) = 1$, by the fact that $\tau \leq (\lceil \log_2 K \rceil + 1)\tau(B_b)$ holds for all $b \in [\lceil \log_2 K \rceil + 1]$. We define good events $\mathcal{E}_b$ for each copy b, and we show a copy b only output wrong answer when $\neg\mathcal{E}_b$ holds. In Line 3 of Algorithm 1, we assign $\delta/(\lceil \log_2 K \rceil + 1)$ to each copy, which guarantees $\Pr(\neg\mathcal{E}_b) < \frac{O(1)\delta}{\lceil \log_2 K \rceil + 1}$ holds for all $b \in [\lceil \log_2 K \rceil + 1]$. A union bound over all copies completes the proof.

To prove Theorem 3, we consider negative and positive instances separately. If the instance ν is negative, we focus on analyzing $\tau(B_{\lceil \log_2 K \rceil + 1})$, similar to the analysis in Theorem 5.3 in Li and Cheung (2025). By the fact that $\mathbb{E}\tau \leq [\lceil \log_2 K \rceil + 1]\mathbb{E}\tau(B_{\lceil \log_2 K \rceil + 1})$, we complete the proof.

If the instance ν is positive, denote $\mu_1 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$. For $j \in [m]$ and $a \in [K]$, define $\tilde{T}_j(a) = \frac{\log(4K) + \log_2 \frac{1}{\Delta^2_{j,0}}}{\max\{\Delta^2_{a,j}, \Delta^2_{a,0}\}}$, $\tilde{T}_j(a;k) = \frac{\log \frac{4K}{\delta_k} + \log_2 \frac{1}{\Delta^2_{j,0}}}{\max\{\Delta^2_{a,j}, \Delta^2_{a,0}\}}$. Target is to prove $\forall j \in [m], \mathbb{E}\tau \leq O(\log K) \cdot O\left(\frac{\log \frac{1}{\delta}}{\Delta^2_{j,0}} + \frac{\log^3 K}{j} \sum_{a=1}^K \tilde{T}_j(a)\right)$. For any given $j \in [m]$, we introduce $b_j = \min\{b \in [\lceil \log_2 K \rceil + 1] : B_b \cap [j] \neq \emptyset\}$, the smallest bracket index that contains at least one arm among the top-j arms. As we adopt a random permutation to create brackets, b_j is a random variable, and we have $\mathbb{E}\tau = \sum_{b=1}^{\lceil \log_2 K \rceil + 1} \mathbb{E}\tau \mathbb{1}(b_j = b)$. Using $\tau \leq (\lceil \log_2 K \rceil + 1)\tau(B_b)$ for each $b \in [\lceil \log_2 K \rceil + 1]$, we suffice to analyze the upper bound of $O(\log K)\, \mathbb{E}_\nu[\tau(B_b)\mathbb{1}(b_j = b)]$.

To analyze $\tau(B_b)\mathbb{1}(b_j = b)$, we reformulate it as $\sum_{a=1}^K N_a(\tau(B_b))\mathbb{1}(b_j = b)\mathbb{1}(a \in B_b)$. Denote κ_{b_j} as the minimum phase index k, such that the concentration event of copy b_j holds. We prove $N_a\mathbb{1}(b_j = b)\mathbb{1}(a \in B_b) \leq \tilde{T}_j(a;\kappa_{b_j})\mathbb{1}(b_j = b)\mathbb{1}(a \in B_b)$, in the exploration part. The reason is $\tilde{T}_j(a;\kappa_{b_j})$ samplings on a are sufficient to confirm either $\mu_a \leq \mu_j$ or $\mu_a \geq \mu_0$ holds, in the exploration part, with confidence level $1 - \delta_{\kappa_{b_j}}$. Then, the exploitation part captures a candidate arm $\hat{a}$ with mean reward no less than μ_j, spending $O((\log \frac{1}{\delta})/\Delta^2_{j,0})$ sampling times to confirm $\hat{a}$ is indeed qualified, showing that $\tau(B_b)\mathbb{1}(b_j = b) \leq O(\log \frac{1}{\delta})/\Delta^2_{j,0}\mathbb{1}(b_j = b) + \sum_{a=1}^K \tilde{T}_j(a;\kappa_{b_j})\mathbb{1}(b_j = b)\mathbb{1}(a \in B_b)$. Taking expectation on both sides with Lemma 7, 18 bounding the tail probability of b_j, κ_{b_j}, we complete the proof.

4 Conclusion

This paper works on the open problem 1-identification, providing a novel lower bound for positive instances derived by a specialized optimization formulation. We further design an algorithm that is nearly optimal across all instances. Further refining logarithmic factors remains a future research.

References

Lijie Chen and Jian Li. On the optimal sample complexity for best arm identification. *arXiv preprint arXiv:1511.03774*, 2015.

Lijie Chen, Jian Li, and Mingda Qiao. Nearly instance optimal sample complexity bounds for top-k arm selection. In *Artificial Intelligence and Statistics*, pages 101–110. PMLR, 2017.

Rémy Degenne and Wouter M Koolen. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32, 2019.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Pac bounds for multi-armed bandit and markov decision processes. In *Computational Learning Theory: 15th Annual Conference on Computational Learning Theory, COLT 2002 Sydney, Australia, July 8–10, 2002 Proceedings 15*, pages 255–270. Springer, 2002.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, 25, 2012.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR, 2016.

Kevin Jamieson and Robert Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2014.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sebastien Bubeck. On finding the largest mean among many. *arXiv preprint arXiv:1306.3917*, 2013.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439. PMLR, 2014.

Marc Jourdan and Clémence Réda. An anytime algorithm for good arm identification. *arXiv preprint arXiv:2310.10359*, 2023.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, volume 12, pages 655–662, 2012.

Hideaki Kano, Junya Honda, Kentaro Sakamaki, Kentaro Matsuura, and Masashi Sugiyama. Good arm identification via bandit feedback. *Machine Learning*, (2), 2017.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International conference on machine learning*, pages 1238–1246. PMLR, 2013.

Julian Katz-Samuels and Kevin Jamieson. The true sample complexity of identifying good arms. In *International Conference on Artificial Intelligence and Statistics*, pages 1781–1791. PMLR, 2020.

Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Conference on Learning Theory*, pages 228–251. PMLR, 2013.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17:1–42, 2016.

Zitian Li and Wang Chi Cheung. Near optimal non-asymptotic sample complexity of 1-identification. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 34144–34184. PMLR, 13–19 Jul 2025. URL `https://proceedings.mlr.press/v267/li25f.html`.

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *International Conference on Machine Learning*, pages 1690–1698. PMLR, 2016.

Blake Mason, Lalit Jain, Ardhendu Tripathy, and Robert Nowak. Finding all epsilon-good arms in stochastic bandits. *Advances in Neural Information Processing Systems*, 33:20707–20718, 2020.

Subhojyoti Mukherjee, Kolar Purushothama Naveen, Nandan Sudarsanam, and Balaraman Ravindran. Thresholding bandits with augmented ucb. *arXiv preprint arXiv:1704.02281*, 2017.

Riccardo Poiani, Martino Bernasconi, and Andrea Celli. Non-asymptotic analysis of (sticky) track-and-stop. In *International Conference on Learning Representations*, volume 2026, pages 65859–65889, 2026.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, pages 1794–1834. PMLR, 2017.

Tzu-Hsien Tsai, Yun-Da Tsai, and Shou-De Lin. lil'hdoc: an algorithm for good arm identification under small threshold gap. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 78–89. Springer, 2024.

A Detailed Literature Review

A.1 Simplification of History Result

In this section, we only consider Gaussian reward with unit variance. We discuss more details of the upper bounds presented in Table 1. We simplify the result in the history literature with smaller but explicit formulations, for the comparison in Table 1.

A.1.1 Theorem 2 in Poiani et al. (2026)

Theorem 2 in Poiani et al. (2026) presents the Non-asymptotic analysis for algorithm S-TaS. For an instance ν with $r_a^\nu = \mu_a$, $\mu_1 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$ and , they define $\neg a = \{\nu' : r_a^{\nu'} < \mu_0\}$ for $a \in [m]$, and

$$i_F(\nu) = \arg\max_{a\in[K]} \sup_{w\in\Delta_K} \inf_{\nu'\in\neg a} \sum_{a'\in K} \frac{w_{a'}(r_{a'}^{\nu'} - r_{a'}^{\nu})^2}{2}.$$

Then, Theorem 2 in Poiani et al. (2026) considers ϵ_ν such that for all instance ν' such that

$$\|\{r_a^{\nu'}\}_{a=1}^K - \{r_a^{\nu}\}_{a=1}^K\|_\infty \leq \epsilon_\nu \implies i_F(\nu') \subset i_F(\nu) \cup ([K] \setminus i^*(\nu)).$$

Here $i^*(\nu) = [m]$. Let

$$T_\nu = \max\left\{10K^4, \inf\left\{n \in \mathbb{N} : \sqrt{\frac{8D_K \log n}{\sqrt{\sqrt{n} + K^2} - 2K}} \leq \epsilon_\nu\right\}\right\}$$

where $D_K \geq 10$. The upper bound for $\mathbb{E}_\nu \tau$ is

$$\mathbb{E}_\nu \tau \leq 10K^4 + T_0(\delta) + \frac{\pi^2}{24}$$

where

$$T_0(\delta) = \inf\left\{t \in \mathbb{N} : \log(Ct^2/\delta) \leq (t - \sqrt{t} - 1 - T_\nu)\frac{\Delta_{0,1}^2}{2} - g(t)\right\}$$

and $C \geq e\sum_{t=1}^{+\infty}(\frac{e}{K})^K \frac{(\log^2(Ct^2)\log(t))^K}{t^2}$, $g(t) = \tilde{O}(K^2\sqrt{D_K t} + \sqrt{D_K t^{\frac{3}{2}}})$.

To simplify the above result, we can prove $\frac{2\log(1/\delta)}{\Delta_{0,1}^2} + \frac{K^2}{\Delta_{0,1}^4}$ is a lower bound of $10K^4 + T_0(\delta) + \frac{\pi^2}{24}$. From the definition of T_ν, we know

$$\begin{aligned}
T_\nu \geq & \max\left\{10K^4, \inf\left\{n \in \mathbb{N} : \sqrt{\frac{8D_K \log n}{\sqrt{\sqrt{n}+K^2}-2K}} \leq \epsilon_\nu\right\}\right\} \\
\geq & \max\left\{10K^4, \inf\left\{n \in \mathbb{N} : \sqrt{\frac{8D_K}{\sqrt{\sqrt{n}+K^2}-2K}} \leq \epsilon_\nu\right\}\right\} \\
= & \max\left\{10K^4, \inf\left\{n \in \mathbb{N} : \frac{80}{\epsilon_\nu^2} + 2K \leq \sqrt{\sqrt{n}+K^2}\right\}\right\} \\
\geq & \max\left\{10K^4, \inf\left\{n \in \mathbb{N} : \frac{320K}{\epsilon_\nu^2} + 4K^2 \leq \sqrt{n}+K^2\right\}\right\} \\
\geq & \max\left\{10K^4, \frac{90000K^2}{\epsilon_\nu^4}\right\}.
\end{aligned}$$

We can further derive

$$\begin{aligned}
T_0(\delta) \geq & \inf\left\{t \in \mathbb{N} : \log(C/\delta) \leq \left(t - \max\left\{10K^4, \frac{90000K^2}{\epsilon_\nu^4}\right\}\right)\frac{\Delta_{0,1}^2}{2}\right\} \\
= & \frac{2\log(C/\delta)}{\Delta_{0,1}^2} + \max\left\{10K^4, \frac{90000K^2}{\epsilon_\nu^4}\right\}.
\end{aligned}$$

Thus, we can conclude $\frac{2\log(1/\delta)}{\Delta_{0,1}^2} + \frac{K^2}{\epsilon_\nu^4}$ is smaller than the proposed upper bound in Poiani et al. (2026). The remaining work is to show $\epsilon_\nu \leq \Delta_{0,1}$. Denote $a^{\text{sub}} = \arg\max_{a:\mu_a<\mu_1} \mu_a$. We have

- If $\mu_{a^{\text{sub}}} \geq \mu_0$, we assert $\epsilon_\nu \leq \Delta_{1,a^{\text{sub}}} < \Delta_{1,0}$. If not, we can set ν' with $r_{a^{\text{sub}}}^{\nu'} = \mu_1$ while keeping $r_a^{\nu'} = r_a^{\nu'}$ for $a \neq a^{\text{sub}}$. With this construction, we have $a^{\text{sub}} \in i_F(\nu'), a^{\text{sub}} \notin \{1, m+1, m+2, \cdots, K\}$, contradicting to the requirement of ϵ_ν.
- If $\mu_{a^{\text{sub}}} < \mu_0$, we assert $\epsilon_\nu \leq \Delta_{1,0}$. If $\epsilon_\nu > \Delta_{1,0}$, we can take ν' with $r_a^{\nu'} = \begin{cases}\mu_1 - \epsilon_\nu & \text{if } \mu_a = \mu_1 \\ \mu_a & \text{if } \mu_a < \mu_1\end{cases}$ for $a \in [K]$. Easy to validate $\|\{r_a^{\nu'}\}_{a=1}^K - \{r_a^{\nu}\}_{a=1}^K\|_\infty \leq \epsilon_\nu$ and $i_F(\nu') = \mathsf{None} \notin i_F(\nu) \cup ([K] \setminus i^*(\nu))$

Thus, we have shown $\frac{2\log(1/\delta)}{\Delta_{0,1}^2} + \frac{K^2}{\Delta_{0,1}^4}$ is a lower bound of $10K^4 + T_0(\delta) + \frac{\pi^2}{24}$.

A.1.2 Theorem 3 in Kano et al. (2017)

Following Appendix A.2 in Li and Cheung (2025), we can also simplify the Theorem 3 in Kano et al. (2017). By applying the algorithm HDoC, for $\nu \in \mathcal{S}^{\text{pos}}$, we have

$$\mathbb{E}_\nu \tau \leq n_1 + \sum_{a=2}^K \left(\frac{\log(K \max_{j\in[K]} n_j)}{2(\Delta_{1,a}-2\epsilon)^2} + \delta n_a\right) + \frac{K^{2-\frac{\epsilon^2}{(\min_{a\in[K]}\Delta_{0,a}-\epsilon)^2}}}{2\epsilon^2} + \frac{K(5+\log\frac{1}{2\epsilon^2})}{4\epsilon^2},$$

where $n_a = \frac{1}{(\Delta_{0,a}-\epsilon)^2}\log\left(\frac{4\sqrt{K/\delta}}{(\Delta_{0,a}-\epsilon)^2}\log\frac{5\sqrt{K/\delta}}{(\Delta_{0,a}-\epsilon)^2}\right)$, $\epsilon = \min\left\{\min_{a\in[K]}\Delta_{0,a}, \min_{a\in[K-1]}\frac{\Delta_{a,a+1}}{2}\right\}$. From direct calculation, we can observe

$$\begin{aligned}
& n_1 \geq \frac{\log\frac{K}{\delta}}{2\Delta_{0,1}^2}, \quad \frac{K(5+\log\frac{1}{2\epsilon^2})}{4\epsilon^2} \geq \frac{K}{4\epsilon^2} \\
& \log(K \max_{j\in[K]} n_j) \geq \log(\frac{K}{2}\log\frac{1}{\delta}) \geq \log\log\frac{1}{\delta}.
\end{aligned}$$

From the above inequality, we can conclude

$$\begin{aligned}
&\sum_{a=2}^{K} \frac{\log(K \max_{j\in[K]} n_j)}{2(\Delta_{1,a}-2\epsilon)^2} \\
\geq &\sum_{a=2}^{K} \frac{\log\log \frac{1}{\delta}}{2(\Delta_{1,a}-2\epsilon)^2} \\
\geq &\sum_{a=2}^{K} \frac{\log\log \frac{1}{\delta}}{2\Delta_{1,a}^2} \\
\geq &\frac{H_1}{2}\log\log\frac{1}{\delta}.
\end{aligned}$$

Thus, we can conclude

$$\begin{aligned}
&n_1 + \sum_{a=2}^{K}\left(\frac{\log(K \max_{j\in[K]} n_j)}{2(\Delta_{1,a}-2\epsilon)^2} + \delta n_a\right) + \frac{K^{2-\frac{\epsilon^2}{(\min_{a\in[K]}\Delta_{0,a}-\epsilon)^2}}}{2\epsilon^2} + \frac{K(5+\log\frac{1}{2\epsilon^2})}{4\epsilon^2} \\
\geq &\frac{\log\frac{K}{\delta}}{2\Delta_{0,1}^2} + \frac{H_1}{2}\log\log\frac{1}{\delta} + \frac{K}{4\epsilon^2}.
\end{aligned}$$

A.1.3 Thereom 2 in Jourdan and Réda (2023)

Jourdan and Réda (2023) do not explicitly provide upper bound for $\mathbb{E}\tau$. Therorem 2 in Jourdan and Réda (2023) suggests that for a positive instance ν.

$$\mathbb{E}_\nu \tau \leq C^{\text{pos}}(\delta) + \frac{K\pi^2}{6} + 1$$

where

$$\begin{aligned}
C^{\text{pos}}(\delta) =& \sup\left\{t : t \leq \left(\sum_{a:\mu_a\geq\mu_0}\frac{2}{\Delta_{0,a}^2}\right)(\sqrt{c(t,\delta)}+\sqrt{3\log t})^2 + D^{\text{pos}}(\{\mu_a\}_{a=1}^K)\right\} \\
2c(t,\delta) =& \bar{W}_{-1}\left(2\log\frac{K}{\delta} + 4\log\log(e^4 t) + \frac{1}{2}\right) \\
\bar{W}_{-1}(x) =& -W_{-1}(-e^{-x}) \approx x + \log x.
\end{aligned}$$

W_{-1} is the negative branch of the Lambert W function, and $D^{\text{pos}}(\{\mu_a\}_{a=1}^K)$ is further defined in the Appendix F in Jourdan and Réda (2023), as

$$\begin{aligned}
D^{\text{pos}}(\{\mu_a\}_{a=1}^K) =& S_\mu + 1 - \frac{6\log S_\mu}{\Delta_{0,1}^2} \\
S_\mu =& h_1\left(12\sum_{a=1}^{K}\frac{1}{\Delta_{0,a}^2}, n_0 K + 2m\right) \\
h_1(z,y) =& z\bar{W}_{-1}(\frac{y}{z} + \log z) \\
\bar{W}_{-1}(x) =& -W_{-1}(-e^{-x})
\end{aligned}$$

From the definition, we know

$$\begin{aligned}
C^{\text{pos}}(\delta) &\geq \sup\left\{t : t \leq \left(\sum_{a:\mu_a \geq \mu_0} \frac{2}{\Delta_{0,a}^2}\right) c(t,\delta)\right\} \\
&\geq \sup\left\{t : t \leq \left(\sum_{a:\mu_a \geq \mu_0} \frac{2}{\Delta_{0,a}^2}\right) \left(\log\frac{K}{\delta} + 2\log\log(e^4 t) + \frac{1}{4}\right)\right\} \\
&\geq \sup\left\{t : t \leq \left(\sum_{a:\mu_a \geq \mu_0} \frac{2}{\Delta_{0,a}^2}\right) \log\frac{K}{\delta} + 2\log 4\right\} \\
&\geq \left(\sum_{a:\mu_a \geq \mu_0} \frac{2}{\Delta_{0,a}^2}\right) \log\frac{K}{\delta}.
\end{aligned}$$

Thus, we can conclude $\left(\sum_{a:\mu_a \geq \mu_0} \frac{2}{\Delta_{0,a}^2}\right) \log\frac{K}{\delta}$ is a lower bound of $C^{\text{pos}}(\delta) + \frac{K\pi^2}{6} + 1$.

A.1.4 Theorem 8 in Katz-Samuels and Jamieson (2020)

As discussed in the Appendix A.1 in Li and Cheung (2025), proposed Algorithm FWER-FWPD only work on the positive instance with a different definition of $\mathbb{E}\tau$. But we can adapt their result. Apply Algorithm FWER-FWPD on a positive instance ν, Theorem 8 in Katz-Samuels and Jamieson (2020) suggests that

$$\mathbb{E}\tau \leq O\left(\left((\frac{K}{m}-1)\cdot\frac{\log\frac{1}{\delta}}{\Delta^2} + \log\frac{\frac{K}{m}}{\delta}\right)\log\left((\frac{K}{m}-1)\cdot\frac{\log\frac{1}{\delta}}{\Delta^2} + \log\frac{\frac{K}{m}}{\delta}\right)\right),$$

where $\Delta = \min_{a:\mu_a > \mu_0} \Delta_{a,0}$.

Not hard to see

$$\begin{aligned}
&\left((\frac{K}{m}-1)\cdot\frac{\log\frac{1}{\delta}}{\Delta^2} + \log\frac{\frac{K}{m}}{\delta}\right)\log\left((\frac{K}{m}-1)\cdot\frac{\log\frac{1}{\delta}}{\Delta^2} + \log\frac{\frac{K}{m}}{\delta}\right) \\
\geq &\left((\frac{K}{m}-1)\cdot\frac{\log\frac{1}{\delta}}{\Delta^2}\right)\log\left((\frac{K}{m}-1)\cdot\frac{\log\frac{1}{\delta}}{\Delta^2}\right).
\end{aligned}$$

A.2 Literature Review on Other Topics

If the threshold μ_0 is viewed as an additional arm with deterministic reward, then 1-identification can be reduced to a standard Best-Arm Identification (BAI) problem Even-Dar et al. (2002); Gabillon et al. (2012); Jamieson et al. (2013); Kalyanakrishnan et al. (2012); Karnin et al. (2013); Jamieson and Nowak (2014); Chen and Li (2015). However, this reduction is known to be inefficient. Both asymptotic analyses Degenne and Koolen (2019); Garivier and Kaufmann (2016) and non-asymptotic results Li and Cheung (2025); Kano et al. (2017) show that such a formulation incurs unnecessary sampling. Intuitively, 1-identification does not require high-confidence pairwise comparisons among all arms, and therefore does not necessitate the intensive sampling in BAI Algorithms.

Beyond 1-identification, Degenne and Koolen (2019) research on several pure-exploration variants with multiple correct answers. These include identifying all arms in $\{a : \mu_a \geq \mu_1 - \epsilon\}$ Mason et al. (2020), or identifying a single arm from this set Even-Dar et al. (2002); Gabillon et al. (2012); Kalyanakrishnan et al. (2012); Kaufmann and Kalyanakrishnan (2013); Katz-Samuels and Jamieson (2020). As observed in Katz-Samuels and Jamieson (2020), when the gap $\Delta_{0,1}$ is known and $\mu_1 > \mu_0$, the answer sets of ϵ-good arm identification and 1-identification coincide by setting $\epsilon = \Delta_{0,1}$, allowing the same sampling and recommendation rules. When the gaps are unknown, however, the precise relationship between these problems remains unclear. In addition, thresholding bandits Locatelli et al. (2016); Mukherjee et al. (2017) aim to identify arms with mean value above μ_0 under a fixed budget. The correct output might not be unique under their proposed problem setting. Kano et al. (2017) also target at identifying all the arms with mean value above μ_0, but they focus on

the fixed confidence setting. Since Kano et al. (2017) provide an upper bound for each outputting time index, their result can be generalized to 1-identification.

Finally, our algorithmic design builds on the idea of running multiple algorithm copies in parallel, which has appeared in prior work. In particular, Chen and Li (2015); Chen et al. (2017); Katz-Samuels and Jamieson (2020) employ a master procedure to manage multiple executions of a base algorithm. The approach of Katz-Samuels and Jamieson (2020) allocates 2^ℓ rounds to the execution indexed by ℓ, while Chen and Li (2015); Chen et al. (2017) schedule parallel executions according to a global round index $t \in \mathbb{N}$, activating an execution when its index divides t. In contrast, our method adopts a simple round-robin scheduling scheme, which enables a balanced execution of parallel algorithm instances.

A.3 Notation Description

Table 2 presents part of the notation description used in the following appendix.

Table 2: Notation Summary

Symbol	Type	Meaning
K	Integer	Number of arms.
a, i, j	Integer in $[K]$	Arm Index
δ	Real	Tolerance level
$\hat{a}$	NA	Output of Algorithm 1, 3, takes value in $[K] \cup \{\mathsf{None}\}$
$\hat{a}_k$	NA	Rough guess of an exploration period in Algorithm 3, takes value in $[K] \cup \{\mathsf{None}, \mathsf{Not\ Complete}\}$
k	Integer	Phase index in Algorithm 3
$\nu, \nu', \nu'', \tilde{\nu}$	NA	An 1-identification instance
ρ_a	NA	Reward distribution of arm a
$\sigma(\nu)$	NA	Switch the arm positions in instance ν with a permutation σ
μ_a	Real	Mean value of arm a.
m	Integer	Maximum arm index whose mean value is above μ_0, we usually assume $\mu_1 \geq \cdots \geq \mu_m \geq \mu_0 > \mu_{m+1} \geq \cdots \geq \mu_K$
r_a^ν	Real	Mean value of arm a instance ν.
$\Delta_{i,j}$	Real	Mean gap between arm index i and j.
A_t	Integer in $[K]$	Action in round t
B	Subset of $[K]$	A bracket
$b, \tilde{b}$	Integer	Bracket index in $[\lceil \log_2 K \rceil + 1]$
$\tau, \tau(B)$	Random Variable	τ is the termination round of Algorithm 1, $\tau(B)$ is the termination round of an Algorithm 3
$\tau_k^{\text{ee}}(B), \tau_k^{\text{et}}(B)$	Random Variable	The total length of the exploration / exploitation periods at to the end of phase k, of Algorithm 3 with bracket B as input
$\tau, \tau_n, \tilde{\tau}, \tau', \tau_n^j(a)$	Random Variable	Stopping times for analysis
$\mathcal{E}$	Random Event	Random Event
Z_s, Z_t	Random Variable	Realized KL divergence, usually follows $N(\frac{\Delta^2}{2}, \Delta^2)$ as we only consider Gaussian distribution with variance 1 as reward distribution in the proof lower bound
D, D_i	Random Variable	$D := \max_{1 \leq t \leq n} \sum_{s=1}^t \left(Z_s - \frac{\Delta^2}{2} \right)$. See the proof of Lemma 8
$N_a(t)$	Random Variable	Pulling times of arm a up to round t
$T_j(a)$	Real Value	A pulling times threshold dependent on arm a and j
$U(t, \delta)$	Real Value	Half length of confidence interval with pulling times t
$\mathcal{H}^{ee}, \mathcal{H}^{et}, Q$	Random Set	Collected history in Algorithm 3
$N_a^{\text{ee}}(\mathcal{H}^{ee}), N_a^{\text{et}}(\mathcal{H}^{et})$	Random Variable	Pulling times of arm a in history $\mathcal{H}^{ee}, \mathcal{H}^{et}$
$N_a^{\text{ee}}(t)$	Random Variable	Pulling times of arm a at round t of an exploration period
$N_a^{\text{et}}(t)$	Random Variable	Pulling times of arm a at round t of an exploitation period

Symbol	Type	Meaning
$\hat{\mu}_a^{\text{ee}}(\mathcal{H}^{ee}), \hat{\mu}_a^{\text{et}}(\mathcal{H}^{et})$	Random Variable	Empirical mean value of arm a in history $\mathcal{H}^{ee}, \mathcal{H}^{et}$
$\text{UCB}_a^{\text{ee}}(\mathcal{H}^{ee}, \delta)$	Random Variable	Upper Confidence Bound calculated from history $\mathcal{H}^{ee}$
$\text{LCB}_a^{\text{ee}}(\mathcal{H}^{ee}, \delta)$	Random Variable	Lower Confidence Bound calculated from history $\mathcal{H}^{ee}$
$\text{UCB}_a^{\text{et}}(\mathcal{H}^{et}, \delta)$	Random Variable	Upper Confidence Bound calculated from history $\mathcal{H}^{et}$
$\text{LCB}_a^{\text{et}}(\mathcal{H}^{et}, \delta)$,	Random Variable	Lower Confidence Bound calculated from history $\mathcal{H}^{et}$
C	Real Value	Fixed Constant in Algorithm 3, we can take C to any real number greater than 1
$t^{\text{ee}}, t^{\text{et}}$	Integer	Counter of the total pulling times in exploration and exploitation periods, used in Algorithm 3
$\delta_k, \beta_k, \alpha_k$	Real Value	Hyperparameters of Algorithm 3 to control the pulling process
$X_{a,s}^{\text{ee}}$	Random Variable	Realized reward by pulling arm a at the s-th times in the exploration period
$X_{a,s}^{\text{et}}$	Random Variable	Realized reward by pulling arm a at the s-th times in the exploitation period
$\kappa_b^{\text{ee}}, \kappa_B^{\text{ee}}$	Random Variable	Minimum phase index such that the concentration event holds in the exploration period of algorithm copy b. If we only discuss a given bracket B, we do not emphasize its bracket index and may use κ_B^{ee}.
$\kappa_b^{\text{et}}, \kappa_B^{\text{et}}$	Random Variable	Minimum phase index such that the concentration event holds in the exploitation period of algorithm copy b. If we only discuss a given bracket B, we do not emphasize its bracket index and may use κ_B^{et}.
$\hat{\mu}_{a,b}^{ee}(t)$	Random Variable	The empirical mean value of arm a of the first t collected reward in the exploration period of algorithm copy b
$\hat{\mu}_{a,b}^{et}(t)$	Random Variable	The empirical mean value of arm a of the first t collected reward in the exploitation period of algorithm copy b
$\hat{\mu}_{a,t}$	Random Variable	The empirical mean value of arm a of the first t collected reward. We use this notation if a specific bracket and a specific exploration/exploitation period is given, we may turn to use $\hat{\mu}_{a,t}$ instead of $\hat{\mu}_{a,b}^{ee}(t), \hat{\mu}_{a,b}^{et}(t)$.
$\mathcal{A}_B(\mathcal{H}^{\text{ee}}, \delta_k, k)$	Random Set	The available action set when implementing Algorithm 3.
A_t^{ee}	Random Variable	Arbitrary element in $\mathcal{A}_B(\mathcal{H}^{\text{ee}}, \delta_k, k)$, which is action that Algorithm 3 takes during the exploration period.

