2023 NORTH AMERICAN SCHOOL OF INFORMATION THEORY

DISTORTION, REALISM, & LEARNED COMPRESSION

LUCAS THEIS, GOOGLE

files.theis.io/nasit2023_slides.pdf

Overview

Part I

Learned compression I:

Variational auto-encoders (VAEs)

Realism I:

Realism-distortion trade-off

Learned compression II:

Adversarial losses and diffusion

Part II

Realism II:

Channel simulation

Learned compression III:

Diffusion-based compression

Motivation

Learned compression specializes easily to different types of content

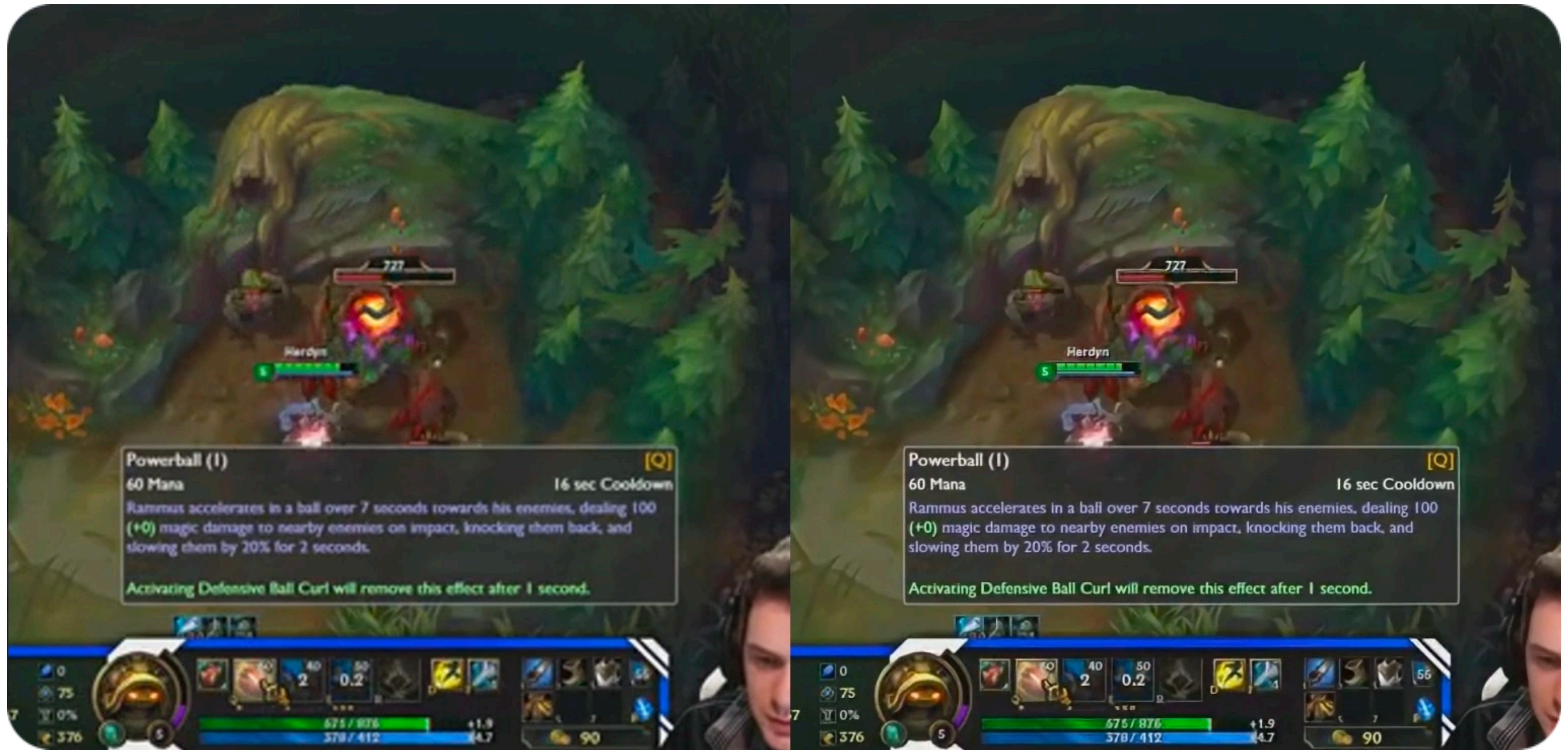
- Video games
- Video calls
- Medical images

...

and new forms of media

- AR/VR
- 360 video
- Light fields

...



Magic Pony Technology

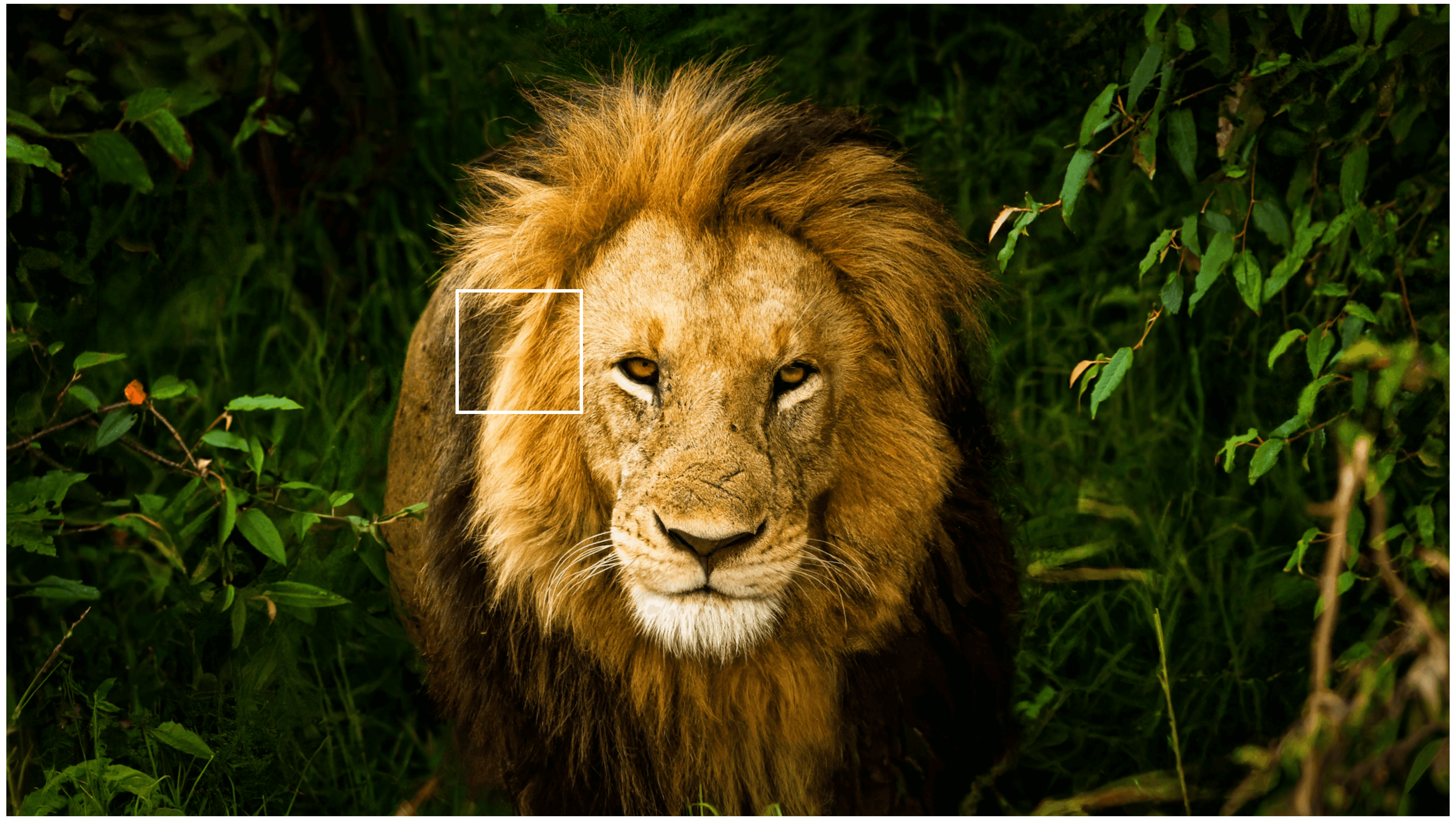
Google Meet

Lytro light field camera

Motivation

Learned compression with *realism constraints* enable extremely low bit-rates...

... but many open questions remain.



