

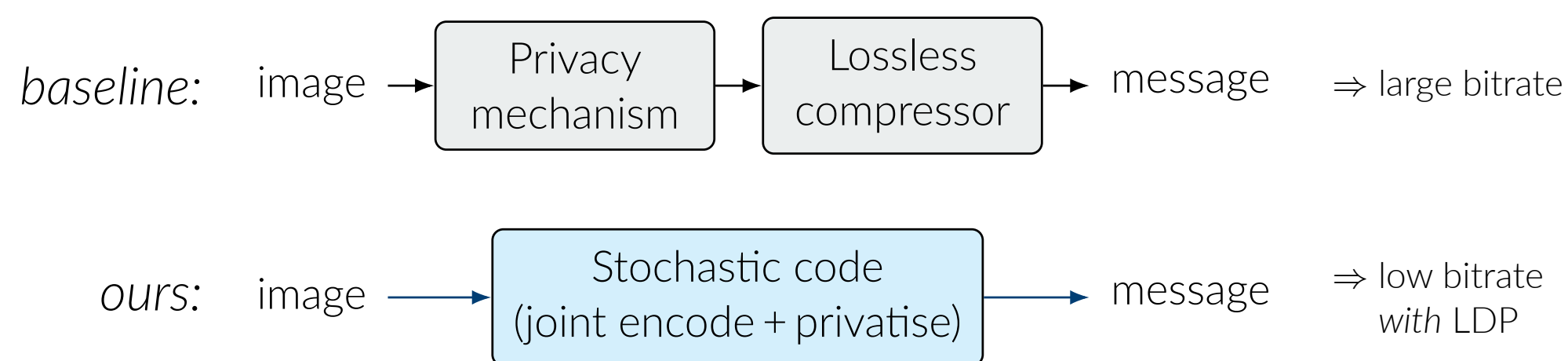
Scalable Differentially Private Data Compression via Diffusion and Stochastic Codes

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 ,

$$\mathbb{P}[M(x) \in S] \leq e^\epsilon \mathbb{P}[M(x') \in S].$$

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 $(Z_i)_{i=1,2,\dots} \stackrel{\text{i.i.d.}}{\sim} Q$.

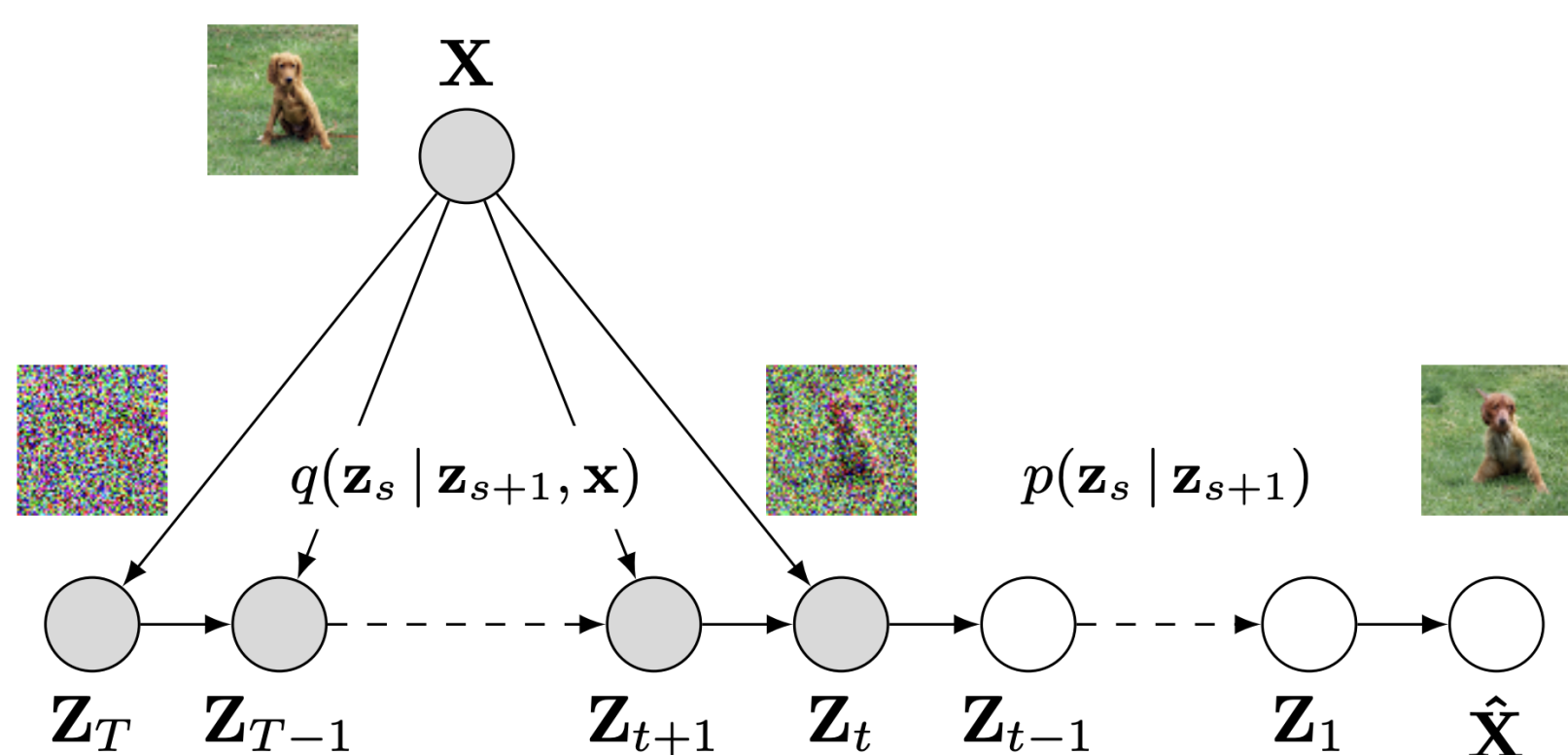
(a) The user knows $(Z_i)_i$, x , $P_{Z|X}$ and performs:

- Generate the Poisson process $(T_i)_i$ with rate 1.
- Compute $\tilde{T}_i \triangleq T_i \cdot \left(\frac{dP_{Z|X}(\cdot|x)}{dQ} (Z_i) \right)^{-1}$.
- Generate $K \in \mathbb{Z}_+$ with $\Pr(K = k) = \tilde{T}_k^{-\alpha} / \left(\sum_{i=1}^{\infty} \tilde{T}_i^{-\alpha} \right)$.
- 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 $t \rightarrow s$, the reverse process $p_\theta(x_s | x_t)$ is the decoder's *proposal*; the true posterior $q(x_s | x_t, x_0)$ 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]:



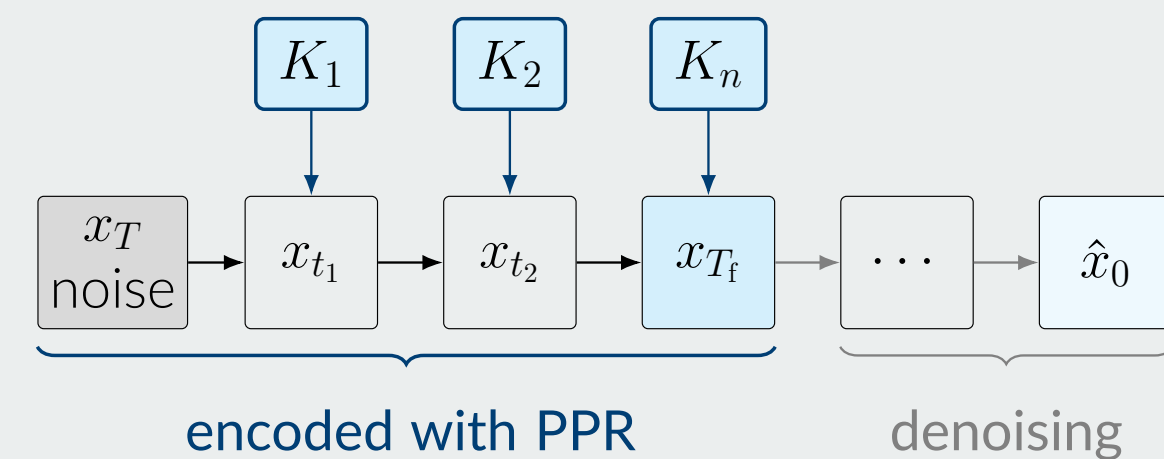
Our Contribution

- Step-limited PPR.** An approximate version of PPR [1] with fixed runtime and $2\alpha\epsilon$ -LDP guarantee, and is implementable.
- Laplace-DiffC.** Diffusion-based compression with moment-matched Laplace denoising targets – pure ϵ -LDP per step, negligible quality cost.
- DP-DiPP.** up to **30×** fewer bits than privatize-then-compress on CIFAR-10.