B Proof of Lower Bound

B.1 Proof of Preparation

The first step is to prove Lemma 4, showing that we suffice restrict attention to symmetric algorithms.

Proof of Lemma 4. Let S be the set of all possible permutations of $[K]$.

Given the alg, we define $\widetilde{\text{alg}}$ as the following. Shuffle all the arms randomly before the pulling process, and then apply the algorithm alg. We firstly show $\widetilde{\text{alg}}$ is indeed symmetric. For any pulling sequence $\{a_t\}_{t=1}^T \in [K]^T$ and any permutation σ, we have

$$\Pr_{\widetilde{\text{alg}},\sigma(\nu')}(\forall t \in [T], A_t = \sigma(a), \tau > t) = \frac{\sum_{\tilde{\sigma}\in S} \Pr_{\text{alg},\tilde{\sigma}(\nu')}(\forall t \in [T], A_t = \tilde{\sigma}(a), \tau > t)}{K!},$$

which means $\widetilde{\text{alg}}$ fits in the definition. Then, we have

$$\begin{aligned}
&\max_{\sigma\in S} \mathbb{E}_{\text{alg},\sigma(\nu')}\tau \\
\geq &\frac{\sum_{\sigma\in S} \mathbb{E}_{\text{alg},\sigma(\nu')}\tau}{K!} \\
\geq &\frac{\sum_{\sigma\in S}\sum_{t=1}^{+\infty} \Pr_{\text{alg},\sigma(\nu')}(\tau \geq t)}{K!} \\
= &\mathbb{E}_{\widetilde{\text{alg}},\nu'}\tau
\end{aligned}$$

The last thing is to show $\widetilde{\text{alg}}$ is also δ-PAC. Since alg is δ-PAC, we know for any permutation σ, we have

$$\Pr_{\text{alg},\sigma(\nu)}(\tau < +\infty, \hat{a} \in i^*(\sigma(\nu))) > 1-\delta.$$

That means

$$\begin{aligned}
&\Pr_{\widetilde{\text{alg}},\sigma(\nu)}(\tau < +\infty, \hat{a} \in i^*(\sigma(\nu))) \\
= &\frac{\sum_{\sigma\in S} \Pr_{\text{alg},\sigma(\nu)}(\tau < +\infty, \hat{a} \in i^*(\sigma(\nu)))}{K!} \\
> &\frac{\sum_{\sigma\in S} 1-\delta}{K!} \\
= &1-\delta.
\end{aligned}$$

We can conclude $\widetilde{\text{alg}}$ is also δ-PAC. □

The next step establishes a technical change-of-measure bound that controls a conditional KL divergence. This lemma is the main device used to argue that a symmetric algorithm cannot reliably distinguish two arms when the total number of samples from the pair is small.

Lemma 8. *Fix a pair $i, j \in [K], i \neq j$ and an arbitrary instance ν', denote $\Delta = |r_i^{\nu'} - r_j^{\nu'}|$. Define*

$$\tau_n = \min\{t : N_i(t) + N_j(t) = n, A_{t+1} \in \{i, j\}\}.$$

Apply a symmetric algorithm, we have

$$\begin{aligned}
&\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i)}{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j)} + \\
&\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j)}{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i)} \\
\leq &\sqrt{2n\Delta^2 \log \frac{1}{\Pr_{\nu'}(\tau_n < \tau)}} \Pr_{\nu'}(\tau_n < \tau) + \sqrt{2\pi n\Delta^2} \Pr_{\nu'}(\tau_n < \tau) + \frac{1}{2} n\Delta^2 \Pr_{\nu'}(\tau_n < \tau).
\end{aligned}$$

An observation is given a fixed sample path and a symmetric algorithm, the set $\{t : N_i(t) + N_j(t) = n, A_{t+1} \in \{i, j\}\} =: S_{ij}(n)$ contains at most one element. If $|S_{ij}(n)| = 1$, τ_n will be the unique element, and $\tau_n < \tau$ holds; if $|S_{ij}(n)| = 0$, then $\tau_n = \infty$ and hence $\tau_n \geq \tau$. Moreover,

$$\frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i) \log \frac{\Pr_{\nu'}(\tau_n<\tau, A_{\tau_n+1}=i)}{\Pr_{\nu'}(\tau_n<\tau, A_{\tau_n+1}=j)} + \Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j) \log \frac{\Pr_{\nu'}(\tau_n<\tau, A_{\tau_n+1}=j)}{\Pr_{\nu'}(\tau_n<\tau, A_{\tau_n+1}=i)}}{\Pr_{\nu'}(\tau_n < \tau)}$$

is the kl-divergence between two Bernoulli distribution with probability $\Pr_{\nu'}(A_{\tau_n+1} = i|\tau_n < \tau)$, $\Pr_{\nu'}(A_{\tau_n+1} = j|\tau_n < \tau)$. From the definition of τ_n, not hard to see $\Pr_{\nu'}(A_{\tau_n+1} = i|\tau_n < \tau) + \Pr_{\nu'}(A_{\tau_n+1} = j|\tau_n < \tau) = 1$.

Proof of Lemma 8. Denote ν'' as the instance which switch the position of arm i and arm j in ν', i.e. $r_a^{\nu''} = \begin{cases} r_j^{\nu'} & a = i \\ r_i^{\nu'} & a = j \\ r_a^{\nu'} & \text{else} \end{cases}$. We first derive

$$\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i)}{\Pr_{\nu''}(\tau_n < \tau, A_{\tau_n+1} = i)} \leq \mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau, A_{\tau_n+1} = i) \left(\sum_{s=1}^{N_i(\tau_n)+N_j(\tau_n)} Z_s \right) \tag{13}$$

$$\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j)}{\Pr_{\nu''}(\tau_n < \tau, A_{\tau_n+1} = j)} \leq \mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau, A_{\tau_n+1} = j) \left(\sum_{s=1}^{N_i(\tau_n)+N_j(\tau_n)} Z_s \right) \tag{14}$$

where $Z_s|Z_1, Z_2, \cdots, Z_{s-1} \sim N(\frac{\Delta^2}{2}, \Delta^2)$. Since these two inequalities are symmetric, we suffice to prove the first (13). Take $\mathcal{E} = \{\tau_n < \tau, A_{\tau_n+1} = i\}$. Notice that $\mathcal{E}$ is τ_n-measurable, as both $\{\tau_n < \tau\}, \{A_{\tau_n+1} = i\}$ are τ_n-measurable. From the Lemma 18 in Kaufmann et al. (2016),

$$\begin{aligned}
\Pr_{\nu''}(\mathcal{E}) =& \mathbb{E}_{\nu'} \mathbb{1}(\mathcal{E}) \exp\left(- \sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} - \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s} \right) \\
=& \mathbb{E}_{\nu'} \left[\mathbb{1}(\mathcal{E}) \exp\left(- \sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} - \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s} \right) |\mathcal{E} \right] \Pr_{\nu'}(\mathcal{E}) + \\
& \mathbb{E}_{\nu'} \left[\mathbb{1}(\mathcal{E}) \exp\left(- \sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} - \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s} \right) |\neg\mathcal{E} \right] \Pr_{\nu'}(\neg\mathcal{E}) \\
=& \mathbb{E}_{\nu'} \left[\exp\left(- \sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} - \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s} \right) |\mathcal{E} \right] \Pr_{\nu'}(\mathcal{E}) \\
\geq& \exp\left(\mathbb{E}_{\nu'} \left[- \sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} - \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s} |\mathcal{E} \right] \right) \Pr_{\nu'}(\mathcal{E}),
\end{aligned}$$

where $\hat{KL}_{i,s} := \log \frac{\exp(-\frac{(X_{i,s}-r_i^{\nu'})^2}{2})}{\exp(-\frac{(X_{i,s}-r_i^{\nu''})^2}{2})} = \frac{(2X_{i,s}-r_i^{\nu'}-r_j^{\nu'})(r_i^{\nu'}-r_j^{\nu'})}{2}$, $X_{i,s} \sim N(r_i^{\nu'}, 1)$, $\hat{KL}_{j,s} := \log \frac{\exp(-\frac{(X_{j,s}-r_j^{\nu'})^2}{2})}{\exp(-\frac{(X_{j,s}-r_j^{\nu''})^2}{2})} = \frac{(2X_{j,s}-r_j^{\nu'}-r_i^{\nu'})(r_j^{\nu'}-r_i^{\nu'})}{2}$, $X_{j,s} \sim N(r_j^{\nu'}, 1)$. The last line is by the Jensen's Inequality. Following the last line, we have

$$\begin{aligned}
\log \frac{\Pr_{\nu'}(\mathcal{E})}{\Pr_{\nu''}(\mathcal{E})} \leq& \mathbb{E}_{\nu'} \left[\sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} + \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s} |\mathcal{E} \right] \\
=& \frac{\mathbb{E}_{\nu'} \left[\mathbb{1}(\mathcal{E}) \left(\sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} + \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s} \right) \right]}{\Pr_{\nu'}(\mathcal{E})},
\end{aligned}$$

and we can conclude $\Pr_{\nu'}(\mathcal{E}) \log \frac{\Pr_{\nu'}(\mathcal{E})}{\Pr_{\nu''}(\mathcal{E})} \leq \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})\left(\sum_{s=1}^{N_i(\tau_n)} \hat{KL}_{i,s} + \sum_{s=1}^{N_j(\tau_n)} \hat{KL}_{j,s}\right)\right]$. From the above calculation, we know $\hat{KL}_{i,s}, \hat{KL}_{j,s}$ both follow $N(\frac{(r_j^{\nu'} - r_i^{\nu'})^2}{2}, (r_j^{\nu'} - r_i^{\nu'}))$, and we can rewrite the conclusion as

$$\Pr_{\nu'}(\mathcal{E}) \log \frac{\Pr_{\nu'}(\mathcal{E})}{\Pr_{\nu''}(\mathcal{E})} \leq \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E}) \sum_{s=1}^{N_i(\tau_n)+N_j(\tau_n)} Z_s\right]$$

where $Z_s | Z_1, Z_2, \cdots, Z_{s-1} \sim N(\frac{\Delta^2}{2}, \Delta^2)$. The above inequality suggests that we have proved (13) and (14).

By symmetry of the algorithm, we can conclude $\Pr_{\nu''}(\tau_n < \tau, A_{\tau_n+1} = i) = \Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j)$ and $\Pr_{\nu''}(\tau_n < \tau, A_{\tau_n+1} = j) = \Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i)$. Sum up inequality (13) and (14), we get

$$\begin{aligned}
&\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i)}{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j)} + \\
&\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j)}{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i)} \\
=&\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = i)}{\Pr_{\nu''}(\tau_n < \tau, A_{\tau_n+1} = i)} + \\
&\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j) \log \frac{\Pr_{\nu'}(\tau_n < \tau, A_{\tau_n+1} = j)}{\Pr_{\nu''}(\tau_n < \tau, A_{\tau_n+1} = j)} \\
\leq&\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) \left(\sum_{s=1}^{N_i(\tau_n)+N_j(\tau_n)} Z_s\right).
\end{aligned}$$

The remaining work is to prove

$$\begin{aligned}
&\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) \left(\sum_{s=1}^{N_i(\tau_n)+N_j(\tau_n)} Z_s\right) \\
\leq&\sqrt{2n\Delta^2 \log \frac{1}{\Pr_{\nu'}(\tau_n < \tau)}} \Pr_{\nu'}(\tau_n < \tau) + \sqrt{2\pi n\Delta^2} \Pr_{\nu'}(\tau_n < \tau) + \frac{1}{2} n\Delta^2 \Pr_{\nu'}(\tau_n < \tau).
\end{aligned}$$

Let $D := \max_{1 \leq t \leq n} \sum_{s=1}^{t} \left(Z_s - \frac{\Delta^2}{2}\right)$. Since $\{\tau_n < \tau\}$ implies $N_i(\tau_n) + N_j(\tau_n) \leq n$, we have

$$\begin{aligned}
&\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) \left(\sum_{s=1}^{N_i(\tau_n)+N_j(\tau_n)} Z_s\right) \\
\leq&\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) \left(\max_{1 \leq t \leq n} \sum_{s=1}^{n} Z_s\right) \\
=&\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) \left(\max_{1 \leq t \leq n} \sum_{s=1}^{n} Z_s - \frac{\Delta^2}{2} + \frac{\Delta^2}{2}\right) \\
\leq&\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) \left(\max_{1 \leq t \leq n} \sum_{s=1}^{n} Z_s - \frac{\Delta^2}{2}\right) + \mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) \frac{n\Delta^2}{2} \\
=&\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) D + \frac{n\Delta^2}{2} \Pr_{\nu'}(\tau_n < \tau).
\end{aligned}$$

The remaining work is to prove $\mathbb{E}_{\nu'} \mathbb{1}(\tau_n < \tau) D \leq \sqrt{2n\Delta^2 \log \frac{1}{\Pr_{\nu'}(\tau_n < \tau)}} \Pr_{\nu'}(\tau_n < \tau) + \sqrt{2\pi n\Delta^2} \Pr_{\nu'}(\tau_n < \tau)$. To simplify the notation, we take event $\mathcal{E}_n = \{\tau_n < \tau\}$ here and use

the layer-cake representation,

$$
\begin{aligned}
&\mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E}_n)D\right] \\
\leq&\mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E}_n)\mathbb{1}(D\geq 0)D\right] \\
=&\int_0^{+\infty}\Pr_{\nu'}\Big(\mathbb{1}(\mathcal{E}_n)\mathbb{1}(D\geq 0)D>s\Big)ds \\
\leq&\int_0^{+\infty}\Pr_{\nu'}\Big(\mathcal{E}_n, D>s\Big)ds \\
\leq&\int_0^{+\infty}\min\left\{\Pr_{\nu'}(\mathcal{E}_n), \Pr_{\nu'}\Big(D>s\Big)\right\}ds
\end{aligned}
$$

From Lemma 32, we can conclude $\Pr_{\nu'}\left(D\geq s\right)\leq\exp(-\frac{s^2}{2n\Delta^2})$. Plugging this tail bound yields

$$
\begin{aligned}
\mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E}_n)D\right]\leq&\int_0^{+\infty}\min\left\{\Pr_{\nu'}(\mathcal{E}_n), \exp(-\frac{s^2}{2n\Delta^2})\right\}ds \\
=&\int_0^{\sqrt{2n\Delta^2\log\frac{1}{\Pr_{\nu'}(\mathcal{E}_n)}}}\Pr_{\nu'}(\mathcal{E}_n)ds+\int_{\sqrt{2n\Delta^2\log\frac{1}{\Pr_{\nu'}(\mathcal{E}_n)}}}^{+\infty}\exp(-\frac{s^2}{2n\Delta^2})ds \\
=&\sqrt{2n\Delta^2\log\frac{1}{\Pr_{\nu'}(\mathcal{E}_n)}}\Pr_{\nu'}(\mathcal{E}_n)+\frac{\sqrt{2\pi}\sqrt{n\Delta^2}}{\sqrt{2\pi}}\int_{\sqrt{2\log\frac{1}{\Pr_{\nu'}(\mathcal{E}_n)}}}^{+\infty}\exp(-\frac{s^2}{2})ds \\
\leq&\sqrt{2n\Delta^2\log\frac{1}{\Pr_{\nu'}(\mathcal{E}_n)}}\Pr_{\nu'}(\mathcal{E}_n)+\sqrt{2\pi n\Delta^2}\exp(-\frac{2\log\frac{1}{\Pr_{\nu'}(\mathcal{E}_n)}}{2}) \\
=&\sqrt{2n\Delta^2\log\frac{1}{\Pr_{\nu'}(\mathcal{E}_n)}}\Pr_{\nu'}(\mathcal{E}_n)+\sqrt{2\pi n\Delta^2}\Pr_{\nu'}(\mathcal{E}_n).
\end{aligned}
$$

The second last step is by the equality $\int_t^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}ds\leq e^{-\frac{t^2}{2}}$ for $t>0$. The reason is

$$
\int_t^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}ds=\int_0^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(s+t)^2}{2}}ds\leq e^{-\frac{t^2}{2}}\int_0^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}ds\leq e^{-\frac{t^2}{2}}.
$$

We complete the proof. □

Besides Lemma 8, we further introduce a Lemma which is similar, but focuses on the single side.

Lemma 9. *Consider two instances ν', ν'', which only differ in arm $a_1, a_2, \cdots, a_M$, i.e. $r_a^{\nu'}\neq r_a^{\nu''}$ holds for all $a\in\{a_1,a_2,\cdots,a_M\}$ and $r_a^{\nu'}=r_a^{\nu''}$ holds for all $a\notin\{a_1,a_2,\cdots,a_M\}$. Denote $\Delta_i=|r_{a_i}^{\nu'}-r_{a_i}^{\nu''}|$, and $\tilde{\tau}$ is an arbitrary stopping time. If event $\mathcal{E}\in\mathcal{F}_{\tilde{\tau}}$ fulfills $\mathcal{E}\subset\{N_{a_i}(\tilde{\tau})\leq n_i,\forall i\in[M]\}$ with some $\{n_i\}_{i=1}^M\in\mathbb{N}^M$, we can conclude*

$$
\log\frac{\Pr_{\nu'}(\mathcal{E})}{\Pr_{\nu''}(\mathcal{E})}\leq\sum_{i=1}^M\sqrt{2n_i\Delta_i^2\log\frac{1}{\Pr_{\nu'}(\mathcal{E})}}+\sqrt{2\pi n_i\Delta_i^2}+\frac{1}{2}n_i\Delta_i^2
$$

Proof of Lemma 9. From the Lemma 18 in Kaufmann et al. (2016), we have

$$\begin{aligned}
\Pr_{\nu''}(\mathcal{E}) =& \mathbb{E}_{\nu'} \mathbb{1}(\mathcal{E}) \exp\left(-\sum_{i=1}^{M}\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right) \\
=& \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E}) \exp\left(-\sum_{i=1}^{M}\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right)|\mathcal{E}\right] \Pr_{\nu'}(\mathcal{E}) + \\
& \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E}) \exp\left(-\sum_{i=1}^{M}\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right)|\neg\mathcal{E}\right] \Pr_{\nu'}(\neg\mathcal{E}) \\
=& \mathbb{E}_{\nu'}\left[\exp\left(-\sum_{i=1}^{M}\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right)|\mathcal{E}\right] \Pr_{\nu'}(\mathcal{E}) \\
\geq& \exp\left(\mathbb{E}_{\nu'}\left[-\sum_{i=1}^{M}\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}|\mathcal{E}\right]\right) \Pr_{\nu'}(\mathcal{E}),
\end{aligned}$$

where $\hat{KL}_{i,s} := \log \frac{\exp(-\frac{(X_{i,s}-r_{a_i}^{\nu'})^2}{2})}{\exp(-\frac{(X_{i,s}-r_{a_i}^{\nu''})^2}{2})} = \frac{(2X_{i,s}-r_{a_i}^{\nu'}-r_{a_i}^{\nu''})(r_{a_i}^{\nu'}-r_{a_i}^{\nu''})}{2}$, $X_{i,s} \sim N(r_{a_i}^{\nu'}, 1)$. Following the last line, we have

$$\begin{aligned}
\log \frac{\Pr_{\nu'}(\mathcal{E})}{\Pr_{\nu''}(\mathcal{E})} \leq& \mathbb{E}_{\nu'}\left[\sum_{i=1}^{M}\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}|\mathcal{E}\right] \\
=& \frac{\mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})\sum_{i=1}^{M}\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right]}{\Pr_{\nu'}(\mathcal{E})}.
\end{aligned}$$

and we can conclude $\Pr_{\nu'}(\mathcal{E}) \log \frac{\Pr_{\nu'}(\mathcal{E})}{\Pr_{\nu''}(\mathcal{E})} \leq \sum_{i=1}^{M} \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right]$. We turn to find an upper bound for each $\mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right]$. Since $\mathcal{E} \Rightarrow N_{a_i}(\tilde{\tau}) < n_i$, we have

$$\begin{aligned}
& \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})\sum_{s=1}^{N_i(\tilde{\tau})} \hat{KL}_{i,s}\right] \\
\leq& \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})\left(\max_{1\leq t\leq n_i}\sum_{s=1}^{t} \hat{KL}_{i,s} - \frac{\Delta_i^2}{2}\right)\right] + \frac{n_i\Delta_i^2}{2}\Pr_{\nu'}(\mathcal{E})
\end{aligned}$$

Denote $D_i := \max_{1\leq t\leq n_i}\sum_{s=1}^{t}\left(Z_s - \frac{\Delta_i^2}{2}\right)$, $Z_s|Z_1, Z_2, \cdots, Z_{s-1} \sim N(\frac{\Delta_i^2}{2}, \Delta_i^2)$, we have

$$\begin{aligned}
& \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})D_i\right] \\
\leq& \mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})\mathbb{1}(D\geq 0)D_i\right] \\
=& \int_0^{+\infty} \Pr_{\nu'}\Big(\mathbb{1}(\mathcal{E})\mathbb{1}(D_i\geq 0)D_i > s\Big)ds \\
\leq& \int_0^{+\infty} \Pr_{\nu'}\Big(\mathcal{E}, D_i > s\Big)ds \\
\leq& \int_0^{+\infty} \min\left\{\Pr_{\nu'}(\mathcal{E}), \Pr_{\nu'}\Big(D_i > s\Big)\right\}ds
\end{aligned}$$

From Lemma 32, we can conclude $\Pr_{\nu'}\left(D_i \geq s\right) \leq \exp(-\frac{s^2}{2n_i\Delta_i^2})$. Plug in the inequality, we get

$$\begin{aligned}
\mathbb{E}_{\nu'}\left[\mathbb{1}(\mathcal{E})D_i\right] \leq & \int_0^{+\infty} \min\left\{\Pr_{\nu'}(\mathcal{E}), \exp(-\frac{s^2}{2n_i\Delta_i^2})\right\} ds \\
= & \int_0^{\sqrt{2n_i\Delta_i^2 \log\frac{1}{\Pr_{\nu'}(\mathcal{E})}}} \Pr_{\nu'}(\mathcal{E}) ds + \int_{\sqrt{2n_i\Delta_i^2 \log\frac{1}{\Pr_{\nu'}(\mathcal{E})}}}^{+\infty} \exp(-\frac{s^2}{2n_i\Delta^2}) ds \\
= & \sqrt{2n_i\Delta_i^2 \log\frac{1}{\Pr_{\nu'}(\mathcal{E})}} \Pr_{\nu'}(\mathcal{E}) + \frac{\sqrt{2\pi}\sqrt{n_i\Delta_i^2}}{\sqrt{2\pi}} \int_{\sqrt{2\log\frac{1}{\Pr_{\nu'}(\mathcal{E})}}}^{+\infty} \exp(-\frac{s^2}{2}) ds \\
\leq & \sqrt{2n_i\Delta_i^2 \log\frac{1}{\Pr_{\nu'}(\mathcal{E})}} \Pr_{\nu'}(\mathcal{E}) + \sqrt{2\pi n_i\Delta_i^2} \exp(-\frac{2\log\frac{1}{\Pr_{\nu'}(\mathcal{E})}}{2}) \\
= & \sqrt{2n_i\Delta_i^2 \log\frac{1}{\Pr_{\nu'}(\mathcal{E})}} \Pr_{\nu'}(\mathcal{E}) + \sqrt{2\pi n_i\Delta_i^2} \Pr_{\nu'}(\mathcal{E}).
\end{aligned}$$

In summary, we have proved

$$\begin{aligned}
& \Pr_{\nu'}(\mathcal{E}) \log\frac{\Pr_{\nu'}(\mathcal{E})}{\Pr_{\nu''}(\mathcal{E})} \\
\leq & \sum_{i=1}^{M} \sqrt{2n_i\Delta^2 \log\frac{1}{\Pr_{\nu'}(\mathcal{E})}} \Pr_{\nu'}(\mathcal{E}) + \sqrt{2\pi n_i\Delta^2} \Pr_{\nu'}(\mathcal{E}) + \frac{n_i\Delta_i^2}{2} \Pr_{\nu'}(\mathcal{E}),
\end{aligned}$$

which is equivalent to the conclusion we need to prove. □

B.2 Proof of Theorem 5

Following the sketch proof in section 2, we prove the following lemmas one by one. The first step is to build a "probabilistic constant lower bound". Without further description, we only consider symmetric algorithms in this section.

