A Notation

Table 2 reports a summary on the notation used throughout the paper.

Table 2: Notation

Symbol	Meaning
K	Number of arms
M	Number of fidelity
$\delta \in (0,1)$	Confidence level
τ	Stopping time of an algorithm
$\rho(\tau), c(\tau)$	Sample/Cost complexity of an algorithm
$\hat{I}(\tau)$	Arm recommended by the algorithm when it stops
ν	Bandit model
$\nu_{i,m}$	Distribution of arm i-th at fidelity m within bandit model ν
σ^2	Pseudo-variance of sub-gaussian distributions
$\mu_{i,m}$	Mean of arm i-th at fidelity m
$\xi_{i,m} := \mu_{i,m} - \mu_{i,M}$	Bias of fidelity m for arm i
$\xi_m \geq \max_{i \in [K} \lvert \xi_{i,m} \rvert$	Upper bound on the maximum bias introduced by fidelity m
λ_m	Cost for querying an arm at fidelity m
I_t, m_t, R_t	Arm/Fidelity/Reward selected/observed by the agent at round $t \in \mathbb{R}$
$\Delta_i := \mu_1 - \mu_i$	Arm gap
S	IISE's set of active arms
$\hat{\mu}_{i,m,t}$	Empirical estimation of the mean of arm i-th at fidelity m using t samples
$U(t,\delta,\xi)$	Bound on the empirical mean used in the IISE elimination rule
$B(t,\delta)$	Anytime confidence interval with confidence δ and t samples
α_m	Thresholds used in IISE to decide when to switch to fidelity $m+1$
m_i	Smallest fidelity for which IISE ensures that arm i is eliminated
$\mathcal{Y}$	MDP's state space
$\mathcal{A}$	MDP's action space
$p = \{p_h\}_{h \geq 1}$	MDP's set of transition kernel
$r = \{r_h\}_{h \geq 1}$	MDP's set of reward function
η	MDP's discount factor
Λ	MDP's branching factor
ϵ	Required precision in nearly-optimal identification
$\mathcal{K}_\epsilon$	Set of ϵ-optimal arms
$\gamma_m := \max_{i,j \in [K]} \{\xi_{i,m} - \xi_{j,m}\}$	Maximum bias variation of fidelity m
ϕ_m	Upper-bound on maximum bias variation for fidelity m

B Lower Bound

Theorem 1. *Consider a multi-fidelity bandit model ν with Gaussian distributions $\nu_{i,m} = \mathcal{N}(\mu_{i,m}, \sigma^2)$ such that $|\mu_{i,m} - \mu_{i,M}| \leq \xi_m$ for every $i \in [K]$ and $m \in [M]$. Then, for any δ-correct algorithm and $\delta \leq 0.15$, it holds that:*

$$\mathbb{E}\left[c(\tau)\right] \geq \left[\min_{m \in \mathcal{M}_1} \frac{\lambda_m}{\mathrm{KL}(\nu_{1,m}, \overline{\nu}_{2,m})} + \sum_{i=2}^{K} \min_{m \in \mathcal{M}_i} \frac{\lambda_m}{\mathrm{KL}(\nu_{i,m}, \overline{\nu}_{1,m})} \right] \log\left(\frac{1}{2.4\delta}\right),$$

where $\mathrm{KL}(p,q)$ is the Kullback-Leibler divergence between distributions p and q, $\overline{\nu}_{2,m} = \mathcal{N}(\mu_{2,M} + \xi_m, \sigma^2)$, $\overline{\nu}_{1,m} = \mathcal{N}(\mu_{1,M} - \xi_m, \sigma^2)$, $\mathcal{M}_1 := \{m \in [M] : \mu_{1,m} > \mathbb{E}_{x \sim \overline{\nu}_{2,m}}[x]\}$ and $\mathcal{M}_i := \{m \in [M] : \mathbb{E}_{x \sim \overline{\nu}_{1,m}}[x] > \mu_{i,m}\}$ for $i > 1$.

Proof. In our proof, we consider a multi-fidelity bandit model ν and create a new instance ν' by modifying the reward distribution at fidelity M of a given $i \in [K]$. The goal is creating a new bandit model ν' in which the optimal arm is different. We notice that, due to the modification of a considered arm at fidelity M, we need to modify also distributions at fidelity $m < M$, so that the new instance ν' satisfies the constraint on the maximum bias introduced by approximators $m < M$. Then, we show that any δ-correct algorithm requires a certain amount of cost to distinguish between the two problems. In the rest of this proof, we denote with $\mathbb{E}_\nu, \mathbb{P}_\nu, \mathbb{E}_{\nu'}, \mathbb{P}_{\nu'}$ expectations and probabilities in bandit models ν and ν' respectively.

We begin with the construction of the alternative instance ν'. For all $i \in [K]$, we can always build an alternative model ν' in which we modify only distributions related to arm i (i.e., $\nu_{i,1}, \nu_{i,2}, \dots, \nu_{i,M}$). We denote the distributions of the modified arms at a generic fidelity m with $\nu'_{i,m}$, and their mean with $\mu'_{i,m}$. We split the construction of ν' into two parts: $i \neq 1$ and $i = 1$.

Focus on $i \neq 1$. We can set $\nu'_{i,M} = \mathcal{N}(\mu'_{i,M}, \sigma^2)$ where $\mu'_{i,M} = \mu_{1,M} + c_{i,M}$ with $c_{i,M} > 0$. Then, for $\epsilon_{i,M}$ that satisfies $\epsilon_{i,M} \geq c_{i,M}(c_{i,M} + 2\Delta_i)$, it holds that:

$$(\mu'_{i,M} - \mu_{i,M})^2 \leq (\mu_{1,M} - \mu_{i,M})^2 + \epsilon_{i,M}.$$

Moreover, since we are considering Gaussian distributions with the same variance σ^2,[6] it follows that:

$$\mathrm{KL}(\nu_{i,M}, \nu'_{i,M}) \leq \mathrm{KL}(\nu_{i,M}, \nu_{1,M}) + \frac{\epsilon_{i,M}}{2\sigma^2}.$$

Now, since we have modified $\nu_{i,M}$, we need to modify arm i at all the fidelity $m < M$ in which condition $|\mu_{i,m} - \mu'_{i,M}| \leq \xi_m$ is no longer respected. This is equivalent to $\mu_{i,m} \notin [\mu'_{i,M} - \xi_m, \mu'_{i,M} + \xi_m]$, that, in turn, reduces to $\mu_{i,m} < \mu_{1,M} + c_{i,M} - \xi_m$. In particular, we modify arm $\nu_{i,m}$ with $\nu'_{i,m} = \mathcal{N}(\mu'_{i,m}, \sigma^2)$, where $\mu'_{i,m} = \mu'_{i,M} - \xi_m$. Notice that the new arm now satisfies the precision condition on fidelity $|\mu'_{i,m} - \mu'_{i,M}| \leq \xi_m$ and $\mathrm{KL}(\nu_{i,m}, \nu'_{i,m}) = \frac{(\mu_{i,m} - \mu_{1,M} - c_{i,M} + \xi_m)^2}{2\sigma^2}$. Finally, notice that the optimal arm in ν' is arm i.

Now, consider arm $i = 1$ and focus on fidelity M. We set $\nu'_{1,M} = \mathcal{N}(\mu'_{1,M}, \sigma^2)$ where $\mu'_{1,M} = \mu_{2,M} - c_{1,M}$ with $c_{1,M} > 0$. Then, for $\epsilon_{1,M}$ that satisfies $\epsilon_{1,M} \geq c_{1,M}(c_{1,M} + 2\Delta_i)$ we have:

$$(\mu_{1,M} - \mu'_{1,M})^2 \leq (\mu_{1,M} - \mu_{2,M})^2 + \epsilon_{1,M}.$$

Moreover, by exploiting the fact that the distributions are Gaussians, we have:

$$\mathrm{KL}(\nu_{1,M}, \nu'_{1,M}) \leq \mathrm{KL}(\nu_{1,M}, \nu_{2,M}) + \frac{\epsilon_{1,M}}{2\sigma^2}.$$

Now, since we have modified $\nu_{1,M}$, we need to modify distributions of arm 1 at all the fidelity $m < M$ in which condition $|\mu_{1,m} - \mu'_{1,M}| \leq \xi_m$ is no longer respected. This is equivalent to $\mu_{1,m} \notin [\mu_{2,M} - c_{1,M} - \xi_m, \mu_{2,M} - c_{1,M} + \xi_m]$, that, in turn, reduces to $\mu_{1,m} > \mu_{2,M} - c_{1,M} + \xi_m$. In particular, we build an arm $\nu'_{1,m} = \mathcal{N}(\mu'_{1,m}, \sigma^2)$ where $\mu'_{1,m} = \mu_{2,M} - c_{1,M} + \xi_m$. Notice that the precision condition on the fidelity is now satisfied for the new arm. Moreover, we have that $\mathrm{KL}(\nu_{1,m}, \nu'_{1,m}) = \frac{(\mu_{1,m} - \mu_{2,M} - \xi_m + c_{1,M})^2}{2\sigma^2}$. Finally, notice that the optimal arm in ν' is no longer arm 1.