HFD: 0.0562 bpp

VVC: 0.0687 bpp

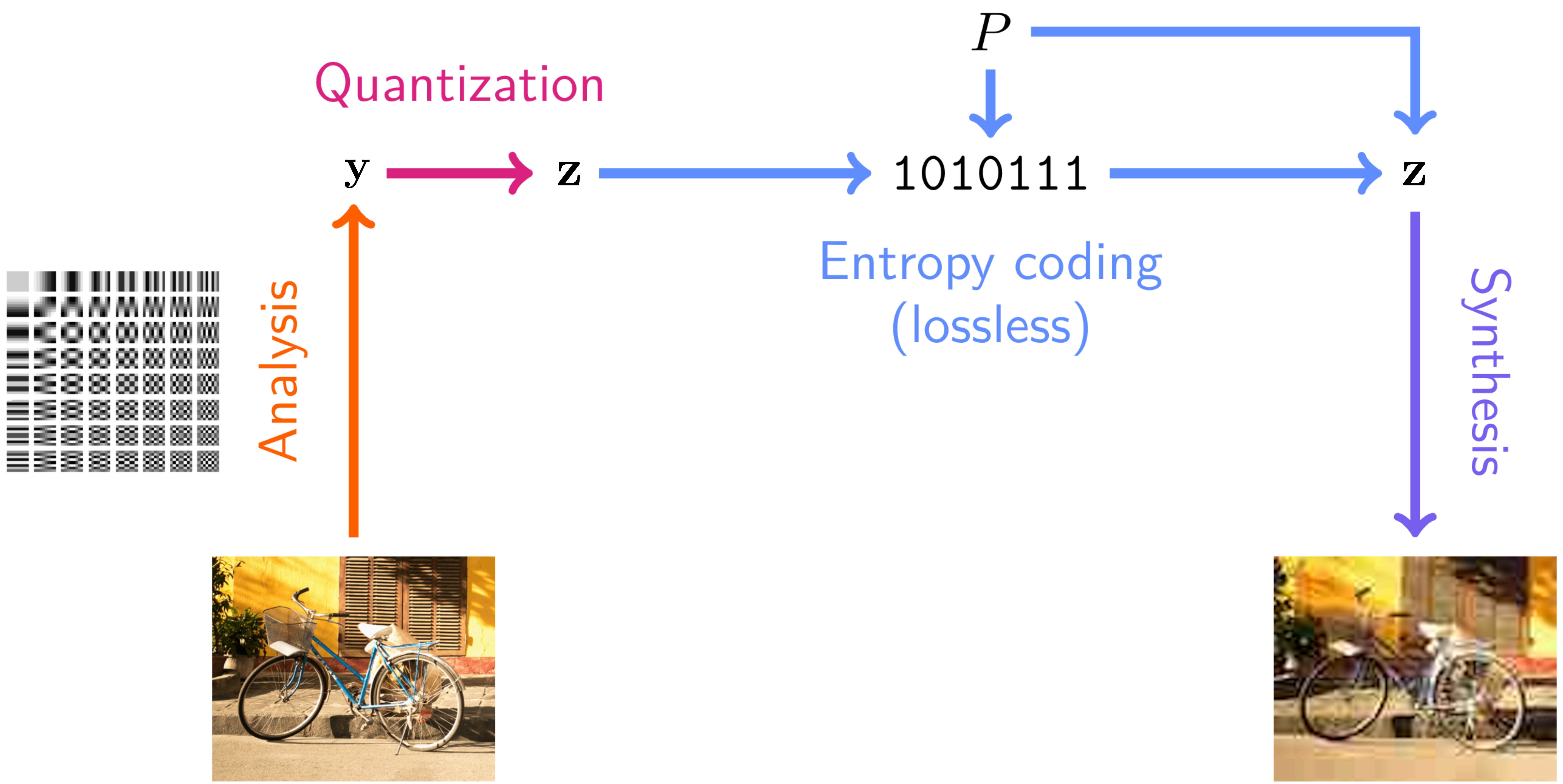
JPEG: 0.1251 bpp

Hoogeboom et al. (2023)

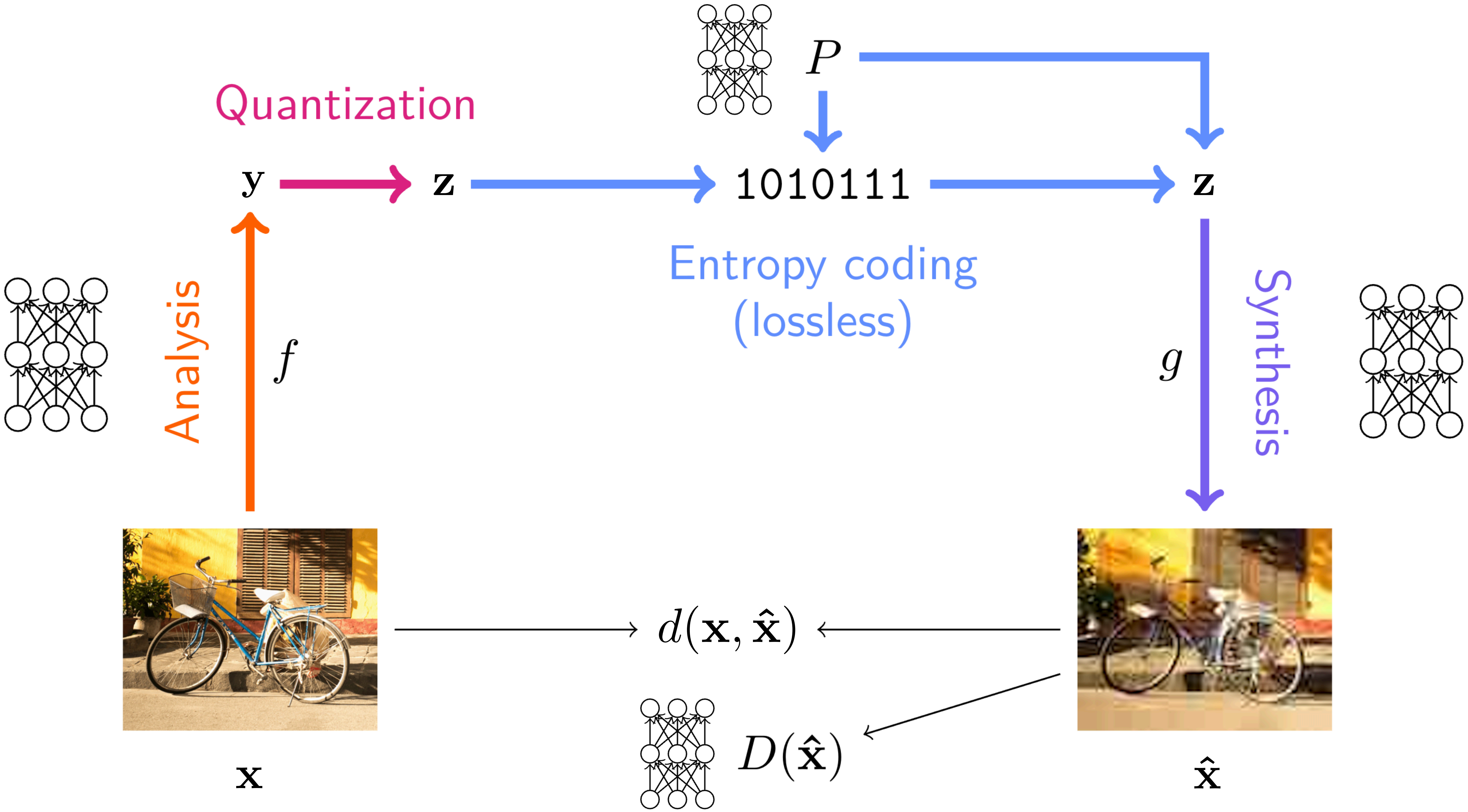
LEARNED COMPRESSION I:

Variational auto-encoders (VAEs)