Lemma 10. *Apply a δ-PAC and symmetric algorithm to a positive instance ν, If $\max_{i,j\in[K]} \Delta_{i,j} < 1$ and $\delta < 10^{-8}$, $\frac{1}{8m} \geq \sqrt{\delta}$, all holds, we have*

$$\sum_{j=1}^{m} \Pr_{\nu}\left(\hat{a} = j, N_j(\tau) > \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\right) \geq \frac{1}{2}$$

Proof of Lemma 10. Recall the notation $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$. The δ-PAC assumption implies

$$\Pr_{\nu}(\tau < +\infty, \hat{a} \in [m] \cup \{a : \mu_a = \mu_0\}) > 1 - \delta$$

Before the proof of the Lemma 10, we first show a claim

$$\Pr_{\nu}(\tau < +\infty, \hat{a} \in \{a : \mu_a = \mu_0\}) < 4\delta \tag{15}$$

This claim is to show $\Pr_{\nu}(\tau < +\infty, \hat{a} \in [m]) \geq 1 - \Theta(\delta)$. If $\{a : \mu_a = \mu_0\} = \emptyset$, (15) is immediate. We therefore assume $\{a : \mu_a = \mu_0\} \neq \emptyset$ and argue by contradiction. If $\Pr_{\nu}(\tau < +\infty, \hat{a} \in \{a : \mu_a = \mu_0\}) \geq 4\delta$ holds, then there exists a (possibly δ-dependent) constant $T_0(\delta)$, such that

$$\Pr_{\nu}(\tau < T_0, \hat{a} \in \{a : \mu_a = \mu_0\}) \geq 2\delta. \tag{16}$$

The reason is $\{\tau < +\infty, \hat{a} \in \{a : \mu_a = \mu_0\}\} = \cup_{T=1}^{+\infty}\{\tau < +T, \hat{a} \in \{a : \mu_a = \mu_0\}\}$. Since $\{\tau < +T, \hat{a} \in \{a : \mu_a = \mu_0\}\}$ is non-decreasing regarding T, we have

$$\begin{aligned}
& \lim_{T\to+\infty} \Pr_{\nu}(\tau < +T, \hat{a} \in \{a : \mu_a = \mu_0\}) \\
= & \Pr_{\nu}(\tau < +\infty, \hat{a} \in \{a : \mu_a = \mu_0\}) \\
\geq & 4\delta.
\end{aligned}$$

The last line implies there exists a (possibly δ-dependent) constant $T_0(\delta)$, such that $\Pr_\nu(\tau < T_0, \hat{a} \in \{a : \mu_a = \mu_0\}) \geq 2\delta$.

We define $\nu'(\epsilon)$ with $r_a^{\nu'(\epsilon)} = \begin{cases} \mu_0 - \epsilon & \mu_a = \mu_0 \\ r_a^\nu & \text{else} \end{cases}$, where ϵ is an arbitrary positive number. We know $i^*(\nu'(\epsilon)) = [m]$. By δ-PAC assumption,

$$\begin{aligned} &\Pr_{\nu'(\epsilon)}(\tau < +\infty, \hat{a} \in [m]) > 1 - \delta \\ \Rightarrow &\Pr_{\nu'(\epsilon)}(\tau = +\infty \text{ or } \hat{a} \notin [m]) < \delta \\ \Rightarrow &\Pr_{\nu'(\epsilon)}(\hat{a} \in \{a : \mu_a = \mu_0\}) < \delta \\ \Rightarrow &\Pr_{\nu'(\epsilon)}(\tau < T_0, \hat{a} \in \{a : \mu_a = \mu_0\}) < \delta. \end{aligned}$$

Apply Lemma 9 by taking event $\mathcal{E} = \{\tau < T_0, \hat{a} \in \{a : \mu_a = \mu_0\}\}$, we have

$$\begin{aligned} \log \frac{\Pr_\nu(\mathcal{E})}{\Pr_{\nu'(\epsilon)}(\mathcal{E})} &\leq \sum_{a:\mu_a=\mu_0} \sqrt{2T_0\epsilon^2 \log \frac{1}{\Pr_{\nu'}(\mathcal{E})}} + \sqrt{2\pi T_0 \epsilon^2} + \frac{1}{2} T_0 \epsilon^2 \\ \log \frac{\Pr_\nu(\mathcal{E})}{\Pr_{\nu'(\epsilon)}(\mathcal{E})} &\geq \log \frac{2\delta}{\delta} = \log 2 \end{aligned}$$

The second line is by (16) and $\Pr_{\nu'(\epsilon)}(\tau < T_0, \hat{a} \in \{a : \mu_a = \mu_0\}) < \delta$. Notice that we take $\epsilon \to 0$ in the right hand side of the first line, we can conclude $\log 2 \leq 0$, which is a contradiction.

With the claim (15) established, we deduce

$$\begin{aligned} &\Pr_\nu(\tau < +\infty, \hat{a} \in [m] \cup \{a : \mu_a = \mu_0\}) > 1 - \delta \\ \Leftrightarrow &\Pr_\nu(\tau < +\infty, \hat{a} \in [m]) + \Pr_\nu(\tau < +\infty, \hat{a} \in \{a : \mu_a = \mu_0\}) > 1 - \delta \\ \Rightarrow &\Pr_\nu(\tau < +\infty, \hat{a} \in [m]) > 1 - \delta - 4\delta \\ \Rightarrow &\Pr_\nu(\tau = +\infty \text{ or } \hat{a} \notin [m]) < 5\delta \\ \Rightarrow &\Pr_\nu(\hat{a} \notin [m]) < 5\delta \end{aligned}$$

Now, we are ready to prove the Lemma. We prove the conclusion by contradiction. If the Lemma doesn't hold, we have

$$
\begin{aligned}
&\Pr_{\nu}\left(\exists j\in[m],\hat{a}=j,N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\right)<\frac{1}{2}\\
\Leftrightarrow&\Pr_{\nu}\left(\cap_{j=1}^m\Big(\{\hat{a}\neq j\}\cup\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\right)\geq\frac{1}{2}\\
\Rightarrow&\Pr_{\nu}\left(\cap_{j=1}^m\{\hat{a}\neq j\}\right)+\Pr_{\nu}\left(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\right)\geq\frac{1}{2}\\
\Rightarrow&5\delta+\Pr_{\nu}\left(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\right)\geq\frac{1}{2}\\
\Rightarrow&\Pr_{\nu}\left(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\right)\geq\frac{1}{4}\\
\Rightarrow&\Pr_{\nu}\left(\Big(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\cap\{\hat{a}\in[m]\}\right)+\Pr_{\nu}\left(\{\hat{a}\notin[m]\}\right)\geq\frac{1}{4}\\
\Rightarrow&\Pr_{\nu}\left(\Big(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\cap\{\hat{a}\in[m]\}\right)+5\delta\geq\frac{1}{4}\\
\Rightarrow&\Pr_{\nu}\left(\Big(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\cap\{\hat{a}\in[m]\}\right)\geq\frac{1}{8}\\
\Rightarrow&\sum_{j'=1}^m\Pr_{\nu}\left(\Big(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\cap\{\hat{a}=j'\}\right)\geq\frac{1}{8}
\end{aligned}
$$

From the last step, we know there exists j_0, such that

$$
\Pr_{\nu}\left(\Big(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\cap\{\hat{a}=j_0\}\right)\geq\frac{1}{8m}\geq\sqrt{\delta}.
$$

We can consider instance $\nu_{j_0}^{(0)}$ whose reward of arm j_0 is $\mu_0-\epsilon$ for some $\epsilon>0$, others remain the same as ν. By the Lemma 9, we have

$$
\begin{aligned}
&\log\frac{\sqrt{\delta}}{\delta}\\
\leq&\log\frac{\Pr_{\nu}\left(\Big(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\cap\{\hat{a}=j_0\}\right)}{\Pr_{\nu_{j_0}^{(0)}}\left(\Big(\cap_{j=1}^m\{N_j(\tau)\leq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\}\Big)\cap\{\hat{a}=j_0\}\right)}\\
\leq&\sqrt{2\frac{\log\frac{1}{\delta}}{100\Delta_{j_0,0}^2}(\Delta_{j_0,0}+\epsilon)^2\log\frac{1}{\Pr_{\nu'}(\mathcal{E})}}+\sqrt{2\pi\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}(\Delta_{j_0,0}+\epsilon)^2}+\frac{1}{2}\frac{\log\frac{1}{\delta}}{100\Delta_{j_0,0}^2}(\Delta_{j_0,0}+\epsilon)^2\\
\leq&\sqrt{2\frac{\log\frac{1}{\delta}}{100\Delta_{j_0,0}^2}(\Delta_{j_0,0}+\epsilon)^2\log\frac{1}{\sqrt{\delta}}}+\sqrt{2\pi\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}(\Delta_{j_0,0}+\epsilon)^2}+\frac{1}{2}\frac{\log\frac{1}{\delta}}{100\Delta_{j_0,0}^2}(\Delta_{j_0,0}+\epsilon)^2.
\end{aligned}
$$

Taking $\epsilon\to0$, the last line suggests that

$$
\frac{1}{2}\log\frac{1}{\delta}\leq\sqrt{\frac{\log\frac{1}{\delta}}{100}\log\frac{1}{\delta}}+\sqrt{2\pi\frac{\log\frac{1}{\delta}}{100}}+\frac{1}{2}\frac{\log\frac{1}{\delta}}{100},
$$

which leads to a contradiction, as $\delta<10^{-8}$. □

We next quantify the algorithm's inability to distinguish two arms when the total arm pulls of the pair is small.

Lemma 11. *Assume $\delta < \frac{1}{e}$. For any $j \in [m]$, $a \in [K], a \neq j$, define*

$$\begin{aligned}\tau_n^j(a) =& \min\{t : N_j(t) + N_a(t) = n, A_{t+1} \in \{j, a\}\}\\ T_j(a) =& \lceil \frac{1}{200\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}\left(1 + \log \frac{1}{\Pr_\nu(N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a}=j)}\right)} \rceil.\end{aligned}$$

Then, for $n = 0, 1, 2, \cdots, T_j(a) - 1$, we have

$$\begin{aligned}&\Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = j) \log \frac{\Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = j)}{\Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = a)} +\\ &\Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = a) \log \frac{\Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = a)}{\Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = j)}\\ <& 0.4 \Pr_\nu(\tau_n^j(a) < \tau)\end{aligned}$$

Dividing $\Pr_\nu(\tau_n^j(a) < \tau)$ on both sides, the conclusion is equivalent to

$$\mathrm{kl}\Big(\Pr_\nu(A_{\tau_n^j(a)+1} = a | \tau_n^j(a) < \tau), 1 - \Pr_\nu(A_{\tau_n^j(a)+1} = a | \tau_n^j(a) < \tau)\Big) \leq 0.4$$

where $\mathrm{kl}(x, y) = x \log \frac{x}{y} + (1 - x) \log \frac{1-x}{1-y}$. The reason is $\Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = j) + \Pr_\nu(\tau_n^j(a) < \tau, A_{\tau_n^j(a)+1} = a) = \Pr_\nu(\tau_n^j(a) < \tau)$.

Proof of Lemma 11. The main idea is to utilize Lemma 8. Take $\Delta = \Delta_{a,j}$. From the definition, we know $T_j(a) \geq 1$. For $n = 1, 2, \cdots, T_j(a) - 1$, we have

$$\begin{aligned}&\sqrt{2n\Delta^2 \log \frac{1}{\Pr_{\nu'}(\tau_n < \tau)}}\\ \leq& \sqrt{2n\Delta^2 \log \frac{1}{\Pr_\nu(N_a(\tau) + N_j(\tau) > n)}}\\ \leq& \sqrt{2\Big(\frac{2}{200\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\} \log \frac{e}{\Pr_\nu(N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a}=j)}}\Big)\Delta_{a,j}^2 \log \frac{1}{\Pr_\nu(N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a} = j)}}\\ \leq& \sqrt{\frac{1}{50}}.\end{aligned}$$

The third line is by the fact that $\{N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a} = j\} \subset \{N_a(\tau) + N_j(\tau) > n\}$ holds for all $n = 0, 1, \cdots, T_j(a) - 1$. The last step is by the fact that $\frac{\Delta_{a,j}^2}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}} \leq 1$ and

$$\frac{\log \frac{1}{\Pr_\nu(N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a}=j)}}{1 + \log \frac{1}{\Pr_\nu(N_j(\tau) \geq \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a}=j)}} \leq 1.$$

Similarly, we have

$$\begin{aligned}\sqrt{2\pi n\Delta^2} \leq& \sqrt{\frac{\pi}{50}}\\ \frac{n\Delta^2}{2} \leq& \frac{1}{200}.\end{aligned}$$

Not hard to validate that $\sqrt{\frac{1}{50}} + \sqrt{\frac{\pi}{50}} + \frac{1}{200} < 0.4$. Applying Lemma 8, we prove the result. □

The next lemma converts Lemma 11 into a conclusion relating $\mathbb{E}_\nu N_a(\tau)$ to the probability of correctly recommending a given qualified arm with sufficiently many samples.

Lemma 12. *Following the same notation as Lemma 11 and still assume $\delta < \frac{1}{e}$, we have*

$$\mathbb{E}_\nu N_a(\tau) \geq \frac{T_j(a)}{4} \Pr_\nu \left(N_j(\tau) \geq \frac{\log \frac{1}{\delta}}{100\Delta_{j,0}^2}, \hat{a} = j \right)$$

holds for all $j \in [m], a \in [K], a \neq j$.

Proof of Lemma 12. We first claim that if $\tau_{n+1}^j(a) < \tau$, for all t such that $\tau_{n+1}^j(a) < t < \tau_{n+1}^j(a) + 1$, we have $A_t \neq a$. We prove this claim by contradiction. Assume there exists t such that $\tau_n^j(a)+1 < t < \tau_{n+1}^j(a) + 1$, $A_t = a$. Then, from the definition of $\tau_n^j(a)$, we know $N_j(\tau_n^j(a)) + N_a(\tau_n^j(a)) = n, N_j(\tau_n^j(a) + 1) + N_a(\tau_n^j(a) + 1) = n + 1$. Since $t > \tau_n^j(a) + 1$, which means $t - 1 \geq \tau_n^j(a) + 1$, we can conclude

$$\begin{aligned}
&N_j(t) + N_a(t) \\
=&\mathbb{1}(A_t = j) + \mathbb{1}(A_t = a) + N_j(t-1) + N_a(t-1) \\
=&1 + N_j(t-1) + N_a(t-1) \\
\geq&1 + N_j(\tau_n^j(a) + 1) + N_a(\tau_n^j(a) + 1) \\
=&n + 2.
\end{aligned}$$

However, $t \leq \tau_{n+1}^j(a)$ suggests that $N_j(t) + N_a(t) \leq N_1(\tau_{n+1}^j(a)) + N_a(\tau_{n+1}^j(a)) = n+1$, which leads to a contradiction. The claim follows.

Using the claim, we can rewrite

$$\begin{aligned}
&\mathbb{E}_\nu N_a(\tau) \\
=&\mathbb{E} \sum_{t=1}^{+\infty} \mathbb{1}(A_t = a)\mathbb{1}(\tau \geq t) \\
=&\mathbb{E} \sum_{n=0}^{+\infty} \mathbb{1}(A_{\tau_n^j(a)+1} = a)\mathbb{1}(\tau > \tau_n^j(a)) \\
\geq& \sum_{n=0}^{T_j(a)-1} \Pr_\nu \Big(A_{\tau_n^j(a)+1} = a, \tau > \tau_n^j(a) \Big). \qquad (17)
\end{aligned}$$

From Lemma 11, we know

$$\mathrm{kl}\left(\Pr_\nu \Big(A_{\tau_n^j(a)+1} = a | \tau > \tau_n^j(a) \Big), 1 - \Pr_\nu \Big(A_{\tau_n^j(a)+1} = a | \tau > \tau_n^j(a) \Big) \right) < 0.4$$

where $\mathrm{kl}(x, y) = x \log \frac{x}{y} + (1 - y) \log \frac{1-y}{1-x}$. From Pinsker's Inequality, we can further conclude

$$\begin{aligned}
&2 \left(1 - 2 \Pr_\nu \Big(A_{\tau_n^j(a)+1} = a | \tau > \tau_n^j(a) \Big) \right)^2 < 0.4 \\
\Rightarrow& \frac{1 - \sqrt{0.2}}{2} < \Pr_\nu \Big(A_{\tau_n^j(a)+1} = a | \tau > \tau_n^j(a) \Big) \\
\Rightarrow& \Pr_\nu \Big(A_{\tau_n^j(a)+1} = a, \tau > \tau_n^j(a) \Big) > \frac{1}{4} \Pr_\nu \Big(\tau > \tau_n^j(a) \Big).
\end{aligned}$$

Combine the last line and (17), we get

$$\begin{aligned}
&\mathbb{E}_\nu N_a(\tau)\\
\geq&\sum_{n=0}^{T_j(a)-1}\Pr_\nu\left(A_{\tau_n^j(a)+1}=a,\tau>\tau_n^j(a)\right)\\
\geq&\frac{1}{4}\sum_{n=0}^{T_j(a)-1}\Pr_\nu\left(\tau>\tau_n^j(a)\right)\\
\geq&\frac{1}{4}\sum_{n=0}^{T_j(a)-1}\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right)\\
=&\frac{T_j(a)}{4}\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right),
\end{aligned}$$

which is the conclusion we wish to prove. □

Finally, we use Lemma 12 to lower bound $\mathbb{E}_\nu\tau$ by the optimal value of an optimization problem, as stated in Lemma 6.

Proof of Lemma 6. For any $j\in[m]$, by Lemma 12, we have

$$\begin{aligned}
&\mathbb{E}\tau\\
=&\mathbb{E}N_j(\tau)+\sum_{a\neq j}\mathbb{E}\sum_{n=0}^{+\infty}\mathbb{1}(A_{\tau_n^j(a)+1}=a)\mathbb{1}(\tau>\tau_n^j(a))\\
\geq&\mathbb{E}N_j(\tau)\mathbb{1}(\hat{a}=j,N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2})+\sum_{a\neq j}\mathbb{E}\sum_{n=0}^{T_j(a)-1}\mathbb{1}(A_{\tau_n^j(a)+1}=a)\mathbb{1}(\tau>\tau_n^j(a))\\
\geq&\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\Pr(\hat{a}=j,N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2})+\sum_{a\neq j}\frac{1}{4}T_j(a)\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right)\\
\geq&\frac{1}{100\Delta_{j,0}^2}\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right)+\sum_{a\neq j}\frac{1}{4}T_j(a)\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right)\\
\geq&\sum_{a=1}^{K}\frac{1}{4}T_j(a)\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right).
\end{aligned}$$

The third step is by the Lemma 12. Since the last line holds for any $j\in[m]$,

$$\begin{aligned}
&\mathbb{E}\tau\\
\geq&\max_{1\leq j\leq m}\sum_{a=1}^{K}\frac{1}{4}T_j(a)\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right)\\
\geq&\max_{1\leq j\leq m}\sum_{a=1}^{K}\frac{\Pr_\nu\left(N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j\right)}{800\max\{\Delta_{j,0}^2,\Delta_{a,j}^2\}\left(1+\log\frac{1}{\Pr_\nu(N_j(\tau)\geq\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2},\hat{a}=j)}\right)}
\end{aligned}\tag{18}$$

Meanwhile, we can also derive

$$\begin{aligned}
\mathbb{E}\tau\geq&\sum_{j=1}^{m}\mathbb{E}\tau\mathbb{1}(\hat{a}=j,N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2})\\
\geq&\sum_{j=1}^{m}\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\Pr_\nu(\hat{a}=j,N_j(\tau)>\frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2})
\end{aligned}\tag{19}$$

Merge (18), (19), and notice that Lemma 10 asserts

$$\sum_{j=1}^{m} \Pr_{\nu}\left(\hat{a}=j, N_j(\tau) > \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\right) \geq \frac{1}{2},$$

we know $v = 800\mathbb{E}\tau, p_j = \Pr_{\nu}\left(\hat{a}=j, N_j(\tau) > \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\right), \forall j \in [m]$ must be a feasible solution to the following optimization problem

$$\begin{array}{ll} \min & v \\ s.t. & v \geq \sum_{a=1}^{K} \frac{p_j}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}\left(1+\log\frac{1}{p_j}\right)}, \forall j \in [m] \\ & v \geq \sum_{j=1}^{m} \frac{p_j \log\frac{1}{\delta}}{\Delta_{j,0}^2} \\ & \sum_{j=1}^{m} p_j \geq \frac{1}{2} \end{array} \tag{20}$$

The next step is to show $\frac{1}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}} \geq \frac{1}{4\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}}$ holds for $j \in [m], a \in [K]$. Indeed, if $\mu_a \geq \mu_j$, we have $\frac{1}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}} \geq \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}}$. If $\mu_j > \mu_a > \mu_0$, we have $\frac{1}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}} = \frac{1}{\Delta_{j,0}^2} > \frac{1}{4\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}}$. If $\mu_0 \geq \mu_a$, we have $\frac{1}{\max\{\Delta_{j,0}^2, \Delta_{a,j}^2\}} = \frac{1}{\Delta_{j,a}^2} > \frac{1}{4\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}}$. Plug in the inequality in (20), we know we know $v = 3200\mathbb{E}\tau, p_j = \Pr_{\nu}\left(\hat{a}=j, N_j(\tau) > \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\right), \forall j \in [m]$ must be a feasible solution to the following optimization problem

$$\begin{array}{ll} \min & v \\ s.t. & v \geq \sum_{a=1}^{K} \frac{p_j}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}\left(1+\log\frac{1}{p_j}\right)}, \forall j \in [m] \\ & v \geq \sum_{j=1}^{m} \frac{p_j \log\frac{1}{\delta}}{\Delta_{j,0}^2} \\ & \sum_{j=1}^{m} p_j \geq \frac{1}{2} \end{array} \tag{21}$$

Since $\mathbb{E}\tau$ must be non-negative and $0 \leq \Pr_{\nu}\left(\hat{a}=j, N_j(\tau) > \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\right) \leq 1, \forall j \in [m]$, we can add two more constraints to (21), showing that $v = 3200\mathbb{E}\tau, p_j = \Pr_{\nu}\left(\hat{a}=j, N_j(\tau) > \frac{\log\frac{1}{\delta}}{100\Delta_{j,0}^2}\right), \forall j \in [m]$ must be a feasible solution to the following optimization problem

$$\begin{array}{ll} \min & v \\ s.t. & v \geq \sum_{a=1}^{K} \frac{p_j}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}\left(1+\log\frac{1}{p_j}\right)}, \forall j \in [m] \\ & v \geq \sum_{j=1}^{m} \frac{p_j \log\frac{1}{\delta}}{\Delta_{j,0}^2} \\ & \sum_{j=1}^{m} p_j \geq \frac{1}{2} \\ & v \geq 0 \\ & 0 \leq p_j \leq 1, \forall j \in [m] \end{array} \tag{22}$$

which is exactly (2). That means the optimal value of (2) times $\frac{1}{3200}$ is a lower bound of $\mathbb{E}\tau$. □

Given Lemma 6, the last step is to prove a lower bound of the optimal value of (2).

Lemma 13. *Denote opt as the optimal value of problem (2). We have*

$$opt \geq \Omega\left(\min_{1\leq j\leq m} \frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{\frac{1}{j}}{\log^2(m+1)} \sum_{a=1}^{K} \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}}\right)$$

Proof of Lemma 13. We first show function $f(p) = \frac{p}{1+\log\frac{1}{p}}$ is convex for $p \in (0,1]$. Easy to calculate

$$\begin{aligned}\frac{\partial \frac{p}{1+\log\frac{2}{p}}}{\partial p} =& \frac{1+\log\frac{2}{p} - p(-\frac{1}{p})}{(1+\log\frac{2}{p})^2}\\ =& \frac{1}{1+\log\frac{2}{p}} + \frac{1}{(1+\log\frac{2}{p})^2}\\ \frac{\partial^2 \frac{p}{1+\log\frac{2}{p}}}{\partial p^2} =& \left(-\frac{1}{(1+\log\frac{2}{p})^2} - \frac{2}{(1+\log\frac{2}{p})^3}\right)\left(-\frac{1}{p}\right)\\ =& \left(\frac{1}{(1+\log\frac{2}{p})^2} + \frac{2}{(1+\log\frac{2}{p})^3}\right)\frac{1}{p}.\end{aligned}$$

It is straight forward to see $\frac{\partial^2 \frac{p}{1+\log\frac{2}{p}}}{\partial p^2} > 0$ holds for all $p \in (0,1]$, suggesting that function $\frac{p}{1+\log\frac{1}{p}}$ is convex, further indicating (2) is a convex optimization problem.

Introduce dual variables $\alpha \geq 0, \gamma \geq 0, \{\beta_j\}_{j=1}^m \in \mathbb{R}_+^m$ and Lagrangian function

$$\begin{aligned}&L(\alpha,\gamma,\{\beta_j\}_{j=1}^m;\{p_j\}_{j=1}^m,v)\\ =&v + \left(\sum_{j=1}^m \frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2}p_j - v\right)\alpha +\\ &\sum_{j=1}^m\left(\sum_{a=1}^K \frac{p_j}{\max\{\Delta_{a,0}^2,\Delta_{a,j}^2\}\left(1+\log\frac{1}{p_j}\right)} - v\right)\beta_j + (\frac{1}{2} - \sum_{j=1}^m p_j)\gamma\\ =&v(1-\alpha-\sum_{j=1}^m\beta_j) + \sum_{j=1}^m\left(\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2}p_j + \beta_j\sum_{a=1}^K\frac{p_j}{\max\{\Delta_{a,0}^2,\Delta_{a,j}^2\}\left(1+\log\frac{1}{p_j}\right)} - \gamma p_j\right) + \frac{\gamma}{2}.\end{aligned}$$

Take $\alpha^0 = \frac{1}{2}$, $\beta_j^0 = \frac{\frac{1}{2j}}{2\sum_{s=1}^m\frac{1}{2s}}$, $\gamma^0 = \min\limits_{1\leq j\leq m}\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2}\frac{1}{2} \cdot + \frac{\frac{1}{2j}\sum_{a=1}^K\frac{1}{\max\{\Delta_{a,0}^2,\Delta_{a,j}^2\}}}{(1+\log(m+1))(1+\log 200+\log m)}$. Since (2) is a convex optimization problem, by the strong duality, we can assert

$$\begin{aligned}\text{opt} =& \max_{\alpha\geq 0,\gamma\geq 0,\{\beta_j\}_{j=1}^m\in\mathbb{R}_+^m}\min_{\{p_j\}_{j=1}^m\in[0,1]^m,v\geq 0} L(\alpha,\gamma,\{\beta_j\};\{p_j\},v)\\ \geq& \min_{\{p_j\}_{j=1}^m\in[0,1]^m,v\geq 0} L(\alpha^0,\gamma^0,\{\beta_j^0\};\{p_j\},v).\end{aligned}$$

In the following, we suffice to prove

$$\begin{aligned}&\min_{\{p_j\}\in[0,1]^m,v\geq 0} L(\alpha^0,\gamma^0,\{\beta_j^0\}_{j=1}^m;\{p_j\}_{j=1}^m,v)\\ \geq&\Omega\left(\min_{1\leq j\leq m}\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{\frac{1}{j}}{\log^2(m+1)}\sum_{a=1}^K\frac{1}{\max\{\Delta_{a,0}^2,\Delta_{a,j}^2\}}\right).\end{aligned} \tag{23}$$

We start the proof of (23) by claiming

$$\min_{1\geq p_j\geq 0}\left(\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2}\alpha^0\cdot p_j + \frac{p_j}{1+\log\frac{2}{p_j}}\sum_{a=1}^K\frac{1}{\max\{\Delta_{a,0}^2,\Delta_{a,j}^2\}}\cdot\beta_j^0 - \gamma^0 p_j\right) \geq -\frac{\gamma^0}{100m} \tag{24}$$

holds for all $j \in [m]$. Define $\bar{L}_j(p) = \frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2}\alpha^0\cdot p + \frac{p}{1+\log\frac{2}{p}}\sum_{a=1}^K\frac{1}{\max\{\Delta_{a,0}^2,\Delta_{a,j}^2\}}\cdot\beta_j^0 - \gamma^0 p$ and $p_j^* = \arg\min\limits_{1\geq p\geq 0}\bar{L}_j(p)$. Not hard to validate $\bar{L}_j(p)$ is a convex function for $p\in[0,1]$, and there are two cases.

