Supplementary Materials

A Simulation Details of Figure 1

In Figure 1, we report the cost complexities of Algorithm 1 with EXPLORE-A and Algorithm 1 with EXPLORE-B (let $\tilde{\mu}_1^{(M)} = 0.95$ and $\tilde{\mu}_2^{(M)} = 0.75$). We set the confidence parameter δ as $0.05, 0.1, 0.15, 0.2, 0.25$ respectively in comparing the performance of both procedures. For each simulation, we run 100 trials, plot their cost complexities' mean as markers and their deviation as shaded regions. We present the parameters of `MF-MAB` instances of Figures 1a and 1b in Tables 2 and 3 respectively. We note that it is more difficult to find the optimal fidelity m_k^* in the `MF-MAB` instances for Figure 1b because (1) there are more fidelities choices in this instance than that of Figure 1a; (2) the value of $\tilde{\Delta}_k^{(m)}/\sqrt{\lambda^{(m)}}$ are closer in the second instance than that of Figure 1a.

Table 2: Figure 1a's `MF-MAB` with $K = 5$ arms and $M = 3$ fidelities

Parameters	$\mu_1^{(m)}$	$\mu_2^{(m)}$	$\mu_3^{(m)}$	$\mu_4^{(m)}$	$\mu_5^{(m)}$	$\zeta^{(m)}$	$\lambda^{(m)}$
$m = 1$	0.70	0.75	0.50	0.50	0.30	0.30	1
$m = 2$	0.80	0.775	0.60	0.55	0.45	0.15	1.1
$m = 3$	0.90	0.80	0.70	0.60	0.50	0	1.2

Table 3: Figure 1b's `MF-MAB` with $K = 5$ arms and $M = 5$ fidelities

Parameters	$\mu_1^{(m)}$	$\mu_2^{(m)}$	$\mu_3^{(m)}$	$\mu_4^{(m)}$	$\mu_5^{(m)}$	$\zeta^{(m)}$	$\lambda^{(m)}$
$m = 1$	0.83	0.82	0.76	0.82	0.70	0.10	1
$m = 2$	0.84	0.83	0.80	0.80	0.72	0.08	1.1
$m = 3$	0.85	0.85	0.80	0.82	0.74	0.06	1.2
$m = 4$	0.85	0.86	0.80	0.80	0.76	0.04	1.3
$m = 5$	0.90	0.88	0.86	0.84	0.80	0	1.4

B A Third Fidelity Selection Procedure: EXPLORE-C

Besides the EXPLORE-A and -B procedures, here we consider a third naïve and conservative idea for fidelity selection that one should start from low risk (cost), gradually increase the risk (cost) as the learning task needs, and stop when finding the optimal arm. To decide when to increase the fidelity for exploring an arm k, we use the arm's confidence radius $\beta(N_{k,t}^{(m)}, t, \delta)$ at fidelity m as a measure of the amount of information left in this fidelity, and when the fidelity m's confidence radius is less than the error upper bound $\zeta^{(m)}$ at this fidelity, we increase the fidelity by 1 for higher accuracy, or formally, the fidelity is selected as follows,

$$m_{k,t} \leftarrow \min \left\{ m \,\middle|\, \beta(N_{k,t}^{(m)}, t, \delta) \geqslant \zeta^{(m)} \right\}.$$

Algorithm 3 EXPLORE-C Procedures

procedure EXPLORE-C(k)

$m_{k,t} \leftarrow \min \left\{ m \,\middle|\, \beta(N_{k,t}^{(m)}, t, \delta) \geqslant \zeta^{(m)} \right\}$

Pull $(k, m_{k,t})$, observe reward, and update corresponding statistics

Theorem B.1 (Cost complexity upper bounds for Algorithm 1 with EXPLORE-C)**.** *Given Assumption 3.3 and $L \geqslant 4KM$, Algorithm 1 outputs the optimal arm with a probability at least $1 - \delta$. The cost complexity of Algorithm 1 with* EXPLORE-C *are upper bounded as follows,*

$$\mathbb{E}[\Lambda] = O\left(H^{\ddagger} \log\left(\frac{L(H^{\ddagger} + Q)}{\lambda^{(1)}\delta} \right) + Q \log\left(\frac{L(H^{\ddagger} + Q)}{\lambda^{(1)}\delta} \right) \right), \tag{10}$$

where, letting $m_k^\ddagger$ denote the smallest fidelity for arm k such that $\Delta_k^{(m)} > 2\zeta^{(m)}$, or formally,

$$m_k^\ddagger := \min\{m : \Delta_k^{(m)} > 2\zeta^{(m)}\},$$

and we denote

$$H^\ddagger := \sum_{k\in\mathcal{K}} \frac{\lambda^{(m_k^\ddagger)}}{(\Delta_k^{(m_k^\ddagger)})^2}, \quad Q := \sum_{k\in\mathcal{K}} \sum_{m=1}^{m_k^\ddagger-1} \frac{\lambda^{(m)}}{(\zeta^{(m)})^2}.$$

Remark B.2 (EXPLORE-C *vs.* EXPLORE-A and -B)**.** The EXPLORE-C procedure does not require additional knowledge as the other two. It is a one-size-fits-all option. If with some addition information of a specific scenario, e.g., the exact or approximated reward means of top two arms, one can use the EXPLORE-A or B.

C Proofs for Best Arm Identification with Fixed Confidence

C.1 Proof of Theorem 3.1

Lemma C.1 (Kaufmann et al. [21], Lemma 1)**.** *Let ν and ν' be two bandit models with K arms such that for all k, the distributions ν_k and ν'_k are mutually absolutely continuous. For any almost-surely finite stopping time σ with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ where $\mathcal{F}_t = \sigma(I_1, X_1, \dots, I_t, X_t)$,*

$$\sum_{k=1}^{K} \mathbb{E}_\nu[N_k(\sigma)]\,\mathrm{KL}(\nu_k, \nu'_k) \geqslant \sup_{\mathcal{E}\in\mathcal{F}_\sigma} \mathrm{kl}(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E})),$$

where $\mathrm{kl}(x, y)$ is the binary relative entropy.

In `MF-MAB` model, regarding each arm-fidelity (k, m)-pair as an individual arm, we can extend Lemma C.1 to multi-fidelity case as follows,

$$\sum_{k=1}^{K}\sum_{m=1}^{M} \mathbb{E}_\nu[N_k^{(m)}(\sigma)]\,\mathrm{KL}(\nu_k^{(m)}, {\nu'}_k^{(m)}) \geqslant \sup_{\mathcal{E}\in\mathcal{F}_\sigma} \mathrm{kl}(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E})). \tag{11}$$

Next, we construct instances ν and ν'. We set the reward distributions $\nu = (\nu_k^{(m)})_{(k,m)\in\mathcal{K}\times\mathcal{M}}$ as Bernoulli and the reward means fulfill $\mu_1^{(M)} > \mu_2^{(M)} \geqslant \mu_3^{(M)} \geqslant \dots \geqslant \mu_K^{(M)}$, where $\mu_k^{(m)} = \mathbb{E}_{X\sim\nu_k^{(m)}}[X]$. We let ${\nu'}_k^{(m)}$ be the same to $\nu_k^{(m)}$ for all k and m, except for that an arm $\ell \neq 1$. We set arm ℓ's reward means on fidelities $m \in \mathcal{M}_k$ to be ${\nu'}_\ell^{(m)} = \nu_1^{(M)} - \zeta^{(m)} + \epsilon$. So, in instance ν', the optimal arm is ℓ and its true reward mean ${\mu'}_\ell^{(M)}$ is slightly greater than $\mu_1^{(M)}$. This implies for the event $\mathcal{E} = \{\text{output arm } 1\}$ and any algorithm π that can find the optimal arm with a confidence $1-\delta$, $\mathbb{P}_{\nu,\pi}(\mathcal{E}) \geqslant 1-\delta$ and $\mathbb{P}_{\nu',\pi}(\mathcal{E}) \leqslant \delta$. Then, from Eq.(11), we have

$$\begin{aligned}\sum_{m\in\mathcal{M}_k} \mathbb{E}_\nu[N_\ell^{(m)}(\sigma)]\,\mathrm{KL}(\nu_\ell^{(m)}, {\nu'}_\ell^{(m)}) &\geqslant \sup_{\mathcal{E}\in\mathcal{F}_\sigma} \mathrm{kl}(\mathbb{P}_{\nu,\pi}(\mathcal{E}), \mathbb{P}_{\nu',\pi}(\mathcal{E}))\\ &\geqslant \mathrm{kl}(1-\delta, \delta)\\ &\geqslant \log\frac{1}{2.4\delta}.\end{aligned}$$

We rewrite the above inequality as follows,

$$\sum_{m\in\mathcal{M}_k} \lambda^{(m)}\mathbb{E}_\nu[N_\ell^{(m)}(\sigma)] \cdot \frac{\mathrm{KL}(\nu_\ell^{(m)}, {\nu'}_\ell^{(m)})}{\lambda^{(m)}} \geqslant \log\frac{1}{2.4\delta}.$$

Therefore, for the arm ℓ, our aim is to minimize its cost complexity with a constraint as follows,

$$\min_{\mathbb{E}[N_\ell^{(m)}],\forall m} \sum_{m=1}^{M} \lambda^{(m)}\mathbb{E}_\nu[N_\ell^{(m)}(\sigma)]$$

$$\text{such that} \sum_{m\in\mathcal{M}_k} \lambda^{(m)}\mathbb{E}_\nu[N_\ell^{(m)}(\sigma)] \cdot \frac{\mathrm{KL}(\nu_\ell^{(m)}, {\nu'}_\ell^{(m)})}{\lambda^{(m)}} \geqslant \log\frac{1}{2.4\delta}.$$

Note that the above is a linear programming (LP) and its optimum is reached at one of its polyhedron constraint's vertex—only one $\mathbb{E}[N_\ell^{(m)}]$ is positive and all others are equal to zero.

$$
\begin{aligned}
\min_{\mathbb{E}[N_\ell^{(m)}],\forall m} \sum_{m=1}^{M} \lambda^{(m)} \mathbb{E}_\nu[N_\ell^{(m)}(\sigma)] &\overset{(a)}{\geqslant} \min_{m\in\mathcal{M}_k} \frac{\lambda^{(m)}}{\mathrm{KL}(\nu_\ell^{(m)}, {\nu'}_\ell^{(m)})} \log\frac{1}{2.4\delta} \\
&= \min_{m\in\mathcal{M}_k} \frac{\lambda^{(m)}}{\mathrm{KL}(\nu_\ell^{(m)}, \nu_1^{(M)} - \zeta^{(m)} + \epsilon)} \log\frac{1}{2.4\delta} \\
&\overset{(b)}{\geqslant} \min_{m\in\mathcal{M}_k} \frac{\lambda^{(m)}}{(1+\varepsilon)\,\mathrm{KL}(\nu_\ell^{(m)}, \nu_1^{(M)} - \zeta^{(m)})} \log\frac{1}{2.4\delta}
\end{aligned}
$$

where the inequality (a) is due to the property of LP we mentioned above, and the inequality (b) is because of the continuity of KL-divergence.