Denote with $\mathcal{M}_i(c_{i,M})$ the set of modified fidelity for arm i; that is $\mathcal{M}_1(c_{1,M}) := \{m \in [M] : \mu_{1,m} > \mu_{2,M} - c_{1,M} + \xi_m\}$, and $\mathcal{M}_i(c_{i,M}) := \{m \in [M] : \mu_{i,m} < \mu_{1,M} + c_{i,M} - \xi_m\}$ for $i \neq 1$. Moreover, for all $i \neq 1$ and for each fidelity $m \in \mathcal{M}_i(c_{i,M})$, define $\underline{\nu}_{i,m}(c_{i,M}) = \mathcal{N}(\mu_{1,M} + c_{i,M} - \xi_m, \sigma^2)$. For $i = 1$, and each fidelity $m \in \mathcal{M}_1(c_{1,M})$, define $\underline{\nu}_{1,m}(c_{1,M}) = \mathcal{N}(\mu_{2,M} - c_{1,M} + \xi_m, \sigma^2)$.[7]

We introduce the event $\mathcal{I} := \{\widehat{I}(\tau) = 1\}$. Then, for any δ-correct algorithm it holds that $\mathbb{P}_\nu(\mathcal{I}) \geq 1 - \delta$ and $\mathbb{P}_{\nu'}(\mathcal{I}) \leq \delta$. Lemma 1 of [24] applied to stopping time τ in the alternative instance ν' in which arm i is modified leads to:

$$\sum_{m \in \mathcal{M}_i(c_{i,M})} \mathbb{E}_\nu[T_{i,m}(\tau)] \mathrm{KL}(\nu_{i,m}, \nu'_{i,m}) \geq \log\left(\frac{1}{2.4\delta}\right). \tag{10}$$

[6] We recall that the KL divergence between two Gaussian distributions $p = \mathcal{N}(\mu_p, \sigma^2)$ and $q = \mathcal{N}(\mu_q, \sigma^2)$ is given by $\mathrm{KL}(p,q) = \frac{(\mu_p - \mu_q)^2}{2\sigma^2}$.

[7] For all i, in the limit condition $c_{i,M} \to 0$, we adopt the following abbreviations: $\mathcal{M}_i(c_{i,M}) \to \mathcal{M}_i$, $\underline{\nu}_{i,m}(c_{i,M}) \to \overline{\nu}_{1,m}$, and $\underline{\nu}_{1,m}(c_{1,M}) \to \overline{\nu}_{2,m}$.

Now, focus on the left-hand term of Equation (10):

$$\sum_{m\in\mathcal{M}_i(c_{i,M})} \mathbb{E}_\nu[T_{i,m}(\tau)]\mathrm{KL}(\nu_{i,m},\nu'_{i,m}) = \sum_{m\in\mathcal{M}_i(c_{i,M})} \lambda_m \mathbb{E}_\nu[T_{i,m}(\tau)] \frac{\mathrm{KL}(\nu_{i,m},\nu'_{i,m})}{\lambda_m}$$
$$\leq \max_{m\in\mathcal{M}_i(c_{i,M})} \frac{\mathrm{KL}(\nu_{i,m},\nu'_{i,m})}{\lambda_m} \sum_{k\in\mathcal{M}_i(c_{i,M})} \lambda_k \mathbb{E}_\nu[T_{i,k}(\tau)].$$

Plugging this result into Equation (10), we obtain:

$$\sum_{m\in\mathcal{M}_i(c_{i,M})} \lambda_m \mathbb{E}_\nu[T_{i,m}(\tau)] \geq \min_{m\in\mathcal{M}_i(c_{i,M})} \frac{\lambda_m \log\left(\frac{1}{2.4\delta}\right)}{\mathrm{KL}(\nu_{i,m},\nu'_{i,m})}.$$

Now, we can further lower bound the right-hand term using the definition of the alternative instance ν' that we discussed at the beginning of the proof. For $i \neq 1$, we obtain:

$$\sum_{m\in\mathcal{M}_i(c_{i,M})} \lambda_m \mathbb{E}_\nu[T_{i,m}(\tau)]$$
$$\geq \min\left\{ \frac{\lambda_M \log\left(\frac{1}{2.4\delta}\right)}{\mathrm{KL}(\nu_{i,M},\nu_{1,M}) + \frac{\epsilon_{i,M}}{2\sigma^2}}, \min_{m\in\mathcal{M}_i(c_{i,M})\setminus\{M\}} \frac{\lambda_m \log\left(\frac{1}{2.4\delta}\right)}{\mathrm{KL}(\nu_{i,m},\underline{\nu}_{i,m}(c_{i,M}))} \right\}.$$

For $i = 1$, similarly:

$$\sum_{m\in\mathcal{M}_1(c_{1,M})} \lambda_m \mathbb{E}_\nu[T_{1,m}(\tau)]$$
$$\geq \min\left\{ \frac{\lambda_M \log\left(\frac{1}{2.4\delta}\right)}{\mathrm{KL}(\nu_{1,M},\nu_{2,M}) + \frac{\epsilon_{1,M}}{2\sigma^2}}, \min_{m\in\mathcal{M}_1(c_{1,M})\setminus\{M\}} \frac{\lambda_m \log\left(\frac{1}{2.4\delta}\right)}{\mathrm{KL}(\nu_{1,m},\underline{\nu}_{1,m}(c_{1,M}))} \right\}.$$

Now, letting $c_{i,M} \to 0$ and $\epsilon_{i,M} \to 0$ for all $i \in [K]$ and summing over the arms, we conclude the proof.

□

C Upper Bound

C.1 Proof of Theorem 2

We begin our proofs with the standard concentration results. For completeness, we report the Hoeffding inequality [4] for σ^2-subgaussian random variables.

Lemma 1 (Hoeffding inequality)**.** *Let $X_1, X_2, \dots, X_n$ be independent σ^2-subgaussian random variables with mean $\mu_1, \mu_2, \dots, \mu_n$ respectively. Let $\overline{\mu}_n \coloneqq \frac{1}{n}\sum_{i=1}^n \mu_i$ and $\hat{\mu}_n \coloneqq \frac{1}{n}\sum_{i=1}^n X_i$. Then, $\forall \epsilon > 0$, it holds that:*

$$\mathbb{P}(|\hat{\mu}_n - \overline{\mu}_n| > \epsilon) \leq 2\exp\left(-\frac{n\epsilon^2}{2\sigma^2}\right).$$

We now continue with bounding the probability of the failure event.

Lemma 2 (Failure event)**.** *Consider $\delta \in (0,1)$ and $B(t,\delta) \coloneqq \sqrt{\frac{2\sigma^2\log\left(\frac{4t^2}{\delta}\right)}{t}}$, and let:*

$$\mathcal{E} \coloneqq \bigcup_{i=1}^{K}\bigcup_{m=1}^{M}\bigcup_{t=1}^{+\infty} \left\{ |\hat{\mu}_{i,m,t} - \mu_{i,m}| \geq B\left(t, \frac{\delta}{KM}\right) \right\}.$$

Then, $\mathbb{P}(\mathcal{E}) \leq \delta$.

Proof. By the union bound we can write:

$$\mathbb{P}(\mathcal{E}) \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \mathbb{P}\left(\bigcup_{t=1}^{+\infty} \left\{ |\hat{\mu}_{i,m,t} - \mu_{i,m}| \geq B\left(t, \frac{\delta}{MK}\right) \right\} \right).$$

Now, focus on a single term within the summation. Recall that for all $i \in [K]$ and $m \in [M]$, $\hat{\mu}_{i,m,t}$ denotes the expectation with t samples of $\mu_{i,m}$. Applying a union bound together with Lemma 1, we obtain:

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{t=1}^{+\infty} \left\{ |\hat{\mu}_{i,m,t} - \mu_{i,m}| \geq B\left(t, \frac{\delta}{MK}\right) \right\} \right) &\leq \sum_{t=1}^{+\infty} 2\exp\left(-\frac{t}{2\sigma^2} \frac{2\sigma^2 \log\left(\frac{4t^2KM}{\delta}\right)}{t} \right) \\
&\leq \sum_{t=1}^{+\infty} \frac{\delta}{2t^2KM} \leq \frac{\delta}{KM}.
\end{aligned}$$

Thus, $\mathbb{P}(\mathcal{E}) \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \frac{\delta}{KM} = \delta$, which concludes the proof.

□

Now, we focus on the properties of Algorithm 1. We begin by showing that the optimal arm remains in the active set with high probability. Denote with $\overline{\mathcal{E}}$ the complementary of event $\mathcal{E}$.

Lemma 3 (Non-elimination of the optimal arm)**.** *With probability at least $1-\delta$, the optimal arm remains in the active set S until termination.*

Proof. We notice that, considering a generic active fidelity $m \in [M]$ and $t \geq 1$, arm $i \neq 1$ is removed from S if and only if there exists $j \in S$ and $j \neq i$ for which it holds that:

$$\hat{\mu}_{j,m,t} - U\left(t, \frac{\delta}{KM}, \xi_m\right) \geq \hat{\mu}_{i,m,t} + U\left(t, \frac{\delta}{KM}, \xi_m\right).$$

However, due to Lemma 2, we know that with probability at least $1-\delta$ the following conditions hold:

$$\begin{aligned}
\hat{\mu}_{j,m,t} &\leq \mu_{j,m} + B\left(t, \frac{\delta}{KM}\right) \leq \mu_{j,M} + U\left(t, \frac{\delta}{KM}, \xi_m\right), \\
\hat{\mu}_{i,m,t} &\geq \mu_{i,m} - B\left(t, \frac{\delta}{KM}\right) \geq \mu_{i,M} - U\left(t, \frac{\delta}{KM}, \xi_m\right).
\end{aligned}$$

Plugging these inequalities into the elimination condition, we obtain that i is eliminated with probability at $1-\delta$ only if it exists $j \neq i$ for which:

$$\mu_{j,M} \geq \hat{\mu}_{j,m,t} - U\left(t, \frac{\delta}{KM}, \xi_m\right) \geq \hat{\mu}_{i,m,t} + U\left(t, \frac{\delta}{KM}, \xi_m\right) \geq \mu_{i,M}$$

which is never satisfied for $i = 1$, thus concluding the proof.

□

We now proceed with a technical lemma that will be useful in bounding the duration of a given phase m and the maximum number of samples that are necessary to discard a sub-optimal arm $i \neq 1$.

Lemma 4 (Technical Lemma)**.** *If:*

$$t \geq \frac{128\sigma^2}{n^2} \log\left(\frac{128\sigma^2}{n^2\delta}\right),$$

then $n \geq 4B(t, \delta)$ is satisfied.