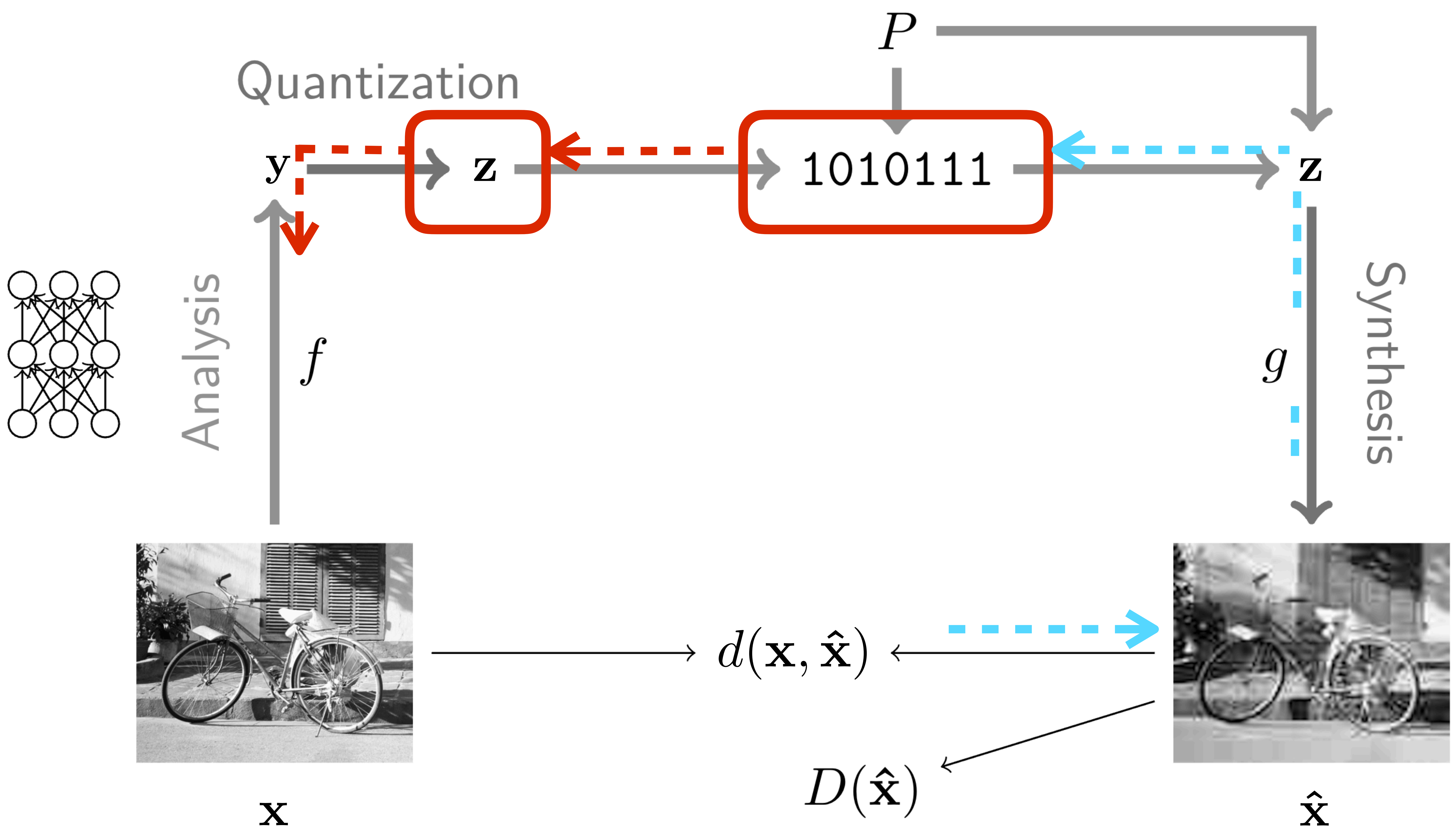
JPEG


Quantization
Entropy coding
(lossless)
Analysis
Synthesis
P
y
z
1010111
z

Learned compression

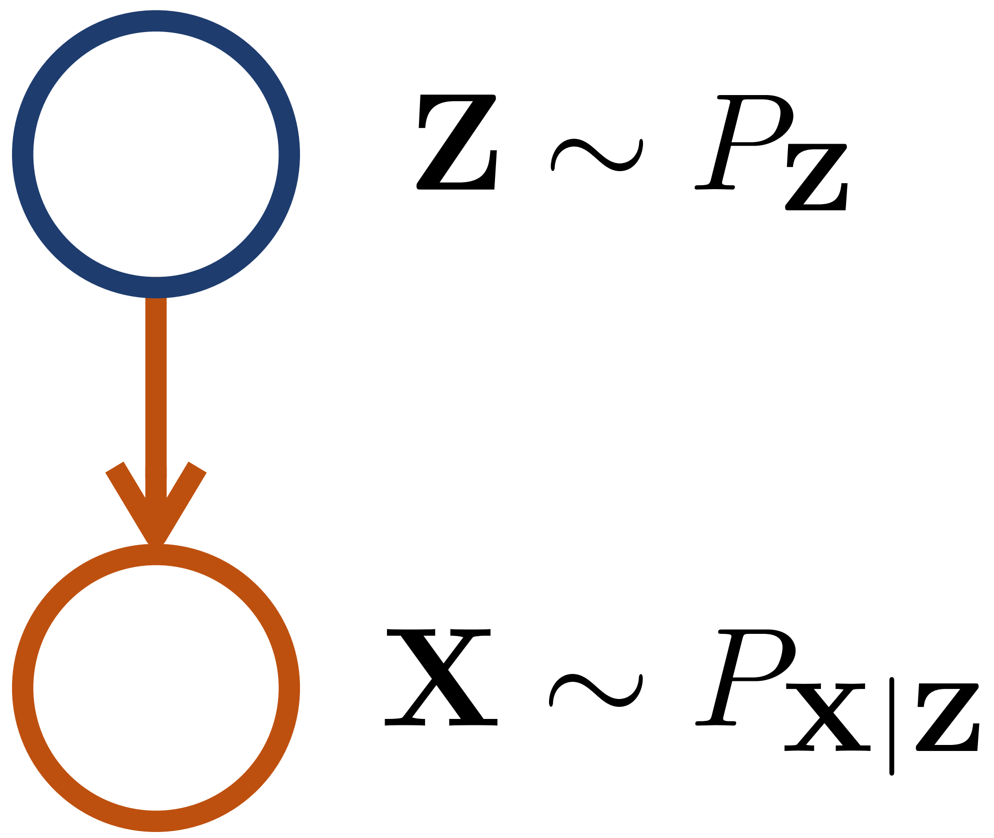


Backpropagation



Variational autoencoders (VAEs)

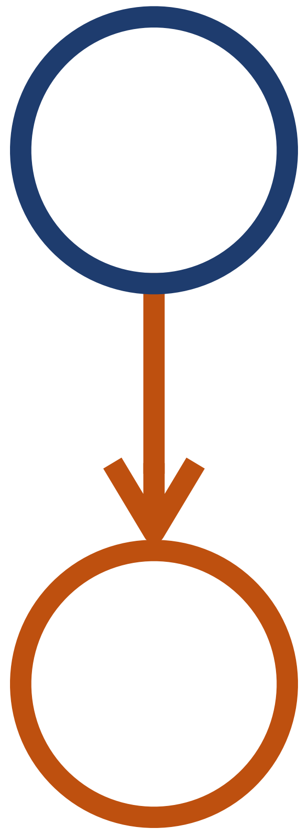
$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z})\, d\mathbf{z}$$



Rezende et al. (2014); Kingma et al. (2014)

Variational autoencoders (VAEs)

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z}) \, d\mathbf{z}$$

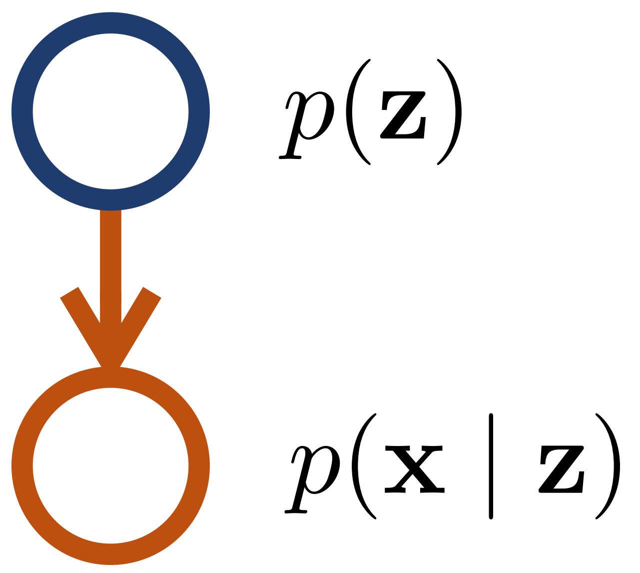


$$\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_n)$$

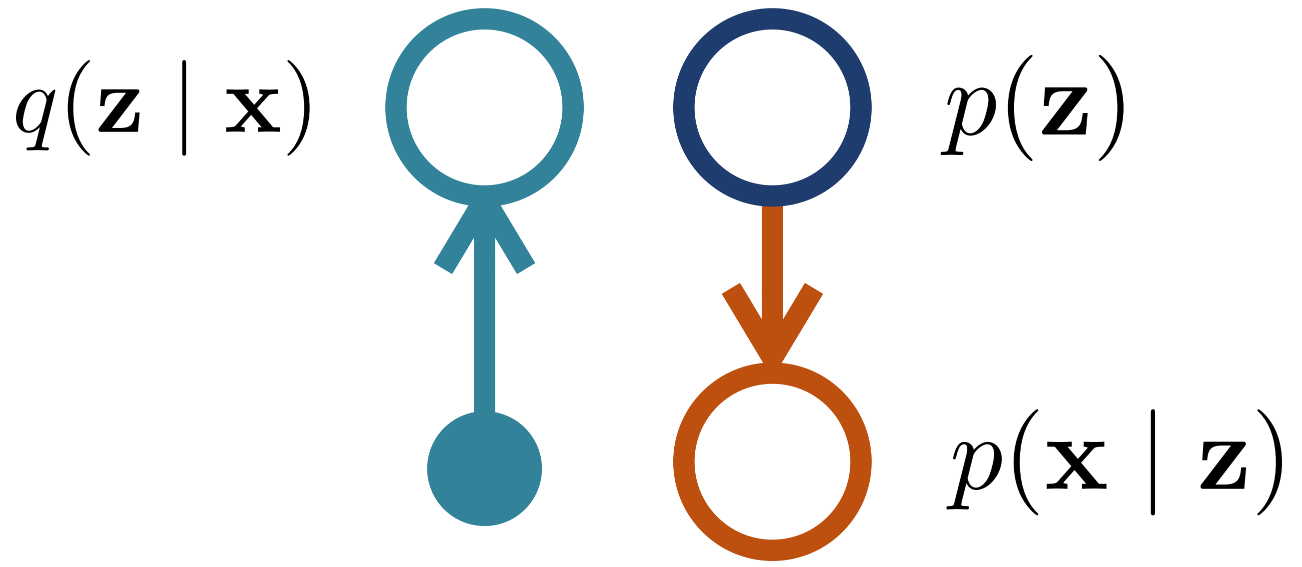
$$\mathbf{X} \sim \mathcal{N}(g_\theta(\mathbf{Z}), \lambda^{-1}\mathbf{I}_m)$$

Rezende et al. (2014); Kingma et al. (2014)

Maximum likelihood



$$\log p_\theta(\mathbf{x}) = \log \int p_\theta(\mathbf{x}, \mathbf{z})\, d\mathbf{z}$$



Evidence lower bound (ELBO)

$$\log p_\theta(\mathbf{x}) \geq \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} \, d\mathbf{z}$$

Evidence lower bound (ELBO)

$$
\begin{aligned}
\log p_\theta(\mathbf{x}) &= \int q(\mathbf{z} \mid \mathbf{x}) \log p_\theta(\mathbf{x}) \, d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \left(p_\theta(\mathbf{x}) \frac{p_\theta(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \frac{q(\mathbf{z} \mid \mathbf{x})}{q(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} \, d\mathbf{z} + \int q(\mathbf{z} \mid \mathbf{x}) \log \left(\frac{q(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} \, d\mathbf{z} + D_{\mathrm{KL}}[q(\mathbf{z} \mid \mathbf{x}) \| p_\theta(\mathbf{z} \mid \mathbf{x})] \\
&\geq \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} \, d\mathbf{z}
\end{aligned}
$$