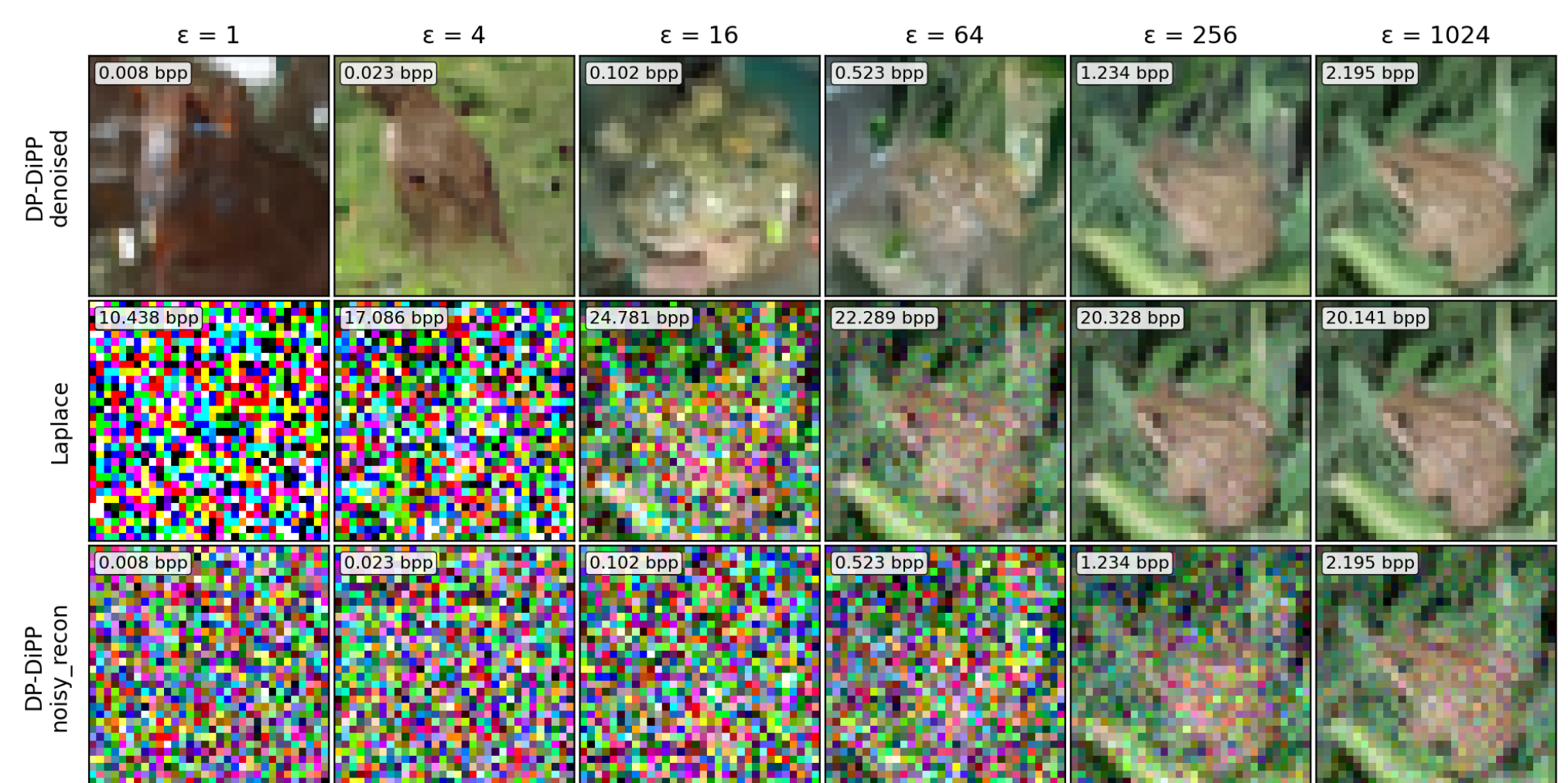
DP-DiPP: guided diffusion via privatised indices

Run a DDPM in reverse from x_T (noise) toward x_0 . At each encoding step $t \rightarrow s$, 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, $2\alpha\epsilon$ -LDP per step