To bound the optimal arm 1's cost complexity, we use the same ν as above and construct another instance ν''. The instance ν'''s reward means are the same to ν except for arm 1 whose reward means for fidelity $m \in \mathcal{M}_1$ are set as ${\mu''}_1^{(m)} = \mu_2^{(m)} + \zeta^{(m)} - \epsilon$. Then, with similar procedure as the above, we obtain

$$
\min_{\mathbb{E}[N_1^{(m)}],\forall m} \sum_{m=1}^{M} \lambda^{(m)} \mathbb{E}_\nu[N_1^{(m)}(\sigma)] \geqslant \min_{m\in\mathcal{M}_1} \frac{\lambda^{(m)}}{(1+\varepsilon)\,\mathrm{KL}(\nu_1^{(m)}, \nu_2^{(M)} + \zeta^{(m)})} \log\frac{1}{2.4\delta}.
$$

Summing up the above costs leads to the lower bound as follows, and letting the ϵ goes to zeros concludes the proof.

$$
\mathbb{E}[\Lambda] \geqslant \left(\min_{m\in\mathcal{M}_1} \frac{\lambda^{(m)}}{(1+\varepsilon)\,\mathrm{KL}(\nu_1^{(m)}, \nu_2^{(M)} + \zeta^{(m)})} + \sum_{k\neq 1} \min_{m\in\mathcal{M}_k} \frac{\lambda^{(m)}}{(1+\varepsilon)\,\mathrm{KL}(\nu_k^{(m)}, \nu_1^{(M)} - \zeta^{(m)})} \right) \log\frac{1}{2.4\delta}.
$$

C.2 Proof of Theorem 3.4

Notation. Denote the threshold $c = \frac{\mu_1^{(M)} + \mu_2^{(M)}}{2}$ as the average of the optimal and best suboptimal arms' reward means. Denote $\mathcal{A}_t \coloneqq \{k \in \mathcal{K} : \mathtt{LCB}_{k,t} > c\}$ and $\mathcal{B}_t \coloneqq \{k \in \mathcal{K} : \mathtt{UCB}_{k,t} < c\}$ as the above and below sets which respectively contain arms whose rewards are clearly higher or lower than the threshold with high probability, and let $\mathcal{C}_t \coloneqq \mathcal{K} \setminus (\mathcal{A}_t \cup \mathcal{B}_t)$ as the complement of both sets' union. Then, we define two events as follows

$$
\begin{aligned}
\mathtt{TERM}_t &\coloneqq \{\mathtt{LCB}_{\ell_t,t} > \mathtt{UCB}_{u_t,t}\}, \\
\mathtt{CROS}_t &\coloneqq \{\exists k \neq 1 : k \in \mathcal{A}_t\} \cup \{1 \in \mathcal{B}_t\}.
\end{aligned}
$$

The $\mathtt{TERM}_t$ event corresponds to the complement of the main while loop condition in the LUCB algorithm. When the $\mathtt{TERM}_t$ event happens, the LUCB algorithm terminates. The $\mathtt{CROS}_t$ event means there exists a suboptimal arm whose $\mathtt{LCB}_{k,t}$ is greater than c or that the optimal arm 1's $\mathtt{UCB}_{1,t}$ is less than c, both of which means that at least one arm's reward mean confidence interval incorrectly crosses the threshold c.

Step 1. Prove $\neg\mathtt{TERM}_t \cap \neg\mathtt{CROS}_t \implies (\ell_t \in \mathcal{C}_t) \cup (u_t \in \mathcal{C}_t)$. We show this statement by contradiction case by case. That is, the negation of $(\ell_t \in \mathcal{C}_t) \cup (u_t \in \mathcal{C}_t)$ cannot happen when $\neg\mathtt{TERM}_t \cap \neg\mathtt{CROS}_t$.

Case 1: $(\ell_t \in \mathcal{A}_t) \cap (u_t \in \mathcal{A}_t) \cap \neg\mathtt{TERM}_t$

$\implies (\ell_t \in \mathcal{A}_t) \cap (u_t \in \mathcal{A}_t) \implies |\mathcal{A}_t| \geqslant 2 \implies \exists k \neq 1 : k \in \mathcal{A}_t \implies \mathtt{CROS}_t,$

Case 2: $(\ell_t \in \mathcal{B}_t) \cap (u_t \in \mathcal{A}_t) \cap \neg\mathtt{TERM}_t$

$\implies \mathtt{UCB}_{\ell_t,t} < c < \mathtt{LCB}_{u_t,t} < \mathtt{UCB}_{u_t,t} \implies \emptyset$ (contradicts the selection of ℓ_t and u_t),

Case 3: $(\ell_t \in \mathcal{A}_t) \cap (u_t \in \mathcal{B}_t) \cap \neg\mathtt{TERM}_t$

$\implies \{\mathtt{LCB}_{\ell_t,t} > c > \mathtt{UCB}_{u_t,t}\} \cap \neg\mathtt{TERM}_t \implies \emptyset,$

Case 4: $(\ell_t \in \mathcal{B}_t) \cap (u_t \in \mathcal{B}_t) \cap \neg\mathtt{TERM}_t$

$\implies (\ell_t \in \mathcal{B}_t) \cap (u_t \in \mathcal{B}_t) \implies |\mathcal{B}_t| = K \implies 1 \in \mathcal{B}_t \implies \mathtt{CROS}_t.$

Step 2. Prove $\mathbb{P}(\texttt{CROS}_t) \leqslant \frac{KM\delta}{Lt^3}$. For any suboptimal arm $k \neq 1$, we bound the probability that the arm k is in $\mathcal{A}_t$ as follows,

$$\begin{aligned}
\mathbb{P}(k \in \mathcal{A}_t) &= \mathbb{P}(\texttt{LCB}_{k,t} > c) = \mathbb{P}\left(\max_{m\in\mathcal{M}} \texttt{LCB}_{k,t}^{(m)} > c\right) \leqslant \sum_{m\in\mathcal{M}} \mathbb{P}\left(\texttt{LCB}_{k,t}^{(m)} > c\right) \\
&= \sum_{m\in\mathcal{M}} \mathbb{P}\left(\hat{\mu}_{k,t}^{(m)} - \zeta^{(m)} - \beta(N_{k,t}^{(m)}, t) > c\right) \\
&= \sum_{m\in\mathcal{M}} \mathbb{P}\left(\hat{\mu}_{k,t}^{(m)} - \mu_k^{(m)} + (\mu_k^{(m)} - \zeta^{(m)} - c) > \beta(N_{k,t}^{(m)}, t)\right) \\
&\overset{(a)}{\leqslant} \sum_{m\in\mathcal{M}} \mathbb{P}\left(\hat{\mu}_{k,t}^{(m)} - \mu_k^{(m)} > \beta(N_{k,t}^{(m)}, t)\right) \leqslant \sum_{m\in\mathcal{M}} \sum_{n=1}^{t} \mathbb{P}(\hat{\mu}_{k,t}^{(m)} - \mu_k^{(m)} > \beta(n, t)) \\
&\leqslant \sum_{m\in\mathcal{M}} \sum_{n=1}^{t} \exp(-n(\beta(n,t))^2) = \sum_{m\in\mathcal{M}} \sum_{n=1}^{t} \frac{\delta}{Lt^4} \\
&\leqslant \frac{M\delta}{Lt^3},
\end{aligned}$$

where the inequality (a) is due to that $\mu_k^{(m)} - \zeta^{(m)} \leqslant \mu_k^{(M)} < c$. With similar derivation, we have $\mathbb{P}(1 \in \mathcal{B}_t) \leqslant \frac{M\delta}{Lt^3}$. Noticing that $\mathbb{P}(\texttt{CROS}_t) \leqslant \sum_{k\neq 1} \mathbb{P}(k \in \mathcal{A}_t) + \mathbb{P}(1 \in \mathcal{B}_t)$, we have $\mathbb{P}(\texttt{CROS}_t) \leqslant \frac{KM\delta}{Lt^3}$.

Step 3. Prove $\mathbb{P}\left(\exists k \in \mathcal{K} : (N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap (k \in \texttt{Mid}_t)\right) \leqslant \frac{16\delta \sum_{k\in\mathcal{K}} \Delta_k^{-2}}{Lt^4}$, where $N_{k,t}^* := \frac{\log(Lt^4/\delta)}{(\Delta_k^{(\tilde{m}_k^*)})^2}$. For any fixed suboptimal arm $k \neq 1$ (with $\mu_k^{(M)} < c$), we have

$$\begin{aligned}
&\mathbb{P}\left((N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap (k \in \texttt{Mid}_t)\right) \\
&= \mathbb{P}\left((N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap (k \notin \mathcal{A}_t \cup \mathcal{B}_t)\right) \\
&\leqslant \mathbb{P}\left((N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap (\texttt{UCB}_{k,t} > c)\right) \\
&= \mathbb{P}\left((N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap \left(\min_{m\in\mathcal{M}} \hat{\mu}_{k,t}^{(m)} + \zeta^{(m)} + \beta(N_{k,t}^{(m)}, t) > c\right)\right) \\
&\leqslant \mathbb{P}\left((N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap (\hat{\mu}_{k,t}^{(m_k^*)} + \zeta^{(m_k^*)} + \beta(N_{k,t}^{(m_k^*)}, t) > c)\right) \\
&\leqslant \mathbb{P}\left((N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap (\hat{\mu}_{k,t}^{(m_k^*)} - \mu_k^{(m_k^*)} > (c - \mu_k^{(m_k^*)} - \zeta^{(m_k^*)}) - \beta(N_{k,t}^{(m_k^*)}, t))\right) \\
&\overset{(a)}{\leqslant} \mathbb{P}\left((N_{k,t}^{(m_k^*)} > 16N_{k,t}^*) \cap \left(\hat{\mu}_{k,t}^{(m_k^*)} - \mu_k^{(m_k^*)} > \frac{\Delta_k^{(\tilde{m}_k^*)}}{2} - \beta(N_{k,t}^{(\tilde{m}_k^*)}, t)\right)\right) \\
&\overset{(b)}{\leqslant} \sum_{\tau > 16N_{k,t}^*} \mathbb{P}\left(\hat{\mu}_{k,t(\tau)}^{(m_k^*)} - \mu_k^{(m_k^*)} > \frac{\Delta_k^{(\tilde{m}_k^*)}}{4}\right) \quad \left(\text{denote } \hat{\mu}_{k,t(\tau)}^{(m_k^*)} \text{ as the empirical mean of } \tau \text{ observations}\right) \\
&\leqslant \sum_{\tau > 16N_{k,t}^*} \exp\left(-\frac{\tau(\Delta_k^{(\tilde{m}_k^*)})^2}{16}\right) \leqslant \int_{\tau > 16N_{k,t}^*} \exp\left(-\frac{\tau(\Delta_k^{(\tilde{m}_k^*)})^2}{16}\right) d\tau \leqslant \frac{16\delta}{(\Delta_k^{(\tilde{m}_k^*)})^2 Lt^4},
\end{aligned}$$

where inequality (a) is due to Eq.(6) and inequality (b) is due to $\beta(\tau, t) < \frac{\Delta_k}{4}$ for $\tau > 16N_{k,t}^*$.