Evidence lower bound (ELBO)

$$
\begin{aligned}
\log p_\theta(\mathbf{x}) &= \int q(\mathbf{z} \mid \mathbf{x}) \log p_\theta(\mathbf{x})\, d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \left(p_\theta(\mathbf{x}) \frac{p_\theta(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \frac{q(\mathbf{z} \mid \mathbf{x})}{q(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z} + \int q(\mathbf{z} \mid \mathbf{x}) \log \left(\frac{q(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z} + D_{\mathrm{KL}}[q(\mathbf{z} \mid \mathbf{x}) \| p_\theta(\mathbf{z} \mid \mathbf{x})] \\
&\geq \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z}
\end{aligned}
$$

Evidence lower bound (ELBO)

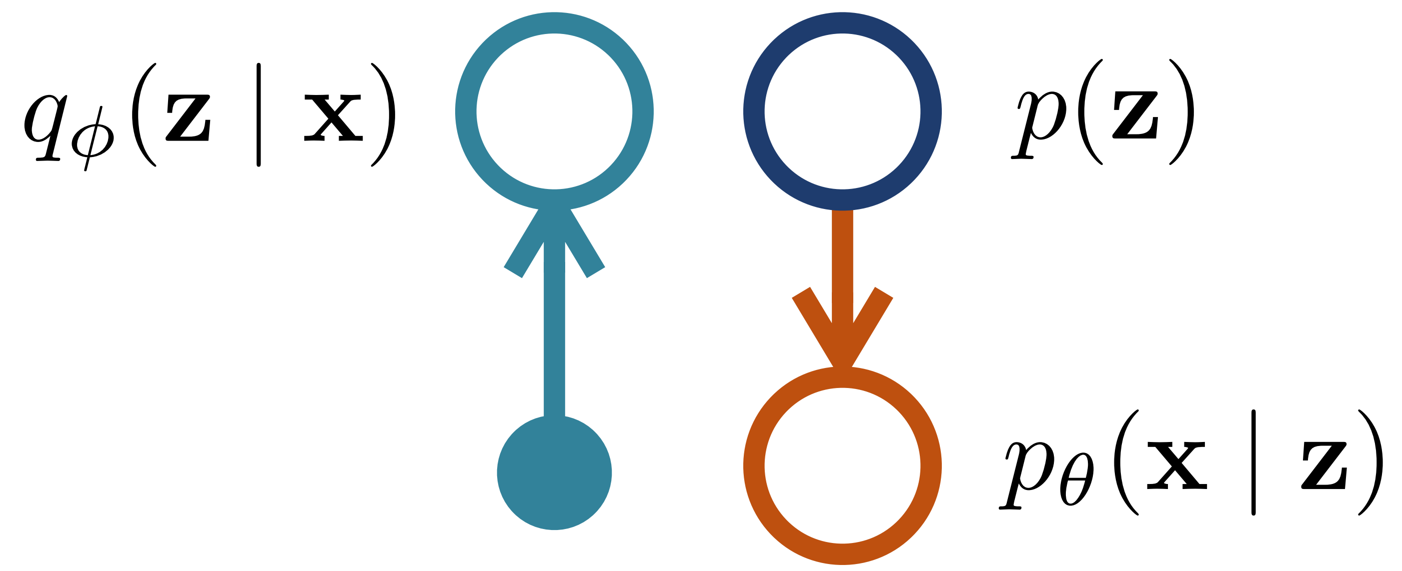
$$
\begin{aligned}
\log p_\theta(\mathbf{x}) &= \int q(\mathbf{z} \mid \mathbf{x}) \log p_\theta(\mathbf{x})\, d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \left(p_\theta(\mathbf{x}) \frac{p_\theta(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \frac{q(\mathbf{z} \mid \mathbf{x})}{q(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z} + \int q(\mathbf{z} \mid \mathbf{x}) \log \left(\frac{q(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z} + D_{\mathrm{KL}}[q(\mathbf{z} \mid \mathbf{x}) \| p_\theta(\mathbf{z} \mid \mathbf{x})] \\
&\geq \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z}
\end{aligned}
$$

Evidence lower bound (ELBO)

$$\begin{aligned}
\log p_\theta(\mathbf{x}) &= \int q(\mathbf{z} \mid \mathbf{x}) \log p_\theta(\mathbf{x})\, d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \left(p_\theta(\mathbf{x}) \frac{p_\theta(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \frac{q(\mathbf{z} \mid \mathbf{x})}{q(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z} + \int q(\mathbf{z} \mid \mathbf{x}) \log \left(\frac{q(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z} + D_{\mathrm{KL}}[q(\mathbf{z} \mid \mathbf{x}) \| p_\theta(\mathbf{z} \mid \mathbf{x})] \\
&\geq \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})}\, d\mathbf{z}
\end{aligned}$$

Evidence lower bound (ELBO)

$$
\begin{aligned}
\log p_\theta(\mathbf{x}) &= \int q(\mathbf{z} \mid \mathbf{x}) \log p_\theta(\mathbf{x}) \, d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \left(p_\theta(\mathbf{x}) \frac{p_\theta(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \frac{q(\mathbf{z} \mid \mathbf{x})}{q(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} \, d\mathbf{z} + \int q(\mathbf{z} \mid \mathbf{x}) \log \left(\frac{q(\mathbf{z} \mid \mathbf{x})}{p_\theta(\mathbf{z} \mid \mathbf{x})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} \, d\mathbf{z} + D_{\mathrm{KL}}[q(\mathbf{z} \mid \mathbf{x}) \| p_\theta(\mathbf{z} \mid \mathbf{x})] \\
&\geq \int q(\mathbf{z} \mid \mathbf{x}) \log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q(\mathbf{z} \mid \mathbf{x})} \, d\mathbf{z} \qquad \text{(Gibbs' inequality)}
\end{aligned}
$$

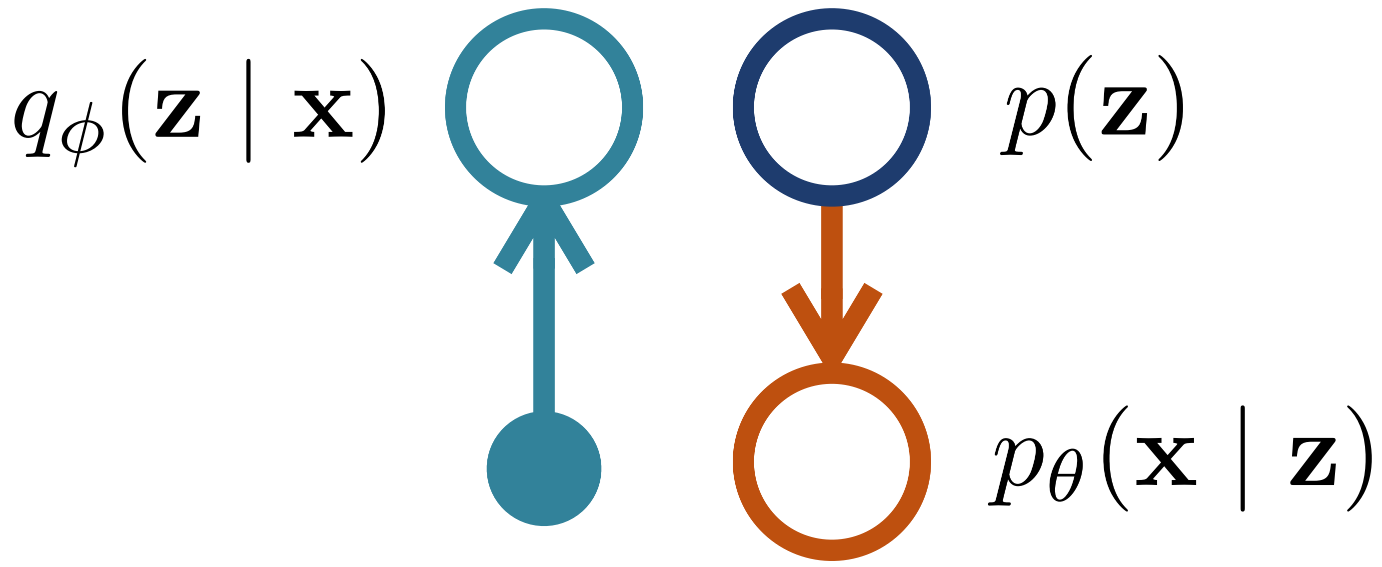


Evidence lower bound (ELBO)

$$\ell(\phi, \theta) = -\mathbb{E}_{q_\phi}\left[\log \frac{p_\theta(\mathbf{x}, \mathbf{Z})}{q_\phi(\mathbf{Z} \mid \mathbf{x})}\right]$$

$$= \mathbb{E}_{q_\phi}\left[-\log p_\theta(\mathbf{x} \mid \mathbf{Z})\right] + \mathbb{E}_{q_\phi}\left[-\log p(\mathbf{Z})\right] - \mathbb{E}_{q_\phi}\left[-\log q_\phi(\mathbf{Z} \mid \mathbf{x})\right]$$