Proof. First of all, rewrite $n \geq 4B(t,\delta) = 4\sqrt{\frac{2\sigma^2 \log\left(\frac{4t^2}{\delta}\right)}{t}}$ in the following way:

$$tn^2 \geq 32\sigma^2 \log\left(\frac{4t^2}{\delta}\right) = 64\sigma^2 \log\left(\frac{2t}{\sqrt{\delta}}\right).$$

Then, we proceed by contradiction and analyze: $tn^2 < 64\sigma^2 \log\left(\frac{2t}{\sqrt{\delta}}\right)$. The proof follows from a direct application of Lemma 12 of [18]. In particular, due to Lemma 12 of [18], if $tn^2 < 64\sigma^2 \log\left(\frac{2t}{\sqrt{\delta}}\right)$, then, the following inequality holds as well:

$$t < \frac{1}{n^2}\left(64\sigma^2 \log\left(\frac{2}{n^4\sqrt{\delta}}\left(64\sigma^2\sqrt{\frac{2}{\sqrt{\delta}}}\right)^2\right)\right).$$

With some basic manipulations, we obtain:

$$t < \frac{128\sigma^2}{n^2} \log\left(\frac{128\sigma^2}{n^2\delta}\right).$$

Then, as soon as $t \geq \frac{128\sigma^2}{n^2} \log\left(\frac{128\sigma^2}{n^2\delta}\right)$, $n \geq 4B(t,\delta)$ holds, which concludes the proof.

□

The next result is crucial to Theorem 2 and provides guarantees on when (i.e., at which fidelity), at worst, sub-optimal arms will be eliminated, and the number of samples required to take this decision in that given phase.

Lemma 5 (Arm elimination)**.** *Consider $\alpha_m > 0$ for all $m \in [M-1]$ and $\alpha_M = 0$. Suppose $|S| > 2$ at the beginning of a given phase $m \in [M]$. Consider $i \in [K]$ such that $i \neq 1$. With probability at least $1-\delta$:*

- *if $i \in S$ and $\Delta_i \geq 4\xi_m + \alpha_m$, then, arm i will be removed from S during phase m using, at most $\left\lceil \frac{128\sigma^2}{(\Delta_i - 4\xi_m)^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i - 4\xi_m)^2\delta}\right)\right\rceil$ samples at arm i;*
- *if $i \notin S$ and $m > 1$, then $\Delta_i < \min_{\overline{m}<m}\{4\xi_{\overline{m}} + \alpha_{\overline{m}}\}$.*

Proof. Let us focus on $i \in S$, $i \neq 1$, and $\Delta_i \geq 4\xi_m + \alpha_m$. First of all, we notice that with probability at least $1-\delta$ the optimal arm remains in S until termination (Lemma 3). Therefore, one of the events that eliminate arm i from S is given by:

$$\hat{\mu}_{1,m,t} - U\left(t, \frac{\delta}{KM}, \xi_m\right) \geq \hat{\mu}_{i,m,t} + U\left(t, \frac{\delta}{KM}, \xi_m\right). \tag{11}$$

Moreover, due to Lemma 2, with probability at least $1-\delta$, the following holds:

$$\hat{\mu}_{1,m,t} \geq \mu_{1,m} - B\left(t, \frac{\delta}{KM}\right) \geq \mu_{1,M} - U\left(t, \frac{\delta}{KM}, \xi_m\right),$$
$$\hat{\mu}_{i,m,t} \leq \mu_{i,m} + B\left(t, \frac{\delta}{KM}\right) \leq \mu_{i,M} + U\left(t, \frac{\delta}{KM}, \xi_m\right).$$

Plugging these results into Equation (11), we obtain:

$$\begin{aligned}\hat{\mu}_{1,m,t} - U\left(t, \frac{\delta}{KM}, \xi_m\right) &\geq \mu_{1,M} - 2U\left(t, \frac{\delta}{KM}, \xi_m\right)\\ &\geq \mu_{i,M} + 2U\left(t, \frac{\delta}{KM}, \xi_m\right)\\ &\geq \hat{\mu}_{i,m,t} + U\left(t, \frac{\delta}{KM}, \xi_m\right).\end{aligned}$$

It follows that event of Equation (11) is guaranteed to occur if

$$\mu_{1,M} - 2U\left(t, \frac{\delta}{KM}, \xi_m\right) \geq \mu_{i,M} + 2U\left(t, \frac{\delta}{KM}, \xi_m\right),$$

which, in turn, is equivalent to $\Delta_i - 4\xi_m \geq 4B\left(t, \frac{\delta}{KM}\right)$. Due to Lemma 4, this inequality is satisfied as soon as:

$$t \geq \left\lceil \frac{128\sigma^2}{(\Delta_i - 4\xi_m)^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i - 4\xi_m)^2\delta}\right)\right\rceil.$$

What remains to prove is that the condition that guarantees to eliminate i from S (i.e., $\Delta_i - 4\xi_m \geq 4B\left(t, \frac{\delta}{KM}\right)$) activates before switching to the next phase, that happens when $\alpha \geq 4B\left(t, \frac{\delta}{KM}\right)$. However, since $\Delta_i \geq 4\xi_m + \alpha_m$, this is always true.

Focus now on $i \notin S$ and $m > 1$. The proof follows as a direct consequence of the previous claim.

□

We are now ready to prove Theorem 2.

Theorem 2. *If $\alpha_m > 0$ for every $m \in [M-1]$ and $\alpha_M = 0$, then, with probability at least $1-\delta$, IISE returns the optimal arm 1 with cost complexity $c(\tau)$ upper bounded by:*

$$c(\tau) \leq O\left(\sum_{i=2}^{K} \frac{\lambda_{m_i}\sigma^2}{(\Delta_i - 4\xi_{m_i})^2} \log\left(\frac{\sigma^2 MK}{(\Delta_i - 4\xi_{m_i})^2\delta}\right) + \sum_{m<m_i} \frac{\lambda_m \sigma^2}{\alpha_m^2} \log\left(\frac{\sigma^2 MK}{\alpha_m^2 \delta}\right)\right),$$

where m_i is the smallest $m \in [M]$ for which $\Delta_i \geq 4\xi_m + \alpha_m$ holds.

Proof. Consider $i \neq 1$. Due to Lemma 5, using fidelity m_i, arm i will be discarded from the active set with probability at least $1-\delta$, using a number of samples upper bounded by:

$$T_{i,m_i}(\tau) \leq \left\lceil \frac{128\sigma^2}{(\Delta_i - 4\xi_{m_i})^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i - 4\xi_{m_i})^2\delta}\right)\right\rceil.$$

Since m_i, by definition, is the first phase at which Lemma 5 guarantees elimination of arm i, we have $T_{i,k}(\tau) = 0$ for $k > m_i$. At this point, to derive the total cost required by IISE to discard i, we need to consider also the number of pulls at fidelity $m < m_i$. This is provided by Lemma 4, which yields an upper bound on the maximum duration of a given fidelity. More specifically:

$$T_{i,m}(\tau) \leq \left\lceil \frac{128\sigma^2}{\alpha_m^2} \log\left(\frac{128KM\sigma^2}{\alpha_m^2\delta}\right)\right\rceil.$$

Given that $\alpha_m > 0$ for all $m < M$, we notice how the duration of each phase is finite. Together with Lemma 3, and $\alpha_M = 0$, this guarantees that IISE is δ-correct.

To obtain the upper bound on the cost complexity, it is sufficient to multiply the number of pulls with their related cost. Summing over the arms concludes the proof.

□

C.2 Proofs concerning Assumption 1

We now provide evidence of the theoretical claims made that regards Assumption 1. More specifically:

- First, in Proposition 2, we show that the linear cost increase rate combined with the ξ's decay rate discussed in Section 4 provides sufficient conditions for Assumption 1.
- Then, in Proposition 3, we provide bounds on the maximum bias for the planning problem presented in Section 4.
- Finally, we prove Proposition 1.

Finally, at the end on this Section, we provide final remarks on Assumption 1.

Proposition 2. *If the following conditions hold:*

$$\forall i,j \in [M-1], i<j \qquad \left(\sqrt{\lambda_{i+1}}-\sqrt{\lambda_i}\right)^2 \leq \left(\sqrt{\lambda_{j+1}}-\sqrt{\lambda_j}\right)^2, \tag{12}$$

$$\forall \overline{m} \in [M] \qquad \sum_{m<\overline{m}} \frac{1}{(\xi_m-\xi_{m+1})^2} \leq \frac{1}{(\xi_{\overline{m}})^2}. \tag{13}$$

Then, Assumption 1 is verified.

Proof. The proof follows from simple upper bound arguments. Consider a generic fidelity $\overline{m} \in [M-1]$, and proceed from the left-hand term of Equation (5)

$$\begin{aligned}
\sum_{m<\overline{m}} \min_{k>m} \frac{(\sqrt{\lambda_k}-\sqrt{\lambda_m})^2}{(\xi_m-\xi_k)^2} &\leq \sum_{m<\overline{m}} \frac{(\sqrt{\lambda_{m+1}}-\sqrt{\lambda_m})^2}{(\xi_m-\xi_{m+1})^2} \\
&\leq (\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_{\overline{m}-1}})^2 \sum_{m<\overline{m}} \frac{1}{(\xi_m-\xi_{m+1})^2} \\
&\leq \frac{(\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_{\overline{m}-1}})^2}{(\xi_{\overline{m}})^2} \\
&\leq \frac{(\sqrt{\lambda_{\overline{m}+1}}-\sqrt{\lambda_{\overline{m}}})^2}{(\xi_{\overline{m}})^2},
\end{aligned}$$

where, in the second and last inequality we have used Equation (12), while in third one we have used Equation (13). At this point, denote $k_{\overline{m}} \in \operatorname{argmin}_{k>\overline{m}} \frac{(\sqrt{\lambda_k}-\sqrt{\lambda_{\overline{m}}})^2}{(\xi_{\overline{m}}-\xi_k)^2}$. Then, we proceed in the following way:

$$\frac{(\sqrt{\lambda_{\overline{m}+1}}-\sqrt{\lambda_{\overline{m}}})^2}{(\xi_{\overline{m}})^2} \leq \frac{(\sqrt{\lambda_{k_{\overline{m}}}}-\sqrt{\lambda_{\overline{m}}})^2}{(\xi_{\overline{m}}-\xi_{k_{\overline{m}}})^2} = \min_{k>\overline{m}} \frac{(\sqrt{\lambda_k}-\sqrt{\lambda_{\overline{m}}})^2}{(\xi_{\overline{m}}-\xi_k)^2},$$

where in the second step we have used $\lambda_{k_{\overline{m}}} \geq \lambda_{\overline{m}+1}$, together with the fact that $\xi_{\overline{m}} \geq \xi_{\overline{m}} - \xi_{k_{\overline{m}}} > 0$. □