From Step 3, we obtain that the following equation holds with high probability,

$$N_{k,t}^{(\tilde{m}_k^*)} \leqslant \frac{16}{(\Delta_k^{(\tilde{m}_k^*)})^2} \log\left(\frac{Lt^4}{\delta}\right) \leqslant \frac{64}{(\Delta_k^{(\tilde{m}_k^*)})^2} \log\left(\frac{Lt}{\delta}\right). \tag{12}$$

Next, we respectively present the cost complexity upper bounds for different fidelity selection procedures in Algorithm 2.

C.2.1 Proof for EXPLORE-A's Upper Bound

Step 4 for EXPLORE-A: prove that if the small probability events of Steps 2 and 3 do not happen, then the algorithm terminates with a high probability when Λ is large.

Lemma C.2. *Give reward means $\mu_1^{(M)}$ and $\mu_2^{(M)}$. For a fixed arm k, there exist $\bar{N}_{k,t}$ and $\alpha_k > 0$ such that when $N_{k,t} > \bar{N}_{k,t}$, $N_{k,t} < 2N_{k,t}^{(\tilde{m}_k^*)}$, the number of times of pulling this arm k at fidelities $m\,(\neq \tilde{m}_k^*)$ is $O(\log(\log N_{k,t}))$, or formally,*

$$N_{k,t}^{(m)} \leqslant \frac{8}{\lambda^{(m)}} \left(\frac{\Delta_k^{(\tilde{m}_k^*)}}{\sqrt{\lambda^{(\tilde{m}_k^*)}}} - \frac{\Delta_k^{(m)}}{\sqrt{\lambda^{(m)}}} \right)^{-2} \log N_{k,t},\ \forall m \neq \tilde{m}_k^*. \tag{13}$$

Combine Lemma C.2 with Eq.(12) in Step 3, we have, for any arm k and fidelity $m \neq \tilde{m}_k^*$:

$$N_{k,t}^{(m)} \leqslant \frac{8}{\lambda^{(m)}} \left(\frac{\Delta_k^{(\tilde{m}_k^*)}}{\sqrt{\lambda^{(\tilde{m}_k^*)}}} - \frac{\Delta_k^{(m)}}{\sqrt{\lambda^{(m)}}} \right)^{-2} \log \left(\frac{128}{(\Delta_k^{(\tilde{m}_k^*)})^2} \log \left(\frac{Lt}{\delta} \right) \right) \tag{14}$$

Next, we can upper bound the total cost of the LUCB algorithm (before it terminating) via Eq.(12) and Eq.(14). Specially, we show it is impossible for $\Lambda = C\left(H \log \frac{L(G+H)}{\lambda^{(1)}\delta} + G \log\log \frac{L(G+H)}{\lambda^{(1)}\delta}\right)$ via contradiction. Suppose $\Lambda = C\left(H \log \frac{L(G+H)}{\lambda^{(1)}\delta} + G \log\log \frac{L(G+H)}{\lambda^{(1)}\delta}\right)$, we have the following,

$$\begin{aligned}
\mathbb{E}[\Lambda] &\leqslant \sum_{k\in\mathcal{K}} \sum_{m\in\mathcal{M}} \lambda^{(m)} N_{k,t}^{(m)} \\
&\leqslant \sum_{k\in\mathcal{K}} \lambda^{(\tilde{m}_k^*)} N_{k,t}^{(\tilde{m}_k^*)} + \sum_{k\in\mathcal{K}} \sum_{m\neq\tilde{m}_k^*} \lambda^{(m)} N_{k,t}^{(m)} \\
&\overset{(a)}{\leqslant} 64H \log \frac{Lt}{\delta} + 8G \log\log \frac{Lt}{\delta} + G\log(128H) \\
&\overset{(b)}{\leqslant} 64H \log \frac{L\Lambda}{\lambda^{(1)}\delta} + 8G \log\log \frac{L\Lambda}{\lambda^{(1)}\delta} + G\log(128H) \\
&\overset{(c)}{=} 64H \log \left(\frac{L}{\lambda^{(1)}\delta} C \left(H \log \frac{L(G+H)}{\lambda^{(1)}\delta} + G \log\log \frac{L(G+H)}{\lambda^{(1)}\delta} \right) \right) \\
&\quad + 8G \log\log \left(\frac{L}{\lambda^{(1)}\delta} C \left(H \log \frac{L(G+H)}{\lambda^{(1)}\delta} + G \log\log \frac{L(G+H)}{\lambda^{(1)}\delta} \right) \right) + G\log(128H) \\
&\overset{(d)}{\leqslant} 128(2+\log C) \left(H \log \frac{L(G+H)}{\lambda^{(1)}\delta} + G \log\log \frac{L(G+H)}{\lambda^{(1)}\delta} \right) \\
&\overset{(e)}{<} C \left(H \log \frac{L(G+H)}{\lambda^{(1)}\delta} + G \log\log \frac{L(G+H)}{\lambda^{(1)}\delta} \right),
\end{aligned}$$

where the inequality (a) is due to Eq.(12) and Eq.(14), the inequality (b) is because $t \leqslant \frac{\Lambda}{\lambda^{(1)}}$, the inequality (c) is by the supposition, the inequality (d) is by separately bounding the above first two terms via Eq.(15) and Eq.(16) in the following, and the inequality (e) holds for $C > 1200$. This above inequality contradicts the supposition, and, therefore, we conclude the cost complexity upper bound proof for EXPLORE-A. Similar proof also holds for EXPLORE-B by replacing $\tilde{m}_k^*$ with $m_k^\dagger$.

Next, we provide the upper bounds used in the inequality (d) above:

$$\begin{aligned}
&H\log\left(\frac{L}{\lambda^{(1)}\delta}C\left(H\log\frac{L(G+H)}{\lambda^{(1)}\delta}+G\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)\right)\\
&\leqslant H\log\left(\frac{L}{\lambda^{(1)}\delta}CH\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)+H\log\left(\frac{L}{\lambda^{(1)}\delta}CG\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)\\
&\leqslant H\log C+H\log\left(\frac{LH}{\lambda^{(1)}\delta}\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)+H\log C+H\log\left(\frac{LG}{\lambda^{(1)}\delta}\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)\\
&\leqslant 2H\log C+4H\log\frac{L(G+H)}{\lambda^{(1)}\delta}\\
&\leqslant (4+2\log C)H\log\frac{L(G+H)}{\lambda^{(1)}\delta},
\end{aligned}\tag{15}$$

and

$$\begin{aligned}
&G\log\log\left(\frac{L}{\lambda^{(1)}\delta}C\left(H\log\frac{L(G+H)}{\lambda^{(1)}\delta}+G\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)\right)\\
&\leqslant G\log\log\left(\frac{L}{\lambda^{(1)}\delta}CH\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)+G\log\log\left(\frac{L}{\lambda^{(1)}\delta}CG\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)\\
&\leqslant G\log\log C+G\log\log\left(\frac{LH}{\lambda^{(1)}\delta}\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)\\
&\quad+G\log\log C+G\log\log\left(\frac{LG}{\lambda^{(1)}\delta}\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}\right)\\
&\leqslant 2G\log\log C+4G\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}\\
&\leqslant (4+2\log\log C)G\log\log\frac{L(G+H)}{\lambda^{(1)}\delta}.
\end{aligned}\tag{16}$$

Proof of Lemma C.2.

Claim 1. For any fixed $m \neq \tilde{m}_k^*$, if the following equation holds, then the algorithm will not pull arm k at fidelity m with high probability.

$$N_{k,t}^{(m)} > \frac{8}{\lambda^{(m)}}\left(\frac{\Delta_k^{(\tilde{m}_k^*)}}{\sqrt{\lambda^{(\tilde{m}_k^*)}}}-\frac{\Delta_k^{(m)}}{\sqrt{\lambda^{(m)}}}\right)^{-2}\log N_{k,t}.$$

$$\begin{aligned}
\texttt{f-UCB}_{u_t,t}^{(m)}(\mu_1^{(M)}) &= \frac{1}{\sqrt{\lambda^{(m)}}}\left(\mu_1^{(M)}-\hat{\mu}_{u_t,t}^{(m)}-\zeta^{(m)}+\sqrt{\frac{2\log N_{u_t,t}}{N_{u_t,t}^{(m)}}}\right)\\
&\overset{(a)}{\leqslant} \frac{1}{\sqrt{\lambda^{(m)}}}\left(\mu_1^{(M)}-\mu_{u_t}^{(m)}-\zeta^{(m)}+2\sqrt{\frac{2\log N_{u_t,t}}{N_{u_t,t}^{(m)}}}\right)\\
&\overset{(b)}{\leqslant} \frac{1}{\sqrt{\lambda^{(m_{u_t}^*)}}}\left(\mu_1^{(M)}-\mu_{u_t}^{(m_{u_t}^*)}-\zeta^{(m_{u_t}^*)}\right)\\
&\overset{(c)}{\leqslant} \frac{1}{\sqrt{\lambda^{(\tilde{m}_k^*)}}}\left(\mu_1^{(M)}-\hat{\mu}_{u_t,t}^{(m_{u_t}^*)}-\zeta^{(m_{u_t}^*)}+\sqrt{\frac{2\log N_{u_t,t}}{N_{u_t,t}^{(\tilde{m}_k^*)}}}\right)\\
&= \texttt{f-UCB}_{u_t,t}^{(m_{u_t}^*)}(\mu_1^{(M)}),
\end{aligned}$$

where the inequalities (a) and (c) hold with a probability at least $1-\frac{1}{(N_{k,t})^2}$ respectively (by Hoeffding's inequality), and the inequality (b) holds due to the equation in the claim.

C.2.2 Proof for EXPLORE-B's Upper Bound

As EXPLORE-B of Algorithm 2 also employs the LUCB framework, it shares the first three steps of the proof for Theorem 3.4 in Appendix C.2. Hence, in this part, we focus on the proof of the final cost complexity upper bound.