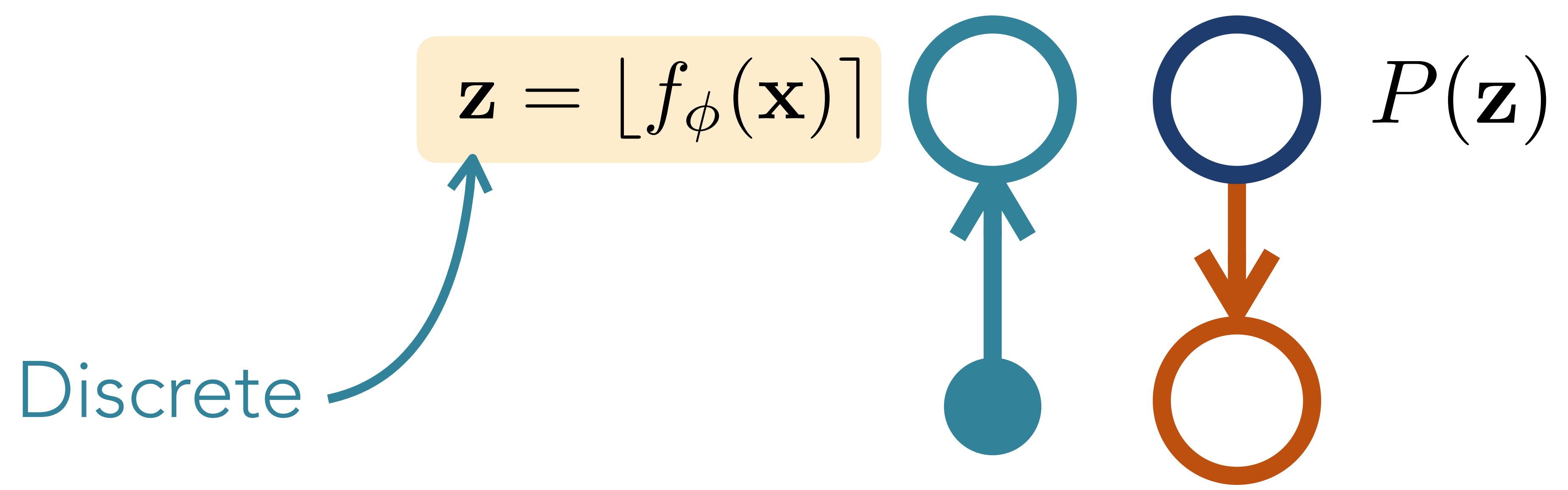
Evidence lower bound (ELBO)



$$\ell(\phi, \theta) = -\mathbb{E}_{q_\phi}\left[\log \frac{p_\theta(\mathbf{x}, \mathbf{Z})}{q_\phi(\mathbf{Z} \mid \mathbf{x})}\right]$$

$$= \mathbb{E}_{q_\phi}\left[-\log p_\theta(\mathbf{x} \mid \mathbf{Z})\right] + \mathbb{E}_{q_\phi}\left[-\log p(\mathbf{Z})\right] - \mathbb{E}_{q_\phi}\left[-\log q_\phi(\mathbf{Z} \mid \mathbf{x})\right]$$

Evidence lower bound (ELBO)



$$p_\theta(\mathbf{x} \mid \mathbf{z}) = \mathcal{N}(\mathbf{x}; g_\theta(\mathbf{z}), \lambda^{-1}\mathbf{I})$$

$$\ell(\phi, \theta) = -\mathbb{E}_{Q_\phi}\left[\log \frac{p_\theta(\mathbf{x} \mid \mathbf{Z}) P(\mathbf{Z})}{Q_\phi(\mathbf{Z} \mid \mathbf{x})}\right]$$

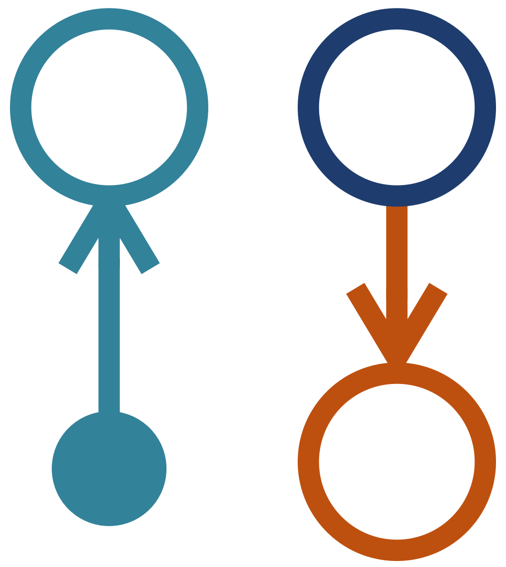
$$= \mathbb{E}_{Q_\phi}[-\log p_\theta(\mathbf{x} \mid \mathbf{Z})] + \mathbb{E}_{Q_\phi}[-\log P(\mathbf{Z})] - \mathbb{E}_{Q_\phi}[-\log Q_\phi(\mathbf{Z} \mid \mathbf{x})]$$

$$= \underbrace{\frac{\lambda}{2}\|\mathbf{x} - g_\theta(\mathbf{z})\|^2}_{\text{Distortion}} \underbrace{- \log P(\mathbf{z})}_{\text{Coding cost}} + \text{const}$$

Evidence lower bound (ELBO)

$$\mathbf{Z} = f_\phi(\mathbf{x}) + \mathbf{U}$$
$$= \mathbf{y} + \mathbf{U}$$

Differentiable



$p(\mathbf{z})$

$$p_\theta(\mathbf{x} \mid \mathbf{z}) = \mathcal{N}(\mathbf{x}; g_\theta(\mathbf{z}), \lambda^{-1}\mathbf{I})$$

$$\ell(\phi, \theta) = -\mathbb{E}_q \left[\log \frac{p_\theta(\mathbf{x}, \mathbf{Z})}{q_\phi(\mathbf{Z} \mid \mathbf{x})} \right]$$

$$= -\mathbb{E}_q\left[-\log p_\theta(\mathbf{x} \mid \mathbf{Z})\right] - \mathbb{E}_q\left[\log p(\mathbf{Z})\right] - \mathbb{E}_q[-\log q_\phi(\mathbf{Z} \mid \mathbf{x})]$$

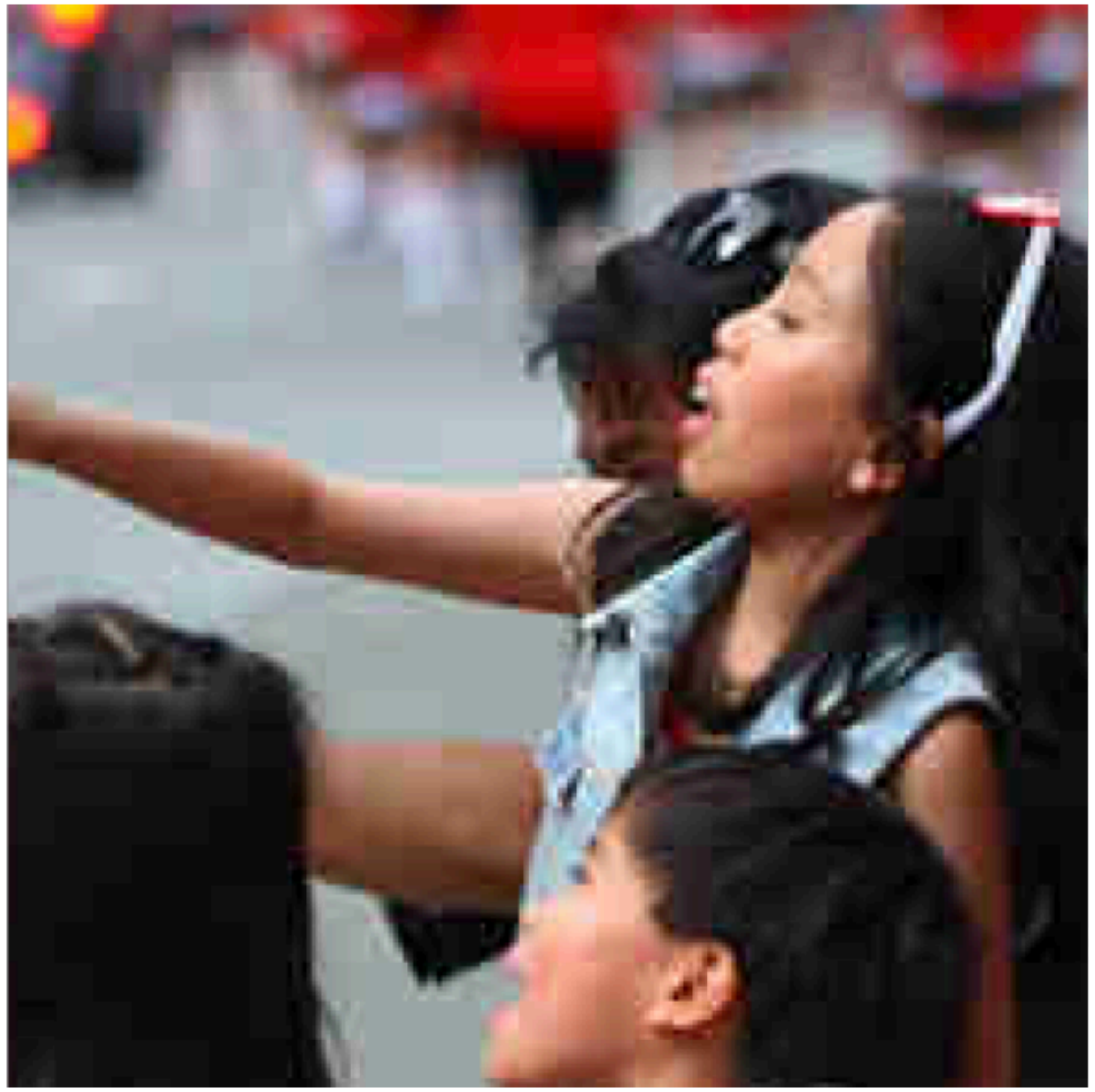
$$= \underbrace{\frac{\lambda}{2}\mathbb{E}[\|\mathbf{x} - g_\theta(\mathbf{y} + \mathbf{U})\|^2]}_{\text{Distortion}} + \underbrace{\mathbb{E}[-\log p(\mathbf{y} + \mathbf{U})]}_{\text{Coding cost}} + \text{const}$$

Ballé et al. (2017)

Input

Photo: GoToVan, 2014

JPEG



$$\mathbf{z} = \lfloor \mathbf{A}\mathbf{x} \rceil$$

Dithered JPEG

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{u}$$

Theis et al. (2017)