If $\frac{\partial \bar{L}_j(p)}{\partial p} \geq 0$ holds for all $0 \leq p \leq 1$, we have

$$\min_{1 \geq p \geq 0} \bar{L}_j(p) = \lim_{p \to 0} \bar{L}_j(p) = 0 > -\frac{\gamma^0}{100m},$$

suggesting that the claim (24) is true.

If $\frac{\partial \bar{L}_j(p)}{\partial p} < 0$ for some p, by the fact that $\bar{L}_j(p)$ is convex, we can assert there exists $p_0 \in (0,1)$ such that $\bar{L}_j(p)$ is decreasing in the interval $(0, p_0)$, and $\bar{L}_j(p_j^*) < 0$. Then, we can prove $p_j^* < \frac{1}{100m}$ by contradiction. If $p_j^* \geq \frac{1}{100m}$, we have

$$\begin{aligned}
&\frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2}\alpha^0 \cdot p_j^* + \frac{p_j^*}{1+\log \frac{2}{p_j^*}} \sum_{a=1}^{K} \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}} \cdot \beta_j^0 - \gamma^0 p_j^* \\
\geq &\frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2}\alpha^0 \cdot p_j^* + \frac{p_j^*}{1+\log 200 + \log m} \sum_{a=1}^{K} \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}} \cdot \beta_j^0 - \gamma^0 p_j^* \\
= &p_j^* \left(\frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2}\alpha^0 + \frac{\sum_{a=1}^{K} \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}} \cdot \beta_j^0}{1+\log 200 + \log m} - \gamma^0 \right) \\
\geq &p_j^* \left(\frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2}\frac{1}{2} + \frac{\sum_{a=1}^{K} \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}} \frac{\frac{1}{2j}}{2\sum_{s=1}^{m} \frac{1}{2s}}}{1+\log 200 + \log m} - \frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2}\frac{1}{2} - \frac{\sum_{a=1}^{K} \frac{1}{2j \max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}}}{(1+\log 200 + \log m)(1+\log(m+1))} \right) \\
> &0.
\end{aligned}$$

The second last line is by the definition of γ^0. The last line is by the fact that $2\sum_{s=1}^{m} \frac{1}{2s} < 1 + \log(m+1)$ for $m \geq 1$. This last line contradicts with the fact that $\bar{L}_j(p_j^*) < 0$. Thus, we can conclude $p_j^* < \frac{1}{100m}$, further

$$\begin{aligned}
&\frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2}\alpha^0 \cdot p_j^* + \frac{p_j^*}{1+\log \frac{2}{p_j^*}} \sum_{a=1}^{K} \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}} \cdot \beta_j^0 - \gamma^0 p_j^* \\
\geq &-\gamma^0 p_j^* \\
\geq &-\frac{\gamma^0}{100m}.
\end{aligned}$$

That means in the case $\frac{\partial \bar{L}_j(p)}{\partial p} < 0$, we can still conclude the claim (24) is true.

Based on the (24), we have

$$\begin{aligned}
&\min_{\{p_j\}_{j=1}^m \in [0,1]^m, v \geq 0} L(\alpha^0, \gamma^0, \{\beta_j^0\}; \{p_j\}, v) \\
= &\min_{v \geq 0} v(1 - \alpha^0 - \sum_{j=1}^{m} \beta_j^0) + \\
&\sum_{j=1}^{m} \min_{1 \geq p_j \geq 0} \left(\frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2} p_j + \beta_j^0 \sum_{a=1}^{K} \frac{p_j}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\} \left(1 + \log \frac{1}{p_j}\right)} - \gamma^0 p_j \right) + \frac{\gamma^0}{2} \\
\geq &\sum_{j=1}^{m} -\frac{\gamma^0}{100m} + \frac{\gamma^0}{2} \\
\geq &\frac{\gamma^0}{4}.
\end{aligned}$$

The second last line is by the fact that $1-\alpha^0-\sum_{j=1}^m \beta_j^0 = 1-\frac{1}{2}-\sum_{j=1}^m \frac{\frac{1}{2j}}{2\sum_{s=1}^m \frac{1}{2s}} = 0$ and the claim (24). From the definition of γ^0, we know

$$\begin{aligned}&\min_{\{p_j\}\in[0,1]^m, v\geq 0} L(\alpha^0, \gamma^0, \{\beta_j^0\}; \{p_j\}, v)\\ \geq&\Omega\left(\min_{1\leq j\leq m} \frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{\frac{1}{j}}{\log^2(m+1)}\sum_{a=1}^K \frac{1}{\max\{\Delta_{a,0}^2, \Delta_{a,j}^2\}}\right),\end{aligned}$$

which also completes the proof. □

C Proof of Upper Bound

C.1 Rigorous Presentation of Algorithm

Introduce

$$U(t,\delta) = \frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2(\lceil\log_2 t\rceil^+)^2}{\delta}}}{t}. \tag{25}$$

where $\lceil x\rceil := \max\{x, 1\}$.

Before presenting the full statement of the algorithm, we first introduce some notations

$$\begin{aligned}&N_a^{\text{ee}}(\mathcal{H}^{ee}) = \sum_{s=1}^{|\mathcal{H}^{ee}|}\mathbb{1}(A_s(\mathcal{H}^{\text{ee}})=a), N_a^{\text{et}}(\mathcal{H}^{et}) = \sum_{s=1}^{|\mathcal{H}^{et}|}\mathbb{1}(A_s(\mathcal{H}^{\text{et}})=a)\\ &\hat{\mu}_a^{\text{ee}}(\mathcal{H}^{ee}) = \frac{\sum_{s=1}^{|\mathcal{H}^{ee}|}\mathbb{1}(A_s(\mathcal{H}^{\text{ee}})=a)X_s(\mathcal{H}^{\text{ee}})}{N_a^{\text{ee}}(\mathcal{H}^{ee})},\\ &\hat{\mu}_a^{\text{et}}(\mathcal{H}^{et}) = \frac{\sum_{s=1}^{|\mathcal{H}^{et}|}\mathbb{1}(A_s(\mathcal{H}^{\text{et}})=a)X_s(\mathcal{H}^{\text{et}})}{N_a^{\text{et}}(\mathcal{H}^{et})}\\ &\text{UCB}_a^{\text{ee}}(\mathcal{H}^{ee},\delta) = \hat{\mu}_a^{\text{ee}}(\mathcal{H}^{ee}) + U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta}{K})\\ &\text{LCB}_a^{\text{ee}}(\mathcal{H}^{ee},\delta) = \hat{\mu}_a^{\text{ee}}(\mathcal{H}^{ee}) - C\cdot U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta}{K})\\ &\text{UCB}_a^{\text{et}}(\mathcal{H}^{et},\delta) = \hat{\mu}_a^{\text{et}}(\mathcal{H}^{et}) + U(N_a^{\text{et}}(\mathcal{H}^{ee}), \frac{\delta}{K})\\ &\text{LCB}_a^{\text{et}}(\mathcal{H}^{et},\delta) = \hat{\mu}_a^{\text{et}}(\mathcal{H}^{et}) - U(N_a^{\text{et}}(\mathcal{H}^{ee}), \frac{\delta}{K})\end{aligned} \tag{26}$$

We treat $\mathcal{H}^{\text{ee}}$ and $\mathcal{H}^{\text{et}}$ as stacks with indexed entries: $A_s(\mathcal{H}^{\text{ee}})$ and $X_s(\mathcal{H}^{\text{ee}})$ denote, respectively, the action and reward stored at index s in $\mathcal{H}^{\text{ee}}$, and analogously for $\mathcal{H}^{\text{et}}$. When $N_a^{\text{ee}}(\mathcal{H}^{\text{ee}}) = 0$ for some $a\in[K]$, we set $\text{UCB}_a^{\text{ee}} = +\infty$ and $\text{LCB}_a^{\text{ee}} = -\infty$; the same convention applies to UCB_a^{et} and LCB_a^{et} when $N_a^{\text{et}}(\mathcal{H}^{\text{et}}) = 0$. In addition, we define

$$\begin{aligned}\mathcal{A}_B(\mathcal{H}^{\text{ee}},\delta,k) = \{&a\in B:\\ &\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}},\delta) = \max_{a'\in B}\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}},\delta),\\ &\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}},\delta) > \mu_0,\\ &N_a^{\text{ee}}(\mathcal{H}^{\text{ee}}) \leq (C+1)^2\beta_k\log\frac{4K}{\delta_k} - 1\},\end{aligned}$$

which is the action set at each round.

With this notation in place, we present Algorithm 3, adapted from Li and Cheung (2025). Our analysis follows closely Section 4 and Appendix B of Li and Cheung (2025); nevertheless, we provide the full algorithmic description and proofs for completeness. In the subsequent subsections, we discuss the algorithm line by line and summarize our modifications relative to the prior literature in Appendix C.1.5.

Algorithm 3: Sequential Exploration Exploitaion (SEE)

Input: Bracket B, Threshold μ_0, Tolerance level δ, Tunable parameter $C > 1$, Arm number K;
Initialize: $\delta_k = \frac{1}{3^k}, \beta_k = 2^k, \alpha_k = 5^k, \forall k \in \mathbb{N}$, History $\mathcal{H}^{\text{ee}} = \mathcal{H}^{\text{et}} = \emptyset$, Temporary Container $Q = \emptyset$. Introduce notations in (26), $t^{\text{ee}} = t^{\text{et}} = 0$;
for *phase index* $k = 1, 2, \cdots$ **do**
  /* Phase k starts */
  while *True* **do**
    if $t^{ee} \geq 1$ *and* $LCB^{ee}_{A^{ee}_t}(\mathcal{H}^{ee}, \delta_k) \geq \mu_0$ **then**
      Take $\hat{a}_k \leftarrow A^{\text{ee}}_t$;
      $\mathcal{H}^{\text{ee}} \leftarrow \mathcal{H}^{\text{ee}} \setminus \{(A^{\text{ee}}_t, X)\}$, $N^{\text{ee}}_{A^{\text{ee}}_t} \leftarrow N^{\text{ee}}_{A^{\text{ee}}_t} - 1$, $Q \leftarrow Q \cup \{(A^{\text{ee}}_t, X)\}$, $\hat{a} \leftarrow A^{\text{ee}}_t$;
      while $N^{et}_{\hat{a}} \leq \frac{(C+3)^2}{(C-1)^2}\beta_k \log \frac{4K\alpha_k}{\delta} - 1$ **do**
        Sample $X \sim \rho_{\hat{a}}$, $\mathcal{H}^{\text{et}} \leftarrow \mathcal{H}^{\text{et}} \cup \{(\hat{a}, X)\}$, $N^{\text{et}}_{\hat{a}} \leftarrow N^{\text{et}}_{\hat{a}} + 1$, $t^{\text{et}} \leftarrow t^{\text{et}} + 1$;
        if $LCB^{et}_{\hat{a}}(\mathcal{H}^{et}, \frac{\delta}{\alpha_k}) > \mu_0$ **then**
          Return $\hat{a}$ as a qualified arm;
        end
      end
      Break; // Enter phase $k+1$
    end
    if $\forall a \in B$, $UCB^{ee}_a(\mathcal{H}^{ee}, \delta_k) \leq \mu_0$ **then**
      if $\delta_k \leq \frac{\delta}{3}$ *and* $|B| = K$ **then**
        Take $\hat{a}_k =$ None and return None;
        /* Only the bracket with size K is allowed to output None */
      else
        Take $\hat{a}_k =$ Not Completed and break; // Enter phase $k+1$
      end
    end
    if $\mathcal{A}_B(\mathcal{H}^{ee}, \delta_k, k) = \emptyset$ **then**
      Take $\hat{a}_k =$ Not Completed and break; // Enter phase $k+1$
    end
    Take arbitrary $A^{\text{ee}}_t \in \mathcal{A}_B(\mathcal{H}^{\text{ee}}, \delta_k, k)$;
    if $\exists (a, X) \in Q$ *such that* $a = A^{ee}_t$ **then**
      $\mathcal{H}^{\text{ee}} \leftarrow \mathcal{H}^{\text{ee}} \cup \{(a, X)\}$, $N^{\text{ee}}_{A^{\text{ee}}_t} \leftarrow N^{\text{ee}}_{A^{\text{ee}}_t} + 1$, $Q \leftarrow Q \setminus \{(a, X)\}$;
    else
      Sample $X \sim \rho_{A^{\text{ee}}_t}$, $\mathcal{H}^{\text{ee}} \leftarrow \mathcal{H}^{\text{ee}} \cup \{(A^{\text{ee}}_t, X)\}$, $N^{\text{ee}}_{A^{\text{ee}}_t} \leftarrow N^{\text{ee}}_{A^{\text{ee}}_t} + 1$, $t^{\text{ee}} \leftarrow t^{\text{ee}} + 1$;
    end
  end
  Take $\tau^{\text{ee}}_k(B) \leftarrow t^{\text{ee}}, \tau^{\text{et}}_k(B) \leftarrow t^{\text{et}}$; // Independent of algorithm design, only for analysis
  /* Phase k ends */
end

C.1.1 From Line 1 to 2: Input and Initialize Parameter

Algorithm 3 takes a bracket B, a threshold μ_0, a tolerance level δ, the number of arms K, tunable parameter C as input, and only pull arms in B. The parameters μ_0 and K coincide with those used by Algorithm 1. When Algorithm 1 calls Algorithm 3, it sets $\delta \leftarrow \delta/(\lceil \log_2 K \rceil + 1)$. We use $C = 1.01$ by default, although any value strictly larger than 1 is admissible.

In addition to the problem parameters, 3 also initialize some parameters to manage the pulling process. Parameter $\{\delta_k = \frac{1}{3^k}, \beta_k = 2^k, \alpha_k = 5^k\}_{k=1}^{+\infty}$ are used to set up the tolerance level and budget during the pulling process. It is possible to specify other values for $\delta_k, \beta_k, \alpha_k$, see Appendix B.1 in Li and Cheung (2025). The containers $\mathcal{H}^{\text{ee}}$, $\mathcal{H}^{\text{et}}$, and Q store samples collected throughout the process. In particular, $\mathcal{H}^{\text{ee}}$ and $\mathcal{H}^{\text{et}}$ are ordered stacks that preserve the arrival order of samples for each arm in B, which allows us to remove and transfer the most recently collected sample when needed. The role of Q is to ensure that LCB_a^{ee} cannot be greater than μ_0 at the beginning of each phase; see Section C.1.4.

C.1.2 Line 3 to Line 34: Overall Idea of Algorithm 3

At the start of Algorithm 3, it splits the arm distributions into two parallel and independent copies. One copy is for exploration period, and another copy is for exploitation period. Correspondingly, reward collected in the exploration periods are stored in $\mathcal{H}^{\text{ee}}$. Reward collected in the exploration periods, which is from Line 6 to Line 14, are stored in $\mathcal{H}^{\text{et}}$.

After specifying the parameters through Line 1 and 2, Algorithm 3 splits the pulling process into multiple phases, noted by the For loop in Line 3. In each phase k, Algorithm 3 consists of two parts. Each phase consists of two components executed within the **while**-loop in Line 4:

- a sequence of stopping/transition checks (Lines 5–25), and
- an *exploration period* (Lines 26–31) that applies a UCB rule (as in Jamieson and Nowak (2014)) to select the next arm to sample.

Although the loop condition is **True**, the stopping/transition checks ensure that the algorithm cannot run indefinitely; see Section C.1.3.

Within phase k, exploration uses tolerance level δ_k to form UCB/LCB bounds, while exploitation uses tolerance level δ/α_k and samples only a single candidate arm. Once $\hat{a}$ takes value in $[K] \cup \{\textsf{None}\}$, Algorithm 3 will terminate. Moreover, each phase enforces explicit sampling budgets: during exploration, each arm $a \in B$ can be sampled at most $(C+1)^2 \beta_k \log \frac{4K}{\delta_k}$ times, while during exploitation the candidate arm is sampled at most $\frac{(C+3)^2}{(C-1)^2} \beta_k \log \frac{4K\alpha_k}{\delta}$ times. Further details appear in Sections C.1.3 and C.1.4.

C.1.3 Line 5 to 25: Conditions on Phase Transition

There are three possible conditions to end the exploration period in a phase k. The first one is Line 5, judge whether the algorithm should enter the exploitation period. Intuitively speaking, the inequality $\text{LCB}^{\text{ee}}_{A_t^{\text{ee}}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_0$ provides a recommended qualified arm with confidence $1 - \Theta(\delta_k)$. Since A_t^{ee} is the action we take in the last round, this condition also guarantees $\text{LCB}^{\text{ee}}_{A_t^{\text{ee}}}(\mathcal{H}^{\text{ee}}, \delta_k) < \mu_0$ holds before the collecting the latest sample of A_t^{ee}. To further confirm whether arm A_t^{ee} is qualified, we need to keep pulling arm A_t^{ee} by collecting samples independent of the exploration periods. Line 10 suggests that the algorithm will terminate and output A_t^{ee}, if the LCB calculated by the tolerance level δ/α_k is above μ_0 during the exploitation process. Before entering the while loop (Line 8) in the exploitation period, Line 7 transfer the latest collected sample of A_t^{ee} in the exploration period into Q. Usage of Q is to guarantee Lemma 23, requiring $\text{LCB}_a^{\text{ee}} < \mu_0$ holds for all $a \in B$ at the start of phase k. Similar to the reason described in the section 4.1 of Li and Cheung (2025), Lemma 23 is required to prove Lemma 24, further Theorem 3. We put more discussion of Q to section C.1.4.

The second condition is Line 16, aiming to output None for negative case. To be specific, if the Upper Confidence Bound(UCB) in the exploration periods are all below μ_0, Algorithm 3 will leave the exploration periods in phase k. Then, the condition in Line 16 determines the next period. If the bracket B is indeed $[K]$ and the tolerance level $\delta_k < \frac{\delta}{3}$, Algorithm 3 will terminate the overall pulling process and output None. In other cases, Algorithm 3 will enter the exploration period in the

next phase $k+1$. In the case $\max_{a\in B}\mu_a < \mu_0$ and $|B| < K$, it is possible that Algorithm 3 never break the for loop in Line 3.

The third condition is Line 23terminates the exploration period when the available action set is empty. This occurs when no arm simultaneously maximizes the exploration UCB over the entire bracket, has UCB above the threshold, and remains within its phase-specific sampling budget. In this case, Algorithm 3 proceeds to the next phase.

C.1.4 Line 26 to 31: Exploration

In phase k, Algorithm 3 use tolerance level δ_k to calculate Upper Confidence Bound(UCB) according to (26) and determine the arm A_t^{ee} by Line 26. Line 27 to 31 determine where to collect sample of arm A_t^{ee}. If Q contains one sample of arm A_t^{ee}, Algorithm 3 will transfer this sample to $\mathcal{H}^{\text{ee}}$ without collecting a new sample from distribution $\nu_{A_t^{\text{ee}}}$. If not, Algorithm 3 samples a new random variable from $\nu_{A_t^{\text{ee}}}$.

As noted above, the purpose of Q is to enforce Lemma 23, which requires $\text{LCB}_a^{\text{ee}} < \mu_0$ for all $a \in B$ at the start of each phase. One might ask whether it suffices to simply discard the most recent exploration sample instead of storing it in Q. **This is not sufficient.** The transfer mechanism preserves the per-arm arrival order of samples, which is needed for the following property.

Lemma 14. *Given an algorithm copy alg taking bracket B as input. Condition on reward sequence in the exploration period $\{X_{a,s}^{ee}\}_{s=1}^{+\infty}$, for each $a \in B$, we can assert $\hat{\mu}_a(\mathcal{H}^{ee}) \in \{\frac{\sum_{s=1}^{t} X_{a,s}^{ee}}{t}\}_{t=1}^{+\infty}$ throughout the pulling process.*

The key point is that the algorithm never permutes the order of $\{X_{a,s}^{\text{ee}}\}_{s\geq 1}$ for any arm $a \in B$; it only temporarily holds out the most recent sample and later reinserts it without changing order.

For the analysis, we introduce the random phase indices

$$\kappa_b^{\text{ee}} = \min\left\{k : \forall t, \forall a \in [K], |\hat{\mu}_{a,b}^{ee}(t) - r_a^{\nu}| < U(t, \frac{\delta_k}{K})\right\}$$

$$\kappa_b^{\text{et}} = \min\left\{k : \forall t, \forall a \in [K], |\hat{\mu}_{a,b}^{et}(t) - r_a^{\nu}| < U(t, \frac{\delta}{K\alpha_k})\right\}.$$

$\hat{\mu}_{a,b}^{ee}(t), \hat{\mu}_{a,b}^{et}(t)$ denote the empirical mean of arm a of the first t collected samples in bracket B_b, corresponding to the exploration and exploitation periods separately. More details are in section C.2. Lemma 14 suggests that for phase $k \geq \kappa_b^{\text{ee}}$, we can guarantee $\text{LCB}_a^{\text{ee}} < \mu_a < \text{UCB}_a^{\text{ee}}$ holds for all $a \in B_b$, which further helps to prove upper bounds of the pulling times for all $a \in B_b$. Simply dropping the latest collected samples is unable to guarantee 14, making it much harder to conduct the analysis.

Besides, we also care about the size of Q during the pulling process, as we can observe the following Lemmas.

Lemma 15. *Given an algorithm copy alg, throughout the pulling process, we have*

$$t^{ee} = |\mathcal{H}^{ee}| + |Q|$$

always hold.

The reason is all the collected samples in the exploration periods are either in $\mathcal{H}^{\text{ee}}$ or Q. Notice that we only transfer sample from $\mathcal{H}^{\text{ee}}$ to Q in Line 7. In this case, A_t^{ee} just get pulled in the last round of exploration. Meanwhile, once Q contains a sample for an arm $a \in B$, condition in Line 27 ensures that Algorithm 3 will not draw a fresh exploration sample for arm a until that held-out sample is transferred back. This yields the following bound.

Lemma 16. *Consider an algorithm copy alg taking bracket B as input, we have $|Q| \leq |B|$ throughout the pulling process*

At the end of this subsection, Lemma 17 is similar to 14, but for the exploitation period.

Lemma 17. *Given an algorithm copy alg taking bracket B as input. Condition on $\{X_{a,s}^{et}\}_{s=1}^{+\infty}$, for each $a \in B$, we can assert $\hat{\mu}_a(\mathcal{H}^{et}) \in \{\frac{\sum_{s=1}^{t} X_{a,s}^{et}}{t}\}_{t=1}^{+\infty}$ throughout the pulling process.*

This is correct, as we collect the samples in the exploitation period sequentially in $\mathcal{H}^{\text{et}}$.

C.1.5 Adaptation of History Literature

Compared to the Algorithm SEE in Li and Cheung (2025), our main adaptation consists of two points. The first adaptation is we place the stopping/transition conditions (Lines 5–25) *before* the exploration step (Lines 26–31). This ordering is required for Lemma 30, which ensures that the algorithm does not sample any arm a whose exploration upper confidence bound satisfies $\mathrm{UCB}_a^{\mathrm{ee}} \le \mu_0$.

The second point is for each phase k, we set up a maximum pulling times of each arm in the exploration period, rather than a budget on the total number of exploration pulls across the bracket (Line 23 and 26). This modification is motivated by the randomization of arms within each bracket: with randomized arm assignments, it becomes difficult to track how a *total* budget is required for the bracket to produce a candidate $\hat{a}_k$.

C.2 Properties of Algorithm 3

This subsection closely follows Appendix B of Li and Cheung (2025), with only minor differences. Since we modify several implementation details (and correspondingly adjust parts of the argument), we include the full proofs for completeness.

In Algorithm 1, we create $\lceil \log_2 K \rceil + 1$ independent copies of Algorithm 3. For each copy index $b \in [\lceil \log_2 K \rceil + 1]$, we further create two independent sampling copies for exploration and exploitation periods separately. Recall the definition

$$\kappa_b^{\mathrm{ee}} = \min\left\{k : \forall t, \forall a \in [K], |\hat{\mu}_{a,b}^{ee}(t) - r_a^{\nu}| < U(t, \delta_k/K)\right\}$$
$$\kappa_b^{\mathrm{et}} = \min\left\{k : \forall t, \forall a \in [K], |\hat{\mu}_{a,b}^{et}(t) - r_a^{\nu}| < U(t, \delta/(K\alpha_k))\right\}.$$

$\hat{\mu}_{a,b}^{ee}(t), \hat{\mu}_{a,b}^{et}(t)$ denote the empirical mean of arm a of the first t collected samples in bracket B_b, corresponding to the exploration and exploitation periods separately. The above definition of $\kappa_b^{\mathrm{ee}}, \kappa_b^{\mathrm{et}}$ are conditioned on the realized reward of $\{X_{a,s}^{\mathrm{ee}}\}_{a=1,s=1}^{K,+\infty}$, $\{X_{a,s}^{\mathrm{et}}\}_{a=1,s=1}^{K,+\infty}$ collected by the algorithm copy b. $\kappa_b^{\mathrm{ee}}, \kappa_b^{\mathrm{et}}$ are the minimum phase index such that concentration event holds. Since B_b is determined by an exogenous random permutation, we know B_b is independent with $\kappa_b^{\mathrm{ee}}, \kappa_b^{\mathrm{et}}$.

In the following, we will analyze the properties of Algorithm 3, a given bracket B. For convenience, given a bracket B with index b, we denote $\kappa_B^{\mathrm{ee}}, \kappa_B^{\mathrm{et}}$ as corresponding $\kappa_b^{\mathrm{ee}}, \kappa_b^{\mathrm{et}}$, without emphasizing the index b. Further, we denote $\tau_k^{\mathrm{ee}}(B), \tau_k^{\mathrm{et}}(B)$ as the value of $t^{\mathrm{ee}}, t^{\mathrm{et}}$ at the end of phase k, i.e. each time we enter Line 33 in Algorithm 3.