Before diving into the maximum bias bounds for the planning problem described in Section 4, let us introduce some notation. We define a policy $\pi_m \coloneqq (\pi_{m,h})_{h=1}^m$ as a sequence of m probability distributions over the action set $\mathcal{A}$. We denote with $\pi_{m,h}$ the h-th element in π_m. More specifically, given $y \in \mathcal{Y}$, $\pi_{m,h}(y)$ represents a probability distributions over $\mathcal{A}$ in state y at depth h. We define the cumulative expected reward associated with policy π_m as $\mathbb{E}_{\pi,p}\left[\sum_{t=1}^m \eta^t r_t\right]$. We define the optimal policy π_m^* as the one that maximizes $\mathbb{E}_{\pi,p}\left[\sum_{t=1}^m \eta^t r_t\right]$. Given this initial setup, we now show how the mean of the arms are defined in our MF-BAI problem. Fix $y_0 \in \mathcal{Y}$, and set, for all $i \in [K]$ and $m \in [M]$:[8]

$$\mu_{i,m} \coloneqq r_0(y_0,i) + \mathbb{E}_{\pi_m^*,p}\left[\sum_{t=1}^m \eta^t r_t\right]. \tag{14}$$

Equation (14) states that the mean of a given initial action i is given by two components. The first one is the immediate reward for taking action i in state y_0 at depth 0. The second one, is the expected cumulative discounted reward that will be collected by an agent that maximizes the cumulative discounted reward over the next m steps. It follows that, selecting the optimal arm within this problem is equivalent to identifying the action that maximizes the the cumulative discounted reward starting from state y_0. Finally, we notice that, to obtain a sample from π_m^* it is sufficient to apply any Monte Carlo algorithm (e.g., depth-first search) to the planning problem and cutting the depth of the search at m. This is equivalent to set reward 0 (i.e., the minimum value of the reward) to all actions at step $\overline{m} > m$.

We are now ready to provide maximum bias bounds on $\mu_{i,m}$.

Proposition 3. *Consider the stochastic planning problem described in Section 4, and consider for all $i \in [K]$ and $m \in [M]$, $\mu_{i,m}$ as specified in Equation* (14)*. Then, for all $m \in [M]$ and $i \in [K]$ it*

[8] In this stochastic planning application, we assume fidelity to be indexed starting from 1.

holds that:

$$|\mu_{i,M} - \mu_{i,m}| \leq \xi_m := \frac{\eta^{m+1} - \eta^{M+1}}{1-\eta}.$$

Proof. We start by considering $|\mu_{i,M} - \mu_{i,m}|$. Due to Equation (14), we have that:

$$\begin{aligned}
|\mu_{i,M} - \mu_{i,m}| &= \mathbb{E}_{\pi^*_M,p}\left[\sum_{t=1}^{M} \eta^t r_t\right] - \mathbb{E}_{\pi^*_m,p}\left[\sum_{t=1}^{m} \eta^t r_t\right] \\
&\leq \mathbb{E}_{\pi^*_m,p}\left[\sum_{t=1}^{m} \eta^t r_t\right] + \max_{\pi} \mathbb{E}_{\pi,p}\left[\sum_{t=m+1}^{M} \eta^t r_t\right] - \mathbb{E}_{\pi^*_m,p}\left[\sum_{t=1}^{m} \eta^t r_t\right] \\
&\leq \sum_{t=m+1}^{M} \eta^t = \frac{\eta^{m+1} - \eta^{M+1}}{1-\eta},
\end{aligned}$$

where in the first inequality we have upper bounded the maximum of π^*_M over the entire depth M into the maximimization of two components (i.e., from 1 to m and from $m+1$ to M), and, in the second inequality we have used $|r_h(\cdot,\cdot)| \leq 1$. □

Proposition 1. *Consider $\Lambda \geq 2$. If $\lambda_m = \Lambda^m$ and $\xi_m = (\eta^{m+1} - \eta^{M+1})/(1-\eta)$ for all $m \in [M]$, then Assumption 1 holds.*

Proof. Consider a generic fidelity $m \in [M-1]$ and plug the definition of λ and ξ within $\min_{k>m} \frac{(\sqrt{\lambda_k} - \sqrt{\lambda_m})^2}{(\xi_m - \xi_k)^2}$:

$$\min_{k>m} \frac{(\sqrt{\lambda_k} - \sqrt{\lambda_m})^2}{(\xi_m - \xi_k)^2} = \left(\frac{1-\eta}{\eta}\right)^2 \min_{k>m} \left(\frac{\sqrt{\Lambda^k} - \sqrt{\Lambda^m}}{\eta^m - \eta^k}\right)^2. \tag{15}$$

The proof follows in two steps. First, we prove that $k = m+1$ is the minimizer of $\min_{k>m} \left(\frac{\sqrt{\Lambda^k} - \sqrt{\Lambda^m}}{\eta^m - \eta^k}\right)^2$, then, we use this result to show that Assumption 1 is satisfied.

Let us proceed with the first step. Consider $l \geq 1$, we need to show that:

$$\frac{(\Lambda^{\frac{m+l}{2}} - \Lambda^{\frac{m}{2}})^2}{(\eta^m - \eta^{l+m})^2} \geq \frac{(\Lambda^{\frac{m+1}{2}} - \Lambda^{\frac{m}{2}})^2}{(\eta^m - \eta^{m+1})^2}. \tag{16}$$

That is equivalent to:

$$\frac{\Lambda^{\frac{l}{2}} - 1}{1 - \eta^l} \geq \frac{\Lambda^{\frac{1}{2}} - 1}{1-\eta}. \tag{17}$$

Let us recall the Bernoulli's inequality. For all $n > 0$ and $x \geq -1$, it states that: $(1+x)^n \geq 1 + nx$. Now, consider:

$$\Lambda^{\frac{l}{2}} - 1 = \left(1 + (\sqrt{\Lambda} - 1)\right)^l - 1 \geq 1 + l(\sqrt{\Lambda} - 1) - 1 = l(\sqrt{\Lambda} - 1),$$

where we applied the Bernoulli inequality since $\sqrt{\Lambda} - 1 \geq -1$. Moreover, since $\eta - 1 \geq -1$, we apply the same reasoning to $\eta^l - 1$:

$$\eta^l - 1 = (1 + (\eta - 1))^l - 1 \geq 1 + l(\eta - 1) - 1 = l(\eta - 1),$$

At this point, we can plug these two results to lower bound the left-hand term of Equation (17):

$$\frac{\Lambda^{\frac{l}{2}} - 1}{1 - \eta^l} \geq \frac{l(\Lambda^{\frac{1}{2}} - 1)}{l(1 - \eta^l)} = \frac{\Lambda^{\frac{1}{2}} - 1}{1 - \eta^l}.$$

Therefore, Equation (16) is always true. More specifically, since it holds for all $l \geq 1$, we have proved that $k = m+1$ is the minimizer of $\min_{k>m} \left(\frac{\sqrt{\Lambda^k} - \sqrt{\Lambda^m}}{\eta^m - \eta^k}\right)^2$.

We now continue the proof to show that Assumption 1 is satisfied. Fix an $\overline{m} \in [M-1]$. We start by plugging Equation (15) into the left-hand term of Equation (5):

$$\sum_{m<\overline{m}} \min_{k>m} \frac{(\sqrt{\lambda_k}-\sqrt{\lambda_m})^2}{(\xi_m-\xi_k)^2} = \left(\frac{1-\eta}{\eta}\right)^2 \sum_{m<\overline{m}} \min_{k>m} \frac{(\Lambda^{\frac{k}{2}}-\Lambda^{\frac{m}{2}})^2}{(\eta^m-\eta^k)^2}$$
$$= \left(\frac{1-\eta}{\eta}\right)^2 \sum_{m<\overline{m}} \frac{(\Lambda^{\frac{m+1}{2}}-\Lambda^{\frac{m}{2}})^2}{(\eta^m-\eta^{m+1})^2}.$$

Thus, we can further rewrite as:

$$\left(\frac{1-\eta}{\eta}\right)^2 \left(\frac{\Lambda^{\frac{1}{2}}-1}{1-\eta}\right)^2 \sum_{m<\overline{m}} \left(\frac{\Lambda^{\frac{1}{2}}}{\eta}\right)^{2m} = \left(\frac{1-\eta}{\eta}\right)^2 \left(\frac{\Lambda^{\frac{1}{2}}-1}{1-\eta}\right)^2 \left(\frac{\left(\frac{\Lambda}{\eta^2}\right)^{\overline{m}}-1}{\frac{\Lambda}{\eta^2}-1}-1\right). \quad (18)$$

At this point, consider Equation (15) with $m=\overline{m}$:

$$\left(\frac{1-\eta}{\eta}\right)^2 \min_{k>\overline{m}} \left(\frac{\sqrt{\Lambda^k}-\sqrt{\Lambda^{\overline{m}}}}{\eta^{\overline{m}}-\eta^k}\right)^2 = \left(\frac{1-\eta}{\eta}\right)^2 \left(\frac{\sqrt{\Lambda^{\overline{m}+1}}-\sqrt{\Lambda^{\overline{m}}}}{\eta^{\overline{m}}-\eta^{\overline{m}+1}}\right)^2$$
$$= \left(\frac{1-\eta}{\eta}\right)^2 \left(\frac{\Lambda^{\frac{1}{2}}-1}{1-\eta}\right)^2 \left(\frac{\Lambda}{\eta^2}\right)^{\overline{m}},$$

and compare it with Equation (18). Define $q=\frac{\Lambda}{\gamma^2}$, we obtain:

$$\frac{q^{\overline{m}}-1}{q-1}-1 \leq q^{\overline{m}}.$$

That is, $2q^{\overline{m}-1}-1 \leq q^{\overline{m}}$, which is always true for $\Lambda \geq 2$, which concludes the proof.

