We thank the reviewers for their insightful comments. Below we respond to the main points raised by the reviewers. The manuscript will be updated to reflect the reviewers' suggestions and our responses below.

Reviewer 1: Our results show that for any privacy budget ε, we only need $\lceil (\log_2 e) \cdot \varepsilon \rceil$ bits of communication budget to achieve the order-optimal estimation error under ε-LDP. This result can be viewed in two different ways. First, as the reviewer suggests, when we use larger ε, we need a larger communication budget to achieve the order-optimal error. Second, as we emphasize in the paper, we cannot improve the error further by using a communication budget larger than this characterized threshold, since the performance will be dominated by the more stringent constraint, i.e. privacy.

Intuition behind our results. The privacy level ε dictates the "noise level" of the channel we induce from each client to the server, and therefore its "capacity", the amount of information that can be supported by this channel. The larger ε, the larger this capacity and therefore increasing the communication budget benefits the estimation task, but only up to a certain threshold, i.e. the capacity of the ε-LDP channel. However, even with this intuition, we believe it is highly surprising that the encoding over this channel can be done in such a way that it is simultaneously optimal from a both privacy and compression perspective. While our algorithms include some elements from previous approaches, they use these elements in novel ways. For example, as pointed out by the reviewer our scheme for frequency estimation builds on Hadamard matrices, which also appears in [1]. However, [1] uses the Hadamard transform as an efficiently computable random rotation of the observation, while our scheme uses the recursive structure of the Hadamard matrix to construct an encoding scheme whose error decreases exponentially with the number of communication bits used.

Reviewer 2: *The $\varepsilon = \Omega(1)$ regime.* The reviewer mentions that the $\varepsilon = \Omega(1)$ regime is of practical interest. We agree that this regime is frequently used in applications because of utility concerns, and in fact, this regime has become even more relevant due to recent results in the shuffled model of DP [Erlingsson et al. (2020)].

Related work and the presentation of tables. We would like to point out that even though [9, 12] (Duchi-Jordan-Wainwright) is not discussed in the "Related Work" section, we discuss their scheme `privUnit` in detail in Section A and Section C, and compare its performance to our schemes. The experiments show that `privUnit`, while optimal from a privacy perspective alone, does not perform well in the presence of communication constraints. In Table 2, the blue (or red) color indicates that the scheme is optimal (or not). We will revise the tables to improve clarity.

Tight upper bound for frequency estimation with ℓ_∞ loss. We conjecture that our upper bound is tight, though this requires proving a matching lower bound, which seems nontrivial. Note that [8] shows that the upper bound is tight for $\varepsilon = O(1)$, but for a general ε, the problem remains open.

Reviewer 3: *Correctness of our main results.* The reviewer doubts the correctness of our main results due to the 1-bit generic scheme in [8, Thm. 4.1]. First, we would like to clarify that the result in [8] works only for $\varepsilon = O(1)$, while in practice $\varepsilon = \Omega(1)$ regime is also of great interest (see comment to R2). In our work, we show that 1 bit is not enough for the $\varepsilon = \Omega(1)$ regime. Moreover, for larger ε and b, it is not enough to repetitively apply the 1-bit scheme in [8] (which is the same as "re-sampling"), since this makes the estimation error decay linearly in b and ε. However, for frequency estimation, we see that by cleverly designing the algorithm, we can achieve *exponential* decay with increasing b, matching the lower bounds.

How much public randomness is necessary? It can be shown that our frequency estimation scheme (RHR) is optimal not only in terms of the estimation error it achieves but also in terms of its use of public randomness. Recall that for frequency estimation with b bits communication and ε LDP constraints, RHR uses $b^* \triangleq \min(b, \lceil \varepsilon \log_2 e \rceil)$ bits communication and $\lceil \log d - b^* \rceil$ bits of public randomness, while [8] always uses $\lceil \log d \rceil$ bits of public randomness. By slightly extending Theorem 4 in [1], one can show that at least $\log d - b^* - 2$ bits of public randomness is required to get a consistent estimator. This implies that RHR is optimal in terms of the amount of shared randomness it needs, up to an additive constant. For mean estimation, our scheme uses $\min(\lceil b^* \log d \rceil, d)$ bits of public randomness, while [8] uses d bits, which leads to a significant difference, e.g., in the case $b^* = 1$ considered in [8]. On the other hand, for discrete distribution estimation (where $X_i \overset{\text{i.i.d.}}{\sim} \boldsymbol{p}$ with $\boldsymbol{p}$ supported on $\{1, ..., d\}$) our scheme requires no public randomness; whereas [8] still needs it. The same conclusion holds for the distributional version of mean estimation (where $X_i \overset{\text{i.i.d.}}{\sim} P$ with P supported on the Euclidean unit ball). As suggested by the reviewer, we will add this discussion about shared randomness in our revised version.

Reviewer 4: We thank the reviewer for spotting the typos. We will fix the typos and proofread the revised version.

Discussion on shared randomness. For the more detailed discussion on shared randomness, please see the response to R3. Indeed, for frequency estimation with communication budget $b < \log d - 2$, [1, Thm. 4] shows that shared randomness is necessary, and we can further prove that our frequency estimation scheme RHR uses *the minimum amount of shared randomness*. On the other hand, under distributional settings where local data is drawn from some unknown but fixed distribution, our schemes for both frequency estimation and mean estimation require no shared randomness.