Example: Linear VAE

$$f(\mathbf{x}) = \mathbf{A}\mathbf{x} \qquad g(\mathbf{z}) = \mathbf{B}\mathbf{z}$$

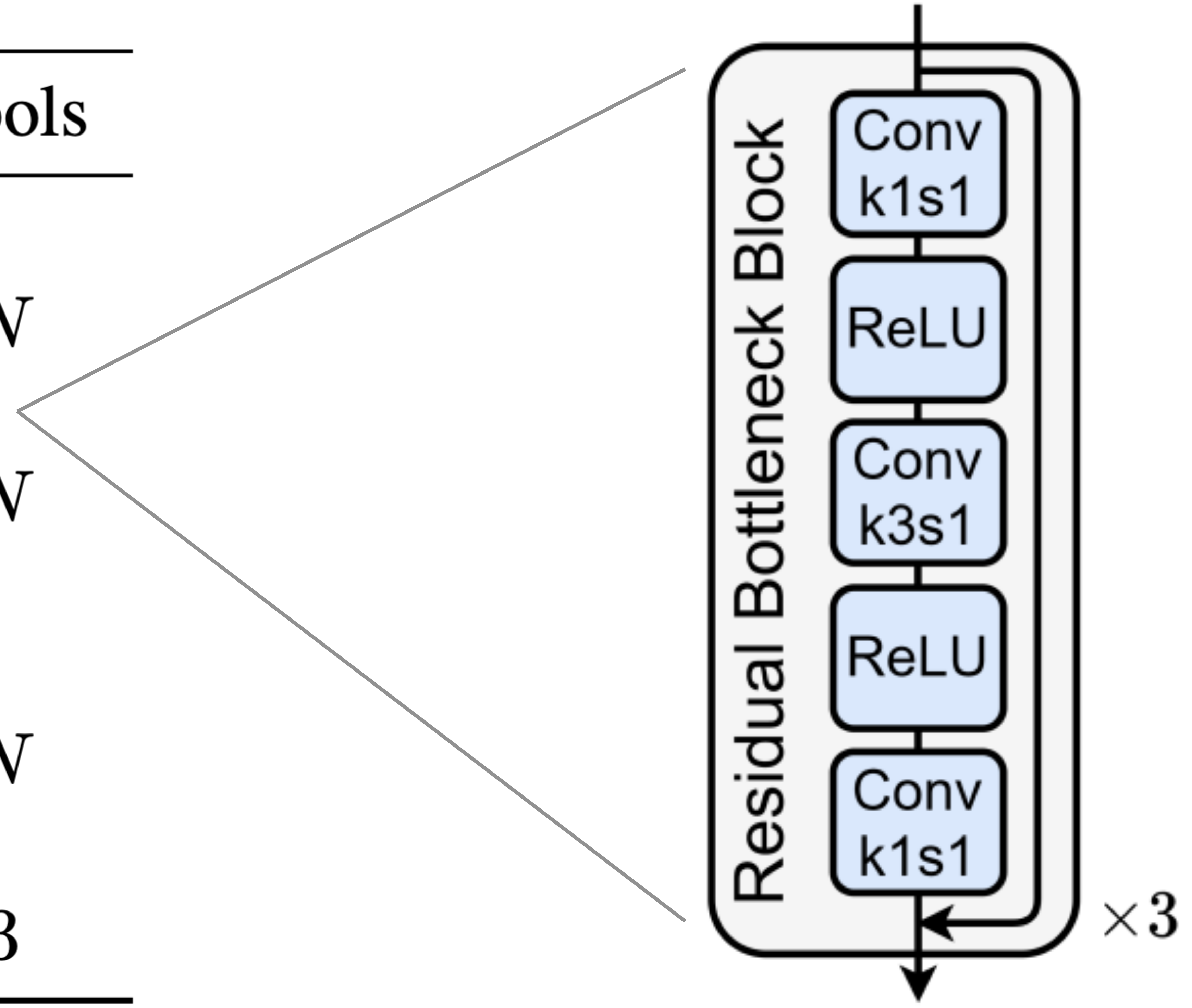
Learned analysis transform


JPEG: 0.1861 bpp
Agustsson et al. (2023): 0.1674 bpp

Example: ELIC (He et al., 2022)

Analyzer f_ϕ	Synthesizer g_θ
in: 3-channel image	in: M-channel symbols
Conv 5×5, s2, N	Attention
ResBottleneck$\times 3$	TConv 5×5, s2, N
Conv 5×5, s2, N	ResBottleneck$\times 3$
ResBottleneck$\times 3$	TConv 5×5, s2, N
Attention	Attention
Conv 5×5, s2, N	ResBottleneck$\times 3$
ResBottleneck$\times 3$	TConv 5×5, s2, N
Conv 5×5, s2, M	ResBottleneck$\times 3$
Attention	TConv 5×5, s2, 3

Table 1. Architecture of ELIC main transform networks.



JPEG/JFIF

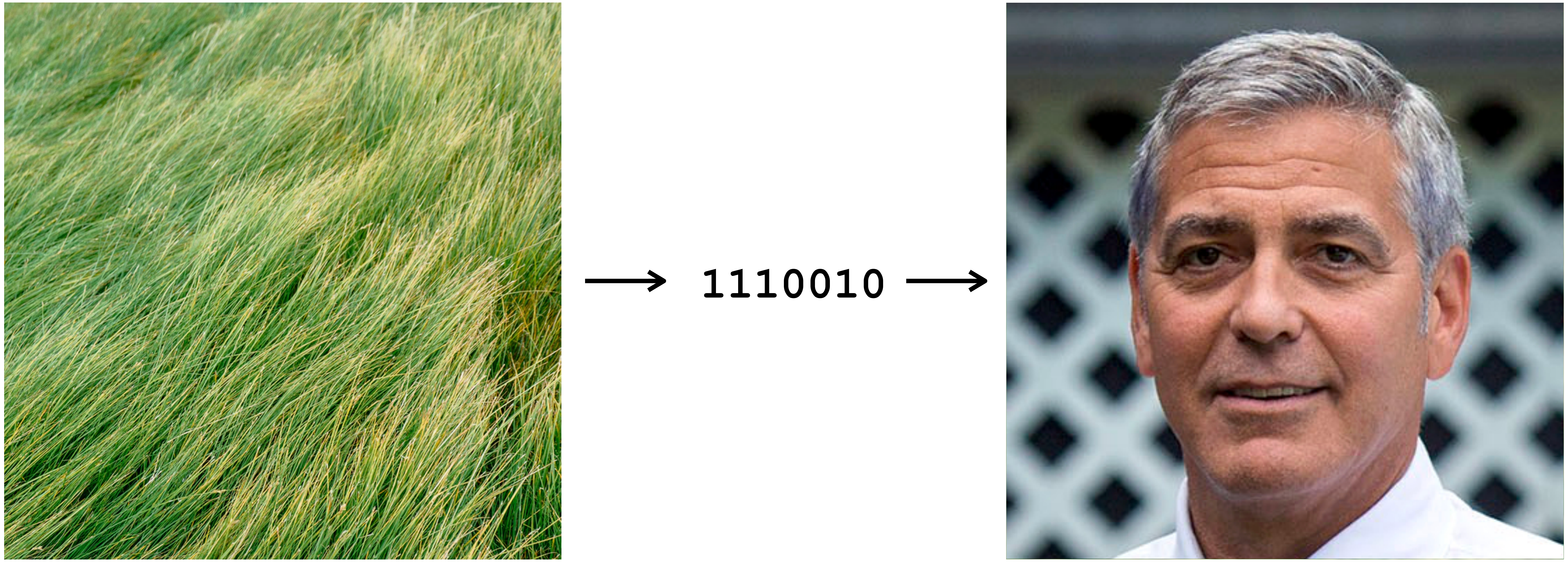
Minnen et al. (2020)

REALISM I:

