This reply is for the reviewer 5.

In the following, we refer to limit-diterministic generalized Büchi automaton as LDGBA and non-generalized one as LDBA.

1. Contributions in our paper.
   Our contributions are as follows.
   - The sparsity of rewards is relaxed comparing with using LDGBA by introducing the augmentation of LDBAs. Therefore, our proposed algorithm is more sample efficient.
   - The recurrent classes of the Markov chain induced by a product MDP $M^{\otimes}$ and a positional policy π on $M^{\otimes}$ are classified as ones that has at least one accepting transition in each accepting set or ones that has no accepting transition in all accepting sets without depending on the policy.
2. Discussions in the section with the simulation.
   If we construct product MDP $M^{\otimes}$ of an MDP and an LDGBA without any augmentation corresponding to a given LTL formula φ, a policy satisfying φ may not be exist depending on $M^{\otimes}$ and φ. We show that an example in which there is not policies satisfying an LTL formula φ when using the corresponding LDGBA without any augmentation.
3. The motivations of using LDBA.
   the motivation of using LDGBA is to relax the sparsity of rewards. the motivation of our augmentation is to circulate all accepting sets without depending on an MDP, an LTL formula, and a policy.
4. Comparing to the size of state space of the product MDP with the augmented LDGBA and one with the non-augmented LDGBA.
5. Does maximizing the collection of the proposed rewards implies the maximization of the satisfaction probability?
   No, it does. It does not generally hold. However, our proposed method can be combined with some research results of attempting to maximize the satisfaction probability such as [1].

References

[1] E. M. Hahn, M. Perez, S. Schewe, F. Somenzi, A. Triverdi, and D. Wojtczak, “Omega-regular objective in model-free reinforcement learning,” *Lecture Notes in Computer Science*, no. 11427, pp. 395–412, 2019.