Lemma 18. *For any B, we have* $\Pr(\kappa_B^{ee} \ge k) \le \frac{\pi^2}{6}\delta_{k-1}$, $\Pr(\kappa_B^{ee} \ge k) \le \frac{\pi^2}{6}\frac{\delta}{\alpha_{k-1}}$

Proof of Lemma 18. By the Lemma 35, we have

$$\begin{aligned}
\Pr(\kappa_B^{\mathrm{ee}} \ge k) \le & \Pr\left(\exists t, \exists a \in [K], |\hat{\mu}_a^{ee}(t) - r_a^{\nu}| \ge \sqrt{\frac{4\log\frac{2K}{\delta_{k-1}} + 8\log\log(2t)}{t}}\right) \\
\le & \sum_{a=1}^{K} \Pr\left(\exists t, |\hat{\mu}_a^{ee}(t) - r_a^{\nu}| \ge \sqrt{\frac{4\log\frac{2K}{\delta_{k-1}} + 8\log\log(2t)}{t}}\right) \\
\le & \sum_{a=1}^{K} \frac{\pi^2\delta_{k-1}}{6K} \\
= & \frac{\pi^2\delta_{k-1}}{6}
\end{aligned}$$

and

$$\begin{aligned}\Pr(\kappa_B^{\text{et}} \geq k) \leq & \Pr\left(\exists t, \exists a \in [K], |\hat{\mu}_a^{et}(t) - r_a^{\nu}| \geq \sqrt{\frac{4\log\frac{2K\alpha_{k-1}}{\delta} + 8\log\log(2t)}{t}}\right) \\ \leq & \sum_{a=1}^{K} \Pr\left(\exists t, |\hat{\mu}_a^{ee}(t) - r_a^{\nu}| \geq \sqrt{\frac{4\log\frac{2K\alpha_{k-1}}{\delta} + 8\log\log(2t)}{t}}\right) \\ \leq & \sum_{a=1}^{K} \frac{\pi^2\delta}{6K} \\ = & \frac{\pi^2\delta}{6\alpha_{k-1}}.\end{aligned}$$

□

Following Lemma 18, we can control the probability of wrong output through the following lemmas.

Lemma 19. *Consider an Algorithm copy b with bracket B, we have* $\Pr(\textit{alg copy } b \textit{ outputs } \hat{a} \textit{ such that } \mu_{\hat{a}} \leq \mu_0) \leq \frac{\pi^2}{6}\frac{\delta}{5}$.

Proof of Lemma 19. By the Line 6 to 14, we know

$$\begin{aligned}&\Pr(\text{alg copy } b \text{ outputs } \hat{a} \text{ such that } \mu_{\hat{a}} < \mu_0) \\ \leq & \Pr\left(\exists a, s.t. \mu_a < \mu_0, t \in \mathbb{N}, \hat{\mu}_a^{\text{et}}(t) - \frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K\alpha_1(\lceil\log_2 t\rceil^+)^2}{\delta}}}{t} \geq \mu_0\right) \\ \leq & \sum_{a:\mu_a<\mu_0} \Pr\left(t \in \mathbb{N}, \hat{\mu}_a^{\text{et}}(t) - \frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K\alpha_1(\lceil\log_2 t\rceil^+)^2}{\delta}}}{t} \geq \mu_a\right) \\ \leq & \sum_{a:\mu_a\leq\mu_0} \frac{\pi^2}{6}\frac{\delta}{\alpha_1 K} \\ \leq & \frac{\pi^2}{6}\frac{\delta}{5}.\end{aligned}$$

The second last line is by the Lemma 35. □

Lemma 20. *Assume $\mu_1 > \mu_0$. Consider an Algorithm copy $\lceil\log_2 K\rceil + 1$, whose allocated bracket is $[K]$. We have* $\Pr(\textit{alg copy } \lceil\log_2 K\rceil + 1 \textit{ outputs } \textsf{None}) \leq \frac{\pi^2}{6}\frac{\delta}{3}$.

Proof of Lemma 20. By the condition in Line 17, we know

$$\begin{aligned}&\Pr(\text{alg copy } \lceil\log_2 K\rceil + 1 \text{ outputs } \textsf{None}) \\ \leq & \Pr(\exists t \in \mathbb{N}, \delta_k \leq \frac{\delta}{3}, \hat{\mu}_1^{\text{ee}}(t) + \frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K(\lceil\log_2 t\rceil^+)^2}{\delta_k}}}{t} \leq \mu_0) \\ \leq & \Pr(\exists t \in \mathbb{N}, \hat{\mu}_1^{\text{ee}}(t) + \frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K(\lceil\log_2 t\rceil^+)^2}{\delta/3}}}{t} \leq \mu_0) \\ \leq & \frac{\pi^2}{6}\frac{\delta}{3}.\end{aligned}$$

The last line is by the Lemma 35. □

The following two Lemmas present almost surely upper bound of pulling times.

Lemma 21. *For any B, and any $a \in B$, we have* $N_a^{ee,B}\left(\tau_k^{ee}(B)\right) \leq (C+1)^2\beta_k\log\frac{4K}{\delta_k}$.

Proof of Lemma 21. We use induction on $k \in \mathbb{N}$. For $k = 1$, by the Line 26 in Algorithm 3, we konw

$$N_a^{\text{ee},B}\left(\tau_1^{\text{ee}}(B)\right) \leq (C+1)^2\beta_1 \log\frac{4K}{\delta_1}.$$

If the conclusion holds for k, then for $k+1$, we can still conclude

$$\begin{aligned}
N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \leq& \max\left\{N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right), (C+1)^2\beta_k \log\frac{4K}{\delta_k}\right\}\\
\leq& \max\left\{(C+1)^2\beta_{k-1}\log\frac{4K}{\delta_{k-1}}, (C+1)^2\beta_k\log\frac{4K}{\delta_k}\right\}\\
=&(C+1)^2\beta_k\log\frac{4K}{\delta_k}.
\end{aligned}$$

The first step is still caused by the Line 26 in Algorithm 3. The second step is from the conclusion at phase index k. The last line shows that the induction holds. □

Lemma 22. *For any B, $\tau_k^{ee}(B) \leq |B|(C+1)^2\beta_k\log\frac{4K}{\delta_k}$, $\tau_k^{et}(B) \leq \frac{2(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K\alpha_k}{\delta}$*

Proof of Lemma 22. Regarding $\tau_k^{\text{ee}}(B) \leq |B|(C+1)^2\beta_k\log\frac{4K}{\delta_k}$, we can directly conclude this from Lemma 21.

Regarding $\tau_k^{\text{et}}(B) \leq \frac{2(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K\alpha_k}{\delta}$, we first show $\tau_k^{\text{et}}(B) - \tau_{k-1}^{\text{et}}(B) \leq \frac{(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K\alpha_k}{\delta}$ holds for any phase index k. By direct calculation, we have

$$\begin{aligned}
&\tau_k^{\text{et}}(B) - \tau_{k-1}^{\text{et}}(B)\\
=&\sum_{a=1}^K\left(N_{\hat{a}_k}^{\text{et}}\left(\tau_k^{\text{et}}(B)\right) - N_{\hat{a}_k}^{\text{et}}\left(\tau_{k-1}^{\text{et}}(B)\right)\right)\mathbb{1}(\hat{a}_k = a)\\
\leq&\sum_{a=1}^K N_{\hat{a}_k}^{\text{et}}\left(\tau_k^{\text{et}}(B)\right)\mathbb{1}(\hat{a}_k = a)\\
\leq&\sum_{a=1}^K\frac{(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K\alpha_k}{\delta}\mathbb{1}(\hat{a}_k = a)\\
\leq&\frac{(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K\alpha_k}{\delta}.
\end{aligned}$$

The second last line is by the loop condition in Line 8 of Algorithm 3. Then we have $\tau_k^{\text{et}}(B) \leq \sum_{k'=1}^k\frac{(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K\alpha_k}{\delta}$. Since $\frac{\frac{(C+3)^2}{(C-1)^2}\beta_{k+1}\log\frac{4K\alpha_{k+1}}{\delta}}{\frac{(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K\alpha_k}{\delta}} \geq 2$ holds for all k, the proof is done. □

The following Lemma shows the effect of the Temporary Container. It can guarantee all the UCBs are below μ_0, at the start of any phase. This is crucial for the analysis.

Lemma 23. *For any B, at the start of each phase k, i.e. each time we enter Line 3 in Algorithm 3, we have $\text{LCB}_a^{ee}(\mathcal{H}^{ee}, \delta_k) < \mu_0$ holds for all $a \in B$.*

Proof of Lemma 23. We prove the Lemma by induction on k. For $k = 1$, this is obviously true, as we denote the LCB_a^{ee} as $-\infty$ if the pulling times of arm a is 0.

If the conclusion holds for the phase $k-1$, we prove the case of k by discussing all the possible values of $\hat{a}_{k-1}$. Since the pulling process enter phase k, we can conclude phase $k-1$ may end with taking $\hat{a}_{k-1}$ at the Line 20, 24 or 6.

If $\hat{a}_{k-1}$ takes Not Completed at the Line 20, we have we $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) < \text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) < \mu_0$, forall $a \in [K]$.

If $\hat{a}_{k-1}$ takes Not Completed at the Line 24, We can also assert $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) < \mu_0, \forall a \in [K]$. We argue by contradiction. Assume there exists $a \in [K]$ such that $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) \geq \mu_0$ at the

end of phase $k-1$. From the induction, since $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) < \mu_0$ at the start of phase $k-1$, the algorithm would enter Line 6 and $\hat{a}_{k-1}$ cannot take value Not Completed. Thus, we can assert $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) < \mu_0, \forall a \in [K]$ holds at the end of phase $k-1$. Further, at the start of phase k, we have $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) < \text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) < \mu_0$.

The remaining work is to prove the conclusion in the case that $\hat{a}_{k-1} \in [K]$ at the Line 6. We can first assert $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) < \mu_0, \forall a \neq \hat{a}_{k-1}$. The reason is from the induction, $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_{k-1}) < \mu_0$ holds for all a at the start of phase $k-1$. Once there exists an arm a whose LCB is above μ_0 after an arm pull, the algorithm will enter Line 6, suggesting that there is at most one arm whose LCB is above μ_0. This observation guarantees that $\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) < \mu_0$.

Then, we turn to analyze $\text{LCB}_{\hat{a}_{k-1}}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k)$ at the start of phase k. Before the execution of Line 7, the following inequalities must hold

$$\frac{X + \sum_{s=1}^{N_{\hat{a}_{k-1}}^{\text{ee}}-1} X_{\hat{a}_{k-1},s}}{N_{\hat{a}_{k-1}}^{\text{ee}}} - C \cdot U\left(N_{\hat{a}_{k-1}}^{\text{ee}}, \frac{\delta_{k-1}}{K}\right) > \mu_0$$
$$\frac{\sum_{s=1}^{N_{\hat{a}_{k-1}}^{\text{ee}}-1} X_{\hat{a}_{k-1},s}}{N_{\hat{a}_{k-1}}^{\text{ee}} - 1} - C \cdot U\left(N_{\hat{a}_{k-1}}^{\text{ee}} - 1, \frac{\delta_{k-1}}{K}\right) \leq \mu_0$$

Since we put the last collected sample X into Q in Line 7, we can assert $\hat{\mu}_{\hat{a}_{k-1}}$ before the start of phase k must be $\frac{\sum_{s=1}^{N_{\hat{a}_{k-1}}^{\text{ee}}-1} X_{\hat{a}_{k-1},s}}{N_{\hat{a}_{k-1}}^{\text{ee}}-1}$. Since $U\left(N_{\hat{a}_{k-1}}^{\text{ee}} - 1, \frac{\delta_k}{K}\right) > U\left(N_{\hat{a}_{k-1}}^{\text{ee}} - 1, \frac{\delta_{k-1}}{K}\right)$, we know before the start of phase k, $\text{LCB}_{\hat{a}_{k-1}}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) < \mu_0$ must hold. □

Following lemma suggests that the exploration period is able to output correct answer, given some concentration event.

Lemma 24. *For B such that $\max_{a\in B} \mu_a \geq \mu_j$ for some $\mu_j > \mu_0$, and phase index $k \geq \kappa_B^{ee}$, Algorithm 3 either takes $\hat{a}_k \in [K]$ such that $\mu_{\hat{a}_k} \geq \omega\mu_j + (1-\omega)\mu_0$, $\omega = \frac{C-1}{C+3}$ or takes $\hat{a}_k =$ Not Completed*

Proof of Lemma 24. From the assumption $k \geq \kappa_B^{\text{ee}}$, by the Lemma 14, we know

$$\hat{\mu}_a(\mathcal{H}^{\text{ee}}) - U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}) \leq \mu_a \leq \hat{\mu}_a(\mathcal{H}^{\text{ee}}) + U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}) \tag{27}$$

holds for $a \in B$ through out the phase k.

We first show that Algorithm 3 never enters Line 17. Denote $a^*(B) = \arg\max_{a\in B} \mu_a$. From the assumption, we know $\mu_{a^*(B)} > \mu_0$. Together with the definition $\text{UCB}_a^{\text{ee}}(\mathcal{H}^{ee}, \delta_k) = \hat{\mu}_a^{\text{ee}}(\mathcal{H}^{ee}) + U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K})$, we know $\text{UCB}_{a^*(B)}^{\text{ee}}(\mathcal{H}^{ee}, \delta_k) > \mu_0$ holds through out the phase k, suggesting that the condition in Line 16 never holds. Meanwhile, to quit the While loop in the Line 4, Algorithm 3 must enter one of the Lines in 17, 6, 24. Then, we can conclude $\hat{a}_k$ only takes value in $[K] \cup \{\text{Not Completed}\}$.

The remaining work is to prove if $\hat{a}_k \in [K]$, we have $\mu_{\hat{a}_k} \geq \omega\mu_j + (1-\omega)\mu_0$, $\omega = \frac{C-1}{C+3}$. For simplicity, we denote τ_k^{ee} as $\tau_k^{\text{ee}}(B)$. From the outputting rule, we have

$$\begin{aligned}
&A_{\tau_k^{\text{ee}}}^{\text{ee}} = \hat{a}_k, \hat{a}_k = \arg\max_{i\in B} \hat{\mu}_{i,N_i^{\text{ee}}(\tau_k^{\text{ee}}-1)} + U(N_i^{\text{ee}}(\tau_k^{\text{ee}} - 1), \frac{\delta_k}{K}) \\
&\hat{\mu}_{\hat{a}_k, N_{\hat{a}}^{\text{ee}}(\tau_k^{\text{ee}})} - C \cdot U(N_{\hat{a}_k}^{\text{ee}}(\tau_k^{\text{ee}}), \frac{\delta_k}{K}) > \mu_0 \\
&\hat{\mu}_{\hat{a}_k, N_{\hat{a}_k}^{\text{ee}}(\tau_k^{\text{ee}})-1} - C \cdot U(N_{\hat{a}_k}^{\text{ee}}(\tau_k^{\text{ee}}) - 1, \frac{\delta_k}{K}) < \mu_0.
\end{aligned} \tag{28}$$

where $\hat{\mu}_{i,N_i^{\text{ee}}(\tau_k^{\text{ee}}-1)}$ denotes the empirical mean value of arm i of the first $N_i^{\text{ee}}(\tau_k^{\text{ee}} - 1)$ samples. The third line is by the Lemma 23, suggesting $\text{LCB}_{\hat{a}_k}^{\text{ee}} < \mu_0$ must hold before the latest collection of the sample.

We consider two cases, $N^{\text{ee}}_{\hat{a}}(\tau^{\text{ee}}_k)=1$ or $N^{\text{ee}}_{\hat{a}}(\tau^{\text{ee}}_k)\geq 2$. If $N^{\text{ee}}_{\hat{a}}(\tau^{\text{ee}}_k)=1$, we have

$$\begin{aligned}
&\mu_0 \leq \hat{\mu}_{\hat{a}_k,1} - C\cdot U(1,\frac{\delta_k}{K})\\
\Rightarrow&\mu_0 \leq \mu_{\hat{a}_k} - (C-1)\cdot U(1,\frac{\delta_k}{K})\\
\Rightarrow&\mu_0 \leq \mu_{\hat{a}_k} - 2(C-1)\\
\Leftrightarrow&\mu_0 + 2(C-1) \leq \mu_{\hat{a}_k}.
\end{aligned}$$

The second last line is from the definition of (25), suggesting that $U(1,\frac{\delta_k}{K})\geq 2$. From the assumption $\{\mu_a\}_{a=1}^K\in[0,1]^K$ and $C+1$, we know $\omega\mu_j+(1-\omega)\mu_0=\frac{C-1}{C+3}\mu_j+(1-\frac{C-1}{C+3})\mu_0<\frac{C-1}{4}+\mu_0<\mu_0+2(C-1)$. We complete the proof.

If $N^{\text{ee}}_{\hat{a}}(\tau^{\text{ee}}_k)\geq 2$, we argue by contradiction. Assume $\mu_{\hat{a}_k}<\omega\mu_j+(1-\omega)\mu_0$. As we take $U(t,\frac{\delta}{K})=\frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 t\rceil,1\}}\log\frac{2K(\lceil\log_2 t\rceil)^2}{\delta}}}{t}$, we get

$$\begin{aligned}
&\hat{\mu}_{\hat{a}_k,N^{\text{ee}}_{\hat{a}}(\tau^{\text{ee}}_k)} - C\cdot U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k),\frac{\delta_k}{K}) > \mu_0\\
\Leftrightarrow&\hat{\mu}_{\hat{a}_k,N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)} - C\frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)} > \mu_0\\
\overset{(27)}{\Rightarrow}&\mu_{\hat{a}_k} - (C-1)\frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)} > \mu_0\\
\Rightarrow&\omega\mu_{a^*(B)}+(1-\omega)\mu_0 - (C-1)\frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)} > \mu_0\\
\Leftrightarrow&\omega(\mu_{a^*(B)}-\mu_0) > (C-1)\frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)}\\
\Leftrightarrow&\frac{2\omega(\mu_{a^*(B)}-\mu_0)}{C-1} > \frac{2\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)}.
\end{aligned}$$

By $\omega=\frac{C-1}{C+3}$, we have $\frac{(1-\omega)}{2}=\frac{1-\frac{C-1}{C+3}}{2}=\frac{4}{2(C+3)}=\frac{2\omega}{C-1}$, we have

$$\frac{(1-\omega)(\mu_{a^*(B)}-\mu_0)}{2} > \frac{2\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)}. \tag{29}$$

Notice that

$$\begin{aligned}
&\frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1}\\
\leq&\frac{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1}\frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)}\\
\leq&\frac{2\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)},
\end{aligned}$$

Together with (29), we can conclude

$$\begin{aligned}
&\hat{\mu}_{\hat{a}_k,N^{\text{ee}}_{\hat{a}}(\tau^{\text{ee}}_k)} - C\cdot U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k),\frac{\delta_k}{K}) > \mu_0\\
\Rightarrow&\frac{(1-\omega)(\mu_{a^*(B)}-\mu_0)}{2} \geq \frac{\sqrt{2\cdot 2^{\max\{\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1\rceil,1\}}\log\frac{2K(\lceil\log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1\rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1}.
\end{aligned} \tag{30}$$

On the other hands,

$$
\begin{aligned}
&\hat{\mu}_{\hat{a}_k, N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)} + U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k - 1), \frac{\delta_k}{K}) \leq \mu_{a^*(B)} \\
\overset{(27)}{\Leftarrow} &\mu_{\hat{a}_k} + 2U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k - 1), \frac{\delta_k}{K}) \leq \mu_{a^*(B)} \\
\Leftarrow &\omega\mu_{a^*(B)} + (1-\omega)\mu_0 + 2U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k - 1), \frac{\delta_k}{K}) \leq \mu_{a^*(B)} \\
\Leftrightarrow &\frac{2\sqrt{2 \cdot 2^{\max\{\lceil \log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1 \rceil, 1\}} \log \frac{2K(\lceil \log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1 \rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k) - 1} \leq (1-\omega)(\mu_{a^*(B)} - \mu_0) \\
\Leftrightarrow &\frac{\sqrt{2 \cdot 2^{\max\{\lceil \log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1 \rceil, 1\}} \log \frac{2K(\lceil \log_2 N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k)-1 \rceil)^2}{\delta}}}{N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k) - 1} \leq \frac{(1-\omega)(\mu_{a^*(B)} - \mu_0)}{2}.
\end{aligned} \tag{31}
$$

Combining (30) and (31), we can conclude

$$
\begin{aligned}
&\hat{\mu}_{\hat{a}_k, N^{\text{ee}}_{\hat{a}_k}(\tau)} - C \cdot U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k), \frac{\delta_k}{K}) > \mu_0 \\
\Rightarrow &\hat{\mu}_{\hat{a}_k, N_{\hat{a}_k}(\tau^{\text{ee}}-1)} + U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k - 1), \frac{\delta_k}{K}) \leq \mu_{a^*(B)} \\
\overset{(27)}{\Rightarrow} &\hat{\mu}_{\hat{a}_k, N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}-1)} + U(N^{\text{ee}}_{\hat{a}_k}(\tau^{\text{ee}}_k - 1), \frac{\delta_k}{K}) < \hat{\mu}_{a^*(B), N^{\text{ee}}_{a^*(B)}(\tau^{\text{ee}}-1)} + U(N^{\text{ee}}_{a^*(B)}(\tau^{\text{ee}}_k - 1), \frac{\delta_k}{K}) \\
\Rightarrow &A^{\text{ee}}_{\tau^{\text{ee}}_k} \neq \hat{a}_k.
\end{aligned}
$$

The last line suggests that under the assumption $\mu_{\hat{a}_k} < \omega\mu_j + (1-\omega)\mu_0$, we can conclude $A^{\text{ee}}_{\tau^{\text{ee}}_k} \neq \hat{a}_k$, which contradicts with (28). And we have proved$\mu_{\hat{a}} \geq \omega\mu_j + (1-\omega)\mu_0$. □

The following Lemma proves an upper bound of pulling times, for large enough phase index k that guarantees the concentration event.

Lemma 25. *For B such that $\max_{a' \in B} \mu_{a'} \geq \mu_j$ for some $\mu_j > \mu_0$, phase index $k \geq \kappa^{ee}_B$, for each arm $a \in B$, we have*

$$
\begin{aligned}
&N^{ee,B}_a\left(\tau^{ee}_k(B)\right) \\
\leq &\max\left\{ N^{ee,B}_a\left(\tau^{ee}_{\kappa^{ee}_B - 1}(B)\right), \frac{29(C+1)^2\left(\log\frac{2K}{\delta_k} + \log\log\frac{24(C+1)^2}{\max\{\Delta^2_{a,j}, \Delta^2_{a,0}\}}\right)}{\max\{\Delta^2_{a,j}, \Delta^2_{a,0}\}} \right\}
\end{aligned}
$$

Proof of Lemma 25. We first prove an intermediate conclusion, which is for each arm $a \in B$, we have

$$
\begin{aligned}
&N^{\text{ee},B}_a\left(\tau^{\text{ee}}_k(B)\right) \\
\leq &\max\left\{ N^{\text{ee},B}_a\left(\tau^{\text{ee}}_{k-1}(B)\right), \frac{29(C+1)^2\left(\log\frac{2K}{\delta_k} + \log\log\frac{24(C+1)^2}{\max\{\Delta^2_{a,j}, \Delta^2_{a,0}\}}\right)}{\max\{\Delta^2_{a,j}, \Delta^2_{a,0}\}} \right\}
\end{aligned} \tag{32}
$$

holds for for any phase index $k \geq \kappa_B^{\text{ee}}$. If (32) holds, we can use induction to conclude

$$
\begin{aligned}
&N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \\
\leq &\max\left\{N_a^{\text{ee},B}\left(\tau_{\kappa_B^{\text{ee}}-1}^{\text{ee}}(B)\right), \frac{29(C+1)^2\left(\log\frac{2K}{\delta_{k-1}}+\log\log\frac{24(C+1)^2}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right. \\
&\left.\frac{29(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{24(C+1)^2}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right\} \\
= &\max\left\{N_a^{\text{ee},B}\left(\tau_{\kappa_B^{\text{ee}}-1}^{\text{ee}}(B)\right), \frac{29(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{24(C+1)^2}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right\},
\end{aligned}
$$

which proves the Lemma. Thus, we suffice to prove the intermediate conclusion (32).

From the assumption $k \geq \kappa_B^{\text{ee}}$, we know (27) still holds. Fix the arm $a \in B$, we prove the claim by discussing the value of μ_a. We consider two cases, $\mu_a \geq \frac{\mu_j+\mu_0}{2}$ or $\mu_a < \frac{\mu_j+\mu_0}{2}$.

For arm a such that $\mu_a \geq \frac{\mu_j+\mu_0}{2}$ holds, we suffice to prove

$$
\begin{aligned}
&N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \\
\leq &\max\left\{N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right), \frac{29(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{1}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2}\right\}.
\end{aligned}
$$

We can conduct direct calculation, as

$$
\begin{aligned}
t \geq &\frac{28(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \\
\overset{\text{Lemma 36}}{\Rightarrow} &(C+1)\sqrt{\frac{4\log\frac{2K(\log_2(2t))^2}{\delta_k}}{t}} \leq \Delta_{a,0} \\
\Leftrightarrow &\mu_a-(C+1)\sqrt{\frac{4\log\frac{2K(\log_2(2t))^2}{\delta_k}}{t}} > \mu_0 \\
\Rightarrow &\mu_a-(C+1)\frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K(\lceil\log_2 t\rceil^+)^2}{\delta_k}}}{t} > \mu_0 \\
\overset{(27)}{\Rightarrow} &\hat{\mu}_{a,t}-C\cdot\frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K(\lceil\log_2 t\rceil^+)^2}{\delta_k}}}{t} > \mu_0.
\end{aligned}
$$

By the Lemma 23, we know at the start of phase k, $\text{LCB}_a^{\text{ee}} < \mu_0$, suggesting that $N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right) \leq \frac{28(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2}$. Then, we can assert

$$
N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) < \frac{28(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2}+1.
$$

If not, the algorithm will output $\hat{a}_k = a$ before $N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \geq \frac{28(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2}+1$ holds.

For arm a such that $\mu_a < \frac{\mu_j+\mu_0}{2}$ holds, we suffice to prove

$$N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \leq \max\left\{N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right), \frac{29(C+1)^2\left(\log\frac{4K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,j}^2}\right)}{\Delta_{a,j}^2}\right\}.$$

We can also conduct direct calculation, with

$$\begin{aligned} t &\geq \frac{28(C+1)^2\left(\log\frac{2K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,j}^2}\right)}{\Delta_{a,j}^2} \\ &\overset{\text{Lemma 36}}{\Rightarrow}(C+1)\sqrt{\frac{4\log\frac{2K(\log_2(2t))^2}{\delta_k}}{t}} \leq \Delta_{a,j} \\ &\overset{C\geq 1}{\Rightarrow}\mu_a + 2\sqrt{\frac{4\log\frac{2K(\log_2(2t))^2}{\delta_k}}{t}} \leq \mu_j \\ &\Rightarrow \mu_a + 2\frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K(\lceil\log_2 t\rceil^+)^2}{\delta_k}}}{t} \leq \mu_j \\ &\overset{(27)}{\Rightarrow}\hat{\mu}_{a,t} + \frac{\sqrt{2\cdot 2^{\lceil\log_2 t\rceil^+}\log\frac{2K(\lceil\log_2 t\rceil^+)^2}{\delta_k}}}{t} \leq \mu_j \end{aligned}$$

If $N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right) \geq \frac{28(C+1)^2\left(\log\frac{4K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,j}^2}\right)}{\Delta_{a,j}^2}+1$, we know $\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_k}{K}) < \mu_j \leq \mu_{a^*(B)} \leq \text{UCB}_{a^*(B)}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_k}{K})$ holds through out the phase k, and arm a will never get pulled. We can assert $N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) = N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right)$.