□

C.2.1 Remarks on Assumption 1

Now, we remark our contributions concerning Assumption 1. The purpose of Assumption 1 is to make the multi-fidelity structure provably convenient. Assumptions, in the multi-fidelity literature, are typical in many settings. In many optimization works [20, 22, 39, 40, 11], assumptions were required on the smoothness/structure of the target function to be optimized at *each* of the different fidelity. In the context of finite armed bandits [21], instead, assumptions were required on the values of ξ to limit the cumulative regret that is introduced by the adversarial behavior that fidelity might have in the unstructured setting of finite-armed bandit (i.e., there is no prior knowledge on the arm space). More specifically, [21], assumed, that for each $m \in [M]$ the following holds:

$$\sum_{k<m} \frac{1}{\xi_k^2} \leq \frac{1}{\xi_m^2}, \quad (19)$$

that states a decay rate on the maximum bias function. We remark that also in [21] a subset of fidelity on which the assumption holds can be pre-selected so that the algorithm enjoys nice theoretical properties.

In our work, we provably confirm the adversarial behavior of fidelity in the fixed-confidence setting (i.e., Theorem 1). However, Assumption 1 significantly differs from Equation (19), since, as already remarked in Section 4, it states a direct relationships between costs and biases. This is as expected since, in MF problems, the goal is to trade-off cheaper but biased samples with more expensive (but more accurate) ones. As a straight consequence, compared to Equation (19), Assumption 1, depending on the values of λ's, is valid with ξ's whose decay is slower (e.g., linear) w.r.t. to the one of Equation (19).

Moreover, in this paper, we further discussed the meaning and relevance of Assumption 1 with two sets of sufficient conditions:

- in Proposition 2, we derived sufficient conditions for Assumption 1 to hold that do not put a direct relationships between costs and biases, but that allows for an easy to grasp interpretation that has been discussed in Section 4;
- in Proposition 1, instead, we have shown that with exponential increasing costs (i.e., $\lambda_m = \Lambda^m$) and exponential decreasing biases (i.e., $\xi_m = (\eta^{m+1} - \eta^{M+1})/(1-\eta)$), Assumption 1 is always satisfied. This is the practical case of the stochastic planning application discussed in Section 4.

Furthermore, as we will see in Lemma 6, Assumption 1, is not necessary in the entire proofs of Theorem 2. Indeed, m_i is the cost minimizer among the fidelity subset for which $\Delta_i > 4\xi_m$ holds, without any regard to Assumption 1. The purpose of Assumption 1 (that is used in the last step of the proofs only) is to limit the cost at previous fidelity that were not useful to discard arm with small enough gaps (which are *unknown* to the agent).

As a final recap of this remark:

- we have discussed why assumptions are necessary (i.e., Theorem 1 and previous literature);
- we have discussed the interpretation of Assumption 1 by deriving two set of sufficient conditions. Moreover, we have analyzed the relationship with typical assumptions of the MF literature (i.e., Section 4), with a particular focus on [21], that studies finite-armed bandits without any structured assumption on the arm space;
- we have discussed that Assumption 1 is not necessary in the entire proof of Theorem 3, but it is necessary only for the last step of the proof (see Lemma 6).

C.3 Proof of Theorem 3

The proof follows in two steps. First, we prove that m_i is the minimizer of the identification cost.[9] Then, we make a direct use of Assumption 1 to prove Theorem 3.

Lemma 6 (Cost minimizer)**.** *Consider $i \neq 1$. Choose $\alpha_M = 0$ and α_m as in Equation* (8)*. Define m_i as the smallest $m \in [M]$ for which $\Delta_i \geq 4\xi_m + \alpha_m$ holds. Then:*

$$\frac{\lambda_{m_i}}{(\Delta_i - 4\xi_{m_i})^2} = \min_{m\in[M]:\Delta_i > 4\xi_m} \frac{\lambda_m}{(\Delta_i - 4\xi_m)^2}.$$

Proof. We proceed by contradiction. Suppose there exists $\overline{m} \in [M]$ such that $\Delta_i > 4\xi_{\overline{m}}$ for which it holds that:

$$\frac{\lambda_{\overline{m}}}{(\Delta_i - 4\xi_{\overline{m}})^2} < \min_{m\in[M]\setminus\{\overline{m}\}:\Delta_i > 4\xi_m} \frac{\lambda_m}{(\Delta_i - 4\xi_m)^2} \leq \frac{\lambda_{m_i}}{(\Delta_i - 4\xi_{m_i})^2}. \tag{20}$$

We split the proof into two parts. Let us first consider $\overline{m} > m_i$ and consider:

$$\frac{\lambda_{m_i}}{(\Delta_i - 4\xi_{m_i})^2} \leq \min_{M\geq m > m_i} \frac{\lambda_m}{(\Delta_i - 4\xi_m)^2}.$$

That holds if $\Delta_i \geq 4\xi_{m_i} + \alpha_{m_i}$, which we know to be satisfied by the definition of m_i, thus Equation (20) is false.

Now, consider $\overline{m} < m_i$. If Equation (20) holds, then, it would imply $\Delta_i \geq 4\xi_{\overline{m}} + \alpha_{\overline{m}}$. However, we know by definition of m_i, that m_i is the smallest m for which the previous condition holds; thus Equation (20) is false. □

Theorem 3. *Under Assumption 1, selecting the thresholds α_m as in Equation* (8)*, with probability at least $1-\delta$, IISE returns the optimal arm with cost complexity $c(\tau)$ upper bounder by:*

$$c(\tau) \leq \widetilde{O}\left(\sum_{i=2}^{K} \min_{m\in[M]:\Delta_i > 4\xi_m} \frac{\lambda_m \sigma^2}{(\Delta_i - 4\xi_m)^2}\right) \leq \widetilde{O}\left(\sum_{i=2}^{K} \frac{\lambda_M \sigma^2}{\Delta_i^2}\right)$$

[9] Notice that this holds without any regard to Assumption 1.

Proof. First of all, notice that, choosing $\alpha_M = 0$ and $\alpha_m = \max_{M \geq \overline{m} > m} \frac{4(\xi_m - \xi_{\overline{m}})\sqrt{\lambda_m}}{\sqrt{\lambda_{\overline{m}}} - \sqrt{\lambda_m}}$ for $m < M$, satisfies the requirement of Theorem 2, i.e., $\alpha_m > 0$, since $\xi_m > \xi_{\overline{m}}$. Therefore, IISE returns the optimal arm with probability at least $1 - \delta$ with cost complexity upper bounded by:

$$\begin{aligned} c(\tau) &\leq \widetilde{O}\left(\sum_{i=2}^{K}\left(\frac{\lambda_{m_i}\sigma^2}{(\Delta_i - 4\xi_{m_i})^2} + \sum_{m<m_i}\frac{\lambda_m \sigma^2}{\alpha_m^2}\right)\right) \\ &= \widetilde{O}\left(\sum_{i=2}^{K}\left(\frac{\lambda_{m_i}\sigma^2}{(\Delta_i - 4\xi_{m_i})^2} + \sum_{m<m_i}\frac{\lambda_m \sigma^2}{\left(\max_{M\geq\overline{m}>m}\frac{4(\xi_m - \xi_{\overline{m}})\sqrt{\lambda_m}}{\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_m}}\right)^2}\right)\right). \end{aligned}$$

Consider an arm $i \neq 1$ and us focus on the sum at previous fidelity $m < m_i$:

$$\sum_{m<m_i}\frac{\lambda_m \sigma^2}{\left(\max_{M\geq\overline{m}>m}\frac{4(\xi_m - \xi_{\overline{m}})\sqrt{\lambda_m}}{\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_m}}\right)^2} = \sum_{m<m_i}\min_{\overline{m}>m}\frac{\sigma^2(\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_m})^2}{16(\xi_m - \xi_{\overline{m}})^2}.$$

With some basic manipulations and applying Assumption 1, we can rewrite this summation as:

$$\begin{aligned} \sum_{m<m_i}\min_{\overline{m}>m}\frac{\sigma^2(\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_m})^2}{16(\xi_m - \xi_{\overline{m}})^2} &= \sum_{m<m_{i-1}}\min_{\overline{m}>m}\frac{\sigma^2(\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_m})^2}{16(\xi_m - \xi_{\overline{m}})^2} \\ &\quad + \min_{\overline{m}>m_{i-1}}\frac{\sigma^2(\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_{m_{i-1}}})^2}{16(\xi_{m_{i-1}} - \xi_{\overline{m}})^2} \\ &\leq 2\min_{\overline{m}>m_{i-1}}\frac{\sigma^2(\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_{m_{i-1}}})^2}{16(\xi_{m_{i-1}} - \xi_{\overline{m}})^2} \\ &= 2\frac{\sigma^2\lambda_{m_{i-1}}}{((4\xi_{m_{i-1}}+\alpha_{m_{i-1}}) - 4\xi_{m_{i-1}})^2}. \end{aligned}$$

At this point, consider the following general equation of the form:

$$\frac{\lambda_{m_{i-1}}}{(d - 4\xi_{m_{i-1}})^2} \leq \frac{\lambda_{m_i}}{(d - 4\xi_{m_i})^2}, \tag{21}$$

with $d > 4\xi_m$. This is equivalent to:

$$\lambda_{m_{i-1}}^{\frac{1}{2}}(d - 4\xi_{m_i}) \leq \lambda_{m_i}^{\frac{1}{2}}(d - 4\xi_{m_{i-1}}),$$

which in turn can be rewritten as:

$$d \geq \frac{4(\xi_{m_{i-1}}\sqrt{\lambda_{m_i}} - \xi_{m_i}\sqrt{\lambda_{m_{i-1}}})}{\sqrt{\lambda_{m_i}}-\sqrt{\lambda_{m_{i-1}}}} = 4\xi_{m_{i-1}} + \frac{4(\xi_{m_{i-1}} - \xi_{m_i})\sqrt{\lambda_{m_{i-1}}}}{\sqrt{\lambda_{m_i}}-\sqrt{\lambda_{m_{i-1}}}}.$$