Lemma C.3. *If the condition in Line 10 holds, then the committed fidelity $\hat{m}_k^*$ fulfills the following inequality:*

$$2 \cdot \frac{\Delta_k^{(\hat{m}_k^*)}}{\sqrt{\lambda^{(\hat{m}_k^*)}}} \geqslant \frac{\Delta_k^{(\tilde{m}_k^*)}}{\sqrt{\lambda^{(\tilde{m}_k^*)}}}. \tag{17}$$

Proof of Lemma C.3. Eq.(17) is proved as follows,

$$\begin{aligned}
\frac{\Delta_k^{(\tilde{m}_k^*)}/\sqrt{\lambda^{(\tilde{m}_k^*)}}}{\Delta_k^{(\hat{m}_k^*)}/\sqrt{\lambda^{(\hat{m}_k^*)}}} &\leqslant \frac{(\hat{\mu}_*^{(M)} - (\hat{\mu}_k^{(\tilde{m}_k^*)} + \zeta^{(\tilde{m}_k^*)}))/\sqrt{\lambda^{(\tilde{m}_k^*)}} + \sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}}{(\hat{\mu}_*^{(M)} - (\hat{\mu}_k^{(\hat{m}_k^*)} + \zeta^{(\hat{m}_k^*)}))/\sqrt{\lambda^{(\hat{m}_k^*)}} - \sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}} \\
&\overset{(a)}{\leqslant} \frac{(\hat{\mu}_*^{(M)} - (\hat{\mu}_k^{(\hat{m}_k^*)} + \zeta^{(\hat{m}_k^*)}))/\sqrt{\lambda^{(\hat{m}_k^*)}} + \sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}}{(\hat{\mu}_*^{(M)} - (\hat{\mu}_k^{(\hat{m}_k^*)} + \zeta^{(\hat{m}_k^*)}))/\sqrt{\lambda^{(\hat{m}_k^*)}} - \sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}} \\
&\leqslant 1 + \frac{2\sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}}{(\hat{\mu}_*^{(M)} - (\hat{\mu}_k^{(\hat{m}_k^*)} + \zeta^{(\hat{m}_k^*)}))/\sqrt{\lambda^{(\hat{m}_k^*)}} - \sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}} \\
&\overset{(b)}{\leqslant} 1 + \frac{2\sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}}{2\sqrt{\log(2KM/\delta)/\lambda^{(1)} N_{k,t}^{(m)}}} = 2,
\end{aligned}$$

where inequality (a) is due to the definition of $\hat{m}_k^*$, and inequality (b) is due to the condition in Line 10. $\square$

We next upper bound the number of times of $N_{k,t}^{(m)}$ that guarantees that the condition in Line 10 is true. Let us consider the case of exploring arm u_t.

$$\begin{aligned}
\max_{m \in \mathcal{M}} \frac{\hat{\Delta}_{k,t}^{(m)}}{\sqrt{\lambda^{(m)}}} &\geqslant \frac{\hat{\mu}_{k_*}^{(M)} - (\hat{\mu}_k^{(\tilde{m}_k^*)} + \zeta^{(\tilde{m}_k^*)})}{\sqrt{\lambda^{(\tilde{m}_k^*)}}} \\
&\overset{(a)}{\geqslant} \frac{\hat{\mu}_{k_*}^{(M)} - (\mu_k^{(\tilde{m}_k^*)} + \zeta^{(\tilde{m}_k^*)})}{\sqrt{\lambda^{(\tilde{m}_k^*)}}} - \sqrt{\frac{\log(2KM/\delta)}{\lambda^{(1)} N_{k,t}^{(m)}}} \\
&\overset{(b)}{\geqslant} \frac{\Delta_k^{(\tilde{m}_k^*)}}{\sqrt{\lambda^{(\tilde{m}_k^*)}}} - \sqrt{\frac{\log(2KM/\delta)}{\lambda^{(1)} N_{k,t}^{(m)}}},
\end{aligned}$$

where inequality (a) is because that $\hat{\mu}_k^{(\tilde{m}_k^*)} \leqslant \mu_k^{(\tilde{m}_k^*)} + \sqrt{\frac{\log(2KM/\delta)}{N_{k,t}^{(m)}}}$ with a probability of at least $1 - \delta/2KM$ (therefore, with the union bound over all arm-fidelity pairs, the total failure probability of EXPLORE is upper bounded by $\delta/2$), and inequality (b) is because $\hat{\mu}_{k_*}^{(M)} - (\mu_k^{(\tilde{m}_k^*)} + \zeta^{(\tilde{m}_k^*)}) \geqslant \mu_{k_*}^{(M)} - (\mu_k^{(\tilde{m}_k^*)} + \zeta^{(\tilde{m}_k^*)}) = \Delta_k^{(\tilde{m}_k^*)}$.

To make the condition in Line 10 hold, with the above inequality, we need

$$\frac{\Delta_k^{(\tilde{m}_k^*)}}{\sqrt{\lambda^{(\tilde{m}_k^*)}}} - \sqrt{\frac{\log(2KM/\delta)}{\lambda^{(1)} N_{k,t}^{(m)}}} \geqslant 3\sqrt{\frac{\log(2KM/\delta)}{\lambda^{(1)} N_{k,t}^{(m)}}},$$

which, after rearrangement, becomes

$$N_{k,t}^{(m)} > \frac{16\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2} \frac{\log(2KM/\delta)}{\lambda^{(1)}}.$$

It means that if the above inequality holds, than the condition in Line 10 must hold. That is, except for the committed fidelity $\hat{m}_k^*$, we have

$$N_{k,t}^{(m)} \leqslant \frac{16\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2}\frac{\log(2KM/\delta)}{\lambda^{(1)}}, \text{ for any other fidelities } m \neq \hat{m}_k^*.$$

For another thing, Eq.(12) of LUCB's proof guarantees that for the selected fidelity $\hat{m}_k^*$, the number of pulling times is upper bounded as follows,

$$N_{k,t}^{(\hat{m}_k^*)} \leqslant \frac{64}{(\Delta_k^{(\hat{m}_k^*)})^2}\log\left(\frac{Lt}{\delta}\right).$$

Then, we upper bound the total budget of the algorithm as follows,

$$\begin{aligned}
\Lambda &= \sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}}\lambda^{(m)}N_{k,t}^{(m)} \\
&\leqslant \sum_{k\in\mathcal{K}}\frac{64\lambda^{(\hat{m}_k^*)}}{(\Delta_k^{(\hat{m}_k^*)})^2}\log\left(\frac{Lt}{\delta}\right) + \sum_{k\in\mathcal{K}}\sum_{m\neq\hat{m}_k^*}\frac{16\lambda^{(m)}\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2\lambda^{(1)}}\log\left(\frac{2KM}{\delta}\right) \\
&\overset{(a)}{\leqslant} \sum_{k\in\mathcal{K}}\frac{256\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2}\log\left(\frac{Lt}{\delta}\right) + \sum_{k\in\mathcal{K}}\sum_{m\neq\hat{m}_k^*}\frac{16\lambda^{(m)}\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2\lambda^{(1)}}\log\left(\frac{2KM}{\delta}\right) \\
&\leqslant \left(\sum_{k\in\mathcal{K}}\frac{256\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2} + \sum_{k\in\mathcal{K}}\sum_{m\neq\hat{m}_k^*}\frac{16\lambda^{(m)}\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2\lambda^{(1)}}\right)\log\left(\frac{Lt}{\delta}\right) \\
&\leqslant \sum_{m\in\mathcal{M}}\frac{\lambda^{(m)}}{\lambda^{(1)}}\cdot\sum_{k\in\mathcal{K}}\frac{256\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2}\log\left(\frac{Lt}{\delta}\right) \\
&\leqslant \sum_{m\in\mathcal{M}}\frac{\lambda^{(m)}}{\lambda^{(1)}}\cdot\sum_{k\in\mathcal{K}}\frac{256\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2}\log\left(\frac{L\Lambda}{\delta\lambda^{(1)}}\right) \\
&\overset{(b)}{\leqslant} \sum_{m\in\mathcal{M}}\frac{\lambda^{(m)}}{\lambda^{(1)}}\cdot\sum_{k\in\mathcal{K}}\frac{1024\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2}\log\left(\sum_{m\in\mathcal{M}}\frac{\lambda^{(m)}}{\lambda^{(1)}}\cdot\sum_{k\in\mathcal{K}}\frac{256\lambda^{(\tilde{m}_k^*)}}{(\Delta_k^{(\tilde{m}_k^*)})^2}\frac{L}{\lambda^{(1)}\delta}\right) \\
&\leqslant O\left(\sum_{m\in\mathcal{M}}\frac{\lambda^{(m)}}{\lambda^{(1)}}\cdot\tilde{H}\log\left(\sum_{m\in\mathcal{M}}\frac{\lambda^{(m)}}{\lambda^{(1)}}\cdot\tilde{H}\cdot\frac{L}{\lambda^{(1)}\delta}\right)\right),
\end{aligned}$$

where inequality (a) is due to Eq.(17), inequality (b) is due to that $\Lambda \leqslant A\log(B\Lambda) \implies \Lambda \leqslant 4A\log(AB\Lambda)$.

C.2.3 Proof for Explore-C's Upper Bound in Theorem B.1

Step 4 for Explore-C: Prove that if the events of Steps 2 and 3 do not happen, for $\Lambda > O\left(Q\log\left(\frac{KM\sqrt{Q}}{\delta(\lambda^{(1)})^2}\right)\right)$, the algorithm terminates with a probability at least $O(1-\delta/\Lambda^2)$. Denote $\bar{T} \coloneqq \lceil\frac{\Lambda}{2\lambda^{(M)}}\rceil$ and two events E_1, E_2 as follows,

$$\begin{aligned}
E_1 &\coloneqq \{\exists t \geqslant \bar{T} : \texttt{CROS}_t\}, \\
E_2 &\coloneqq \{\exists t \geqslant \bar{T}, k\in\mathcal{K} : (n_{k,t}^{(m_k^*)} > 16n_{k,t}^*) \cup (k \in \texttt{Mid}_t)\}.
\end{aligned}$$

We first upper bound the number of rounds after $\bar{T}$ as follows,

$$\begin{aligned}
\sum_{t \geqslant \bar{T}} \lambda^{(m_t)} \mathbb{1}\{\neg \texttt{TERM}_t\} &\overset{(a)}{=} \sum_{t \geqslant \bar{T}} \lambda^{(m_t)} \mathbb{1}\{\neg \texttt{TERM}_t \cap \neg \texttt{CROS}_t\} \\
&\overset{(b)}{\leqslant} \sum_{t \geqslant \bar{T}} \lambda^{(m_t)} \mathbb{1}\{(\ell_t \in \texttt{Mid}_t) \cup (u_t \in \texttt{Mid}_t)\} \\
&\leqslant \sum_{t \geqslant \bar{T}} \sum_{k \in \mathcal{K}} \lambda^{(m_t)} \mathbb{1}\{((k = \ell_t) \cup (k = u_t)) \cap (k \in \texttt{Mid}_t)\} \\
&\overset{(c)}{\leqslant} \sum_{t \geqslant \bar{T}} \sum_{k \in \mathcal{K}} \lambda^{(m_t)} \mathbb{1}\left\{((k = \ell_t) \cup (k = u_t)) \cap (n_{k,t}^{(m_k^*)} \leqslant 16 n_{k,t}^*)\right\} \\
&= \sum_{k \in \mathcal{K}} \sum_{t \geqslant \bar{T}} \lambda^{(m_t)} \mathbb{1}\left\{((k = \ell_t) \cup (k = u_t)) \cap (n_{k,t}^{(m_k^*)} \leqslant 16 n_{k,t}^*)\right\} \\
&\leqslant \sum_{k \in \mathcal{K}} \left(\sum_{\ell=1}^{m_k^*-1} \frac{\lambda^{(\ell)} \log(Lt^4/\delta)}{(\zeta^{(\ell)})^2} + 16 \lambda^{(m_k^*)} n_{k,t}^* \right) \\
&\leqslant \sum_{k \in \mathcal{K}} \left(\sum_{\ell=1}^{m_k^*-1} \frac{\lambda^{(\ell)}}{(\zeta^{(\ell)})^2} + \frac{16 \lambda^{(m_k^*)}}{\Delta_k^2} \right) \log\left(\frac{LT^4}{\delta}\right) \\
&\leqslant 4 \sum_{k \in \mathcal{K}} \left(\sum_{\ell=1}^{m_k^*-1} \frac{\lambda^{(\ell)}}{(\zeta^{(\ell)})^2} + \frac{16 \lambda^{(m_k^*)}}{\Delta_k^2} \right) \log\left(\frac{LT}{\delta}\right) \\
&\leqslant 4 \sum_{k \in \mathcal{K}} \left(\sum_{\ell=1}^{m_k^*-1} \frac{\lambda^{(\ell)}}{(\zeta^{(\ell)})^2} + \frac{16 \lambda^{(m_k^*)}}{\Delta_k^2} \right) \log\left(\frac{L\Lambda}{\lambda^{(1)}\delta}\right)
\end{aligned} \tag{18}$$

where the equation (a) is due to $\neg E_1$, the inequality (b) is due to Step 1, and the inequality (c) is due to $\neg E_2$.