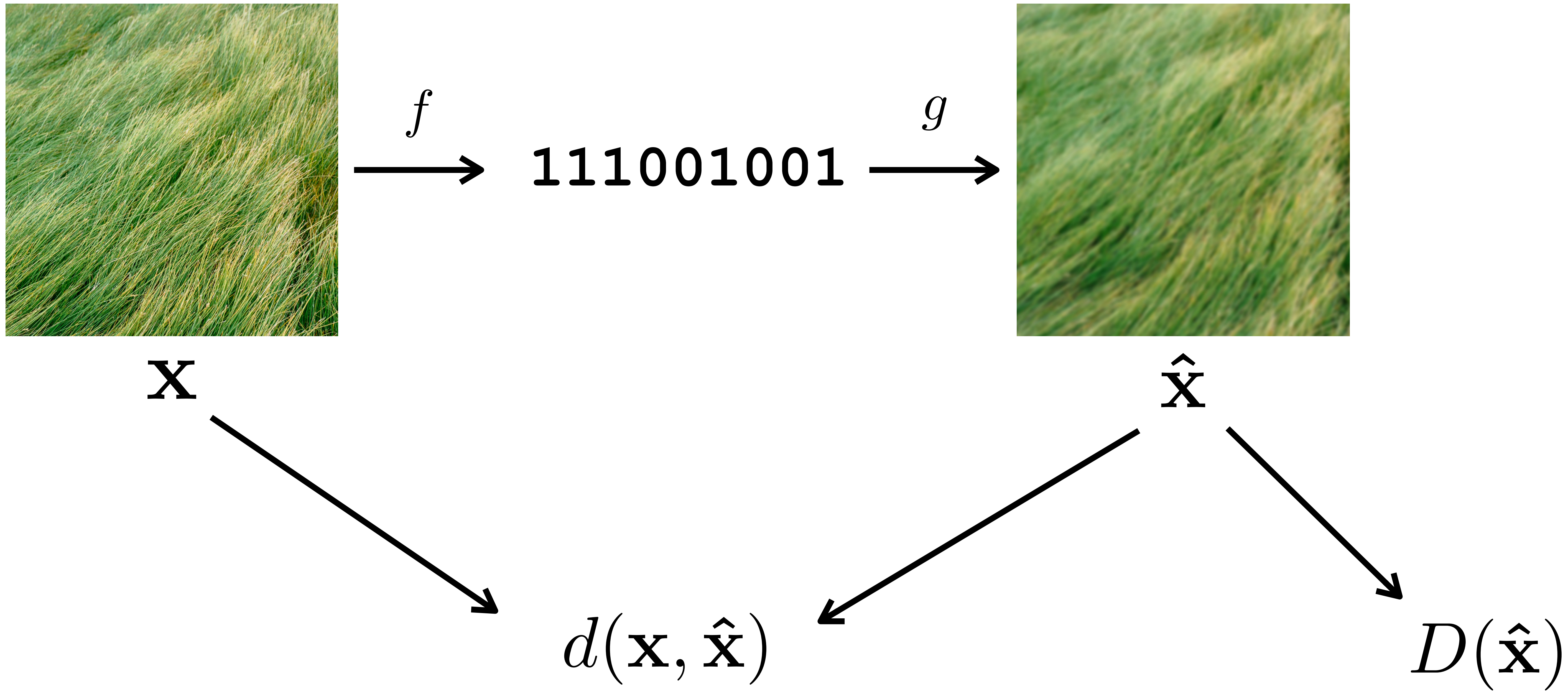
Realism-distortion trade-off

Realism-distortion trade-off

Realism-distortion trade-off



How to measure realism?



Example: $d(\mathbf{x}, \hat{\mathbf{x}}) = \mathbb{E}[||\mathbf{x} - \hat{\mathbf{x}}||^2]$

How to measure realism?

$$P_{\mathbf{x}}, P_{\hat{\mathbf{x}}}$$

How to measure realism?

Divergence:

$$D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}] \geq 0$$

$$D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}] = 0 \Leftrightarrow P_{\mathbf{X}} = P_{\hat{\mathbf{X}}}$$

Example: Total variation distance

$$D_{\mathrm{TV}}[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}] = \sup_{A \in \mathcal{A}} |P_{\mathbf{X}}(A) - P_{\hat{\mathbf{X}}}(A)|$$

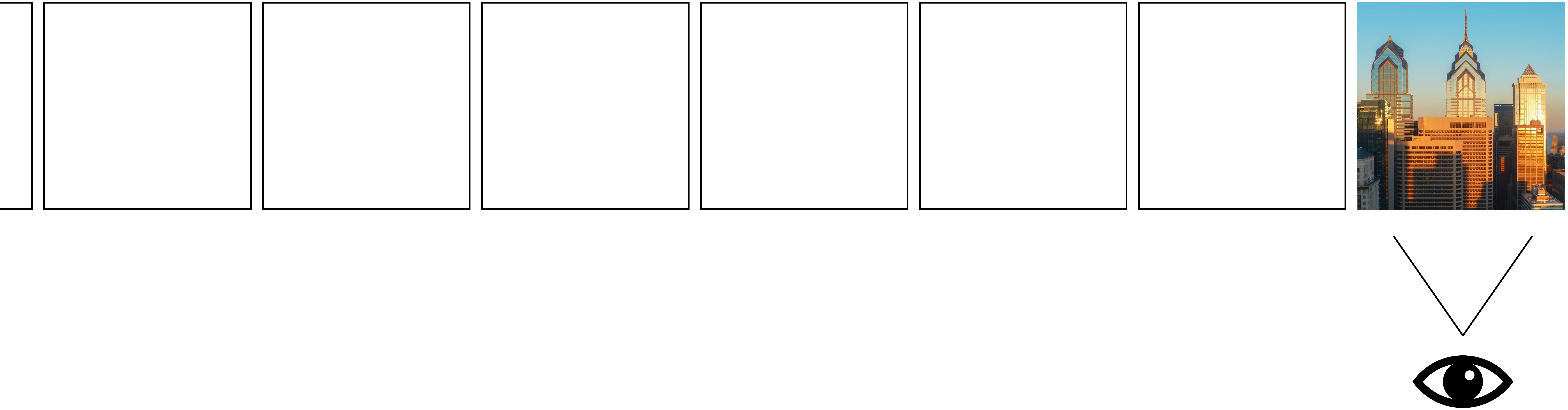
Real? Fake?

e.g., Sriperumbudur et al. (2009)

