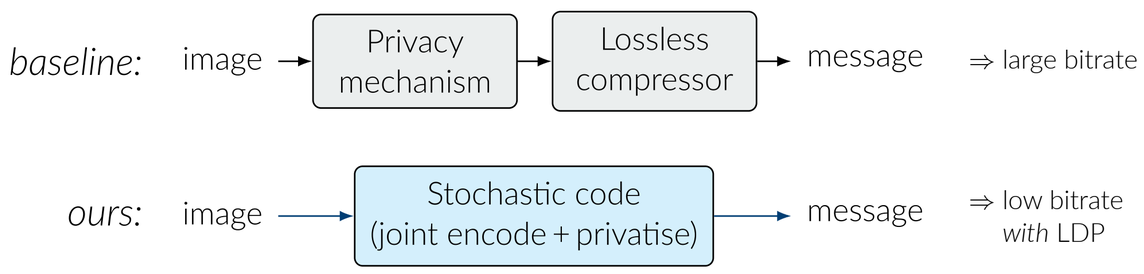
Scalable Differentially Private Data Compression via Diffusion and Stochastic Codes

Gergely Flamich Öykü Sıla Güner Yanxiao Liu Deniz Gündüz

Imperial College London

The Problem

- Compressing high-dimensional images under local-DP guarantees.
- Standard *sequential* pipeline privatises the data and then compresses it.
- Our scheme *jointly* privatises and compresses the data via *stochastic codes*.



Background

Local Differential Privacy

A randomised mechanism M is ε-LDP if for any inputs x, x' and event S,

$$\mathbb{P}\left[M(x) \in S\right] \;\leq\; e^{\varepsilon}\,\mathbb{P}\left[M(x') \in S\right].$$

The smaller ε is, the better the privacy guarantee.

Poisson Private Representation

A *relative entropy code* that **exactly** simulates and compresses arbitrary differential privacy mechanisms, and maintain good privacy guarantees.

Poisson Private Representation (PPR)

Input: private $x \in \mathcal{X}$, (ε, δ)-local DP mechanism $P_{Z|X}$, reference distribution Q, parameter $\alpha > 1$, shared randomness $(Z_i)_{i=1,2,\ldots} \overset{\text{i.i.d.}}{\sim} Q$.

(a) The user knows $(Z_i)_i$, x, $P_{Z|X}$ and performs:

(1) Generate the Poisson process $(T_i)_i$ with rate 1.

(2) Compute $\tilde{T}_i \triangleq T_i \cdot \left(\frac{dP_{Z|X}(\cdot|x)}{dQ}(Z_i)\right)^{-1}$.

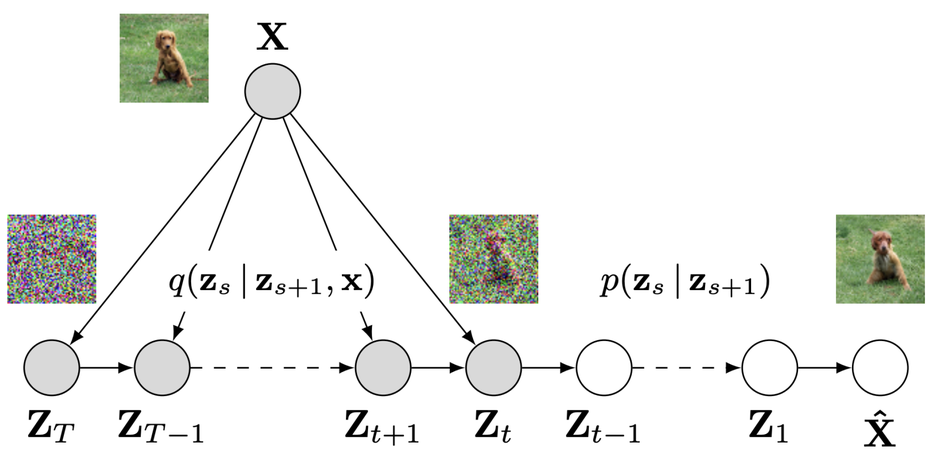
(3) Generate $K \in \mathbb{Z}_+$ with $\Pr(K = k) = \tilde{T}_k^{-\alpha} \,/\, \left(\sum_{i=1}^{\infty} \tilde{T}_i^{-\alpha}\right)$.