Also notice that

$$\Lambda = \sum_{t < \bar{T}} \lambda^{(m_t)} + \sum_{t \geqslant \bar{T}} \lambda^{(m_t)} \mathbb{1}\{\neg \texttt{TERM}_t\} \leqslant \frac{\Lambda}{2} + \sum_{t \geqslant \bar{T}} \lambda^{(m_t)} \mathbb{1}\{\neg \texttt{TERM}_t\},$$

and, combining with Eq.(18), we have,

$$\Lambda \leqslant 8 \sum_{k \in \mathcal{K}} \left(\sum_{\ell=1}^{m_k^*-1} \frac{\lambda^{(\ell)}}{(\zeta^{(\ell)})^2} + \frac{16 \lambda^{(m_k^*)}}{\Delta_k^2} \right) \log\left(\frac{L\Lambda}{\lambda^{(1)}\delta}\right).$$

Solving the above inequality concludes the cost complexity upper bound for EXPLORE-C.

In the end of Step 4, we show that the LUCB algorithm fulfills the fixed confidence requirement. The probability that the algorithm does not terminate after spending Λ budget is upper bounded by $\mathbb{P}(E_1 \cup E_2)$. Based on Steps 2 and 3, it can be upper bounded as follows,

$$\begin{aligned}
\mathbb{P}(E_1 \cup E_2) &\leqslant \sum_{t > \bar{T}} \left(\frac{KM\delta}{Lt^3} + \frac{16\delta \sum_{k \in \mathcal{K}} \Delta_k^{-2}}{Lt^4} \right) \leqslant \Lambda \left(\frac{1}{\lambda^{(1)}} - \frac{1}{2\lambda^{(M)}} \right) \left(\frac{\delta}{(\Lambda/\lambda^{(1)})^3} \right) \sum_{k \in \mathcal{K}} \Delta_k^{-2} \\
&\leqslant \frac{\delta}{\Lambda^2} \sum_{k \in \mathcal{K}} \Delta_k^{-2} \left((\lambda^{(1)})^2 - \frac{(\lambda^{(1)})^3}{2\lambda^{(M)}} \right) \\
&\leqslant \delta.
\end{aligned}$$

$\square$

D Detailed Theorems for Regret Minimization Case

We first present both the problem-independent (worst-case) and problem-dependent regret lower bounds in Section D.1 and then devise an elimination algorithm whose worst-case upper bounds match the worst-case lower bound up to some logarithmic factors and whose problem-dependent upper bound matches the problem-dependent lower bound in a class of `MF-MAB` in Section D.2.

D.1 Regret Lower Bound

We present the problem-independent regret lower bound in Theorem D.1 and the problem-dependent regret lower bound in Theorem D.2. Both proofs are deferred to Appendix E.1 and E.2 respectively.

Theorem D.1 (Problem-independent regret lower bound)**.** *Given budget Λ, the regret of* `MF-MAB` *is lower bounded as follows,*

$$\inf_{Algo} \sup_{\mathcal{I}} \mathbb{E}\left[R(\Lambda)\right] \geqslant \Omega\left(K^{1/3}\Lambda^{2/3}\right),$$

where the inf *is over any algorithms, the* sup *is over any possible* `MF-MAB` *instances $\mathcal{I}$.*

Theorem D.2 (Problem-dependent lower bound)**.** *For any consistent policy that, after spending Λ budgets, fulfills that for any suboptimal arm k (with $\Delta_k^{(M)} > 0$) and any $a > 0$, $\mathbb{E}[N_k^{(\forall m)}(\Lambda)] = o(\Lambda^a)$, its regret is lower bounded by the following inequality,*

$$\liminf_{\Lambda\to\infty} \frac{\mathbb{E}\left[R(\Lambda)\right]}{\log(\Lambda)} \geqslant \sum_{k\in\mathcal{K}} \min_{m:\Delta_k^{(m)}>0} \left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right) \frac{C}{(\Delta_k^{(m)})^2}.$$

The above two lower bound proofs utilize two different regret decomposition as follows,

$$\begin{aligned} R(\Lambda) &\overset{(a)}{\geqslant} \sum_{t=1}^{N} \left(\frac{\lambda^{(m_t)} - \lambda^{(1)}}{\lambda^{(1)}}\mu_1 + \Delta_{I_t}^{(M)}\right) \\ &= \sum_{m=2}^{M} N_{\forall k}^{(m)}(\Lambda)\frac{\lambda^{(m)} - \lambda^{(1)}}{\lambda^{(1)}}\mu_1 + \sum_{k\neq k_1^{(M)}} N_k^{(\forall m)}(\Lambda)\Delta_k^{(M)}; \end{aligned} \tag{19}$$

$$R(\Lambda) = \sum_{k\in\mathcal{K}} \sum_{m=1}^{M} N_k^{(m)}(\Lambda) \left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right), \tag{20}$$

where the inequality (a) is due to $\Lambda > \sum_{t=1}^{N} \lambda^{(m_t)}$, the $N_k^{(m)}(\Lambda)$ is the number of times that arm k is pulled in fidelity m after paying budget Λ. The N with subscript $_{\forall k}$ and superscript $^{(\forall m)}$ mean the pulling times of all arms and all fidelities respectively. Due to the multi-fidelity feedback, both decompositions are different from classic bandits' regret decomposition [24, Lemma 4.5], and, therefore, we need to non-trivially extend the known approaches to our scenario.

To prove the problem-independent lower bound $\Omega(K^{1/3}\Lambda^{2/3})$ in Theorem D.1, we utilize the decomposition in Eq.(19): In the RHS, the first term increases whether one choose fidelity other than the lowest; this is a novel term. The second term of RHS corresponds to the cost of pulling suboptimal arms which also appears in the classic MAB. With this observation, one can construct instance pairs such that it is unavoidable to do exploration at higher fidelities and, therefore, the first term is not negligible. Lastly, one needs to balance the magnitude of the above two terms case-by-case, which together bounds the regret as $\Omega\left(\Lambda^{2/3}\right)$. To prove the problem-dependent lower bound in Theorem E.1, we utilize Eq.(20) to decompose the regret to each arm and bound each of them separately.

D.2 An Elimination Algorithm and Its Regret Upper Bound

In this section, we propose an elimination algorithm for `MF-MAB` based on Auer and Ortner [1]. This algorithm proceeds in phases $p = 0, 1, \ldots$ and maintains a candidate arm set $\mathcal{C}_p$. The set $\mathcal{C}_p$ is initialized as the full arm set $\mathcal{K}$ and the algorithm gradually eliminates arms from the set until there is only one arm remaining. When the candidate arm set contains more than one arms, the algorithm explores arms with the highest fidelity M, and when the set $|\mathcal{C}_p| = 1$, the algorithm exploits the singleton in the set with the lowest fidelity $m = 1$. We present the detail in Algorithm 4.

Algorithm 4 Elimination for `MF-MAB`

```
Input: full arm set 𝒦, budget Λ, and parameter ε
Initialization: phase p ← 0, candidate set 𝒞_p ← 𝒦
while p < log_2 (2/ε) and |𝒞_p| > 1 do
    pull each arm k ∈ 𝒞_p in highest fidelity M such that T_k^(M) = ⌈2^{2p} log (Λ / (2^{2p} λ^(M)))⌉
    Update reward means μ̂_{k,p}^(M) for all arms k ∈ 𝒞_p
    𝒞_{p+1} ← {k ∈ 𝒞_p : μ̂_{k,p}^(M) + 2^{-p+1} > max_{k'∈𝒞_p} μ̂_{k',p}^(M)}        ▷ Elimination
    p ← p + 1
Pull the remaining arms of 𝒞_p in turn in fidelity m = 1 until the budget runs up
```

D.2.1 Analysis Results

We first present the problem-dependent regret upper bound of Algorithm 4 in Theorem D.3. Its full proof is deferred to Appendix E.3.

Theorem D.3 (Problem-Dependent Regret Upper Bound)**.** *For any $\varepsilon > 0$. Algorithm 4's regret is upper bounded as follows,*

$$\begin{aligned}\mathbb{E}[R(\Lambda)] \leqslant & \max_{k:\Delta_k^{(M)}\leqslant\varepsilon} \frac{\Lambda}{\lambda^{(1)}}\Delta_k^{(M)} \\ &+ \sum_{k:\Delta_k^{(M)}>\varepsilon}\left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right)\left(\frac{16}{(\Delta_k^{(M)})^2}\log\frac{\Lambda(\Delta_k^{(M)})^2}{16\lambda^{(M)}} + \frac{48}{(\Delta_k^{(M)})^2} + 1\right) + \frac{64}{\Delta_k^{(M)}}\right) \\ &+ \sum_{k:\Delta_k^{(M)}\leqslant\varepsilon}\left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right)\left(\frac{16}{\varepsilon^2}\log\left(\frac{\Lambda\varepsilon^2}{16\lambda^{(M)}}\right) + \frac{32}{3\varepsilon^2} + 1\right) + \frac{64}{\varepsilon}.\right)\end{aligned} \tag{21}$$

Especially, if letting ε go to zero and budget Λ go to infinity, the above upper bound becomes

$$\limsup_{\Lambda\to\infty}\frac{\mathbb{E}\left[R(\Lambda)\right]}{\log(\Lambda)} \leqslant \sum_{k\in\mathcal{K}}\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right)\frac{16}{(\Delta_k^{(M)})^2}. \tag{22}$$

Letting $\varepsilon = (K\log\Lambda/\Lambda)^{1/3}$ in Eq.(21) of Theorem D.3, one can obtain a problem-independent regret upper bound of Algorithm 4 in Theorem D.4 as follows.

Theorem D.4 (Regret Upper Bound for Algorithm 4)**.** *Letting $\varepsilon = (K\log\Lambda/\Lambda)^{1/3}$, Algorithm 4's regret is upper bounded as follows,*

$$\mathbb{E}[R(\Lambda)] \leqslant O\left(K^{1/3}\Lambda^{2/3}(\log\Lambda)^{1/3}\right).$$

E Proofs for Regret Minimization Results

E.1 Proof of Theorem D.1

Step 1. Construct instances and upper bound KL-divergence Fix a policy π. We construct two `MF-MAB` instances, each with K arms and M fidelities. For the pulling costs of different fidelities, we set $\lambda^{(M)} \leqslant 2\lambda^{(1)}$. For reward feedback, we assume all arms' reward distributions at any fidelity are Bernoulli, and denote $\mu_k^{(m)}(1), \mu_k^{(m)}(2)$ as the reward means of these two instances. Let $\mathbb{P}_1 = \mathbb{P}_{\boldsymbol{\mu}(1),\pi}, \mathbb{E}_1 = \mathbb{E}_{\boldsymbol{\mu}(1),\pi}$ and $\mathbb{P}_2 = \mathbb{P}_{\boldsymbol{\mu}(2),\pi}, \mathbb{E}_2 = \mathbb{E}_{\boldsymbol{\mu}(2),\pi}$ be the probability measures and expectations on the canonical `MF-MAB` model induced by Λ-budget interconnection of π and $\boldsymbol{\mu}_1$ (and $\boldsymbol{\mu}_2$). Denote $k' = \arg\min_{k\in\mathcal{K}}\mathbb{E}_1[N_k^{(m>1)}(\Lambda)]$.