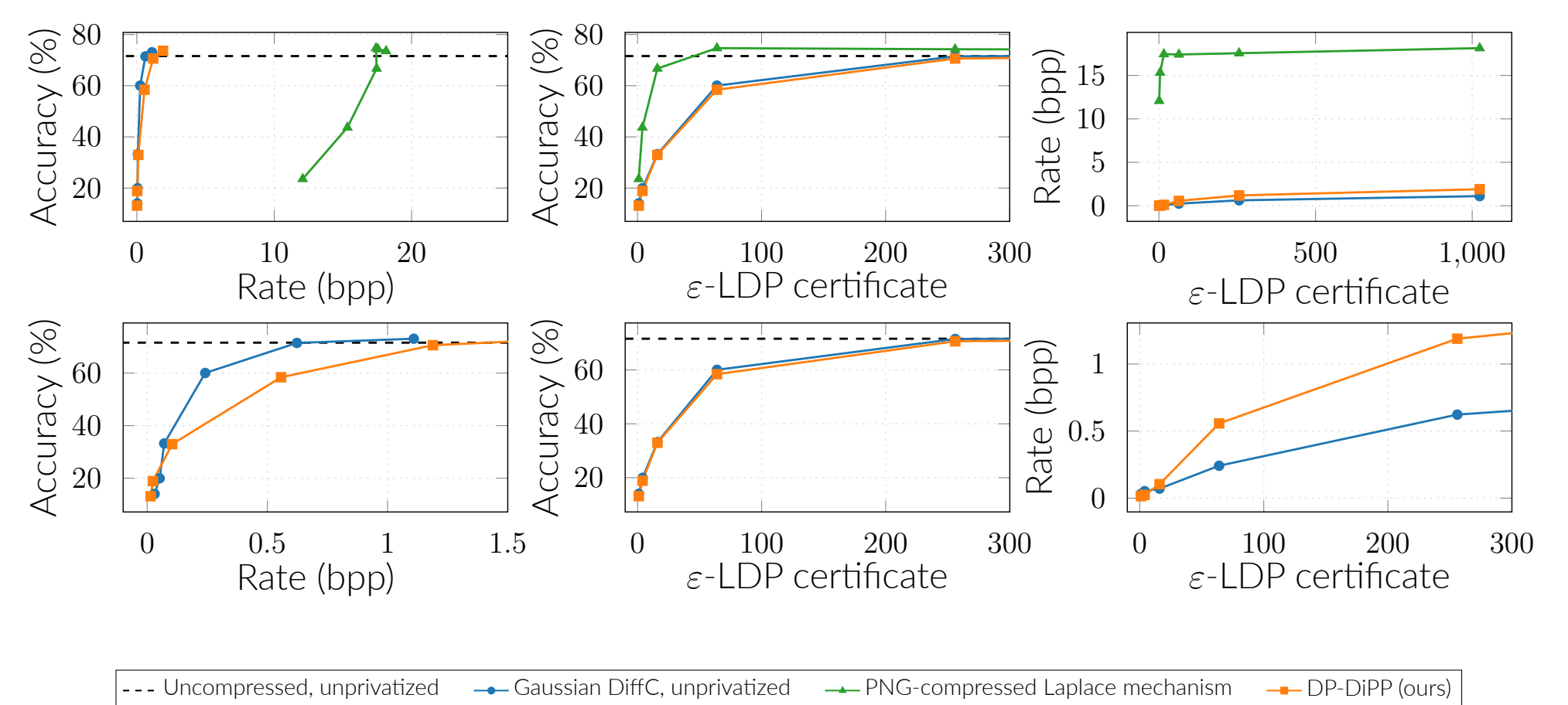
Reconstructions: DP-DiPP vs. Laplace baseline



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 $\epsilon \in \{1, 4, 16, 64, 256\}$, 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 $\epsilon = 64$: DP-DiPP 0.54 bpp vs. Laplace+PNG 21.9 bpp $\Rightarrow$ **10-30×** compression at matched LDP.

References

- Y. Liu et al., “Universal exact compression of differentially private mechanisms,” NeurIPS 2024.
- L. Theis et al., “Lossy compression with Gaussian diffusion,” 2022.
- J. Vonderfecht et al., “Lossy compression with pretrained diffusion models,” ICLR 2025.
- Y. Cao et al., “LDP-Slicing: local differential privacy for images via randomized bit-plane slicing,” CVPR 2026.