(4) Compress and send K.

(b) The server, which knows $(Z_i)_i$ and K, outputs $Z = Z_K$.

DiffC

DiffC [2] reuses a pretrained DDPM as a lossy compressor. At each denoising step $t \to s$, the reverse process $p_\theta(x_s \mid x_t)$ is the decoder's *proposal*; the true posterior $q(x_s \mid x_t, x_0)$ is the encoder's *target*. A stochastic code (originally PFR [4]) transmits one integer per step that picks a target-distributed x_s from candidates drawn under the proposal. The figure below is from [2]:



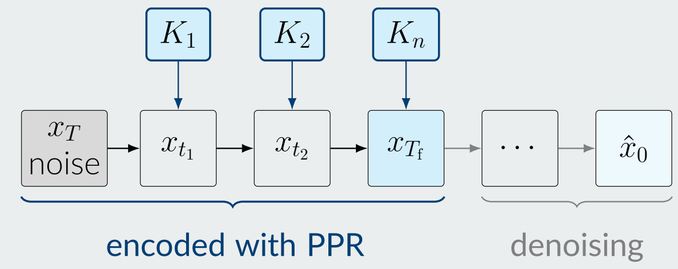
Our Contribution

1. **Step-limited PPR.** An approximate version of PPR [1] with fixed runtime and $2\alpha\varepsilon$-LDP guarantee, and is implementable.
2. **Laplace-DiffC.** Diffusion-based compression with moment-matched Laplace denoising targets — pure ε-LDP per step, negligible quality cost.
3. **DP-DiPP. up to 40× fewer bits** than privatize-then-compress on CIFAR-10.

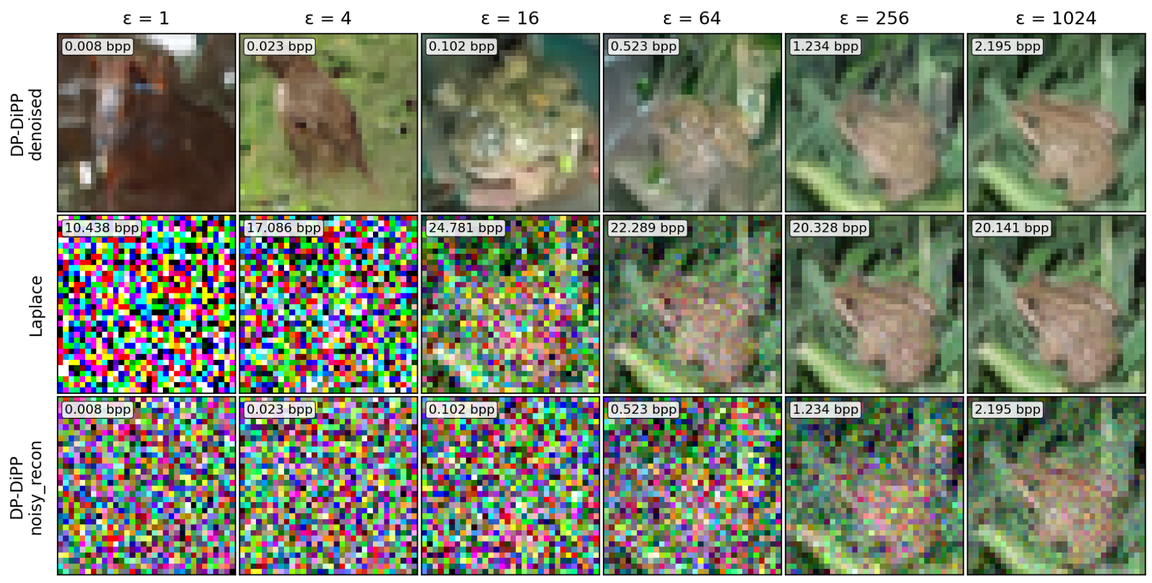
DP-DiPP: guided diffusion via privatised indices