The only difference between both instance pair is in the arm k''s reward mean for fidelities $m > 1$, so that instance 1's optimal arm is arm 1 and instance 2's optimal arm is arm k'. The detailed reward means are listed as follows,

	$m = 1$	$m > 1$
$(\mu_k^{(m)}(1))_{k\in\mathcal{K}}$	$\left(\frac{1}{2}+\Delta, \frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right)$	$\left(\frac{1}{2}+\Delta, \frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right)$
$(\mu_k^{(m)}(2))_{k\in\mathcal{K}}$		$\left(\frac{1}{2}+\Delta, \frac{1}{2}, \dots, \frac{1}{2}, \frac{1}{2}+2\Delta, \frac{1}{2}, \dots, \frac{1}{2}\right)$

Denote the entry (M_t, I_t, X_t) as a tuple of the pulled fidelity, pulled arm, and observed reward random variables at time t, and $\mathcal{H} \coloneqq (M_1, I_1, X_1; M_2, I_2, X_2; \dots; M_N, I_N, X_N)$ as a sequence of applying policy π. We note that N is also a random variable depending on the sequence of fidelities in pulling arms. Next, we calculate the upper bound of the KL-divergence of the above two instances in this sequence.

$$
\begin{aligned}
\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_2) &= \mathbb{E}_1\left[\log\left(\frac{d\mathbb{P}_1}{d\mathbb{P}_2}\right)\right] \\
&= \mathbb{E}_1\left[\log\left(\frac{d\mathbb{P}_1}{d\mathbb{P}_2}(M_1, I_1, X_1; M_2, I_2, X_2; \dots; M_N, I_N, X_N)\right)\right] \\
&= \mathbb{E}_1\left[\mathbb{E}_1\left[\log\left(\frac{d\mathbb{P}_1}{d\mathbb{P}_2}(M_1, I_1, X_1; M_2, I_2, X_2; \dots; M_N, I_N, X_N)\right)\middle| N\right]\right] \\
&\overset{(a)}{=} \mathbb{E}_1\left[\mathbb{E}_1\left[\sum_{t=1}^{N}\log\left(\frac{p_1(X_t|M_t, I_t)}{p_2(X_t|M_t, I_t)}\right)\middle| N\right]\right] \\
&= \mathbb{E}_1\left[\sum_{t=1}^{N}\mathbb{E}_1\left[\log\left(\frac{p_1(X_t|M_t, I_t)}{p_2(X_t|M_t, I_t)}\right)\middle| N\right]\right] \\
&\overset{(b)}{=} \mathbb{E}_1\left[\sum_{t=1}^{N}\mathbb{E}_1\left[\mathrm{KL}\left(P_1^{(M_t,I_t)}, P_2^{(M_t,I_t)}\right)\middle| N\right]\right] \\
&= \sum_{(m,k)\in\mathcal{M}\times\mathcal{K}} \mathbb{E}_1\left[\sum_{t=1}^{N}\mathbb{E}_1\left[\mathbb{1}\{M_t = m, I_t = k\}\,\mathrm{KL}\left(P_1^{(M_t,I_t)}, P_2^{(M_t,I_t)}\right)\middle| N\right]\right] \\
&= \sum_{(m,k)\in\mathcal{M}\times\mathcal{K}} \mathbb{E}_1\left[\mathrm{KL}\left(P_1^{(m,k)}, P_2^{(m,k)}\right)\sum_{t=1}^{N}\mathbb{E}_1\left[\mathbb{1}\{M_t = m, I_t = k\}| N\right]\right] \\
&= \sum_{(m,k)\in\mathcal{M}\times\mathcal{K}} \mathrm{KL}\left(P_1^{(m,k)}, P_2^{(m,k)}\right)\mathbb{E}_1\left[\mathbb{E}_1\left[\sum_{t=1}^{N}\mathbb{1}\{M_t = m, I_t = k\}\middle| N\right]\right] \\
&\overset{(c)}{=} \sum_{(m,k)\in\mathcal{M}\times\mathcal{K}} \mathrm{KL}\left(P_1^{(m,k)}, P_2^{(m,k)}\right)\mathbb{E}_1\left[T_k^{(m)}(\Lambda)\right] \\
&= \mathrm{KL}\left(P_1^{(m,2)}, P_2^{(m,2)}\right)\mathbb{E}_1\left[T_{k'}^{(m>1)}(\Lambda)\right] \\
&\overset{(d)}{\leqslant} \frac{1}{K}\cdot\mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right]\cdot\mathrm{KL}\left(\mathcal{B}\left(\frac{1}{2}\right), \mathcal{B}\left(\frac{1}{2}+2\Delta\right)\right) \\
&\overset{(e)}{\leqslant} \mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right]\cdot\frac{9\Delta^2}{K},
\end{aligned}
\tag{23}
$$

where the equation (a) is due to

$$
\begin{aligned}
&\frac{d\mathbb{P}_1}{d(\rho\times\rho\times\lambda)^N}(m_1,k_1,x_1;m_2,k_2,x_2;\ldots;m_N,k_N,x_N)\\
&=p_{\boldsymbol{\mu}_1,\pi}(m_1,k_1,x_1;m_2,k_2,x_2;\ldots;m_N,k_N,x_N)\\
&=\prod_{t=1}^{N}\pi_t(m_t,k_t|m_1,k_1,x_1;\ldots;m_{t-1},k_{t-1},x_{t-1})p_{\boldsymbol{\mu}_1}(x_t|m_t,k_t),
\end{aligned}
$$

the equation (b) is due to

$$
\begin{aligned}
\mathbb{E}_1\left[\log\left(\frac{p_1(X_t|M_t,I_t)}{p_2(X_t|M_t,I_t)}\right)\middle| N\right] &= \mathbb{E}_1\left[\mathbb{E}_1\left[\log\left(\frac{p_1(X_t|M_t,I_t)}{p_2(X_t|M_t,I_t)}\right)\middle| M_t,I_t\right]\middle| N\right]\\
&= \mathbb{E}_1\left[\mathrm{KL}\left(P_1^{(M_t,I_t)},P_2^{(M_t,I_t)}\right)\middle| N\right],
\end{aligned}
$$

the equation (c) is due to the tower property as well, the inequality (d) is because arm k' is the arm with the smallest number of pulled with fidelity $m>1$, and the inequality (e) is by calculating the KL-divergent between two Bernoulli distributions.

Step 2. Lower bound the regret We first note that the regret defined in Eq.(9) for instance 1 can be decomposed as follows,

$$
\begin{aligned}
\mathbb{E}_1[R(\Lambda)] &\overset{(a)}{\geqslant} \sum_{t=1}^{N}\left(\frac{\lambda^{(m_t)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1+\Delta_{I_t}^{(M)}\right)\\
&=\sum_{m=2}^{M}T_{\forall k}^{(m)}(\Lambda)\frac{\lambda^{(m)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1+\sum_{k\neq k_1^{(M)}}T_k^{(\forall m)}(\Lambda)\Delta_k^{(M)}
\end{aligned}
$$

where inequality (a) is due to $\Lambda>\sum_{t=1}^{N}\lambda^{(m_t)}$, the $T_k^{(m)}(\Lambda)$ is the number of times that arm k is pulled in fidelity m (given total budget Λ).

Then, we lower bound the summation of the regrets under both instances,

$$
\begin{aligned}
&\mathbb{E}_1[R(\Lambda)] + \mathbb{E}_2[R(\Lambda)] \\
&= \mathbb{E}_1\left[\sum_{m=2}^{M} T_{\forall k}^{(m)}(\Lambda)\frac{\lambda^{(m)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1 + \sum_{k\neq k_1^{(M)}} T_k^{(\forall m)}(\Lambda)\Delta_k^{(M)}\right] \\
&\quad + \mathbb{E}_2\left[\sum_{m=2}^{M} T_{\forall k}^{(m)}(\Lambda)\frac{\lambda^{(m)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1 + \sum_{k\neq k_1^{(M)}} T_k^{(\forall m)}(\Lambda)\Delta_k^{(M)}\right] \\
&\geqslant \mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\frac{\lambda^{(2)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1 + \sum_{k\neq k_1^{(M)}} T_k^{(\forall m)}(\Lambda)\Delta_k^{(M)}\right] \\
&\quad + \mathbb{E}_2\left[T_{\forall k}^{(m>1)}(\Lambda)\frac{\lambda^{(2)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1 + \sum_{k\neq k_1^{(M)}} T_k^{(\forall m)}(\Lambda)\Delta_k^{(M)}\right] \\
&\overset{(a)}{\geqslant} \mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right]\frac{\lambda^{(2)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1 \\
&\quad + \Delta\min\left\{\frac{\Lambda}{\lambda^{(M)}} - \frac{\Lambda}{2\lambda^{(1)}}, \frac{\Lambda}{2\lambda^{(1)}}\right\}\left(\mathbb{P}_1\left(T_1^{(\forall m)} \leqslant \frac{\Lambda}{2\lambda^{(1)}}\right) + \mathbb{P}_2\left(T_1^{(\forall m)} > \frac{\Lambda}{2\lambda^{(1)}}\right)\right) \\
&\overset{(b)}{\geqslant} \mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right]\frac{\lambda^{(2)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1 + \frac{\Lambda\Delta}{2}\min\left\{\frac{1}{\lambda^{(M)}} - \frac{1}{2\lambda^{(1)}}, \frac{1}{2\lambda^{(1)}}\right\}\exp(-\operatorname{KL}(\mathbb{P}_1,\mathbb{P}_2)) \\
&\overset{(c)}{\geqslant} \mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right]\frac{\lambda^{(2)}-\lambda^{(1)}}{\lambda^{(1)}}\mu_1 \\
&\qquad + \frac{\Lambda\Delta}{2}\min\left\{\frac{1}{\lambda^{(M)}} - \frac{1}{2\lambda^{(1)}}, \frac{1}{2\lambda^{(1)}}\right\}\exp\left(-2\Delta^2 K^{-1}\mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right]\right),
\end{aligned}
$$

where inequality (a) uses the $\lambda^{(M)} \leqslant 2\lambda^{(1)}$ condition, inequality (b) is by Bretagnolle-Huber inequality [24, Theorem 14.2], and inequality (c) is by Eq.(23).

