We would like to thank all the reviewers for their detailed reviews and valuable comments. We will address all the comments in the revised version of the paper. We will now discuss some common issues raised by the reviewers and then move to specific comments.

Several reviewers mentioned that our paper improves approximation guarantees for k-means++ by a constant factor. We want to point out that we improve the bi-criteria approximation ratio for k-means++ very substantially in the regime where the number of additional centers is small. This regime is important because it is the one to which practical heuristics for determining k (like the elbow method) might lead to. More specifically, when the number of additional centers is $\Delta = \frac{k}{\log k}$, our approximation guarantee is $O(\log \log k)$ while the k-means++ bound by Arthur and Vassilvitskii (2007) and the bi-criteria bounds by Aggarwal, Deshpande, and Kannan (2009) and Wei (2016) give only an $O(\log k)$ approximation. Thus, in this regime of parameters, our paper provides approximation guarantees that are substantially stronger than previously known.

As was pointed out by Reviewer 2, the bounds on the approximation factor for *non-bi-criteria* k-means++ due to Arthur and Vassilvitskii (2007) are tight up to a constant factor. Their upper bound is $8 \ln k + 2$, and their lower bound is $2 \ln k$. Since k-means++ and k-means$\|$ are extensively used in practice, we believe it is really important to narrow down the gap between upper and lower bounds even further. Our paper does so by improving the upper bound from $8 \ln k + 2$ to $5 \ln k + 2$. Moreover, our results (specifically, Lemma 4.1) can be used to get similar improvements for many other papers on k-means++ and its variants. We also show that our bound of 5 for Lemma 4.1 is tight.

Finally, let us mention that our paper not only gives better approximation guarantees for k-means$\|$ than the paper by Bahmani, Moseley, Vattani, Kumar, and Vassilvitskii (2012) but also provides a simpler analysis.

Reviewer 1: We ran some experiments with $\ell \cdot T = k$. The performance was similar to k-means++. Thank you for pointing us to the "k-means++: Few More Steps Yield Constant Approximation" paper. It is a very interesting paper, and we will cite it in the revised version. We currently cite Aggarwal, Deshpande, and Kannan (2009) in the introduction. We will cite this paper in other relevant places as well (including the martingale analysis).

Reviewer 3: Thank you for the detailed suggestions. We will reorganize the paper to improve its readability.

Reviewer 4: Thank you for the detailed comments. We agree that a more unified analysis for k-means++ and k-means$\|$ would be nice to have. But one bottleneck for achieving this is that although the marginal distributions for picking individual points are the same for each round in k-means++ and k-means$\|$ when $\ell = 1$ and $T = k$, the joint distributions are quite different. In each round, k-means++ picks *exactly* one center whereas k-means$\|$ can pick any number of centers.

Regarding the paper by Rozhon (2020): We got to know about this work only after the list of accepted papers for ICML 2020 came out (which was very close to the NeurIPS deadline). So we did not get much time to compare our work with that paper. We will certainly do this in the revised version of our paper. In general, the guarantees proved in that paper for k-means$\|$ are orthogonal to our guarantees.

We will add references to the NP-hardness and APX-hardness of k-means results.