If $N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right) < \frac{28(C+1)^2\left(\log\frac{4K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,j}^2}\right)}{\Delta_{a,j}^2}+1$, we can assert $N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \leq \frac{29(C+1)^2\left(\log\frac{4K}{\delta_k}+\log\log\frac{24(C+1)^2}{\Delta_{a,j}^2}\right)}{\Delta_{a,j}^2}$ as the arm a never gets pulled once $\text{UCB}_a^{\text{ee}} \leq \mu_{a^*(B)} \leq \text{UCB}_{a^*(B)}^{\text{ee}}$ holds. We complete the proof. □

The next Lemma shows that when phase index k is large enough, the exploration part of the Algorithm 3 is able to output a qualified arm, for a positive instance.

Lemma 26. *For B such that $\max_{a'\in B}\mu_{a'} \geq \mu_j > \mu_0$, and phase index $k \geq \max\{\kappa_B^{ee}, \lceil\log_2\frac{112(C+1)^2}{\Delta_{j,0}^2}\rceil\}$, Algorithm 3 will take $\hat{a}_k \in [K]$ such that $\mu_{\hat{a}_k} \geq \omega\mu_j + (1-\omega)\mu_0$, $\omega = \frac{C-1}{C+3}$.*

Proof of Lemma 26. The condition $k \geq \kappa_B^{\text{ee}}$ and Lemma 24 imply once the Algorithm 3 enters the branch at the Line 5 with A_t^{ee} satisfying $\text{LCB}_{A_t^{\text{ee}}}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_0$, $\mu_{A_t^{\text{ee}}} \geq \omega\mu_j + (1-\omega)\mu_0$ must hold.

Thus, we suffice to show during the pulling process of phase k, $\text{LCB}_{A_t^{\text{ee}}}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_0$ must hold for some t^{ee}, A_t^{ee}. We prove this conclusion by contradiction. Assume this $\text{LCB}_{A_t^{\text{ee}}}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_0$ never holds during the pulling process of phase k. There are two possible cases. The first case is the Algorithm 3 fulfills the condition at the Line 16 ahead of the Line 5. The second case is the Algorithm 3 fulfills the condition at the Line 23 ahead of the Line 5. We are going to show both cases will lead to a contradiction.

For the first case, $k \geq \kappa_B^{\text{ee}}$ implies

$$\max_{a'\in B}\text{UCB}_{A_t^{\text{ee}}}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \max_{a'\in B}\mu_{a'} \geq \mu_j > \mu_0,$$

throughout the pulling process of phase k. That means the condition at the Line 16 will never hold during the the pulling process of phase k, which is a condition.

For the second case, denote a_U^* as the arm with highest UCB when $\mathcal{A}_B(\mathcal{H}^{\text{ee}}, \delta_k, k) = \emptyset$, i.e.

$$a_U^* = \arg\max_{a' \in B} \text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k).$$

Again, $k \geq \kappa_B^{\text{ee}}$ implies $\text{UCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_j > \mu_0$. By the definition of $\mathcal{A}_B(\mathcal{H}^{\text{ee}}, \delta_k, k)$ and the assumption $\mathcal{A}_B(\mathcal{H}^{\text{ee}}, \delta_k, k) = \emptyset$, we know $N_{a_U^*} > (C+1)^2 \beta_k \log \frac{4K}{\delta_k} - 1$. Notice that

$$\begin{aligned}
& k \geq \lceil \log_2 \frac{112(C+1)^2}{\Delta_{j,0}^2} \rceil \\
\Rightarrow & \beta_k \geq \frac{112(C+1)^2}{\Delta_{j,0}^2}, \log \frac{1}{\delta_k} \geq \log 3 \log_2 \frac{112(C+1)^2}{\Delta_{j,0}^2} \\
\Rightarrow & \frac{1}{2}\beta_k \log \frac{4K}{\delta_k} \geq \frac{56(C+1)^2 \log \frac{2K}{\delta_k}}{\Delta_{j,0}^2}, \frac{1}{2}\beta_k \log \frac{4K}{\delta_k} \geq \frac{56(C+1)^2 \log\log \frac{24(C+1)^2}{\Delta_{j,0}^2}}{\Delta_{j,0}^2} \\
\Rightarrow & \beta_k \log \frac{4K}{\delta_k} \geq \frac{56(C+1)^2 \left(\log \frac{2K}{\delta_k} + \log\log \frac{24(C+1)^2}{\Delta_{j,0}^2} \right)}{\Delta_{j,0}^2} \\
\Rightarrow & (C+1)^2 \beta_k \log \frac{4K}{\delta_k} - 1 \geq \frac{28(C+1)^2 \left(\log \frac{2K}{\delta_k} + \log\log \frac{24(C+1)^2}{\Delta_{j,0}^2} \right)}{\Delta_{j,0}^2}
\end{aligned}$$

The last line suggests that $N_{a_U^*} \geq \frac{28(C+1)^2 \left(\log \frac{2K}{\delta_k} + \log\log \frac{24(C+1)^2}{\Delta_{j,0}^2} \right)}{\Delta_{j,0}^2}$, which means

$$(C+1)\sqrt{\frac{4 \log \frac{2K(\log_2 2N_{a_U^*})^2}{\delta_k}}{N_{a_U^*}}} < \Delta_{j,0}, \tag{33}$$

by the Lemma 36. Recall the definition

$$\begin{aligned}
\text{UCB}_a^{\text{ee}}(\mathcal{H}^{ee}, \delta_k) &= \hat{\mu}_a^{\text{ee}}(\mathcal{H}^{ee}) + U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}), \forall a \\
\text{LCB}_a^{\text{ee}}(\mathcal{H}^{ee}, \delta_k) &= \hat{\mu}_a^{\text{ee}}(\mathcal{H}^{ee}) - C \cdot U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}), \forall a,
\end{aligned}$$

We know $\text{LCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) = \text{UCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) - (C+1)U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K})$. We can conclude

$$\begin{aligned}
& \text{UCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_j \\
\Rightarrow & \text{LCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_j - (C+1)U(N_{a_U^*}, \frac{\delta_k}{K}) \\
U(N_{a_U^*}, \tfrac{\delta_k}{K}) \leq \sqrt{\frac{4 \log \frac{2K(\log_2 2N_{a_U^*})^2}{\delta_k}}{N_{a_U^*}}} \Rightarrow & \text{LCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \geq \mu_j - (C+1)\sqrt{\frac{4 \log \frac{2K(\log_2 2N_{a_U^*})^2}{\delta_k}}{N_{a_U^*}}} \\
\overset{(33)}{\Rightarrow} & \text{LCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) > \mu_j - \Delta_{j,0} = \mu_0.
\end{aligned}$$

Meanwhile, Lemma 23 implies $\text{LCB}_{a_U^*}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) < \mu_0$ at the start of phase k. Notice that $\text{LCB}_{a_U^*}^{\text{ee}}$ only changes if we take $A_t^{\text{ee}} = a_U^*$. We can conclude there must be a round during phase k, taking $A_t^{\text{ee}} = a_U^*$ and after executing the Line 27 to 31 of Algorithm 3, the updated $\text{LCB}_{a_U^*}^{\text{ee}}$ becomes not smaller than μ_0. If this happens, the Algorithm 3 will first fulfill the condition at the Line 5, which leads to a contradiction to the second case. □

The next Lemma shows the exploitation part Algorithm 3 is able to output a qualified arm, if the input is a qualified arm.

Lemma 27. *For B such that $\max_{a' \in B} \mu_{a'} \geq \mu_j$, and phase index $k \geq \max\{\kappa_B^{et}, \lceil \log_2 \frac{448(C+3)^2}{(C-1)^2 \Delta_{j,0}^2} \rceil\}$. If Algorithm 3 enters Line 6 with $\mu_{\hat{a}_k} \geq \omega \mu_j + (1-\omega)\mu_0$, $\omega = \frac{C-1}{C+3}$, Algorithm 3 will terminate and output $\hat{a}_k$.*

Proof of Lemma 27. From the condition $k \geq \kappa_B^{\text{et}}$, by the Lemma 17, we know

$$\hat{\mu}_a^{\text{et}}(\mathcal{H}^{\text{et}}) - U(N_a^{\text{et}}(\mathcal{H}^{et}), \frac{\delta}{K\alpha_k}) \leq \mu_a \leq \hat{\mu}_a^{\text{et}}(\mathcal{H}^{\text{et}}) + U(N_a^{\text{et}}(\mathcal{H}^{et}), \frac{\delta}{K\alpha_k})$$

holds for $a \in B$ through out the phase k.

By similar calculation in the proof of Lemma 25, we know

$$\begin{aligned}
t &\geq \frac{112\left(\log \frac{2K\alpha_k}{\delta} + \log\log \frac{96}{\omega^2 \Delta_{j,0}^2}\right)}{\omega^2 \Delta_{j,0}^2} \\
\Rightarrow t &\geq \frac{112\left(\log \frac{2K\alpha_k}{\delta} + \log\log \frac{96}{\Delta_{\hat{a}_k,0}^2}\right)}{\Delta_{\hat{a}_k,0}^2} \\
\Rightarrow \hat{\mu}_{\hat{a}_k,t}^{\text{et}} &- \frac{\sqrt{2 \cdot 2^{\lceil \log_2 t \rceil^+} \log \frac{2K\alpha_k(\lceil \log_2 t \rceil^+)^2}{\delta}}}{t} > \mu_0.
\end{aligned}$$

Here $\hat{\mu}_{\hat{a}_k,t}^{\text{et}}$ denotes the empirical mean of first t samples collected in the exploitation period. Notice that

$$\begin{aligned}
& k \geq \lceil \log_2 \frac{448(C+3)^2}{(C-1)^2 \Delta_{j,0}^2} \rceil \\
\Rightarrow & \beta_k \geq \frac{448}{\Delta_{j,0}^2}, \log \alpha_k \geq \log \frac{96(C+3)^2}{(C-1)^2 \Delta_{j,0}^2} \log 5 \\
\Rightarrow & \beta_k \log \frac{4K\alpha_k}{\delta} \geq \frac{448 \log \frac{2K\alpha_k}{\delta}}{\Delta_{j,0}^2}, \beta_k \log \frac{4K\alpha_k}{\delta} \geq \frac{448 \log\log \frac{96}{\omega^2 \Delta_{j,0}^2}}{\Delta_{j,0}^2} \\
\Rightarrow & \beta_k \log \frac{4K\alpha_k}{\delta} \geq \frac{224\left(\log \frac{2K\alpha_k}{\delta} + \log\log \frac{96}{\omega^2 \Delta_{j,0}^2}\right)}{\Delta_{j,0}^2} \\
\Rightarrow & \frac{(C+3)^2}{(C-1)^2} \beta_k \log \frac{4K\alpha_k}{\delta} - 1 \geq \frac{112\left(\log \frac{2K\alpha_k}{\delta} + \log\log \frac{96}{\omega^2 \Delta_{j,0}^2}\right)}{\omega^2 \Delta_{j,0}^2}.
\end{aligned}$$

The last line suggests that Algorithm 3 would output $\hat{a}_k$ before meeting the stop condition $N_{\hat{a}}^{\text{et}} > \frac{(C+3)^2}{(C-1)^2} \beta_k \log \frac{4K\alpha_k}{\delta} - 1$ in Line 8. □

The next Lemma shows the overall Algorithm 3 is able to output a qualified arm, and terminate after a large enough phase k.

Lemma 28. *For B such that $\max_{a' \in B} \mu_{a'} \geq \mu_j$, Algorithm 3 must terminate no later than the end of phase $\max\{\kappa_B^{ee}, \kappa_B^{et}, \lceil \log_2 \frac{448(C+3)^2}{(C-1)^2 \Delta_{j,0}^2} \rceil\}$.*

Proof of Lemma 28. This is a direct conclusion from Lemma 26, 27. □

The following lemmas concern about the property of algorithm when ν is negative. The logic is the roughly the same as the positive part. First, we are going to show when the concentration event holds, Algorithm 3 will not output a wrong answer.

Lemma 29. *Consider the largest bracket B, i.e. $B = [K]$. If $\max_{a' \in B} \mu_{a'} = \max_{a' \in [K]} \mu_{a'} < \mu_0$, then for phase index $k \geq \kappa_B^{ee}$, Algorithm 3 either takes $\hat{a}_k =$ None or takes $\hat{a}_k =$ Not Completed*

Proof of Lemma 29. From the assumption $k \geq \kappa_B^{\text{ee}}$, by the Lemma 14, we know

$$\hat{\mu}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}) - U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}) \leq \mu_a \leq \hat{\mu}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}) + U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}) \tag{34}$$

holds for $a \in B$ through out the phase k.

The analysis is similar to the Lemma 24. To quit the While loop in the Line 4, Algorithm 3 must enter one of the Lines in 17, 6, 24. By the assumption $\mu_a < \mu_0$ holds for all $a \in [K]$ and (29), we know

$$\begin{aligned}
&\text{LCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \delta_k) \\
=&\hat{\mu}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}) - C \cdot U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}) \\
\overset{C>1}{\leq}& \hat{\mu}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}) - U(N_a^{\text{ee}}(\mathcal{H}^{ee}), \frac{\delta_k}{K}) \\
\leq&\mu_a \\
<&\mu_0,
\end{aligned}$$

holds through out the phase k. The last line also suggests that the condition in Line 5 will never be satisfied in phase k. Thus, we can conclude $\hat{a}_k = \textsf{None}$ or $\textsf{Not Completed}$. □

The following lemma suggests that when the concentration event holds, we can find an upper bound of the pulling times.

Lemma 30. *For B such that $|B| = K$ and $\max_{a' \in B} \mu_{a'} < \mu_0$, phase index $k \geq \kappa_B^{ee}$, for each arm $a \in B$, we have*

$$\begin{aligned}
&N_a^{ee,B}\left(\tau_k^{ee}(B)\right) \\
\leq &\max\left\{ N_a^{ee,B}\left(\tau_{\kappa_B^{ee}-1}^{ee}(B)\right), \cdot\frac{113\left(\log\frac{4K}{\delta_k} + \log\log\frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \right\}
\end{aligned}$$

Proof of Lemma 30. Similar to the idea of Lemma 25, we first turn to prove an intermediate conclusion

$$\begin{aligned}
&N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \\
\leq &\max\left\{ N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right), \frac{113\left(\log\frac{4K}{\delta_k} + \log\log\frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \right\}
\end{aligned} \tag{35}$$

holds for all $k \geq \kappa_B^{\text{ee}}$ and $a \in B$. If the above conclusion holds, we can use induction to conclude

$$\begin{aligned}
&N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \\
\leq &\max\left\{ N_a^{\text{ee},B}\left(\tau_{\kappa_B^{\text{ee}}-1}^{\text{ee}}(B)\right), \frac{113\left(\log\frac{4K}{\delta_{k-1}} + \log\log\frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2}, \right. \\
&\left. \frac{113\left(\log\frac{4K}{\delta_k} + \log\log\frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \right\} \\
= &\max\left\{ N_a^{\text{ee},B}\left(\tau_{\kappa_B^{\text{ee}}-1}^{\text{ee}}(B)\right), \frac{113\left(\log\frac{4K}{\delta_k} + \log\log\frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \right\}.
\end{aligned}$$

which proves the Lemma. Thus, we suffice to prove the intermediate conclusion (35).

We can conduct direct calculation for each arm a.

$$
\begin{aligned}
t &\geq \frac{28 \cdot 2^2 \left(\log \frac{2K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \\
&\overset{\text{Lemma 36}}{\Rightarrow} 2\sqrt{\frac{4\log \frac{2K(\log_2(2t))^2}{\delta_k}}{t}} \leq \Delta_{a,0} \\
&\Leftrightarrow \mu_a + 2\sqrt{\frac{4\log \frac{2K(\log_2(2t))^2}{\delta_k}}{t}} \leq \mu_0 \\
&\Rightarrow \mu_a + 2 \cdot \frac{\sqrt{2 \cdot 2^{\lceil \log_2 t \rceil^+} \log \frac{2K(\lceil \log_2 t \rceil^+)^2}{\delta_k}}}{t} < \mu_0 \\
&\overset{(27)}{\Rightarrow} \hat{\mu}_{a,t} + \frac{\sqrt{2 \cdot 2^{\lceil \log_2 t \rceil^+} \log \frac{2K(\lceil \log_2 t \rceil^+)^2}{\delta_k}}}{t} < \mu_0.
\end{aligned}
$$

The last line is by the assumption $k \geq \kappa_B^{\text{ee}}$, which means (27) still holds.

If $N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right) \geq \frac{112\left(\log \frac{4K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} + 1$, we know $\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_k}{K}) < \mu_0$ holds at the start of phase k. If $a = \arg\max_{a' \in B} \text{UCB}_{a'}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_k}{K})$ holds during the phase k, Algorithm 3 will turn to phase $k+1$ as the condition in the Line 16 holds. Thus, we can conclude Algorithm 3 will never pull arm a during the phase k. That means if $N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right) \geq \frac{112\left(\log \frac{4K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} + 1$, $N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) = N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right)$.

If $N_a^{\text{ee},B}\left(\tau_{k-1}^{\text{ee}}(B)\right) < \frac{112\left(\log \frac{4K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} + 1$. We can also assert $\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_k}{K}) \leq \mu_0$ will hold if $N_a^{\text{ee},B} \geq \frac{112\left(\log \frac{4K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} + 1$ holds in the futuer rounds. Similar to the above argument, if $\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_k}{K}) \leq \mu_0$, $a = \arg\max_{a'} \text{UCB}_{a'}^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_k}{K})$ both holds, the algorithm would turn to phase $k+1$ or terminate because of meeting the condition in the Line 16. That means once $N_a^{\text{ee},B} \geq \frac{112\left(\log \frac{4K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} + 1$ holds, Algorithm 3 will never pull arm a in the remaining phase k. We can assert $N_a^{\text{ee},B}\left(\tau_k^{\text{ee}}(B)\right) \leq \frac{113\left(\log \frac{4K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2}$. □

Based on the above Lemmas, we can conclude Algorithm 3 must terminate after a large enough phase k.

Lemma 31. *For B such that $|B| = K$ and $\max_{a' \in B} \mu_{a'} < \mu_0$, Algorithm 3 must terminate no later than the end of phase $\max\{\kappa_B^{ee}, \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil, \lceil \log_3 \frac{3}{\delta} \rceil\}$.*

Proof of Lemma 31. From the calculation in the proof of Lemma 30, we know

$$
\begin{aligned}
t &\geq \frac{112\left(\log \frac{2K}{\delta_k} + \log\log \frac{96}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \\
&\Rightarrow \hat{\mu}_{a,t} + \frac{\sqrt{2 \cdot 2^{\lceil \log_2 t \rceil^+} \log \frac{2K(\lceil \log_2 t \rceil^+)^2}{\delta_k}}}{t} > \mu_0
\end{aligned}
$$

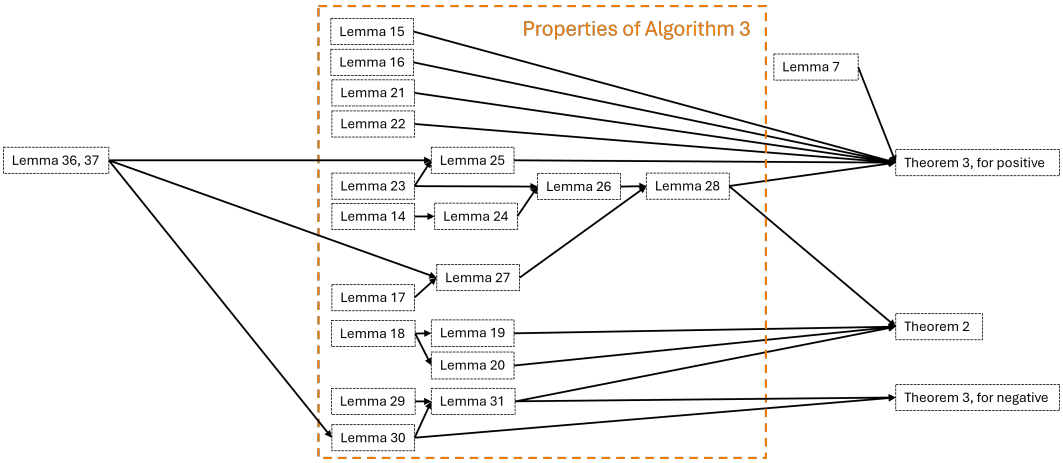

Figure 1: Overall Idea of Full Proof of Theorem 2 and 3

for arm $a \in B = [K]$. Notice that for any $a \in [K]$, we have

$$
\begin{aligned}
& k \geq \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil \\
\overset{\mu_0 > \mu_1 \geq \mu_a}{\Rightarrow} & \beta_k \geq \frac{226}{\Delta_{a,0}^2}, \log \frac{1}{\delta_k} \geq \frac{96}{\Delta_{a,0}^2} \log 3 \\
\Rightarrow & \beta_k \log \frac{4K}{\delta_k} \geq \frac{226 \log \frac{2K}{\delta_k}}{\Delta_{a,0}^2}, \beta_k \log \frac{4K}{\delta_k} \geq \frac{226 \log \log \frac{96}{\Delta_{a,0}^2}}{\Delta_{a,0}^2} \\
\Rightarrow & \beta_k \log \frac{4K}{\delta_k} \geq \frac{113 \left(\log \frac{2K}{\delta_k} + \log \log \frac{96}{\Delta_{a,0}^2} \right)}{\Delta_{a,0}^2} \\
\overset{C \geq 1}{\Rightarrow} & (C+1)^2 \beta_k \log \frac{4K}{\delta_k} - 1 \geq \frac{113 \left(\log \frac{2K}{\delta_k} + \log \log \frac{96}{\Delta_{a,0}^2} \right)}{\Delta_{a,0}^2}.
\end{aligned}
$$

That means at the end of phase $\max\{\kappa_B^{\text{ee}}, \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil, \lceil \log_3 \frac{3}{\delta} \rceil\}$, we must have

$$
\text{UCB}_a^{\text{ee}}(\mathcal{H}^{\text{ee}}, \frac{\delta_{\max\{\kappa_B^{\text{ee}}, \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil, \log_3 \frac{3}{\delta}\}}}{K}) \leq \mu_0
$$

holds for all $a \in [K]$. Also, not hard to see $\delta_{\max\{\kappa_B^{\text{ee}}, \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil, \lceil \log_3 \frac{3}{\delta} \rceil\}} \leq \delta_{\lceil \log_3 \frac{3}{\delta} \rceil} < \frac{\delta}{3}$. We can conclude Algorithm 3 must terminate by the condition in Line 17, 18. □

C.3 Sketch Proof of Theorem 2, 3

The full proof of Theorem 2, 3 requires multiple lemmas to support.

Sketch Proof of Theorem 2. The first step is prove $\Pr(\tau(B_{\lceil \log_2 K \rceil + 1}) = +\infty) = 0$, whose main idea is very similar to the proof of Theorem 5.2 in Li and Cheung (2025). As in Li and Cheung (2025), denote $\kappa_b^{\text{ee}}, \kappa_b^{\text{et}}$ as the minimum phase indices after which the exploration and exploitation concentration events of copy b hold, we can prove $\Pr(\tau(B_{\lceil \log_2 K \rceil + 1}) = +\infty) \leq \Pr(\max\{\kappa_{\lceil \log_2 K \rceil + 1}^{\text{ee}}, \kappa_{\lceil \log_2 K \rceil + 1}^{\text{et}}\} = +\infty) = 0$.

Given the first step, we are able to conclude $\Pr(\tau < +\infty) = 1$, by the fact that $\tau \leq (\lceil \log_2 K \rceil + 1)\tau(B_b)$ holds for all $b \in [\lceil \log_2 K \rceil + 1]$. Then, to prove Algorithm 1 is δ-PAC, we suffice to show $\Pr_\nu(\hat{a} = \text{None} \text{ or } \{\hat{a} \in [K], \mu_{\hat{a}} < \mu_0\}) < \delta$ for a positive instance ν, and $\Pr_\nu(\hat{a} \in [K]) < \delta$ for a negative instance ν. The two cases are proved similarly. In Line 3 of Algorithm 1, we assign

$\delta/(\lceil \log_2 K \rceil + 1)$ to each copy. We are able to prove $\Pr(\text{alg}_b \text{ outputs incorrect } \hat{a}) < \frac{O(1)\delta}{\lceil \log_2 K \rceil + 1}$ holds for all $b \in [\lceil \log_2 K \rceil + 1]$. A union bound over all copies completes the proof. □

Sketch Proof of Theorem 3. We consider negative and positive instances separately.

If the instance ν is negative, we focus on analyzing $\tau(B_{\lceil \log_2 K \rceil + 1})$, which is very similar to the negative case of Theorem 5.3 in Li and Cheung (2025). By the fact that $\mathbb{E}\tau \leq [\lceil \log_2 K \rceil + 1]\mathbb{E}\tau(B_{\lceil \log_2 K \rceil + 1})$, we complete the proof.

If the instance ν is positive, denote $\mu_1 \geq \cdots \geq \mu_m > \mu_0 \geq \mu_{m+1} \geq \cdots \geq \mu_K$. For $j \in [m]$ and $a \in [K]$, define $\tilde{T}_j(a) = \frac{\log(4K) + \log_2 \frac{1}{\Delta_{j,0}^2}}{\max\{\Delta_{a,j}^2, \Delta_{a,0}^2\}}$, $\tilde{T}_j(a;k) = \frac{\log \frac{4K}{\delta_k} + \log_2 \frac{1}{\Delta_{j,0}^2}}{\max\{\Delta_{a,j}^2, \Delta_{a,0}^2\}}$. We show that for every $j \in [m]$, $\mathbb{E}\tau \leq O(\log K) \cdot O\left(\frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{\log^3 K}{j} \sum_{a=1}^K \tilde{T}_j(a)\right)$. In addition, the constant hidden by the Big O notation is independent of j. For any given $j \in [m]$, we introduce $b_j = \min\{b \in [\lceil \log_2 K \rceil + 1] : B_b \cap [j] \neq \emptyset\}$, the smallest bracket index that contains at least one arm among the top-j arms. As we adopt a random permutation to create brackets, b_j is a random variable, and we have $\mathbb{E}\tau = \sum_{b=1}^{\lceil \log_2 K \rceil + 1} \mathbb{E}\tau \mathbb{1}(b_j = b)$. Using $\tau \leq (\lceil \log_2 K \rceil + 1)\tau(B_b)$ for each $b \in [\lceil \log_2 K \rceil + 1]$, it suffices to bound $O(\log K)\, \mathbb{E}_\nu[\tau(B_b)\mathbb{1}(b_j = b)]$.