Step 3. Obtain the $\Omega(K^{\frac{1}{3}}\Lambda^{\frac{2}{3}})$ lower bound We show that $\mathbb{E}_1[R(\Lambda)] + \mathbb{E}_2[R(\Lambda)] \geqslant \Omega(K^{\frac{1}{3}}\Lambda^{\frac{2}{3}})$ for any possible quantity of $\mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right]$ via categorized discussion as follows,

- Case 1: If $\mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right] = 0$, then the last formula becomes $C\Lambda\Delta$. Letting Δ be a constant (via sup), we have a $\Omega(\Lambda)$ lower bound, which means $\Omega(K^{\frac{1}{3}}\Lambda^{\frac{2}{3}})$ is also valid.
- Case 2: If $\mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right] \geqslant K^{\frac{1}{3}}\Lambda^{\frac{2}{3}}$, then we have the $\Omega(K^{\frac{1}{3}}\Lambda^{\frac{2}{3}})$ lower bound from the first term.
- Case 3: If $0 < \mathbb{E}_1\left[T_{\forall k}^{(m>1)}(\Lambda)\right] < K^{\frac{1}{3}}\Lambda^{\frac{2}{3}}$, we choose $\Delta = K^{\frac{1}{3}}\Lambda^{-\frac{1}{3}}$ and also obtain the $\Omega(K^{\frac{1}{3}}\Lambda^{\frac{2}{3}})$ lower bound.

E.2 Proof of Theorem D.2

In this proof, we prove that for any arm $k \in \mathcal{K}$, the total regret due to this arm k is lower bounded as follows,

$$
\liminf_{\Lambda\to\infty} \frac{\mathbb{E}\left[R_k(\Lambda)\right]}{\log(\Lambda)} \geqslant \min_{m:\Delta_k^{(m)}>0}\left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right)\frac{C}{(\Delta_k^{(m)})^2}.
$$

Then, noticing that $R(\Lambda) = \sum_{k\in\mathcal{M}} R_k(\Lambda)$ concludes the proof.

We construct instances ν and ν'. We set the reward distributions $\nu = (\nu_k^{(m)})_{(k,m)\in\mathcal{K}\times\mathcal{M}}$ as Bernoulli and the reward means fulfill $\mu_1^{(M)} > \mu_2^{(M)} \geqslant \mu_3^{(M)} \geqslant \ldots \geqslant \mu_K^{(M)}$, where $\mu_k^{(m)} = \mathbb{E}_{X\sim\nu_k^{(m)}}[X]$. We let ${\nu'}_k^{(m)}$ be the same to $\nu_k^{(m)}$ for all k and m, except for that an arm $\ell \neq 1$. We set arm ℓ's reward means on fidelities $m \in \mathcal{M}_k$ to be ${\nu'}_\ell^{(m)} = \nu_1^{(M)} - \zeta^{(m)} + \epsilon$. One can verify that the reward means of arm k under instance v' fulfill the condition that $|{\mu'}_k^{(m)} - {\mu'}_k^{(M)}| \leqslant \zeta^{(m)}$ for any fidelity m. Notice that ${\mu'}_k^{(M)} > \mu_{k_*}^{(M)}$. Hence, under instance $\mathcal{I}'$, the optimal arm is k. We denote $\mathbb{P}, \mathbb{E}$ and $\mathbb{P}', \mathbb{E}'$ as the probability measures and expectations for instances $\mathcal{I}$ and $\mathcal{I}'$ respectively.

Next, we employ the (extended) key inequality from Garivier et al. [12] as follows,

$$\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}} \mathbb{E}[N_k^{(m)}(\Lambda)]\,\mathrm{KL}(v_k^{(m)}, {v'}_k^{(m)}) \geqslant \mathrm{kl}(\mathbb{E}[Z], \mathbb{E}'[Z]).$$

Let $Z = \lambda^{(M)} N_\ell^{(\forall m)}(\Lambda)/\Lambda$ in the above inequality, we have

$$\begin{aligned}
&\sum_{m:\Delta_\ell^{(m)}>0} \mathbb{E}[N_k^{(m)}(\Lambda)]\,\mathrm{KL}(v_\ell^{(m)}, {v'}_\ell^{(m)})\\
&\geqslant \mathrm{kl}\left(\frac{\mathbb{E}[N_\ell^{(\forall m)}(\Lambda)]}{\Lambda}, \frac{\mathbb{E}'[N_\ell^{(\forall m)}(\Lambda)]}{\Lambda}\right) \qquad (24)\\
&\overset{(a)}{\geqslant} \left(1 - \frac{\lambda^{(M)}\mathbb{E}[N_\ell^{(\forall m)}(\Lambda)]}{\Lambda}\right)\log\frac{\Lambda}{\Lambda - \lambda^{(M)}\mathbb{E}'[N_\ell^{(\forall m)}(\Lambda)]} - \log 2
\end{aligned}$$

where inequality (a) is due to that for all $(p, q) \in [0, 1]^2$,$\mathrm{kl}(p, q) \geqslant (1-p)\log(1/(1-q)) - \log 2$.

Notice that the regret attributed to any arm k can be decomposed and lower bounded as follows,

$$R_\ell(\Lambda) \geqslant \sum_{m=1}^{M} N_\ell^{(m)}(\Lambda)\left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_\ell^{(M)}\right),$$

with the constraint in Eq.(24). Since this is a linear programming, we know its solution is reached at its vertex. Therefore, we lower bound the regret as follows,

$$\begin{aligned}
R_\ell(\Lambda) \geqslant &\min_{m:\Delta_k^{(m)}>0}\left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_\ell^{(M)}\right)\frac{1}{\mathrm{KL}(v_\ell^{(m)}, {v'}_\ell^{(m)})}\\
&\times\left(\left(1 - \frac{\lambda^{(M)}\mathbb{E}[N_\ell^{(\forall m)}(\Lambda)]}{\Lambda}\right)\log\frac{\Lambda}{\Lambda - \lambda^{(M)}\mathbb{E}'[N_\ell^{(\forall m)}(\Lambda)]} - \log 2\right),
\end{aligned} \qquad (25)$$

Notice that the policy is consistent, that is, $\mathbb{E}[N_\ell^{(\forall m)}(\Lambda)] = o(T^a)$ and $\mathbb{E}'[N_k^{(\forall m)}(\Lambda)] = o(\Lambda^a)$ for any $a \in (0, 1]$ and any suboptimal arm $k \neq \ell$. We have $\frac{\lambda^{(M)}\mathbb{E}[N_\ell^{(\forall m)}(\Lambda)]}{\Lambda} = o(1)$ and

$$\Lambda - \lambda^{(M)}\mathbb{E}'[N_\ell^{(\forall m)}(\Lambda)] \leqslant \lambda^{(M)}\sum_{k\neq\ell}\mathbb{E}'[N_k^{(\forall m)}(\Lambda)] = o(\Lambda^a).$$

Dividing both sides of Eq.(25) by Λ, and letting Λ go to infinity and a go to 1, we have

$$\begin{aligned}
\liminf_{\Lambda\to\infty}\frac{\mathbb{E}[R_\ell(\Lambda)]}{\Lambda} &\geqslant \min_{m:\Delta_\ell^{(m)}>0}\left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_\ell^{(M)}\right)\frac{1}{\mathrm{KL}(v_\ell^{(m)}, {v'}_\ell^{(m)})}\\
&\geqslant \min_{m:\Delta_\ell^{(m)}>0}\left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_\ell^{(M)}\right)\frac{1}{\mathrm{KL}(v_\ell^{(m)}, v_1^{(M)} - \zeta^{(m)} + \epsilon)}
\end{aligned}$$

To bound the optimal arm 1's sample cost, we use the same ν as above and construct another instance ν''. The instance ν'''s reward means are the same to ν except for arm 1 whose reward means for fidelity $m \in \mathcal{M}_1$ are set as ${\mu''}_1^{(m)} = \mu_2^{(m)} + \zeta^{(m)} - \epsilon$. Then, with similar procedure as the above, we obtain

$$\liminf_{\Lambda\to\infty}\frac{\mathbb{E}[R_1(\Lambda)]}{\Lambda} \geqslant \min_{m:\Delta_k^{(m)}>0}\left(\frac{\lambda^{(m)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_1^{(M)}\right)\frac{1}{\mathrm{KL}(v_1^{(m)}, v_2^{(m)} + \zeta^{(m)} - \epsilon)}$$

E.3 Proof of Theorem D.4

We first prove a problem dependent regret upper bound as follows and then convert this bound to the problem independent regret upper bound presented in Theorem D.4.

Denote $\Delta_k^{(M)} = \mu_1^{(M)} - \mu_k^{(M)}$, and, especially, $\Delta_1^{(M)} = 0$.

Theorem E.1 (Problem-Dependent Regret Upper Bound)**.** *For any* $\varepsilon > 0$*. Algorithm 4's regret is upper bounded as follows,*

$$\mathbb{E}[R(\Lambda)] \leqslant \sum_{k:\Delta_k^{(M)}>\varepsilon} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right)\left(\frac{16}{(\Delta_k^{(M)})^2}\log\frac{\Lambda(\Delta_k^{(M)})^2}{16\lambda^{(M)}} + \frac{48}{(\Delta_k^{(M)})^2} + 1\right) + \frac{64}{\Delta_k^{(M)}}\right)$$
$$+ \sum_{k:\Delta_k^{(M)}\leqslant\varepsilon} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right)\left(\frac{16}{\varepsilon^2}\log\left(\frac{\Lambda\varepsilon^2}{16\lambda^{(M)}}\right) + \frac{32}{3\varepsilon^2} + 1\right) + \frac{64}{\varepsilon}\right) + \max_{k:\Delta_k^{(M)}\leqslant\varepsilon}\frac{\Lambda}{\lambda^{(1)}}\Delta_k^{(M)}. \tag{26}$$

Proof of Theorem E.1. By Hoeffding's inequality, we have

$$\mathbb{P}\left(\hat{\mu}_{k,p}^{(M)} > \mu_k^{(M)} + 2^{-p}\right) \leqslant \exp\left(-\frac{2^{-2p}}{2\times\frac{1}{2\times 2^{2p}\log(\Lambda/2^{2p}\lambda^{(M)})}}\right) = \frac{2^{2p}\lambda^{(M)}}{\Lambda}$$
$$\mathbb{P}\left(\hat{\mu}_{k,p}^{(M)} < \mu_k^{(M)} - 2^{-p}\right) \leqslant \exp\left(-\frac{2^{-2p}}{2\times\frac{1}{2\times 2^{2p}\log(\Lambda/2^{2p}\lambda^{(M)})}}\right) = \frac{2^{2p}\lambda^{(M)}}{\Lambda}.$$

That is, the empirical mean $\hat{\mu}_k^{(M)}$ is within the confidence interval $(\mu_k^{(M)} - 2^{-p}, \mu_k^{(M)} + 2^{-p})$ with high probability.

Choose any $\varepsilon > \frac{e}{\Lambda}$. Let $\mathcal{K}' = \{k \in \mathcal{K} | \Delta_k^{(M)} > \varepsilon\}$. Denote $p_k := \min\{p : 2^{-p} < \frac{\Delta_k^{(M)}}{2}\}$. From p_k's definition, we have the following inequality

$$2^{p_k} < \frac{4}{\Delta_k^{(M)}} < 2^{p_k+1}. \tag{27}$$

We also note that the cost of pulling a suboptimal arm $k \in \mathcal{K}'$ at highest fidelity M is upper bounded as $\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}$, where the factor $\frac{\lambda^{(M)}}{\lambda^{(1)}}$ is because the budget paying to pull an arm at fidelity M can be used to the the arm at fidelity 1 for fractional times.