Therefore, Equation (21) holds for $d \geq 4\xi_{m_{i-1}} + \frac{4(\xi_{m_{i-1}} - \xi_{m_i})\sqrt{\lambda_{m_{i-1}}}}{\sqrt{\lambda_{m_i}}-\sqrt{\lambda_{m_{i-1}}}}$. However, due to the definition of the thresholds as in Equation (8), we know that:

$$\alpha_{m_{i-1}} = \max_{M\geq\overline{m}>m_{i-1}}\frac{4(\xi_{m_{i-1}} - \xi_{\overline{m}})\sqrt{\lambda_{m_{i-1}}}}{\sqrt{\lambda_{\overline{m}}}-\sqrt{\lambda_{m_{i-1}}}} \geq \frac{4(\xi_{m_{i-1}} - \xi_{m_i})\sqrt{\lambda_{m_{i-1}}}}{\sqrt{\lambda_{m_i}}-\sqrt{\lambda_{m_{i-1}}}}.$$

Therefore, Equation (21) holds also for $d = 4\xi_{m_{i-1}} + \alpha_{m_{i-1}}$. It follows that:

$$2\frac{\sigma^2\lambda_{m_{i-1}}}{((4\xi_{m_{i-1}}+\alpha_{m_{i-1}}) - 4\xi_{m_{i-1}})^2} \leq 2\frac{\sigma^2\lambda_{m_i}}{((4\xi_{m_{i-1}}+\alpha_{m_{i-1}}) - 4\xi_{m_i})^2}.$$

At this point, due to Lemma 5, we know that $\Delta_i < 4\xi_{m_{i-1}} + \alpha_{m_{i-1}}$, otherwise, arm i would have already been discarded at phase m_{i-1}. Moreover, $\Delta_i > 4\xi_{m_i}$ due to the definition of m_i. Conse-

Algorithm 2 Near-optimal Iterative Imprecise Successive Elimination (IISE).

Require: Multi-fidelity bandit model ν, confidence δ, thresholds $\{\alpha_m\}_{m=1}^M$, bounds $\{\xi_m\}_{m=1}^M$, accuracy ϵ

1: $S \leftarrow [K]$
2: $m \leftarrow 1$ and $t \leftarrow 0$
3: **while** $|S| > 1$ **do**
4:     **if** $\epsilon - 4\xi_m \geq 4B\left(t, \frac{\delta}{KM}\right)$ **then**
5:         **return** S
6:     **end if**
7:     **if** $\alpha_m \geq 4B\left(t, \frac{\delta}{KM}\right)$ **then**
8:         $m \leftarrow m+1$ and $t \leftarrow 0$
9:     **end if**
10:     Pull all arms in S at fidelity m and $t \leftarrow t+1$
11:     Update $\hat{\mu}_{j,m,t}$ for all $j \in [S]$
12:     $S \leftarrow S \setminus \left\{ i \in S : \exists j \in S : \hat{\mu}_{j,m,t} - U\left(t, \frac{\delta}{KM}, \xi_m\right) \geq \hat{\mu}_{i,m,t} + U\left(t, \frac{\delta}{KM}, \xi_m\right) \right\}$
13: **end while**
14: **return** S

quently:

$$2\frac{\sigma^2 \lambda_{m_i}}{((4\xi_{m_{i-1}} + \alpha_{m_{i-1}}) - 4\xi_{m_i})^2} \leq 2\frac{\sigma^2 \lambda_{m_i}}{(\Delta_i - 4\xi_{m_i})^2}.$$

Applying the previous reasoning for each $i \neq 1$, we can rewrite the upper bound on the cost complexity of IISE as:

$$c(\tau) \leq \widetilde{O}\left(\sum_{i=2}^{K} \frac{\lambda_{m_i}\sigma^2}{(\Delta_i - 4\xi_{m_i})^2}\right).$$

What is left is proving that m_i satisfies the minimum condition on all $m \in [M]$ such that $\Delta_i > 4\xi_m$, which is guaranteed by Lemma 6 (indeed, notice that, given our threshold choice, $\Delta_i \geq 4\xi_m + \alpha_m$ implies $\Delta_i > 4\xi_m$). Therefore:

$$\frac{\lambda_{m_i}}{(\Delta_i - 4\xi_{m_i})^2} = \min_{m \in [M]: \Delta_i > 4\xi_m} \frac{\lambda_m}{(\Delta_i - 4\xi_m)^2},$$

with which we can further the cost complexity:

$$c(\tau) \leq \widetilde{O}\left(\sum_{i=2}^{K} \frac{\lambda_{m_i}\sigma^2}{(\Delta_i - 4\xi_{m_i})^2}\right) \leq \widetilde{O}\left(\sum_{i=2}^{K} \min_{m \in [M]: \Delta_i > 4\xi_m} \frac{\lambda_m \sigma^2}{(\Delta_i - 4\xi_m)^2}\right) \leq \widetilde{O}\left(\sum_{i=2}^{K} \frac{\lambda_M \sigma^2}{\Delta_i^2}\right),$$

which concludes the proof. □

C.4 Pseudocode and Analysis of Near-Optimal Identification

We now provide the pseudocode and the analysis of a modification of IISE that allows identifying ϵ-optimal arms. Algoritm 2 reports the pseudocode. As we can appreciate, the only modification required is the new if branch at line $4-6$. Everything else is left unchanged. Due to this observation, we notice that Lemma 2 [10] and Lemma 3 still hold. What changes, instead, is the arm elimination lemma (Lemma 5).

Lemma 7 (Near-optimal arm elimination)**.** *Consider $\alpha_m > 0$ for all $m \in [M-1]$ and $\alpha_M = 0$. Suppose that Algorithm 2 has not terminated at the beginning of a given phase $m \in [M]$. Consider $i \in [K]$ such that $i \neq 1$. With probability at least $1-\delta$:*

- *if $i \in S$ and $\Delta_i \geq \max\{\epsilon, 4\xi_m + \alpha_m\}$, then, arm i will be removed from S during phase m using, at most $\left\lceil \frac{128\sigma^2}{(\Delta_i - 4\xi_m)^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i - 4\xi_m)^2\delta}\right)\right\rceil$ samples at arm i;*
- *if $i \notin S$ and $m > 1$, then $\Delta_i < \min_{\overline{m} < m}\{4\xi_{\overline{m}} + \alpha_{\overline{m}}\}$.*

[10] Log factors could be optimized since, in ϵ-optimal arm identification problems, anytime-confidence intervals are not needed. For the sake of simplicity of exposition of the proofs, we continue to use Lemma 2.

Moreover, if m is such that $\epsilon \geq 4\xi_m + \alpha_m$, then, Algorithm 2 will terminate in, at most, $\left\lceil \frac{128\sigma^2}{(\epsilon-4\xi_m)^2} \log\left(\frac{128KM}{(\epsilon-4\xi_m)^2\delta}\right)\right\rceil$ rounds in phase m.

Proof. We begin by focusing on the last claim. From Lemma 4, the ϵ-stopping condition at Line 4 of Algorithm 2 activates as soon as:

$$t \geq \left\lceil \frac{128\sigma^2}{(\epsilon-4\xi_m)^2} \log\left(\frac{128KM\sigma^2}{(\epsilon-4\xi_m)^2\delta}\right)\right\rceil,$$

provided that $\epsilon > 4\xi_m$ and $\epsilon \geq 4\xi_m + \alpha_m$. These conditions are satisfied by hypothesis.

For what concerns on $i \in S$, $i \neq 1$, and $\Delta_i \geq \max\{\epsilon, 4\xi_m + \alpha_m\}$, the proof follows from a direct extension of Lemma 5. The only difference is that we need to prove that the condition that guarantees to remove i from S (i.e., $\Delta_i - 4\xi_m \geq 4B\left(t, \frac{\delta}{KM}\right)$) activates sooner w.r.t. both the time at which the phase change happens (this is already proved in Lemma 5), and also the time at which Algorithm 2 terminates due to Line 4. This second point directly follows from the fact that $\Delta_i \geq \epsilon$ by assumption.

Finally, if $i \notin S$ and $m > 1$, the proof follows as a direct consequence of the previous claim.

□

Theorem 4. *If $\alpha_m > 0$ for every $m \in [M-1]$ and $\alpha_M = 0$, then, with probability at least $1-\delta$, Algorithm 2 returns a subset $S \in [K]$ such that for all $i \in S$, it holds that $\mu_i \geq \mu_1 - \epsilon$. Its cost complexity $c(\tau)$ is upper bounded by:*

$$c(\tau) \leq O\Bigg(\sum_{i \notin \mathcal{K}_\epsilon} \frac{\lambda_{m_i}\sigma^2}{(\Delta_i - 4\xi_{m_i})^2} \log\left(\frac{\sigma^2 MK}{(\Delta_i - 4\xi_{m_i})^2\delta}\right) + \sum_{m<m_i} \frac{\lambda_m\sigma^2}{\alpha_m^2} \log\left(\frac{\sigma^2 MK}{\alpha_m^2\delta}\right) + \tag{22}$$

$$\sum_{i \in \mathcal{K}_\epsilon} \frac{\lambda_{m_i}\sigma^2}{(\epsilon - 4\xi_{m_i})^2} \log\left(\frac{\sigma^2 MK}{(\epsilon - 4\xi_{m_i})^2\delta}\right) + \sum_{m<m_i} \frac{\lambda_m\sigma^2}{\alpha_m^2} \log\left(\frac{\sigma^2 MK}{\alpha_m^2\delta}\right)\Bigg), \tag{23}$$

where $\mathcal{K}_\epsilon := \{i \in [K] : \mu_i > \mu_1 - \epsilon\}$. For $i \notin \mathcal{K}_\epsilon$, m_i is defined as the smallest phase for which $\Delta_i \geq 4\xi_m + \alpha_m$ holds; for $i \in \mathcal{K}_\epsilon$, instead, it is defined as the first phase at which $\epsilon \geq 4\xi_m + \alpha_m$ holds.

