We thank the reviewers for their thorough reading of the paper and many insightful and useful comments. Below we outline how we will address the reviewer's comments.

There were two main concerns raised by the reviewers.

1- Running time dependency on the aspect ratio Δ (max distance over minimum distance).

We believe that in most practical settings the logarithm of the aspect ratio is rather small. The assumption of bounded aspect ratio allows a clean presentation of the result. The dependency can, for example, be removed (using ideas from prior works) if we have a very rough estimate of the optimum solution (e.g., within a factor n or even n^{10}). Indeed, in that case, we can obtain an instance in which each coordinate of each point is an integer in range $[1,\text{poly}(n)]$ by losing a factor $1 + 1/n$ in the approximation guarantee, see [1]. This bounds $\log \Delta = O(\log nd)$. In practice this can be achieved very efficiently. We have provided explanations about this in Appendix (line 749-761).

2- Improvement is only for sufficiently large value of k.

We believe that the case of many centers is of similar importance as the case of small k. For example, an important application of k-means is vector quantization, where we need large k to quantize large datasets. See, for example [Cartesian K-Means, Norouzi, Fleet, CVPR'13] for some further discussion.

3- Empirical Evaluation on more datasets.

We will provide experimental evaluation on more data sets. We have run all the algorithms on Census dataset ($n = 2,458,285; d = 29$) as well. The results are very similar to Song dataset. The quality of the solution is comparable with the baselines ($2 - 3\%$ worse than k-means++ and almost the same as Afkmc2). Our algorithm is noticeably faster from $k = 1000$ and is $1 - 2$ order of magnitude faster than the baselines for large values of k. We will add this to the camera ready version.

Reviewer 2: Regarding the quality decrease: We only saw a larger decrease in the quality of the solution for one dataset out of three (and for small k).

Regarding the large value of k, please refer to the previous discussion.

We will address the near-linear time and the typos in the camera ready version (thanks for the comments).

Reviewer 3: It seems there is a misunderstanding here. FASTkmeans++ is theoretically **faster** than RejectionSampling but it does not come with a theoretical approximation guarantee. In the experiments, it sometimes turns our that Rejection sampling algorithm is slightly faster than Fast k-means++ due the random nature of the algorithm.

For Census dataset, the average cluster size for $k = 5000$ is around 500 and our algorithm is 1-2 orders of magnitude faster. Additionally, we also refer to our discussion of cluster sizes/number of centers above.

Reviewer 4: We disagree with the statement that the algorithms are quite involved. We agree that their theoretical analysis is complex, but the implementation is rather simple. We have also submitted the code and will add the code to github after the paper is accepted. The running time of lemma 4.1 does not depend on k. The total opening time for all the centers is what mentioned there. We will add Algorithm 1 in this lemma.

Memory requirement is $O(nd + n\log n + n\log\Delta)$, we will add that.

We will make Corollary 5.5 more precise.

Corollary 4.3 does not provide any approximation guarantee and the running time follows from the description of the algorithm. Notice that Lemma 4.2 only states the probability of sampling a point and this does not result in any approximation guarantee for Fast k-means++ algorithm. Indeed it is not clear if the presented Fast k-means++ algorithm has any approximation guarantee, only Rejection sampling algorithm has. We remark that after embedding the point into multiple trees, we do not have the triangle inequality, therefore one cannot simply use the arguments of the proof of the noisy k-means++ algorithm from previous work here to prove an approximation guarantee.

We will also discuss the number of trees selected in more detail. The selection becomes clearer in the proof of Lemma 3.1 in the Appendix but we will add intuition in the camera ready version for this choice.

We will address the remaining editorial comments in the camera ready version.

[1] Better Guarantees for k-Means and Euclidean k-Median by Primal-Dual Algorithms. Sara Ahmadian, Ashkan Norouzi-Fard, Ola Svensson, Justin Ward.