The rest of this proof consists of two steps. In the first step, we assume that all empirical means are in their corresponding confidence intervals at each phases, and show the algorithm can *properly* eliminate all suboptimal arms in $\mathcal{K}'$—the arm is eliminated in or before the phase p_k. In the second step, we upper bound the regret if there are any empirical estimates lying outside their corresponding confidence intervals.

Step 1. If all arms' empirical means lie in confidence intervals. That is, any suboptimal arm k is eliminated in or before the phase p_k. Because, if $k \in \mathcal{C}_{p_k}$, we have

$$\begin{aligned}\hat{\mu}_k^{(M)} &\leqslant \mu_k^{(M)} + 2^{-p_k} = \mu_1^{(M)} - \Delta_k^{(M)} + 2^{-p_k} \\ &\overset{(a)}{\leqslant} \mu_1^{(M)} - 4\times 2^{-p_k-1} + 2^{-p_k} = \mu_1^{(M)} - 2^{-p_k} < \hat{\mu}_1^{(M)} \leqslant \max_{k\in\mathcal{C}_{p_k}} \hat{\mu}_k^{(M)},\end{aligned}$$

where (a) is due to Eq.(27). That is, if this arm k haven't been eliminated before phase p_k, it must be eliminated in this phase. Therefore, the total pulling times of this arm k at highest fidelity M is upper bounded as follows,

$$T_k^{(M)} \leqslant \left\lceil 2^{2p_k}\log\frac{\Lambda}{2^{2p_k}\lambda^{(M)}}\right\rceil \overset{(a)}{\leqslant} \frac{16}{(\Delta_k^{(M)})^2}\log\left(\frac{(\Delta_k^{(M)})^2\Lambda}{16\lambda^{(M)}}\right) + 1, \tag{28}$$

where (a) is due to Eq.(27).

We then handle the number of times of pulling arms k with $\Delta_k^{(M)} \leqslant \varepsilon$ in fidelity M. Although these arms' total pulling times, eliminated in or before phase p_k, are also upper bounded by Eq.(28), their corresponding phases p_k is greater than $\log_2 \frac{2}{\varepsilon}$ and, therefore, cannot be reached. So, these arms' (including the optimal arm 1's) total pulling times in fidelity M is upper bounded by

$$\frac{16}{\varepsilon^2} \log\left(\frac{\Lambda\varepsilon^2}{16\lambda^{(M)}}\right) + 1.$$

Therefore, the cost due to pulling arms at fidelity M is upper bounded as follows,

$$\sum_{k\in\mathcal{K}} \left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right)\left(\frac{16}{\left(\max\left\{\varepsilon, \Delta_k^{(M)}\right\}\right)^2}\log\left(\frac{\left(\max\left\{\varepsilon, \Delta_k^{(M)}\right\}\right)^2\Lambda}{16\lambda^{(M)}}\right)+1\right). \tag{29}$$

After the elimination process, arms with $\Delta_k^{(M)} \leqslant \varepsilon$ may remain in the candidate arm set $\mathcal{C}_p$ and are exploited in turn at fidelity $m = 1$. As some of them are not the optimal arm, this additional cost can be upper bounded as follows,

$$\max_{k:\Delta_k^{(M)}\leqslant\varepsilon} \frac{\Lambda}{\lambda^{(1)}}\Delta_k^{(M)}. \tag{30}$$

Step 2. There are two cases that some arms are eliminated improperly: for an suboptimal arm k, either

2.1 The suboptimal arm k is *not* eliminated in (or before) the phase p_k, and the optimal arm 1 is in $\mathcal{C}_{p_k}$ in phase p_k; or

2.2 The suboptimal arm k is eliminated in (or before) the phase p_k, and the optimal arm 1 is *not* in $\mathcal{C}_{p_k}$ in phase p_k.

Case 2.1's happening means that arm k is not eliminated in or before phase p_k, which can only happen when the arm's empirical mean $\hat{\mu}_k^{(M)}$ lies outside its corresponding confidence interval. This event in or before phase p_k is with a probability no greater than $2 \times \frac{2^{2p_k}\lambda^{(M)}}{\Lambda}$. Since this event may happen to any suboptimal arm $k \in \mathcal{K}'$, then the regret of this case is upper bounded by

$$\begin{aligned}\sum_{k\in\mathcal{K}'} \frac{2^{2p_k+1}\lambda^{(M)}}{\Lambda}\frac{\Lambda}{\lambda^{(M)}}\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right) &= \sum_{k\in\mathcal{K}'} 2^{2p_k+1}\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right) \\ &\overset{(a)}{\leqslant} \sum_{k\in\mathcal{K}'} \frac{32}{(\Delta_k^{(M)})^2}\left(\frac{\lambda^{(M)}}{\lambda^{(1)}}\mu_1^{(M)} - \mu_k^{(M)}\right),\end{aligned} \tag{31}$$

where (a) is due to Eq.(27).

If Case 2.2 happens, the optimal arm 1 is not in the candidate arm set $\mathcal{C}_{p_k}$ in phase p_k. We denote p_1 as the phase that the optimal arm 1 is eliminated, and it is at this phase that some arms' empirical means lie outside their confidence interval so that this mis-elimination happens. The probability of this event is upper bounded by $2 \times \frac{2^{2p_1}\lambda^{(M)}}{\Lambda}$. We assume that arms k with $p_k < p_1$ are eliminated in or before phase p_i properly; otherwise, the regret is counted in Case 2.1. Therefore, the optimal arm 1 eliminated in phase p_1 should be eliminated by an arm k with $p_k \geqslant p_1$. Consequently, the maximal per time slot regret in Case 2.2 is among arms with $p_k \geqslant p_1$. Denote $p_\varepsilon := \min\{p | 2^{-p} < \frac{\varepsilon}{2}\}$. We bound the cost of Case 2.2 as follows,

$$
\begin{aligned}
&\sum_{p_1=0}^{\max_{i\in\mathcal{K}'} p_i} \sum_{k>1: p_k \geqslant p_1} \frac{2^{2p_1+1}\lambda^{(M)}}{\Lambda} \frac{\Lambda}{\lambda^{(M)}} \cdot \max_{k'>1:p_{k'}\geqslant p_1} \left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_{k'}^{(M)} \right) \\
=& \sum_{p_1=0}^{\max_{i\in\mathcal{K}'} p_i} \sum_{k>1: p_k \geqslant p_1} 2^{2p_1+1} \cdot \left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} + \max_{k'>1:p_{k'}\geqslant p_1} \Delta_{k'}^{(M)} \right) \\
\overset{(a)}{\leqslant}& \sum_{p_1=0}^{\max_{i\in\mathcal{K}'} p_i} \sum_{k>1: p_k \geqslant p_1} 2^{2p_1+1} \cdot \left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} + 4 \times 2^{-p_1} \right) \\
\overset{(b)}{\leqslant}& \sum_{k>1} \sum_{p_1=0}^{\min\{p_k, p_\varepsilon\}} 2^{2p_1+1} \cdot \left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} + 4 \times 2^{-p_1} \right) \\
\leqslant& \sum_{k>1} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} \right) \sum_{p_1=0}^{\min\{p_k, p_\varepsilon\}} 2^{2p_1+1} + \sum_{p_1=0}^{\min\{p_k, p_\varepsilon\}} 2^{p_1+3} \right) \\
\leqslant& \sum_{k>1} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} \right) \frac{2^{2\min\{p_k, p_\varepsilon\}+1}}{3} + 2^{\min\{p_k,p_\varepsilon\}+4} \right) \\
\leqslant& \sum_{k\in\mathcal{K}'} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} \right) \frac{2^{2p_k+1}}{3} + 2^{p_k+4} \right) + \sum_{k>1, k\notin\mathcal{K}'} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} \right) \frac{2^{2p_\varepsilon+1}}{3} + 2^{p_\varepsilon+4} \right) \\
\overset{(c)}{\leqslant}& \sum_{k\in\mathcal{K}'} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} \right) \frac{32}{3(\Delta_k^{(M)})^2} + \frac{64}{\Delta_k^{(M)}} \right) + \sum_{k>1, k\notin\mathcal{K}'} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_1^{(M)} \right) \frac{32}{3\varepsilon^2} + \frac{64}{\varepsilon} \right) \\
\leqslant& \sum_{k\in\mathcal{K}'} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_k^{(M)} \right) \frac{32}{3(\Delta_k^{(M)})^2} + \frac{64}{\Delta_k^{(M)}} \right) + \sum_{k>1, k\notin\mathcal{K}'} \left(\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_k^{(M)} \right) \frac{32}{3\varepsilon^2} + \frac{64}{\varepsilon} \right)
\end{aligned}
\tag{32}
$$

where (a) and (c) are due to Eq.(27), and (b) is due to the property of swapping two summations. Summing up the costs in Eq.(29), Eq.(30), Eq.(31), and Eq.(32) concludes the proof. □

Next, we derive the problem-independent regret bound from Eq.(26). When Λ is large, the $O(\log\Lambda)$ and $O(\Lambda)$ terms dominate other terms in Eq.(26). For any given ε, if Λ is large enough, we always have $\varepsilon > 4\sqrt{\frac{\lambda^{(M)}e}{\Lambda}}$, which, with some calculus, guarantees that $\frac{16}{(\Delta_k^{(M)})^2} \log \frac{\Lambda(\Delta_k^{(M)})^2}{16\lambda^{(M)}} < \frac{16}{\varepsilon^2} \log \frac{\Lambda\varepsilon^2}{16\lambda^{(M)}}$ for all $\Delta_k^{(M)} > \varepsilon$. Therefore, we can scale all logarithmic arm pulling times as $\frac{16}{\varepsilon^2} \log \frac{\Lambda\varepsilon^2}{16\lambda^{(M)}}$, and upper bound the pulling cost $\left(\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)} - \mu_k^{(M)} \right)$ by $\frac{\lambda^{(M)}}{\lambda^{(1)}} \mu_1^{(M)}$. We derive the problem-independent regret upper bound as follows,

$$
\begin{aligned}
\mathbb{E}\left[R(\Lambda)\right] &\leqslant \frac{K\mu_1^{(M)} \cdot \lambda^{(M)}}{\lambda^{(1)}} \frac{16}{\varepsilon^2} \log \frac{\Lambda\varepsilon^2}{16\lambda^{(M)}} + \frac{\Lambda}{\lambda^{(1)}}\varepsilon \\
&\leqslant \frac{K\mu_1^{(M)} \cdot \lambda^{(M)}}{\lambda^{(1)}} \frac{16}{\varepsilon^2} \log \frac{\Lambda}{16\lambda^{(M)}} + \frac{\Lambda}{\lambda^{(1)}}\varepsilon \\
&\overset{(a)}{\leqslant} 2 \left(\frac{16K\mu_1^{(M)}\lambda^{(M)}}{\lambda^{(1)}} \log \frac{\Lambda}{16\lambda^{(M)}} \right)^{\frac{1}{3}} \left(\frac{\Lambda}{\lambda^{(1)}} \right)^{\frac{2}{3}},
\end{aligned}
$$

where the equation of (a) holds when $\varepsilon = \left(\frac{K\mu_1^{(M)} \log(\Lambda/16\lambda^{(M)})}{(\Lambda/16\lambda^{(M)})} \right)^{\frac{1}{3}}$.