Moreover, under Assumption 1:

$$c(\tau) \leq \widetilde{O}\left(\sum_{i \notin \mathcal{K}_\epsilon} \min_{m:\Delta_i > 4\xi_m} \frac{\lambda_m\sigma^2}{(\Delta_i - 4\xi_m)^2} + \sum_{i \in \mathcal{K}_\epsilon} \min_{m:\epsilon > 4\xi_m} \frac{\lambda_m\sigma^2}{(\epsilon - 4\xi_m)^2} \right).$$

Proof. Let us focus on the result given under generic thresholds α_m. Consider $i \neq 1$. Due to Lemma 7, m_i is, by definition, the first phase at which Lemma 7 guarantees elimination of arm i. It follows that $T_{i,k}(\tau) = 0$ for $k > m_i$. For what concerns $T_{i,m_i}(\tau)$, instead, Lemma 7 provides, with probability at least $1-\delta$ an upper bound on this quantity that is given by:

$$T_{i,m_i}(\tau) \leq O\left(\frac{128\sigma^2}{(\Delta_i - 4\xi_m)^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i - 4\xi_m)^2\delta}\right)\right),$$

for $i \notin \mathcal{K}_\epsilon$, and:

$$T_{i,m_i}(\tau) \leq O\left(\frac{128\sigma^2}{(\epsilon - 4\xi_m)^2} \log\left(\frac{128KM\sigma^2}{(\epsilon - 4\xi_m)^2\delta}\right)\right),$$

for $i \in \mathcal{K}_\epsilon$.

At this point, to derive the total cost required by Algorithm 2 to discard i, we need to consider also the number of pulls at fidelity $m < m_i$. This is provided by Lemma 4, which yields an upper bound on the maximum duration of a given fidelity. More specifically:

$$T_{i,m}(\tau) \leq O\left(\frac{128\sigma^2}{\alpha_m^2} \log\left(\frac{128KM\sigma^2}{\alpha_m^2\delta}\right)\right).$$

Algorithm 3 Iterative Imprecise Successive Elimination with Maximum Bias Variation (IISE-γ).

Require: Multi-fidelity bandit model ν, confidence δ, thresholds $\{\alpha_m\}_{m=1}^M$, bounds $\{\gamma_m\}_{m=1}^M$

$S \leftarrow [K]$
$m \leftarrow 1$ and $t \leftarrow 0$
while $|S| > 1$ **do**
  if $\alpha_m \geq 4B\left(t, \frac{\delta}{KM}\right)$ **then**
    $m \leftarrow m+1$ and $t \leftarrow 0$
  end if
  Pull all arms in S at fidelity m and $t \leftarrow t+1$
  Update $\hat{\mu}_{j,m,t}$ for all $j \in [S]$
  $S \leftarrow S \setminus \left\{i \in S : \exists j \in S : \hat{\mu}_{j,m,t} - B\left(t, \frac{\delta}{KM}\right) \geq \hat{\mu}_{i,m,t} + B\left(t, \frac{\delta}{KM}\right) + \gamma_m\right\}$
end while
return S

Given that $\alpha_m > 0$ for all $m < M$, we notice how the duration of each phase is finite. Together with Lemma 3, $\alpha_M = 0$, and Lemma 7 this implies that, than all arms $i \in S$ returned by Algorithm 2 are ϵ-optimal, with probability at least $1-\delta$.

To obtain the upper bound on the cost complexity, it is sufficient to multiply the number of pulls with their related cost. Summing over the arms concludes the proof. For what concern the statement under Assumption 1, it easily follow using the same analysis of Theorem 3 applied to Equation (22).

$\square$

D Order-Aware BAI

D.1 IISE-γ: pseudocode and analysis

In this section, we provide pseudocode and analysis of IISE-γ. We notice that everything is left unchanged in the case in which we have knowledge on $\phi_m \geq \gamma_m$. We will highlight this equivalence throughout the proofs. Algorithm 3 reports the pseudo-code: as we can see, the only modification lies in the elimination rules.

We now dive into the analysis of IISE-γ. First of all, notice that Lemma 2 holds unchanged.

Lemma 8 (Order-aware non-elimination of the optimal arm)**.** *With probability at least $1-\delta$, the optimal arm remains in the active set S of IISE-γ until termination.*

Proof. We notice that, considering a generic active fidelity $m \in [M]$ and $t \geq 1$, arm $i \neq 1$ is removed from S if and only if there exists $j \in S$ and $j \neq i$ for which it holds that:

$$\hat{\mu}_{j,m,t} - B\left(t, \frac{\delta}{KM}\right) \geq \hat{\mu}_{i,m,t} + B\left(t, \frac{\delta}{KM}\right) + \gamma_m.$$

However, due to Lemma 2, we know that with probability at least $1-\delta$ the following conditions holds:

$$\hat{\mu}_{j,m,t} \leq \mu_{j,m} + B\left(t, \frac{\delta}{KM}\right) = \mu_{j,M} + B\left(t, \frac{\delta}{KM}\right) + \xi_{j,m},$$
$$\hat{\mu}_{i,m,t} \geq \mu_{i,m} - B\left(t, \frac{\delta}{KM}\right) = \mu_{i,M} - B\left(t, \frac{\delta}{KM}\right) + \xi_{i,m}.$$

Plugging these inequalities into the elimination condition, we obtain:

$$\mu_{j,M} + \xi_{j,m} - \xi_{i,m} \geq \mu_{i,M} + \gamma_m.$$

Since, $\xi_{j,m} - \xi_{i,m} \leq \gamma_m$ by definition[11], we have that:

$$\mu_{j,M} + \gamma_m \geq \mu_{i,M} + \gamma_m,$$

which is never satisfied for $i = 1$, thus concluding the proof. $\square$

[11] Note that this also holds for $\phi_m \geq \gamma_m$.

Lemma 9 (Order-aware arm elimination)**.** *Consider $\alpha_m > 0$ for all $m \in [M-1]$ and $\alpha_M = 0$. Suppose $|S| > 2$ at the beginning of a given phase $m \in [M]$ of IISE-γ. Consider $i \in [K]$ such that $i \neq 1$. With probability at least $1-\delta$:*

- *if $i \in S$ and $\Delta_i \geq 2\gamma_m + \alpha_m$, then, arm i will be removed from S during phase m using, at most $\left\lceil \frac{128\sigma^2}{(\Delta_i - 2\gamma_m)^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i-2\gamma_m)^2\delta}\right)\right\rceil$ samples at arm i;*
- *if $i \notin S$ and $m > 1$, then $\Delta_i < \min_{\overline{m}<m}\{2\gamma_{\overline{m}} + \alpha_{\overline{m}}\}$.*

Proof. Let us focus on $i \in S$, $i \neq 1$, and $\Delta_i \geq 2\gamma_m + \alpha_m$. First of all, we notice that with probability at least $1-\delta$ the optimal arm remains in S until termination (Lemma 8). Therefore, one of the events that eliminate arm i from S is given by:

$$\hat{\mu}_{1,m,t} - B\left(t, \frac{\delta}{KM}\right) \geq \hat{\mu}_{i,m,t} + B\left(t, \frac{\delta}{KM}\right) + \gamma_m. \tag{24}$$

Moreover, due to Lemma 2, with probability at least $1-\delta$, the following holds:

$$\hat{\mu}_{1,m,t} \geq \mu_{1,m} - B\left(t, \frac{\delta}{KM}\right) = \mu_{1,M} - B\left(t, \frac{\delta}{KM}\right) + \xi_{1,m},$$
$$\hat{\mu}_{i,m,t} \leq \mu_{i,m} + B\left(t, \frac{\delta}{KM}\right) = \mu_{i,M} + B\left(t, \frac{\delta}{KM}\right) + \xi_{i,m}.$$

It follows that event of Equation (24) is guaranteed to occur if:

$$\mu_{1,M} - 2B\left(t, \frac{\delta}{KM}\right) + \xi_{1,m} \geq \mu_{i,M} + 2B\left(t, \frac{\delta}{KM}\right) + \xi_{i,m} + \gamma_m,$$

which, can be rewritten as:

$$\Delta_i + (\xi_{1,m} - \xi_{i,m}) - \gamma_m \geq 4B\left(t, \frac{\delta}{KM}\right),$$

or, more strongly:[12]

$$\Delta_i - 2\gamma_m \geq 4B\left(t, \frac{\delta}{KM}\right).$$

Due to Lemma 4, this inequality is satisfied as soon as:

$$t \geq \left\lceil \frac{128\sigma^2}{(\Delta_i - 2\gamma_m)^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i - 2\gamma_m)^2\delta}\right)\right\rceil.$$

What is left, is proving that the condition that guarantees to remove i from the active set S (i.e., $\Delta_i - 2\gamma_m \geq 4B\left(t, \frac{\delta}{KM}\right)$) activates before switching to next phase. For the case $m = M$ this is trivially true. Consider now $m < M$. Line 4 of IISE-γ states that it will proceed to the next phase only if $\alpha_m \geq 4B\left(t, \frac{\delta}{KM}\right)$. It follows that, if $\Delta_i \geq 2\gamma_m + \alpha_m$, then arm i will be eliminated in phase m, which concludes the first part of the proof.

Focus now on $i \notin S$ and $m > 1$. The proof follows as a direct consequence of the previous claim.

□

Now, before moving to the proof of the cost complexity result, we introduce the equivalent of Assumption 1 for the maximum bias variation metric.

Assumption 2 (Costs and Maximum Bias Variation Relationship)**.** *For every fidelity $\overline{m} \in [M-1]$, it holds that:*

$$\sum_{m<\overline{m}} \min_{k>m} \frac{(\sqrt{\lambda_k} - \sqrt{\lambda_m})^2}{(\gamma_m - \gamma_k)^2} \leq \min_{k>\overline{m}} \frac{(\sqrt{\lambda_k} - \sqrt{\lambda_{\overline{m}}})^2}{(\gamma_{\overline{m}} - \gamma_k)^2}. \tag{25}$$

[12]Notice that this passage could be carried out with ϕ_m instead of γ_m. This would lead to a final results whose order is given by $\frac{1}{(\Delta_i - 2\phi_m)^2}$.