Run a DDPM in reverse from x_T (noise) toward x_0. At each encoding step $t \to s$, step-limited PPR transmits index K_t that picks the next noisy state out of n candidates. Stop encoding at T_{f}; finish with unconditional denoising.

bitstream, $2\alpha\varepsilon$-LDP per step



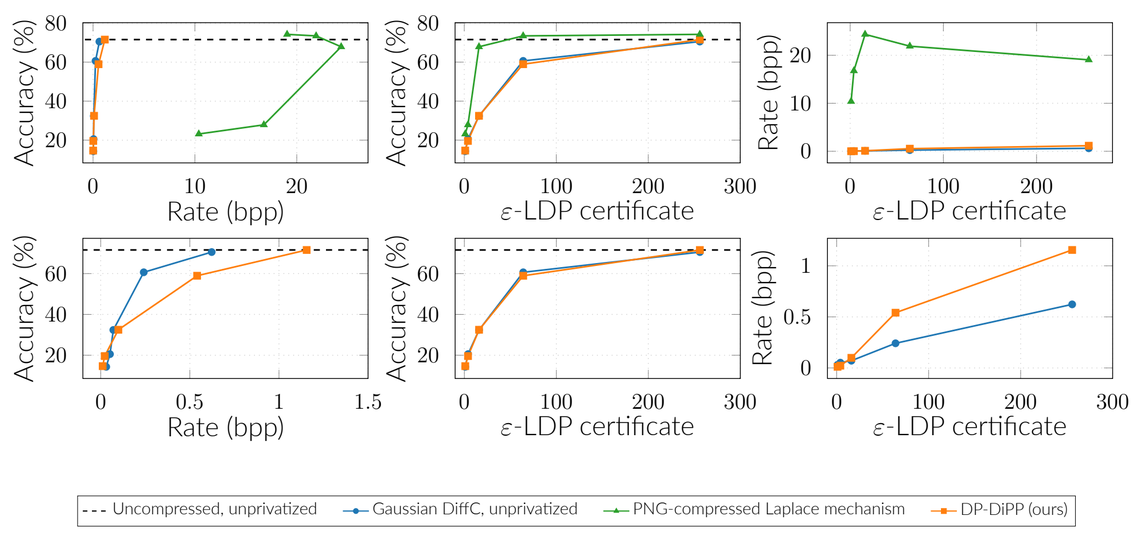
Reconstructions: DP-DiPP vs. Laplace baseline



A single CIFAR-10 image, encoded at six privacy levels. **Top:** DP-DiPP, denoised. **Middle:** privatize-then-compress (Laplace + PNG). **Bottom:** DP-DiPP raw output before the free denoising tail. Numbers in each tile = bits per pixel.

CIFAR-10: rate, privacy, utility

A ResNet-56 is trained directly on privatized+compressed images (no clean data ever seen). Sweep $\varepsilon \in \{1, 4, 16, 64, 256\}$, with $\alpha = 2$.



Top row: full bitrate range. **Bottom row**: zoom on the low-bitrate regime where DP-DiPP operates, Laplace+PNG is off-chart (≥ 10 bpp). At $\varepsilon = 64$: DP-DiPP 0.54 bpp vs. Laplace+PNG 21.9 bpp ⇒ **40×** compression at matched LDP.

References

[1] Y. Liu *et al.*, *"Universal exact compression of differentially private mechanisms,"* NeurIPS 2024.

[2] L. Theis *et al.*, *"Lossy compression with Gaussian diffusion,"* 2022.

[3] J. Vonderfecht *et al.*, *"Lossy compression with pretrained diffusion models,"* ICLR 2025.

[4] Y. Cao *et al.*, *"LDP-Slicing: local differential privacy for images via randomized bit-plane slicing,"* CVPR 2026.

o.guner25@imperial.ac.uk