Let $\tau^{\text{ee}}(B_b), \tau^{\text{et}}(B_b)$ denote the total length of exploration and exploitation periods of algorithm copy alg_b respectively. For any arm $a \in B_b$, denote N_a^{ee,B_b} as the pulling times of arm a in the exploration period. It is evident that $\tau(B_b) = \tau^{\text{ee}}(B_b) + \tau^{\text{et}}(B_b)$. We assert that

$$\mathbb{E}\tau^{\text{et}}(B_b)\mathbb{1}(b_j = b) \leq O\left(\frac{(C+3)^2}{(C-1)^2} \frac{\log \frac{4K}{\delta} + \log_2 \frac{(C+3)^2}{(C-1)^2 \Delta_{j,0}^2}}{\Delta_{j,0}^2}\right) \Pr(b_j = b), \tag{36}$$

$$\mathbb{E}\tau^{\text{ee}}(B_b)\mathbb{1}(b_j = b) \leq \sum_{a=1}^K O(1)\tilde{T}_j(a) \min\{\frac{|B_b|}{K}, \Pr(b_j = b)\} + O(\log K)|B_b| \Pr(b_j = b). \tag{37}$$

Given (36)(37), we can sum up (36)(37) over $b = 1, 2, \cdots, \lceil \log_2 K \rceil + 1$. Lemma 7(section C.4), 33, 34(section D) provides upper bounds for $\sum_{b=1}^{\lceil \log_2 K \rceil + 1} |B_b| \Pr(b_j = b)$ and $\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \min\{\frac{|B_b|}{K}, \Pr(b_j = b)\}$, which complete the proof of Theorem 3. The remaining sketch proof reduces to (36)(37) for each $b \in [\lceil \log_2 K \rceil + 1]$.

To prove (36), we first show the algorithm must terminate no later than the end of phase $L_b^{\text{pos}} = \max\{\kappa_b^{\text{ee}}, \kappa_b^{\text{et}}, \lceil \log_2 \frac{O(1)(C+3)^2}{(C-1)^2 \Delta_{j,0}^2} \rceil\}$, if $b_j = b$. Notice that $\beta_k \log \frac{4K(\lceil \log_2 K \rceil + 1)\alpha_k}{\delta}|_{k = \lceil \log_2 \frac{O(1)(C+3)^2}{(C-1)^2 \Delta_{j,0}^2} \rceil} = O\left(\frac{(C+3)^2}{(C-1)^2} \frac{\log \frac{4K}{\delta} + \log_2 \frac{(C+3)^2}{(C-1)^2 \Delta_{j,0}^2}}{\Delta_{j,0}^2}\right)$, and Lemma 18 shows $\Pr(\kappa_b^{\text{ee}} \geq k) \leq \frac{\pi^2 \delta_{k-1}}{6}$, $\Pr(\kappa_b^{\text{et}} \geq k) \leq \frac{\pi^2 \delta}{6\alpha_{k-1}}$, we can take expectation on both sides of $\tau^{\text{et}}(B_b)\mathbb{1}(b_j = b) \leq \frac{(C+3)^2}{(C-1)^2} \beta_k \log \frac{4K\alpha_k}{\delta}|_{k = L_b^{\text{pos}}} \mathbb{1}(b_j = b)$ to complete the proof of (36).

To prove (37), we first show for each $a \in B_b$, we have

$$N_a^{\text{ee},B_b}\mathbb{1}(b_j = b) \leq O(\tilde{T}_j(a; L_b^{\text{pos}}))(C+1)^2 \mathbb{1}(b_j = b). \tag{38}$$

The intuition is as follows. Given the fact that $j \in B_b$, once the pulling times of arm a is of order $O(1)\tilde{T}_j(a; L_b^{\text{pos}})(C+1)^2$, copy alg_b can assert(with some probability, see Lemma 18, 25) either $\text{UCB}_a < \mu_j$ or $\text{LCB}_a > \mu_0$. In both cases, alg_b stops pulling arm a in the exploration period. By the facts that elements in B_{b_j} are independent of $\kappa_{b_j}^{\text{ee}}, \kappa_{b_j}^{\text{et}}$, we can sum up (38) over $a \in B_b$ and derive

$$\mathbb{E}\tau^{\text{ee}}(B_b)\mathbb{1}(b_j = b) \leq \sum_{a=1}^K O(1)\tilde{T}_j(a) \Pr(a \in B_b, b_j = b) + O(\log K)|B_b| \Pr(b_j = b).$$

Since $\Pr(a \in B_b, b_j = b) \leq \min\{\Pr(a \in B_b), \Pr(b_j = b)\}$, and $\Pr(a \in B_b) = \frac{|B_b|}{K}$, (37) is shown. □

C.4 Proof of Theorem 2 and 3

In this subsection, We present the rigorous proof of Theorem 2 and 3 and the required lemmas. Before illustrating the proof of Theorem 2 and 3, we first prove Lemma 7 for the random permutation, Line 2 of Algorithm 1.

Lemma (restatement of Lemma 7)**.** *Given a qualified arm index $j \in [m]$, denote bracket index $b_j = \min\{b \in [\lceil \log_2 K \rceil + 1] : B_b \cap [j] \neq \emptyset\}$. We have*

$$\Pr(b_j \geq \tilde{b}) \leq \frac{1}{\exp\left(\frac{j|B_{\tilde{b}-1}|}{K}\right)}, \forall \tilde{b} \in [\lceil \log_2 K \rceil + 1]$$

Here we denote $B_0 = \emptyset$ and $|B_0| = 0$. Since we take $|B_{\tilde{b}}| = \min\{2^{\tilde{b}-1}, K\}$ for $\tilde{b} \in [\lceil \log_2 K \rceil + 1]$, we can rewrite the conclusion as

$$\Pr(b_j \geq \tilde{b}) \leq \frac{1}{\exp\left(\frac{j \cdot \lfloor 2^{\tilde{b}-2} \rfloor}{K}\right)}, \forall \tilde{b} \in [\lceil \log_2 K \rceil + 1].$$

Proof of Lemma 7. If $\tilde{b} = 1$, the conclusion is $\Pr(b_j \geq 1) \leq 1$, which is obviously true. We focus on the case that $2 \leq \tilde{b} \leq \lceil \log_2 K \rceil + 1$. We have

$$\begin{aligned}\Pr(b_j \geq \tilde{b}) &= \Pr\left(\{\sigma(i)\}_{i=1}^{\min\{2^{\tilde{b}-1-1}, K\}} \cap [j] = \emptyset\right) \\ &= \Pr\left(\{\sigma(i)\}_{i=1}^{2^{\tilde{b}-2}} \cap [j] = \emptyset\right).\end{aligned}$$

The second line is from the fact that $\tilde{b} \in [\lceil \log_2 K \rceil + 1]$, which concludes $2^{\tilde{b}-2} \leq 2^{\lceil \log_2 K \rceil - 1} \leq 2^{\log_2 K} = K$. We can conclude

$$\begin{aligned}&\Pr\left(\{\sigma(i)\}_{i=1}^{2^{\tilde{b}-2}} \cap [j] = \emptyset\right) \\ =& \prod_{i=1}^{2^{\tilde{b}-2}} \frac{K-j-i+1}{K-i+1} \\ \leq& (\frac{K-j}{K})^{2^{\tilde{b}-2}} \\ =& \left(\left(1 - \frac{j}{K}\right)^{\frac{K}{j}}\right)^{\frac{j \cdot 2^{\tilde{b}-2}}{K}} \\ \leq& \exp\left(-\frac{j \cdot 2^{\tilde{b}-2}}{K}\right).\end{aligned}$$

□

Given the above preparation, we are ready to prove the Theorem 2 and 3. We first prove Theorem 2.

Proof of Theorem 2. We first show $\Pr_\nu(\tau < +\infty) = 1$, by applying Algorithm 1. This conclusion holds no matter ν is positive or negative. Consider the bracket $\lceil \log_2 K \rceil + 1$, which is indeed $[K]$. From Lemma 31 and Lemma 28, we know

$$\begin{aligned}&\Pr_\nu(\tau = +\infty) \\ \leq& \Pr_\nu(\tau(B_{\lceil \log_2 K \rceil + 1}) = +\infty) \\ \leq& \Pr_\nu(\kappa^{\text{ee}}_{\lceil \log_2 K \rceil + 1} = +\infty) + \Pr_\nu(\kappa^{\text{et}}_{\lceil \log_2 K \rceil + 1} = +\infty)\end{aligned}$$

holds for any instance ν. By the Lemma 18, we know $\Pr_\nu(\kappa^{\text{ee}}_{\lceil \log_2 K \rceil + 1} = +\infty) = \Pr_\nu(\kappa^{\text{et}}_{\lceil \log_2 K \rceil + 1} = +\infty) = 0$. That means $\Pr_\nu(\tau = +\infty) = 0$, further $\Pr_\nu(\tau < +\infty) = 1$.

This suggest that $\hat{a}$ is well defined with probability 1, and we can assert $\Pr_\nu(\hat{a} \in [K] \cup \{\textsf{None}\}) = 1$ if ν is positive or negative.

Then, we suffice to show $\Pr_\nu(\mu_{\hat{a}} \le \mu_0 \text{ or } \mu_{\hat{a}} = \textsf{None}) < \delta$ for a positive instance ν and $\Pr_\nu(\hat{a} \in [K]) < \delta$ for a negative instance ν. If ν is positive, we get

$$
\begin{aligned}
&\Pr_\nu(\mu_{\hat{a}} \le \mu_0 \text{ or } \hat{a} = \textsf{None}) \\
\le &\Pr_\nu(\mu_{\hat{a}} \le \mu_0) + \Pr_\nu(\hat{a} = \textsf{None}) \\
\le &\Pr_\nu(\text{alg copy } \lceil \log_2 K \rceil + 1 \text{ outputs } \hat{a} = \textsf{None}) + \\
&\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \Pr_\nu(\text{alg copy } b \text{ outputs } \hat{a}, s.t. \mu_{\hat{a}} \le \mu_0) \\
\le &\frac{\pi^2}{6}\frac{\delta}{3} + \sum_{b=1}^{\lceil \log_2 K \rceil + 1} \frac{\pi^2}{6 \cdot 5} \frac{\delta}{\lceil \log_2 K \rceil + 1} \\
\le &\delta.
\end{aligned}
$$

The second last line is by Lemma 19 and 20.

If ν is negative, we get

$$
\begin{aligned}
&\Pr_\nu(\hat{a} \in [K]) \\
\le &\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \Pr_\nu(\text{alg copy } b \text{ outputs } \hat{a}, s.t. \mu_{\hat{a}} \le \mu_0) \\
\le &\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \frac{\pi^2}{6 \cdot 5} \frac{\delta}{\lceil \log_2 K \rceil + 1} \\
< &\delta.
\end{aligned}
$$

The second last line is by Lemma 19. □

To prove Theorem 3, we split the proof into two cases, ν is positive or negative.

Proof of Theorem 3, when ν is positive. Fix arbitrary $j \in [m]$ and follow the notation in Lemma 7. Denote $b_j = \min\{b \in [\lceil \log_2 K \rceil + 1] : B_b \cap [j] \neq \emptyset\}$. We have $\tau = \sum_{b=1}^{\lceil \log_2 K \rceil + 1} \tau \mathbb{1}(b_j = b)$. Then, we can turn to find the upper bound for each $\tau \mathbb{1}(b_j = b)$.

From the algorithm design, we know

$$
\begin{aligned}
\tau \mathbb{1}(b_j = b) \le &(\lceil \log_2 K \rceil + 1) \tau(B_b) \mathbb{1}(b_j = b) \\
\le &(\lceil \log_2 K \rceil + 1) \Big(\tau^{\text{ee}}(B_b) \mathbb{1}(b_j = b) + \tau^{\text{et}}(B_b) \mathbb{1}(b_j = b) \Big).
\end{aligned}
$$

Then, we are going to derive an upper bound for $\tau^{\text{ee}}(B_b)\mathbb{1}(b_j = b)$ and $\tau^{\text{et}}(B_b)\mathbb{1}(b_j = b)$ separately.

For $\tau^{\text{ee}}(B_b)\mathbb{1}(b_j=b)$, we have

$$
\begin{aligned}
&\tau^{\text{ee}}(B_b)\mathbb{1}(b_j=b)\\
\overset{\text{Lemma 15}}{\leq}\ &\mathbb{1}(b_j=b)\Big(|\mathcal{H}^{\text{ee}}|+|Q|\Big)\\
\overset{\text{Lemma 16,25,28}}{\leq}\ &\mathbb{1}(b_j=b)\Bigg(|B_b|+\sum_{a\in B_b}N_a^{\text{ee},B_b}\Big(\tau^{\text{ee}}_{\kappa_b^{\text{ee}}-1}(B_b)\Big)+\\
&\qquad\frac{29(C+1)^2\left(\log\frac{2K}{\delta_{\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\lceil\log_2\frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil\}}}+\log\log\frac{24(C+1)^2}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\Bigg)\\
\overset{\text{Lemma 21}}{\leq}\ &\underbrace{\mathbb{1}(b_j=b)(C+1)^2|B_b|\left(\beta_{\kappa_b^{\text{ee}}-1}\log\frac{4K}{\delta_{\kappa_b^{\text{ee}}-1}}+1\right)}_{\dagger}+\\
&\underbrace{\sum_{a=1}^{K}\frac{29(C+1)^2\left(\log\frac{4K\log\frac{24(C+1)^2}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}}{\delta_{\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\lceil\log_2\frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil\}}}\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\mathbb{1}(b_j=b,a\in B_b)}_{\ddagger}.
\end{aligned}
$$

For $\tau^{\text{et}}(B_b)\mathbb{1}(b_j=b)$, we have

$$
\begin{aligned}
&\tau^{\text{et}}(B_b)\mathbb{1}(b_j=b)\\
\overset{\text{Lemma 22}}{\leq}\ &\mathbb{1}(b_j=b)\sum_{k=1}^{\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\lceil\log_2\frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil\}}\frac{(C+3)^2}{(C-1)^2}\beta_k\log\frac{4K(\lceil\log_2 K\rceil+1)\alpha_k}{\delta}\\
&\leq\underbrace{\frac{2(C+3)^2}{(C-1)^2}\mathbb{1}(b_j=b)\beta_k\log\frac{4K(\lceil\log_2 K\rceil+1)\alpha_k}{\delta}\Big|_{k=\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\lceil\log_2\frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil\}}}_{\diamondsuit}.
\end{aligned}
$$

The last line is by the fact that $\frac{\beta_k\log\frac{4K(\lceil\log_2 K\rceil+1)\alpha_k}{\delta}}{\beta_{k-1}\log\frac{4K(\lceil\log_2 K\rceil+1)\alpha_{k-1}}{\delta}}\geq 2$.

Then, to bound $\mathbb{E}\tau^{\text{ee}}(B_b)\mathbb{1}(b_j=b)$ and $\mathbb{E}\tau^{\text{et}}(B_b)\mathbb{1}(b_j=b)$, we suffice to consider the upper bound of the expectation of $\dagger,\ddagger,\diamondsuit$. For $\dagger$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}(b_j=b)|B_b|\left(\beta_{\kappa_b^{\text{ee}}-1}\log\frac{4K}{\delta_{\kappa_b^{\text{ee}}-1}}+1\right)\right]\\
\leq&\Pr(b_j=b)|B_b|\mathbb{E}\left[\left(\beta_{\kappa_b^{\text{ee}}-1}\log\frac{4K}{\delta_{\kappa_b^{\text{ee}}-1}}+1\right)\right]\\
\leq&\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)}\min\{2^{b-1},K\}\left(1+\sum_{k=1}^{+\infty}2^{k-1}\log(4K\cdot 3^{k-1})\frac{\pi^2}{6}\frac{1}{3^{k-1}}\right)\\
\leq&O(\log K)\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)}\min\{2^{b-1},K\}.
\end{aligned}
$$

The second line is by the fact that κ_b^{ee} is independent with b_j. The third line is by the Lemma 18, 7. For the term ‡, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{a=1}^{K}\frac{\log\frac{4K}{\delta_{\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\lceil\log_2\frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil\}}}+\log\log\frac{1}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\mathbb{1}(b_j=b)\mathbb{1}(a\in B_b)\right]\\
\leq&\sum_{a=1}^{K}\frac{\log\log\frac{1}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\Pr(a\in B_b,b_j=b)+\\
&\sum_{a=1}^{K}\mathbb{E}\left[\frac{\log(4K)+(\kappa_b^{\text{ee}}+\kappa_b^{\text{et}}+\lceil\log_2\frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil)\log 3}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right]\Pr(a\in B_b,b_j=b)\\
\leq&\sum_{a=1}^{K}\frac{O\left((\log(4K)+\log_2\frac{(C+3)^2}{(C-1)^2\Delta_{j,0}^2})\min\{\Pr(a\in B_b),\Pr(b_j=b)\}\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}+\\
&\sum_{a=1}^{K}\frac{\min\{\Pr(a\in B_b),\Pr(b_j=b)\}}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\cdot(\sum_{k=1}^{+\infty}k\frac{\pi^2}{6}\frac{1}{3^{k-1}}+\sum_{k=1}^{+\infty}k\frac{\pi^2}{6}\frac{\delta}{5^{k-1}}).
\end{aligned}
$$

The second step is by the fact that $\frac{1}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\leq\frac{4}{\Delta_{j,0}^2}$ holds for all $a,j\in[K]$ and the third step is by the Lemma 18.

From Lemma 7, we know $\Pr(b_j=b)\leq\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)}$. What's more, we know $\Pr(a\in B_b)=\frac{|B_b|}{K}=\frac{\min\{2^{b-1},K\}}{K}$ holds for all $a\in[K]$. Thus, we can conclude

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{a=1}^{K}\frac{\log\frac{4K}{\delta_{\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\log_2\frac{\Theta(1)}{\Delta_{j,0}^2}\}}}+\log\log\frac{1}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\mathbb{1}(b_j=b)\mathbb{1}(a\in B_b)\right]\\
\leq&\min\{\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)},\frac{\min\{2^{b-1},K\}}{K}\}\sum_{a=1}^{K}\frac{O\left((\log(4K)+\log_2\frac{(C+3)^2}{(C-1)^2\Delta_{j,0}^2})\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}+\\
&\min\{\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)},\frac{\min\{2^{b-1},K\}}{K}\}\sum_{a=1}^{K}\frac{O(1)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\\
\leq&\min\{\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)},\frac{\min\{2^{b-1},K\}}{K}\}\sum_{a=1}^{K}\frac{O\left((\log(4K)+\log_2\frac{(C+3)^2}{(C-1)^2\Delta_{j,0}^2})\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}.
\end{aligned}
$$

For the expectation of $\diamondsuit$, we have

$$\begin{aligned}
&\mathbb{E}\mathbb{1}(b_j = b)\beta_{\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\lceil\log_2 \frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil\}} \log \frac{4K(\lceil\log_2 K\rceil+1)\alpha_{\max\{\kappa_b^{\text{ee}},\kappa_b^{\text{et}},\lceil\log_2 \frac{448(C+3)^2}{(C-1)^2\Delta_{j,0}^2}\rceil\}}}{\delta})\\
\leq &\Pr(b_j = b)O\left(\mathbb{E}\beta_{\kappa_b^{\text{ee}}} \log\frac{4K\alpha_{\kappa_b^{\text{ee}}}}{\delta} + \mathbb{E}\beta_{\kappa_b^{\text{et}}} \log\frac{4K\alpha_{\kappa_b^{\text{et}}}}{\delta} + \frac{(C+3)^2}{(C-1)^2}\frac{\log\frac{4K}{\delta} + \log_2\frac{(C+3)^2}{(C-1)^2\Delta_{j,0}^2}}{\Delta_{j,0}^2}\right)\\
\leq &\Pr(b_j = b)O\left(\sum_{k=1}^{+\infty}\frac{\pi^2}{6}\frac{2^k}{3^k}\log\frac{4K\cdot 5^k}{\delta} + \sum_{k=1}^{+\infty}\frac{\pi^2}{6}\frac{2^k\delta}{5^k}\log\frac{4K\cdot 5^k}{\delta} + \frac{\log\frac{4K}{\delta} + \log_2\frac{\Theta(1)}{\Delta_{j,0}^2}\log 5}{\Delta_{j,0}^2}\right)\\
\leq &\Pr(b_j = b)O\left(\frac{(C+3)^2}{(C-1)^2}\frac{\log\frac{4K}{\delta} + \log_2\frac{(C+3)^2}{(C-1)^2\Delta_{j,0}^2}}{\Delta_{j,0}^2}\right)
\end{aligned}$$

Sum up all the upper bounds, we have

$$\begin{aligned}
&\mathbb{E}\tau\\
\leq &O(\log K)\sum_{b=1}^{\lceil\log_2 K\rceil+1} O(\log K)\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)}\min\{2^{b-1},K\}+\\
&O(\log K)\sum_{a=1}^{K}\sum_{b=1}^{\lceil\log_2 K\rceil+1}\frac{O\left((\log(4K)+\log_2\frac{(C+3)^2}{(C-1)^2\Delta_{j,0}^2})\right)}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\min\{\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)},\frac{\min\{2^{b-1},K\}}{K}\}+\\
&O(\log K)\sum_{b=1}^{\lceil\log_2 K\rceil+1}\frac{(C+3)^2}{(C-1)^2}\Pr(b_j = b)O\left(\frac{\log\frac{4K}{\delta} + \log_2\frac{(C+3)^2}{(C-1)^2\Delta_{j,0}^2}}{\Delta_{j,0}^2}\right)
\end{aligned}$$

By the Lemma 33, 34, we can conclude

$$\begin{aligned}
\sum_{b=1}^{\lceil\log_2 K\rceil+1}\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)}\min\{2^{b-1},K\} &\leq \frac{K}{j}\cdot O\Big((\log_2\frac{K}{j})+1\Big)\\
\sum_{b=1}^{\lceil\log_2 K\rceil+1}\min\{\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)},\frac{\min\{2^{b-1},K\}}{K}\} &\leq O\Big(\frac{\log(j+1)}{j}\Big).
\end{aligned}$$

Together with the fact $\sum_{b=1}^{\lceil\log_2 K\rceil+1}\Pr(b_j = b) = 1$ and plug in default value $C = 1.01$ specified by Algorithm 1, we have

$$\begin{aligned}
&\mathbb{E}\tau\\
\leq &O(\log K)\cdot O\left(\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{K\log K\Big((\log_2\frac{K}{j})+1\Big)}{j} + \frac{\log(j+1)}{j}\sum_{a=1}^{K}\frac{\log(4K)+\log_2\frac{1}{\Delta_{j,0}^2}}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right)\\
\leq &O(\log K)\cdot O\left(\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{\log(j+1)\log K\Big((\log_2\frac{K}{j})+1\Big)}{j}\sum_{a=1}^{K}\frac{\log(4K)+\log_2\frac{1}{\Delta_{j,0}^2}}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right)\\
\leq &O(\log K)\cdot O\left(\frac{\log\frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{\log^3 K}{j}\sum_{a=1}^{K}\frac{\log(4K)+\log_2\frac{1}{\Delta_{j,0}^2}}{\max\{\Delta_{a,j}^2,\Delta_{a,0}^2\}}\right).
\end{aligned}$$

The second last line is by the assumption $\Delta_{i,j} < 1$ for all $i, j \in [K] \cup \{0\}$, making that $K \leq \sum_{a=1}^{K} \frac{1}{\max\{\Delta_{a,j}^2, \Delta_{a,0}^2\}}$. Notice the above upper bound holds for all $j \in [m]$, we can conclude

$$\mathbb{E}\tau \leq O(\log K) \cdot O\left(\min_{1 \leq j \leq m} \frac{\log \frac{1}{\delta}}{\Delta_{j,0}^2} + \frac{\log^3 K}{j} \sum_{a=1}^{K} \frac{\log(4K) + \log_2 \frac{1}{\Delta_{j,0}^2}}{\max\{\Delta_{a,j}^2, \Delta_{a,0}^2\}} \right).$$

□

Proof of Theorem 3, when ν is negative. Consider the bracket with size K, denoting this bracket as B. Similar to the proof of positive case, we split $\tau(B) = \tau^{\text{ee}}(B) + \tau^{\text{et}}(B)$ and prove an upper bound for $\mathbb{E}\tau^{\text{ee}}(B)$ and $\mathbb{E}\tau^{\text{et}}(B)$ separately. From Lemma 15, 30 and Lemma 31, we know

$$\begin{aligned}
\tau^{\text{ee}}(B) \leq & |Q| + \sum_{a=1}^{K} N_a^{\text{ee},B}\left(\tau_{\kappa_B^{\text{ee}}-1}^{\text{ee}}(B)\right) + \frac{O\left(\log \frac{4K}{\delta_{\max\{\kappa_B^{\text{ee}}, \kappa_B^{\text{et}}, \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil, \lceil \log_3 \frac{3}{\delta} \rceil\}}} + \log\log \frac{1}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2} \\
= & K + \tau_{\kappa_B^{\text{ee}}-1}^{\text{ee}}(B) + \sum_{a=1}^{K} \frac{O\left(\log \frac{4K}{\delta_{\max\{\kappa_B^{\text{ee}}, \kappa_B^{\text{et}}, \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil, \lceil \log_3 \frac{3}{\delta} \rceil\}}} + \log\log \frac{1}{\Delta_{a,0}^2}\right)}{\Delta_{a,0}^2}.
\end{aligned}$$

Taking expectation on both sides, we have

$$\begin{aligned}
& \mathbb{E}\tau_{\kappa_B^{\text{ee}}-1}^{\text{ee}}(B) \\
\leq & \sum_{k=1}^{+\infty} |K| \beta_{k-1} \log \frac{4K}{\delta_{k-1}} \Pr(\kappa_B^{\text{ee}} = k) \\
\leq & \sum_{k=1}^{+\infty} |K| 2^{k-1} \log(4K \cdot 3^{k-1}) \frac{\pi^2}{6 \cdot 3^{k-1}} \\
\leq & O(K \log K).
\end{aligned}$$

The second last line is by the Lemma 18. In addition, we can also derive

$$\begin{aligned}
& \mathbb{E} \log \frac{1}{\delta_{\max\{\kappa_B^{\text{ee}}, \kappa_B^{\text{et}}, \lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil, \lceil \log_3 \frac{3}{\delta} \rceil\}}} \\
\leq & \mathbb{E} \log \frac{1}{\delta_{\kappa_B^{\text{ee}}}} + \mathbb{E} \log \frac{1}{\delta_{\kappa_B^{\text{et}}}} + \mathbb{E} \log \frac{1}{\delta_{\lceil \log_2 \frac{226}{\Delta_{1,0}^2} \rceil}} + \mathbb{E} \log \frac{1}{\delta_{\lceil \log_3 \frac{3}{\delta} \rceil}} \\
\leq & \sum_{k=1}^{+\infty} k \log 3 \Pr(\kappa_B^{\text{ee}} = k) + \sum_{k=1}^{+\infty} k \log 3 \Pr(\kappa_B^{\text{et}} = k) + O\left(\log_2 \frac{1}{\Delta_{1,0}^2}\right) + \log \frac{9}{\delta} \\
\leq & \sum_{k=1}^{+\infty} \frac{\pi^2}{6} \frac{k}{3^{k-1}} \log 3 \Pr(\kappa_B^{\text{ee}} = k) + \sum_{k=1}^{+\infty} \frac{\pi^2}{6} \frac{k\delta}{5^{k-1}} \log 3 + O\left(\log_2 \frac{1}{\Delta_{1,0}^2}\right) + \log \frac{3}{\delta} \\
\leq & O\left(\log_2 \frac{1}{\Delta_{1,0}^2} + \log \frac{1}{\delta}\right).
\end{aligned}$$

The second last line is still by the Lemma 18. The above two inequalities suggests that

$$\begin{aligned}
\mathbb{E}\tau^{\text{ee}}(B) \leq & O(K \log K) + \sum_{a=1}^{K} O\left(\frac{\log \frac{12K(\lceil \log_2 K \rceil + 1)}{\delta} + \log \frac{1}{\Delta_{0,1}^2}}{\Delta_{a,0}^2} \right) \\
\leq & \sum_{a=1}^{K} O\left(\frac{\log \frac{12K(\lceil \log_2 K \rceil + 1)}{\delta} + \log \frac{1}{\Delta_{0,1}^2}}{\Delta_{a,0}^2} \right).
\end{aligned}$$

Meanwhile, for $\mathbb{E}\tau^{\text{et}}(B)$, we can utilize Lemma 29 to conclude

$$\begin{aligned}\mathbb{E}\tau^{\text{et}}(B) \leq & \mathbb{E}\sum_{k=1}^{\kappa_B^{\text{ee}}-1} O(1)\beta_k \log\frac{4K(\lceil\log_2 K\rceil+1)\alpha_k}{\delta}\\ \leq & \mathbb{E}O(1)\beta_{\kappa_B^{\text{ee}}-1}\log\frac{4K(\lceil\log_2 K\rceil+1)\alpha_{\kappa_B^{\text{ee}}-1}}{\delta}\\ \leq & O(1)\sum_{k=1}^{+\infty}\beta_{k-1}\frac{\pi^2}{6}\delta_{k-1}\log\frac{4K(\lceil\log_2 K\rceil+1)\alpha_{k-1}}{\delta}\\ \leq & O\left(\log\frac{4K}{\delta}\right).\end{aligned}$$

Combining the upper bound of $\mathbb{E}\tau^{\text{ee}}(B)$ and $\mathbb{E}\tau^{\text{et}}(B)$, we have

$$\mathbb{E}\tau \leq O(\log K)\mathbb{E}\tau(B) \leq O(\log K)O\left(\sum_{a=1}^{K}\frac{\log\frac{K}{\delta}+\log\frac{1}{\Delta_{0,1}^2}}{\Delta_{a,0}^2}\right)$$

□

D Technical Lemma

This section is for the proof of inequalities.

