We thank the reviewers for their efforts. We will make sure to fix the typos. We address below the main questions.

Response to Reviewer 1. Thank you for the supportive comments!
1.1 Not too much new insight in the main regret theorems? Our analysis shows that the regret of learning in an FMDP is on the same order as the sum of the regret of learning in each transition component (suppose the reward is known). This relationship is obvious if the transition and reward factorizations are the same, namely $\mathcal{X}[I_i] = \mathcal{X}[J_i]$ for all $i \in [m]$, in which case the FMDP has m independent components. *The remarkable aspect here is that such a relationship holds, even if the transition and reward factorizations differ arbitrarily*. Moreover, there is additional insight in the algorithm design to guarantee such theorems, as discussed in our *Response 2.1* below.

Response to Reviewer 2. Thank you for the thoughtful feedback! We address the key points below for space reasons.
2.1 Lacking in conveying the intuition and insight behind the cross-component bonuses? We thank the reviewer for pointing it out, and are happy to add some of the following discussion in the final paper.

Intuition. *An FMDP is not a simple collection of some independent transition components with corresponding reward components*, for which the sum of component-wise bonuses *suffices* to ensure UCB exploration. Instead, for an FMDP, since the transition and reward factorizations may differ arbitrarily, the naive sum of component-wise bonuses may *not suffice* (discussed around Line 230), and thus we may expect some additional bonuses. A contribution of this paper is to identify that cross-component bonuses (CCBs) can fulfill the role of the needed additional bonuses.

Insight. *The CCBs help stablize early-stage exploration.* Concretely, consider the four phases. **1:** initially all counters are small, and the CCBs encourage exploration of the ith and jth components with large $S_i S_j$. **2:** the counters of some components (say, the ith) grow faster such that $N_i(x) \geq S_i$, and the CCBs (say, with the jth) satisfy $HL\sqrt{S_i/N_i(x) \cdot S_j/N_j(x)} \leq HL\sqrt{S_j/N_j(x)}$, which prioritize exploration of underexplored components until all counters satisfy $N_i(x) \geq S_i$. **3:** the CCBs balance the exploration among different components until all counters satisfy $\max\{N_i(x), N_j(x)\} \geq S_i S_j$. **4:** the component-wise bonuses dominate the overall bonus and exploration stabilizes.

To summarize the insight, in the long run, different growth rates of the counters reflect different importance of the components towards maximizing cumulative rewards, and early on, their growth can suffer large variance. In this sense, the CCBs play a role of *variance reduction*. The intuition and insight here may also be helpful in future research on FMDPs, e.g., the function approximation setting and the structure-agnostic setting.

2.2 Is the term "factored MDP" proper? Although the term "factored MDP" is not our invention, we think that it is indeed accurate. In a factored MDP, in addition to the state space, the state-action space, the transition and the reward can all be factored. In fact, *the transition factorization turns out to be essential in reducing the regret* compared with a nonfactored MDP, since the regret depends on $X[I_i]$, the size of the scope of a transition component.

Response to Reviewer 3. Thank you for the kind and positive comments!
3.1 Not considering function approximation setting? Even for nonfactored MDPs, the regret bounds in the function approximation setting are only established under stringent assumptions, e.g., linear MDPs. Our work can be seen as a concrete first step towards understanding more general settings.

3.2 Assuming an oracle solver for FMDPs? Actually, neither F-UCBVI nor F-EULER assumes an oracle solver. The solver, based on value iteration, is explicitly described in Algorithm 2.

3.3 Other issues? **Intuition:** please see our *Response 2.1* for an intuitive explanation regarding why we need the cross-component bonuses. These bonuses suffice to guarantee minimax optimal regret up to log factors, though it is possible that higher-order interactions could provide other benefits. **Line 200:** what we meant to say is that while the F-EULER algorithm does not assume the knowledge of $\mathcal{Q}_i, \mathcal{R}_i, \mathcal{G}$, its regret guarantees automatically adapt to these quantities. We shall rephrase our expression in the final paper.

Response to Reviewer 4. Thank you for the constructive comments!
4.1 Lacking novelty? We would like to emphasize three noteworthy contributions of this paper. 1) As mentioned in our *Response 1.1*, our main regret theorems show that the regret of learning in an FMDP is on the same order of the sum of those in each transition component. This clean result, however, is nontrivial and even surprising, since the transition and reward factorizations may differ arbitrarily. 2) We identify a correct form (cross-component) of bonuses that are novel and *highlight the key difference towards achieving UCB exploration compared with the nonfactored setting*. Moreover, these cross-component bonuses offer *new insight* (see our *Response 2.1*). The insight and techniques we develop might also be helpful for future research on FMDPs. 3) We are *the first to unveil the intricacy of proving lower bounds* in the factored setting and *to establish nontrivial lower bounds* for some fairly general factored structures.

4.2 Assuming known factorization? Although in some practical problems the factorization is indeed known (see [12] for an example), we admit that this assumption can be a bit restrictive. On the other hand, the primary aim of this paper is to promote our understanding about the fundamental theoretic limits, and in this sense, the result here can be seen as a solid first step towards our understanding about the more general structure-agnostic setting.