The interpretation behind Equation (25) is the same as the one provided in Section 4 of the main text for the maximum biases ξ's. In the case in which $\phi_m \geq \gamma_m$ is available to the learner, the assumption required to make the multi-fidelity provably convenient becomes:

Assumption 3 (Costs and Upper Bounds on Maximum Bias Variation Relationship)**.** *For every fidelity $\overline{m} \in [M-1]$, it holds that:*

$$\sum_{m<\overline{m}} \min_{k>m} \frac{(\sqrt{\lambda_k}-\sqrt{\lambda_m})^2}{(\phi_m-\phi_k)^2} \leq \min_{k>\overline{m}} \frac{(\sqrt{\lambda_k}-\sqrt{\lambda_{\overline{m}}})^2}{(\phi_{\overline{m}}-\phi_k)^2}. \tag{26}$$

Notice that, Assumption 3 holds in the stochastic planning application presented in Section 4; and, moreover, that in that case, we have knowledge on ϕ_m (see Section 5).

We are now ready to state the upper bound on cost complexity of IISE-γ. We present the result for γ, but everything holds with ϕ as well.

Theorem 5. *If $\alpha_m > 0$ for every $m \in [M-1]$ and $\alpha_M = 0$, then, with probability at least $1-\delta$, IISE-γ returns the optimal arm with cost complexity $c(\tau)$ upper bounded by:*

$$c(\tau) \leq O\left(\sum_{i=2}^{K} \frac{\lambda_{m_i}\sigma^2}{(\Delta_i - 2\gamma_{m_i})^2} \log\left(\frac{\sigma^2 MK}{(\Delta_i-2\gamma_{m_i})^2\delta} \right) + \sum_{m<m_i} \frac{\lambda_m \sigma^2}{\alpha_m^2} \log\left(\frac{\sigma^2 MK}{\alpha_m^2 \delta} \right) \right), \tag{27}$$

where m_i is defined as the smallest phase for which $\Delta_i \geq 2\gamma_m + \alpha_m$ holds.

Moreover, under Assumption 2:

$$c(\tau) \leq \widetilde{O}\left(\sum_{i=2}^{K} \min_{m\in[M]:\Delta_i > 2\gamma_m} \frac{\lambda_m \sigma^2}{(\Delta_i - 2\gamma_m)^2} \right).$$

Proof. Let us first focus on proving Equation (27). Consider $i \neq 1$. Due to Lemma 9, using fidelity m_i, arm i will be discarded from the active set with probability at least $1-\delta$, using a number of samples upper bounded by:

$$T_{i,m_i}(\tau) \leq \left\lceil \frac{128\sigma^2}{(\Delta_i-2\gamma_{m_i})^2} \log\left(\frac{128KM\sigma^2}{(\Delta_i - 2\gamma_{m_i})^2\delta} \right) \right\rceil.$$

Since m_i, by definition, is the first phase at which Lemma 9 guarantees elimination of arm i, we have $T_{i,k}(\tau) = 0$ for $k > m_i$. At this point, to derive the total cost required by IISE to discard i, we need to consider also the number of pulls at fidelity $m < m_i$. This is provided by Lemma 4, which yields an upper bound on the maximum duration of a given fidelity. More specifically:

$$T_{i,m}(\tau) \leq \left\lceil \frac{128\sigma^2}{\alpha_m^2} \log\left(\frac{128KM\sigma^2}{\alpha_m^2 \delta} \right) \right\rceil.$$

Given that $\alpha_m > 0$ for all $m < M$, we notice how the duration of each phase is finite. Together with Lemma 8, and $\alpha_M = 0$, this guarantees that IISE-γ is δ-correct.

To obtain the upper bound on the cost complexity, it is sufficient to multiply the number of pulls with their related cost. Summing over the arms concludes the proof of Equation (27). For what concern, instead, the second part of the theorem, it follows directly applying the same proofs of Theorem 3, replacing ξ's with γ's.

□

D.2 Practical relevance of maximum bias variation

In this section, we prove that in the setting highlighted in Section 5 we can use ξ_m as upper bound on γ_m.

Proposition 4. *Let $\mu_{i,M} - \mu_{i,m} \geq 0$ for all $i \in [K]$ and $m \in [M]$. Then, for each $m \in [M]$:*

$$\gamma_m \leq \xi_m.$$

Proof. By definition, we know that $\mu_{i,m} = \mu_{i,M} + \xi_{i,m}$ for all $i \in [K]$ and $m \in [M]$ (with $\xi_{i,m} < 0$).

Then, since fidelity $m < M$ underestimates fidelity M, we can write:

$$\mu_{i,M} - \mu_{i,m} = -\xi_{i,m} \le \max_{i \in [K]} |\xi_{i,m}| \le \xi_m.$$

Now, by definition of γ_m:

$$\gamma_m \coloneqq \max_{i,j \in [K]} \{\xi_{i,m} - \xi_{j,m}\} \le \max_{i \in [K]} |\xi_{i,m}| \le \xi_m,$$

where the first inequality follows from the fact that all $\xi_{i,m}$'s have the same sign.

$\square$

E Experiment Details

We have run the experiments using 100 Intel(R) Xeon(R) Gold 6238R CPU @ 2.20GHz cpus and 256GB of RAM. The total time taken to have all the results was around 48 hours. To speed up the running time, we have relied on a parallel implementation that allows us to compure several independent runs in parallel on different cores.

E.1 Synthetic domains

We provide further details on the experiments on synthetic domains. As highlighted in Section 7, to make the γ settings directly comparable with the one that uses ξ, we generated the arms in such a way that the fidelity index m is left unchanged. The arm generation process is the following one.

First of all, we consider unit-variance Gaussian distributions; their means were generated at random as follows. We first generate means for fidelity M, by sampling their value from a normal distribution. Then, for each fidelity $m \in [M-1]$, we first specify γ_m and a bias term $b_m > 0$. Once this is done, $\mu_{i,m}$ is sampled from a uniform distribution defined on the interval $[\mu_{i,M} - b_m - \frac{\gamma_m}{2}, \mu_{i,M} - b_m + \frac{\gamma_m}{2}]$. Since γ_m holds with the equality for two random arms k, j we fix $\mu_{k,m} = \mu_{k,M} - b_m - \frac{\gamma_m}{2}$ and $\mu_{j,m} = \mu_{j,M} - b_m + \frac{\gamma_m}{2}$ Notice that this generation process provides upper bounds on maximum bias ξ_m given by $b_m + \frac{\gamma_m}{2}$. Finally, we remark that setting $\gamma_m > \gamma_{m+1}$ and $b_m > b_{m+1}$ for all $m \in [M-1]$ guarantees $\xi_m > \xi_{m+1}$, thus the fidelity index is left unchanged.

Synthetic A setting parameters are: $K = 2000$, $M = 4$, $\lambda = [1, 10, 100, 1000]$, $\xi = [1.15, 0.225, 0.015, 0]$, $\gamma = [0.3, 0.05, 0.001, 0]$ and $b = [1.0, 0.2, 0.01, 0]$. For Synthetic B, instead: $K = 1000$, $M = 5$, $\lambda = [16, 64, 256, 1024, 4096]$, $\xi = [1.15, 0.45, 0.105, 0.0105, 0]$, $\gamma = [0.3, 0.1, 0.01, 0.001, 0]$, and $b = [1.0, 0.4, 0.1, 0.01, 0]$.

For what concerns MFE, we set $\alpha_m \coloneqq \xi_m(4.1)$. In both domains and for all algorithms, we used $\delta = 0.001$.

E.2 Yahtzee

The Yahtzee game has been proposed in [3] as a benchmark for sequential decision making problems. The goal of this experiment is showing that IISE can successfully optimize the cost complexity of a given planning algorithm in a stochastic domain. We notice that this is different w.r.t. to testing the efficiency of the planning strategy in its original sense [18], i.e., we proposed a method and an analysis for MF-BAI, not for MF stochastic planning. Indeed, given the simplification of the multi-armed bandit setting, there is structure on the planning problem that our study did not aim at optimizing (e.g., the re-use of experience at different depth m). This is the reason why we are not taking a direct comparison with planning methods, but the most fair baseline is given by Successive Elimination [10].

We now provide a more in depth description of the problem. The game proceeds in 13 rounds. At the beginning of each round, the player will roll 5 dice and observes the result. After that, he can choose to re-roll a subset of the 5 dice; this can be repeated up to 3 times. Once the final combination has been rolled, the player needs to choose a particular move that maps dice combination to scores. As already highlighted in Section 7, typical moves are:

- "Sixes", with score given by the sum of dice with the number 6.

- “Yahtzee”, that assigns 50 points if all the dice show the same number, and 0 otherwise.

The total number of these possible moves is 14; all of them provide scores that are always non-negative. In its original version, the sequentiality within the problem arises from the fact that once a given move has been selected at a certain round t, than, it cannot be selected at rounds $\bar{t} > t$. The goal of the player is to maximize the total sum of points.

In our setup, we consider the following modified version. First of all, it is not possible to re-roll the 5 dice: once the initial combination has been rolled at the beginning of a given round, a move needs to be selected. Secondly, we are interested in the discounted case: the goal of the player is to maximize the sum of discounted of rewards (i.e., we are searching for an ϵ-optimal action). Finally, we consider the variation in which the player can re-select moves that he already played, but all the rewards gathered from that move on will be set to 0.

We apply our algorithm to choose the first move in this Yahtzee variation. In particular, the agent observes the initial combination of 5 dice (which is fixed and known) and needs to understand which move among the 14 available leads to the highest sum of discounted rewards. As problem setup, we set $\eta = 0.8$. The initial and observed dice combination is the dice combination $[1, 1, 1, 1, 1]$. We consider $M = 8$ fidelity with the following meaning: $m = 1$ stands for planning 6-step ahead from the initial position, $m = 2$ stands for planning 7 step ahead from the initial position, and so on up to $M = 8$ which means planning 13-step ahead, that is equal to length of entire game.

In our experiments, to ease the burden of the required computing power, we pre-compute the exact expectation of each of the action in the initial position via backward induction. Then, to make the problem more challenging, when running our experiments we add truncated Gaussian noise to each expectation. In MFE, we set $\alpha_m := \xi_m(4.005)$. For all algorithms we used $\delta = 0.001$. Further details can be found within the codebase.