Lemma 32. *Denote $\{Z_s\}_{s=1}^{+\infty}$ is a group of random variable fulfilling that $Z_s|Z_1, Z_2, \cdots, Z_{s-1} \sim N(0,\Delta^2)$ for some $\Delta > 0$. Then, we have*

$$\Pr(\max_{1\leq t\leq n}\sum_{s=1}^{t} Z_s > z) \leq \exp\left(-\frac{z^2}{2n\Delta^2}\right).$$

Proof of Lemma 32. Notice that for $\lambda = \frac{z}{n\Delta^2}$, we have

$$\begin{aligned}&\Pr\big(\max_{1\leq t\leq n}\sum_{s=1}^{t} Z_s \geq z\big)\\ =&\Pr\big(\max_{1\leq t\leq n}\lambda\big(\sum_{s=1}^{t} Z_s\big) \geq \lambda z\big)\\ =&\Pr\left(\max_{1\leq t\leq n}\prod_{s=1}^{t}\exp\big(\lambda\big(Z_s\big)\big) \geq \exp\left(\lambda z\right)\right).\end{aligned}$$

Easy to see $\{\prod_{s=1}^{t}\exp\big(\lambda\big(Z_s\big)\big)\}_{t=1}^{+\infty}$ is a submartingale, as

$$\begin{aligned}&\mathbb{E}[\exp\big(\lambda\big(Z_t\big)\big)\,|z_1,\cdots,z_{t-1}]\\ \overset{\text{Jensen's Inequality}}{\geq}&\exp\big(\lambda\mathbb{E}[\big(Z_t\big)|Z_1,\cdots,Z_{t-1}]\big)\\ =&\exp(0)=1.\end{aligned}$$

Then, we can conclude

$$\begin{aligned}&\Pr(\max_{1\leq t\leq n}\sum_{s=1}^{t} Z_s > z)\\ \leq&\frac{\mathbb{E}_{\nu'}\prod_{t=1}^{n}\exp(\lambda(Z_t))}{\exp(\lambda z)}\\ =&\frac{\exp(\frac{n\lambda^2\Delta^2}{2})}{\exp(\lambda z)}\\ =&\exp(-\frac{z^2}{2n\Delta^2}).\end{aligned}$$

□

Lemma 33. *For any $K \in \mathbb{N}$, $K \geq 2$, $j \in [K]$ we have*

$$\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \frac{1}{\exp\left(\frac{j \cdot \lfloor 2^{b-2} \rfloor}{K}\right)} \min\{2^{b-1}, K\} \leq \frac{K}{j} \cdot O\Big((\log_2 \frac{K}{j}) + 1\Big)$$

Proof. We can conduct direct calculation.

$$\begin{aligned}
&\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \frac{1}{\exp\left(\frac{j \cdot \lfloor 2^{b-2} \rfloor}{K}\right)} \min\{2^{b-1}, K\} \\
\overset{K \geq 2}{=} & 1 + \frac{2}{\exp(\frac{j}{K})} + \sum_{b=3}^{\lceil \log_2 K \rceil + 1} \frac{1}{\exp\left(\frac{j \cdot \lfloor 2^{b-2} \rfloor}{K}\right)} \min\{2^{b-1}, K\} \\
\leq & 1 + \frac{2}{\exp(\frac{j}{K})} + \sum_{b=3}^{\lceil \log_2 K \rceil + 1} \frac{2^{b-1}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \\
= & 1 + \frac{2}{\exp(\frac{j}{K})} + \frac{2K}{j} \sum_{b=3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \\
\leq & 1 + \frac{2K}{j} + \frac{2K}{j} \sum_{b=3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)}.
\end{aligned}$$

Then, we suffice to prove $\sum_{b=3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \leq O\left((\log_2 \frac{K}{j}) + 1\right)$.

From the fact that $\frac{d \frac{x}{\exp(x)}}{dx} = e^{-x} - xe^{-x} = (1-x)e^{-x}$, we know $\frac{x}{\exp(x)}$ monotonically increases in the interval $(0, 1)$ and decreases in the interval $(1, +\infty)$. We have

$$\begin{aligned}
&\sum_{b=3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \\
\leq & \sum_{b=3}^{\lfloor \log_2 \frac{K}{j} \rfloor + 3} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} + \sum_{b=\lceil \log_2 \frac{K}{j} \rceil + 3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \\
\leq & (\lfloor \log_2 \frac{K}{j} \rfloor + 1) \frac{x}{e^x}|_{x=1} + \sum_{b=\lceil \log_2 \frac{K}{j} \rceil + 3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \\
\leq & (\lfloor \log_2 \frac{K}{j} \rfloor + 1) \frac{1}{e} + \sum_{b=\lceil \log_2 \frac{K}{j} \rceil + 3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)}.
\end{aligned}$$

The second last line is by that function $\frac{x}{e^x}$ achieves its maximum at $x = 1$. The remaining work is to prove $\sum_{b=\lceil \log_2 \frac{K}{j} \rceil + 3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \leq O(1)$.

Notice that for $b \geq \lceil \log_2 \frac{K}{j} \rceil + 3$, we have $\frac{j 2^{b-2}}{K} \geq \frac{j 2^{\log_2 \frac{K}{j} + 1}}{K} \geq 2$. Since we can conclude $\frac{x}{e^x} \geq 2 \cdot \frac{2x}{e^{2x}}$ holds for all $x \geq 2$, we can conclude

$$\frac{\frac{j \cdot 2^{b-1}}{K}}{\exp\left(\frac{j \cdot 2^{b-1}}{K}\right)} \Big/ \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \leq \frac{1}{2}$$

holds for all $b \geq \lceil \log_2 \frac{K}{j} \rceil + 3$. Thus, we have

$$\begin{aligned}
&\sum_{b=\lceil \log_2 \frac{K}{j} \rceil + 3}^{\lceil \log_2 K \rceil + 1} \frac{\frac{j \cdot 2^{b-2}}{K}}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} \\
\leq &\sum_{b=\lceil \log_2 \frac{K}{j} \rceil + 3}^{+\infty} 2^{-(b - \lceil \log_2 \frac{K}{j} \rceil - 3)} \frac{\frac{j \cdot 2^{\lceil \log_2 \frac{K}{j} \rceil + 3 - 2}}{K}}{\exp\left(\frac{j \cdot 2^{\lceil \log_2 \frac{K}{j} \rceil + 3 - 2}}{K}\right)} \\
\leq &\frac{1}{e} \sum_{b=0}^{+\infty} 2^{-b} \\
\leq &O(1).
\end{aligned}$$

That means we complete the proof. □

Lemma 34. *For any $K \in \mathbb{N}$, $K \geq 2$, $j \in [K]$ we have*

$$\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \min\{\frac{1}{\exp\left(\frac{j \cdot \lfloor 2^{b-2} \rfloor}{K}\right)}, \frac{\min\{2^{b-1}, K\}}{K}\} \leq O\Big(\frac{\log(j+1)}{j}\Big).$$

Proof. We can conduct direct calculation.

$$\begin{aligned}
&\sum_{b=1}^{\lceil \log_2 K \rceil + 1} \min\{\frac{1}{\exp\left(\frac{j \cdot \lfloor 2^{b-2} \rfloor}{K}\right)}, \frac{\min\{2^{b-1}, K\}}{K}\} \\
\leq &\frac{1}{K} + \sum_{b=2}^{\lceil \log_2 K \rceil + 1} \min\{\frac{1}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)}, \frac{2}{j} \frac{j \cdot 2^{b-2}}{K}\} \\
\leq &\frac{1}{K} + \sum_{b=2}^{\lceil \log_2 \frac{K \log(j+1)}{j} \rceil + 2} \frac{2}{j} \frac{j \cdot 2^{b-2}}{K} + \sum_{b=\lceil \log_2 \frac{K \log(j+1)}{j} \rceil + 3}^{\lceil \log_2 K \rceil + 1} \frac{1}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)}.
\end{aligned}$$

The last line is by the fact that when $b > \lceil \log_2 \frac{K \log(j+1)}{j} \rceil + 2$, we have

$$\begin{aligned}
\frac{2 \cdot 2^{b-2}}{K} &\geq \frac{2 \log(j+1)}{j}, \\
\frac{1}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)} &\leq \frac{1}{\exp\left(\frac{j}{K} \frac{K}{j} \log(j+1)\right)} = \frac{1}{j+1},
\end{aligned}$$

meaning that $\min\{\frac{1}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)}, \frac{2}{j} \frac{j \cdot 2^{b-2}}{K}\} = \frac{1}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)}$ for $b > \lceil \log_2 \frac{K \log(j+1)}{j} \rceil + 2$.

We prove an upper bound for $\sum_{b=2}^{\lceil \log_2 \frac{K \log(j+1)}{j} \rceil + 2} \frac{2}{j} \frac{j \cdot 2^{b-2}}{K}$ and $\sum_{b=\lceil \log_2 \frac{K \log j}{j} \rceil + 3}^{\lceil \log_2 K \rceil + 1} \frac{1}{\exp\left(\frac{j \cdot 2^{b-2}}{K}\right)}$ separately.

First, we have

$$\begin{aligned}
&\sum_{b=2}^{\lceil \log_2 \frac{K\log(j+1)}{j} \rceil+2} \frac{2}{j}\frac{j\cdot 2^{b-2}}{K}\\
\leq &\frac{2}{j}\cdot 2\cdot \frac{j2^{b-2}}{K}|_{b=\lceil \log_2 \frac{K\log(j+1)}{j} \rceil+2}\\
\leq &\frac{2}{j}\cdot 2\cdot \frac{j2^{1+\log_2 \frac{K\log(j+1)}{j}}}{K}\\
= &\frac{8\log(j+1)}{j}.
\end{aligned}$$

For $\sum_{b=\lceil \log_2 \frac{K\log(j+1)}{j} \rceil+3}^{\lceil \log_2 K\rceil+1} \frac{1}{\exp\left(\frac{j\cdot 2^{b-2}}{K}\right)}$, we first notice that when $b \geq \lceil \log_2 \frac{K\log(j+1)}{j} \rceil + 3$, we have

$$\frac{j2^{b-2}}{K} \geq \frac{j2^{\lceil \log_2 \frac{K\log j}{j} \rceil+3-2}}{K} \geq \frac{j2^{\log_2 \frac{K\log j}{j}+1}}{K} \geq 2.$$

That means, we can conclude for $b \geq \lceil \log_2 \frac{K\log(j+1)}{j} \rceil + 3$, we have $\frac{j2^{b-2}}{K}+1 \leq \frac{j2^{b-2}}{K}\cdot 2 = \frac{j2^{b-1}}{K}$. Take $b_0 := \lceil \log_2 \frac{K\log(j+1)}{j} \rceil + 3$, we can use induction to prove $\exp(\frac{j\cdot 2^{b-2}}{K}) \geq \exp(\frac{j2^{b_0-2}}{K}+b-b_0)$ for all $b \geq b_0$. Indeed, when $b = b_0$, the inequality is evident. If the inequality holds for b, then

$$\begin{aligned}
&\exp(\frac{j\cdot 2^{b+1-2}}{K})\\
=&\exp(\frac{j\cdot 2^{b-2}}{K}\cdot 2)\\
\geq &\exp(\frac{j\cdot 2^{b-2}}{K}+1)\\
\overset{\text{Induction}}{\geq} &\exp(\frac{j\cdot 2^{b_0-2}}{K}+b-b_0+1).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\sum_{b=\lceil \log_2 \frac{K\log(j+1)}{j} \rceil+3}^{\lceil \log_2 K\rceil+1} \frac{1}{\exp\left(\frac{j\cdot 2^{b-2}}{K}\right)}\\
\leq &\sum_{b=\lceil \log_2 \frac{K\log(j+1)}{j} \rceil+3}^{+\infty} \frac{1}{\exp(b-\lceil \log_2 \frac{K\log(j+1)}{j} \rceil-3)}\frac{1}{\exp\left(\frac{j\cdot 2^{\lceil \log_2 \frac{K\log(j+1)}{j} \rceil+3-2}}{K}\right)}\\
\leq &\frac{2}{\exp\left(\frac{j\cdot 2^{\lceil \log_2 \frac{K\log(j+1)}{j} \rceil+3-2}}{K}\right)}\\
\leq &\frac{2}{\exp\left(\frac{j\cdot 2^{\log_2 \frac{K\log(j+1)}{j}+1}}{K}\right)}\\
= &\frac{2}{\exp\left(\frac{j\cdot 2\cdot \frac{K\log(j+1)}{j}}{K}\right)}\\
= &\frac{2}{\exp\left(2\log(j+1)\right)}\\
< &\frac{2}{j}.
\end{aligned}$$

In summary, we have proved

$$\begin{aligned}
&\sum_{b=1}^{\lceil \log_2 K\rceil+1} \min\{\frac{1}{\exp\left(\frac{j\cdot\lfloor 2^{b-2}\rfloor}{K}\right)}, \frac{\min\{2^{b-1},K\}}{K}\}\\
\leq&\frac{1}{K}+\frac{8\log(j+1)}{j}+\frac{2}{j}\\
\leq&O\Big(\frac{\log(j+1)}{j}\Big).
\end{aligned}$$

□

Lemma 35 (Adapted Lemma 3 in Jamieson et al. (2014), Lemma D.3 in Li and Cheung (2025))**.** *Denote $\{X_i\}_{i=1}^{+\infty}$ as i.i.d σ^2-subgaussian random variable with true mean reward $\mu = 0$. For any $\delta \in (0,1)$, we have*

$$\Pr\left(\exists t, |\sum_{s=1}^{t} X_s| \geq \sqrt{2\sigma^2 2^{\lceil \log_2 t\rceil^+}\log\frac{2(\log_2 2^{\lceil \log_2 t\rceil^+})^2}{\delta}}\right) < \frac{\pi^2}{6}\delta.$$

Or equivalently,

$$\Pr\left(\exists t, |\sum_{s=1}^{t} X_s| \geq \sqrt{2\sigma^2 2^{\lceil \log_2 t\rceil^+}\log\frac{2(\lceil \log_2 t\rceil^+)^2}{\delta}}\right) < \frac{\pi^2}{6}\delta.$$

Proof of Lemma 35. Define $u_k = 2^k$, $k \geq 1$. Define $x = \sqrt{2\sigma^2 u_k \log\frac{2(\log_2 u_k)^2}{\delta}}$, $S_t = \sum_{i=1}^{t} X_i$ and the event

$$E_k = \left\{\max_{1\leq t\leq u_k} S_t > \sqrt{2\sigma^2 u_k \log\frac{2(\log_2 u_k)^2}{\delta}}\right\} \cup \left\{\min_{1\leq t\leq u_k} S_t < -\sqrt{2\sigma^2 u_k \log\frac{2(\log_2 u_k)^2}{\delta}}\right\},$$

For $\lambda > 0$, notice that

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda(\sum_{s=1}^{t} X_s))|X_1,\cdots,X_{t-1}\right]\\
=&\exp(\lambda(\sum_{s=1}^{t-1} X_s))\mathbb{E}\exp(\lambda X_t)\\
\geq&\exp(\lambda(\sum_{s=1}^{t-1} X_s))\exp(\mathbb{E}\lambda X_t)\\
=&\exp(\lambda(\sum_{s=1}^{t-1} X_s)).
\end{aligned}$$

Take $\lambda = \frac{x}{u_k\sigma^2}$, we can conclude $\{\exp(\lambda S_t)\}$ is a submartingale. Then,

$$\begin{aligned}
&\Pr\left(\max_{1\leq t\leq u_k} S_t \geq x\right)\\
=&\Pr\left(\max_{1\leq t\leq u_k} \exp(\lambda S_t) > \exp(\lambda x)\right)\\
\overset{*}{\leq}&\frac{\mathbb{E}\exp(\lambda S_{u_k})}{\exp(\lambda x)}\\
\leq&\frac{\exp(\frac{u_k\lambda^2\sigma^2}{2})}{\exp(\lambda x)}\\
\overset{\lambda=\frac{x}{u_k\sigma^2}}{=}&\exp(-\frac{x^2}{2u_k\sigma^2}).
\end{aligned}$$

Step * is by the maximal inequality for the submartingale. Take $x = \sqrt{2\sigma^2 u_k \log \frac{2(\log_2 u_k)^2}{\delta}}$, we have

$$\Pr\left(\max_{1\le t\le u_k} S_t \ge \sqrt{2\sigma^2 u_k \log \frac{2(\log_2 u_k)^2}{\delta}}\right) \le \exp\left(-\log \frac{2(\log_2 u_k)^2}{\delta}\right) = \frac{\delta}{2(\log_2 u_k)^2} = \frac{\delta}{2k^2}$$

For the part of $\Pr\left(\min_{1\le t\le u_k} S_t < -x\right)$, the proof is similar. We can conclude $\Pr(E_k) \le \frac{\delta}{k^2}$ and further $\Pr(\cup_{k=1}^{+\infty} E_k) \le \frac{\pi^2\delta}{6}$.
Thus,

$$\begin{aligned}
&\Pr\left(\exists t, |\sum_{s=1}^{t} X_s| \ge \sqrt{2\sigma^2 \max\{2^{\lceil \log_2 t\rceil}, 2\} \log \frac{2(\log_2 \max\{2^{\lceil \log_2 t\rceil}, 2\})^2}{\delta}}\right) \\
\le &\Pr\left(\exists k, \max_{1\le t'\le u_k} |\sum_{s=1}^{t'} X_s| \ge \sqrt{2\sigma^2 u_k \log \frac{2(\log_2 u_k)^2}{\delta}}\right) \\
\le &\Pr(\cup_{k=1}^{+\infty} E_k) \le \frac{\pi^2\delta}{6}.
\end{aligned}$$

□

Some Comments are as follows.

- We can similarly prove $\Pr\left(\exists t, |\sum_{s=1}^{t} X_s| \ge \sqrt{2\sigma^2 2^{\lceil \log_2 t\rceil^+} \log \frac{2\pi^2(\log_2 2^{\lceil \log_2 t\rceil^+})^2}{6\delta}}\right) < \delta$ holds for all $\delta \in (0,1)$.

- Since $\lceil \log_2 t\rceil^+ \le 1 + \log_2 t$, we have

$$\begin{aligned}
\frac{\pi^2\delta}{6} \ge &\Pr\left(\exists t, |\sum_{s=1}^{t} X_s| \ge \sqrt{2\sigma^2 2^{\max\{\lceil \log_2 t\rceil, 1\}} \log \frac{2(\log_2 2^{\lceil \log_2 t\rceil})^2}{\delta}}\right) \\
\ge &\Pr\left(\exists t, |\sum_{s=1}^{t} X_s| \ge \sqrt{4\sigma^2 t \log \frac{2(\log_2 2t)^2}{\delta}}\right)
\end{aligned}$$

Lemma 36 (Lemma D.2 in Li and Cheung (2025))**.** *For any $\Delta \in (0,1], K \ge 2, \delta \in (0, \frac{1}{2}], C \ge 1$, we can conclude*

$$\begin{aligned}
&t > \frac{28C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{16C^2 \log\left(\log\left(\frac{24C^2}{\Delta^2}\right)\right)}{\Delta^2} \\
\Rightarrow &C\sqrt{\frac{4\log \frac{2K(\log_2 2t)^2}{\delta}}{t}} < \Delta
\end{aligned}$$

Proof. By simple calculation, we can derive

$$
\begin{aligned}
& t > \frac{28C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{16C^2 \log\left(\log\left(\frac{24C^2}{\Delta^2}\right)\right)}{\Delta^2} \\
\Leftrightarrow & 2t > \frac{56C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{32C^2 \log\left(\log\left(\frac{24C^2}{\Delta^2}\right)\right)}{\Delta^2} \\
\Leftrightarrow & 2t > \frac{24C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{32C^2 \log\left(\log\left(\frac{24C^2}{\Delta^2}\right)\right) + 32C^2 \log(\frac{2K}{\delta})}{\Delta^2} \\
\overset{\log(x+y)\leq \log x+\log y, \forall x,y\geq 2}{\Rightarrow} & 2t > \frac{24C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{32C^2 \log\left(\log\left(\frac{24C^2}{\Delta^2}\right) + \frac{2K}{\delta}\right)}{\Delta^2} \\
\Rightarrow & 2t > \frac{24C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{32C^2 \log\left(\log\left(\frac{24C^2}{\Delta^2}\right) + \log\left(\log \frac{2K}{\delta}\right)\right)}{\Delta^2} \\
\Leftrightarrow & 2t > \frac{24C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{32C^2 \log\log\left(\frac{24C^2 \log \frac{2K}{\delta}}{\Delta^2}\right)}{\Delta^2} \\
\overset{\text{Lemma 37, as } 24C^2 \log \frac{2K}{\delta} > e^2}{\Rightarrow} & 2t > \frac{24C^2 \log \frac{2K}{\delta}}{\Delta^2} + \frac{16C^2 \log\log(2t)}{\Delta^2} \\
\Rightarrow & 2t > \frac{8C^2 \log \frac{2K}{\delta} + 16C^2 \log\log_2 e + 16C^2 \log\log(2t)}{\Delta^2} \\
\Leftrightarrow & t > \frac{4C^2 \log \frac{2K}{\delta} + 8C^2 \log \frac{\log(2t)}{\log 2}}{\Delta^2} \\
\Leftrightarrow & t > \frac{4C^2 \log \frac{2K}{\delta} + 8C^2 \log(\log_2 2t)}{\Delta^2} \\
\Leftrightarrow & C\sqrt{\frac{4 \log \frac{2K(\log_2 2t)^2}{\delta}}{t}} < \Delta.
\end{aligned}
$$

□

The last lemma is to solve inequality.

Lemma 37. *For any $b \geq a \geq 0$,*

- *If $b \geq e^2$, we have $x \geq b + 2a \log\log b \Rightarrow x \geq a \log\log(x) + b$.*
- *If $b, a \geq e$, we have $e \leq x \leq b + a \log\log b \Rightarrow x \leq a \log\log(x) + b$.*

The second inequality also implies $x \geq a \log\log(x) + b \Rightarrow x \geq b + a \log\log b$.

Proof of Lemma 37. We prove the first claim. Easy to see $\frac{d(x - a\log\log x - b)}{dx} = 1 - \frac{a}{x \log x} = \frac{x \log x - a}{x \log x}$. Take $x_0 = b + 2a \log\log b$, then $x_0 \log x_0 > x_0 \log(e^2 + 2a \log\log e^2) > x_0 \log e^2 = 2x_0 > a$.

Thus $x - a\log\log(x) - b$ increases in the interval $(x_0, +\infty)$. On the other hand, easy to check

$$\begin{aligned}
&x_0 - a\log\log x_0 - b\\
=&b + 2a\log\log b - a\log\log(b + 2a\log\log b) - b\\
=&2a\log\log b - a\log\log(b + 2a\log\log b)\\
=&a\left(2\log\log b - \log\log(b + 2a\log\log b)\right)\\
=&a\left(\log(\log b)^2 - \log\log(b + 2a\log\log b)\right)\\
=&a\log\frac{(\log b)^2}{\log(b + 2a\log\log b)}\\
\geq&a\log\frac{(\log b)^2}{\log(b + 2b\log\log b)}\\
=&a\log\frac{(\log b)^2}{\log b + \log(1 + 2\log\log b)},
\end{aligned}$$

Notice that

$$\begin{aligned}
&(\log b)^2 - \log b - \log(1 + 2\log\log b)\\
=&\log b(\log b - 1) - \log(1 + 2\log\log b)\\
\overset{b\geq e^2}{\geq}&\log b - \log(1 + 2\log\log b)\\
\geq&\log b - 2\log\log b\\
\overset{x>2\log x,\forall x>0}{>}&0,
\end{aligned}$$

we can conclude $x_0 - a\log\log x_0 - b > 0$ holds for all $x \geq b + 2a\log\log b$.

Then we turn to prove the second claim. As $\frac{d(x - a\log\log x - b)}{dx} = 1 - \frac{a}{x\log x} = \frac{x\log x - a}{x\log x}$, we know there is at most 1 zero point of $\frac{x\log x - a}{x\log x}$ in the interval $(e, b + a\log\log b)$. Thus,

$$\max_{e\leq x\leq b + a\log\log b} x - a\log\log x - b = \max\{x - a\log\log x - b|_{x=e}, x - a\log\log x - b|_{x=b+a\log\log b}\}.$$

Easy to see

$$e - a\log\log e - b = e - b \leq 0,$$

and

$$\begin{aligned}
&b + a\log\log b - a\log\log(b + a\log\log b) - b\\
=&a\log\log b - a\log\log(b + a\log\log b)\\
\leq&a\log\log b - a\log\log(b)\\
=&0.
\end{aligned}$$

That means $\max_{e\leq x\leq b+a\log\log b} x - a\log\log x - b \leq 0$. The second conclusion is done. □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We summarize the improvement over theoretical analysis.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We admit the gap is still up to a polynomial of logarithm factor.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide both sketch idea and full proof.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A]

   Justification: This paper doesn't include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: This paper does not include experiments requiring code.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: This paper does not include experiments

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: This paper does not include experiments

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We fully follow all the ethics requirements.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.