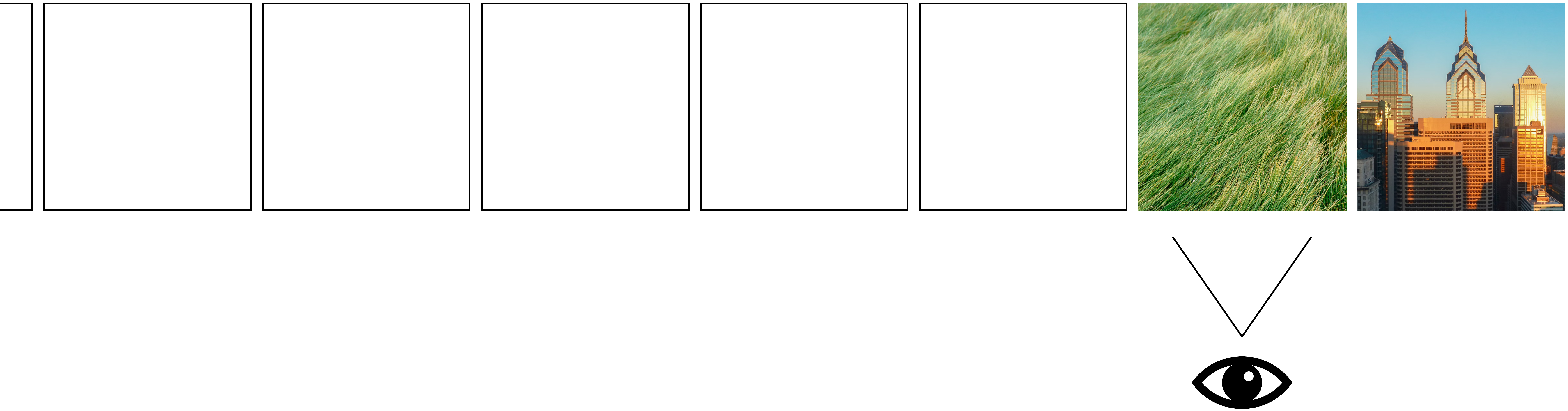
Divergences vs no-reference metrics

$\nu(\tilde{\boldsymbol{X}})$

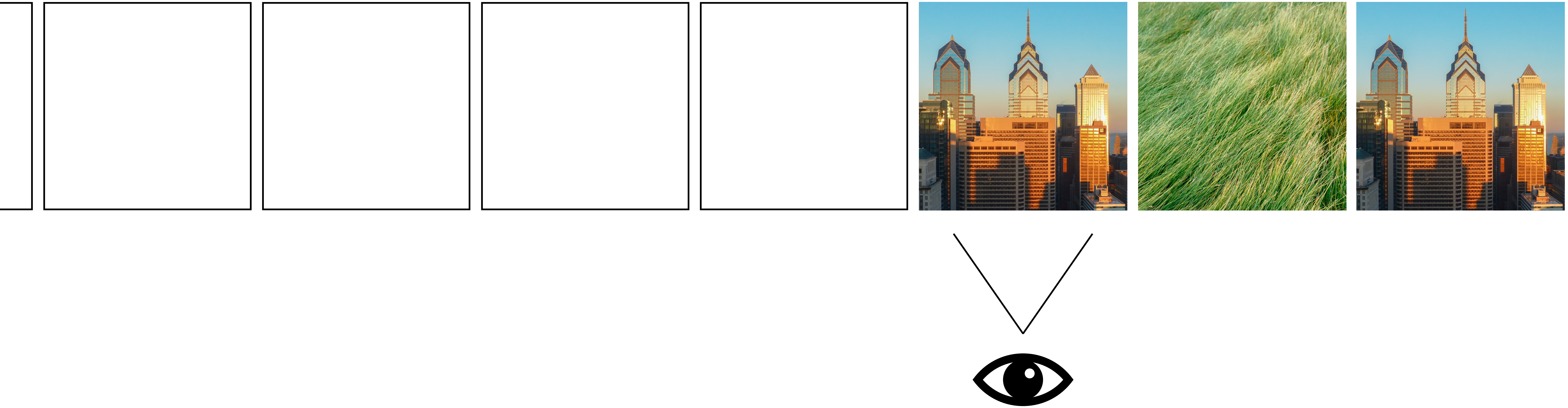
Divergences vs no-reference metrics



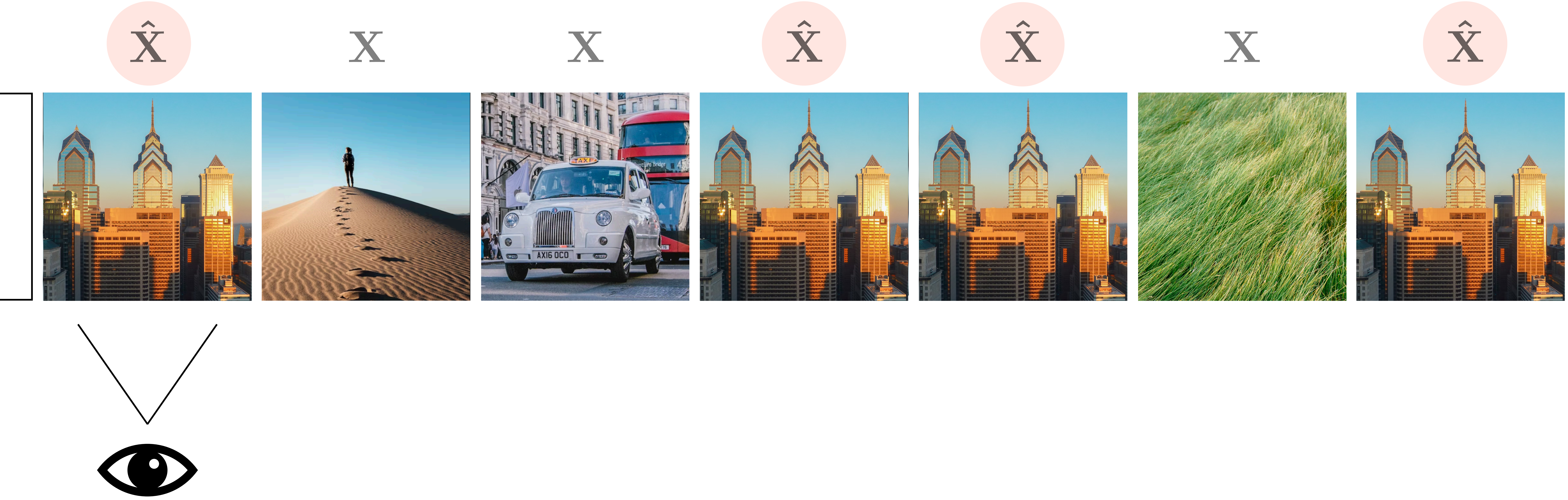
Divergences vs no-reference metrics



Divergences vs no-reference metrics



Divergences vs no-reference metrics



Realism-distortion trade-off

$\mathbf{x}$

Rate

1110010

$\hat{\mathbf{x}}$

$$d(\mathbf{x}, \hat{\mathbf{x}})$$

Distortion

Realism-distortion trade-off

$\mathbf{x}$

Rate

→ 1110010 →

$\hat{\mathbf{x}}$

$$D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}]$$

Realism

Realism-distortion trade-off

$\mathbf{x}$

Rate

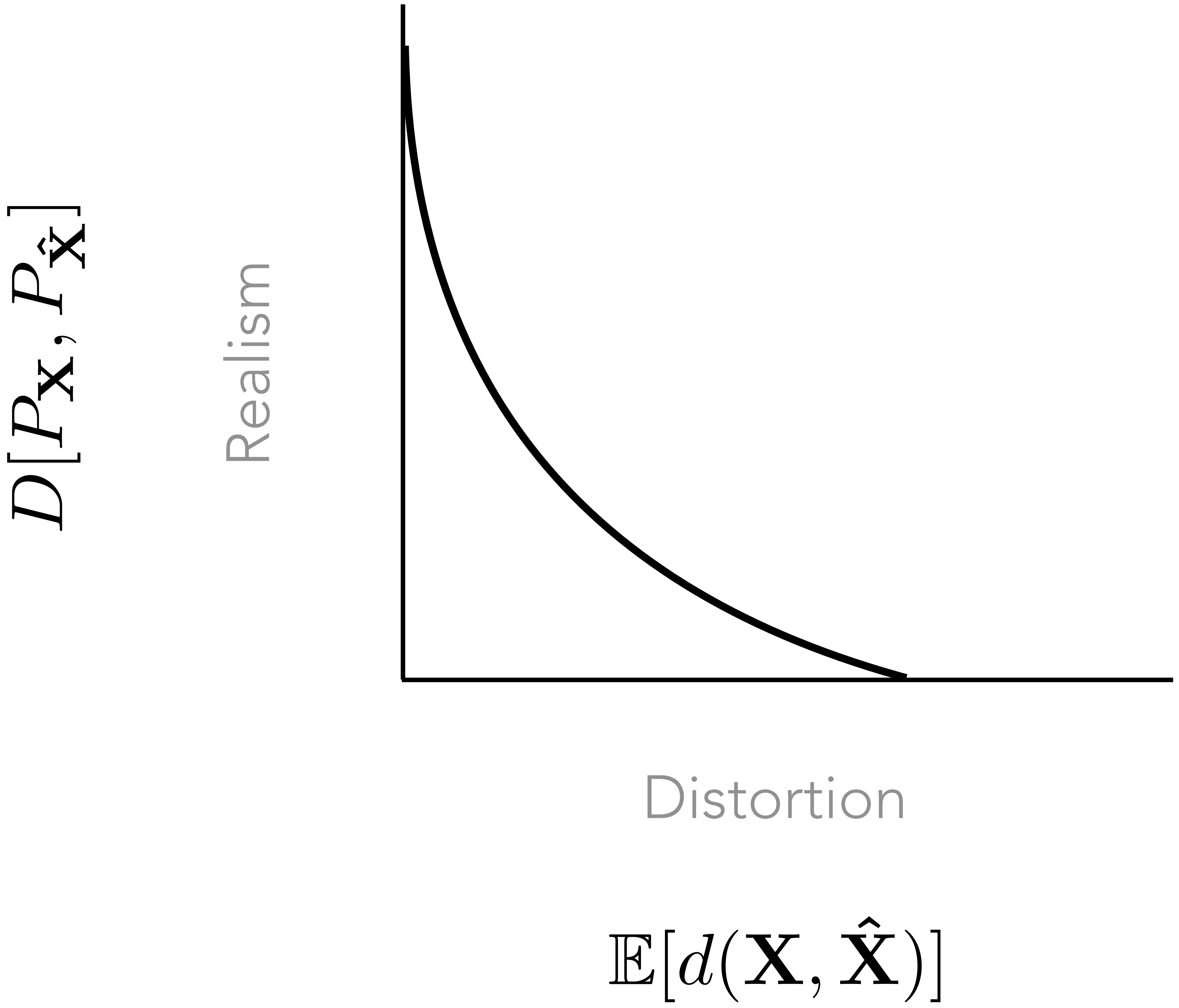
1110010

$\hat{\mathbf{x}}$

$$\mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] + \lambda D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}]$$

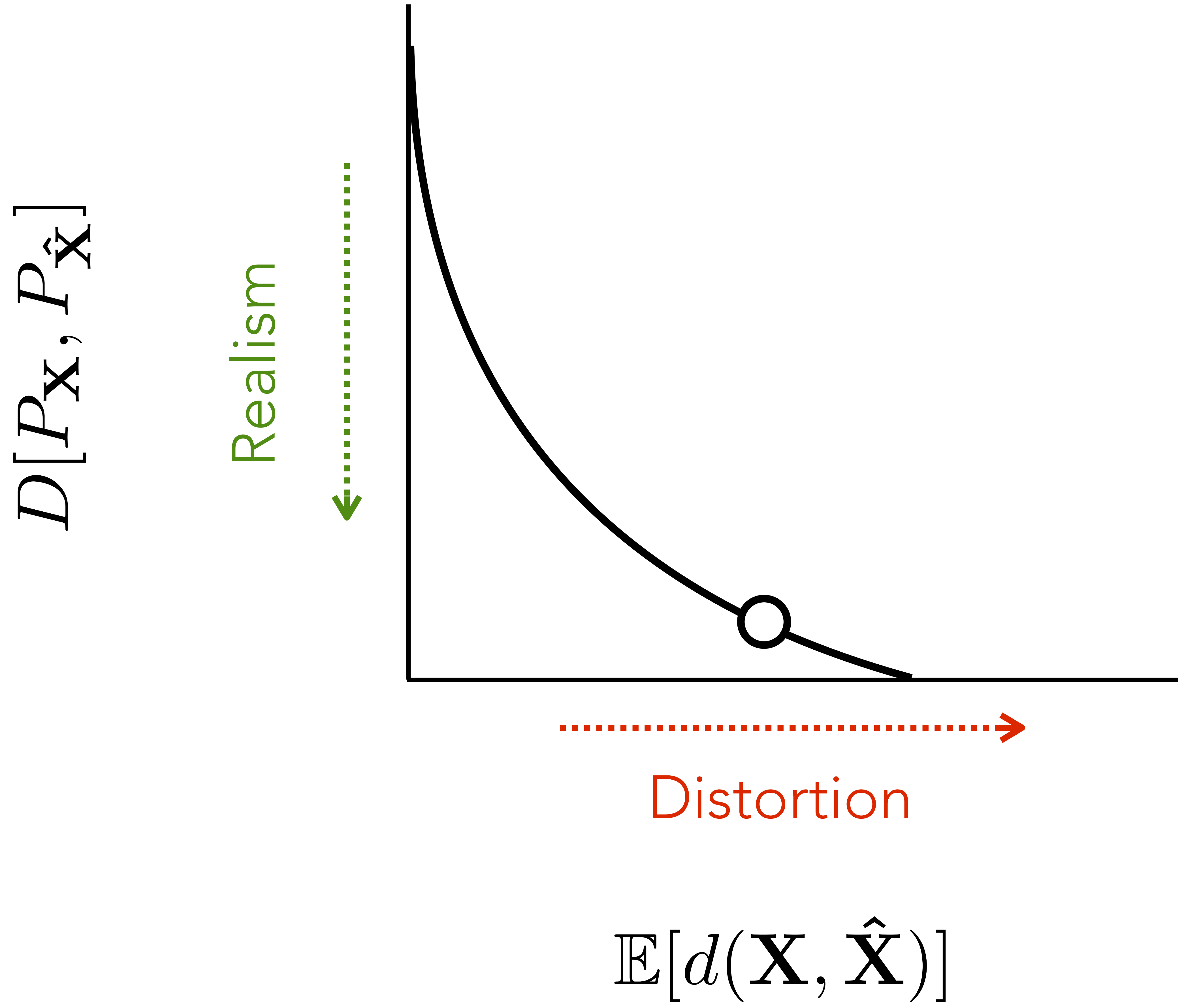
Realism-distortion trade-off

Realism-distortion trade-off



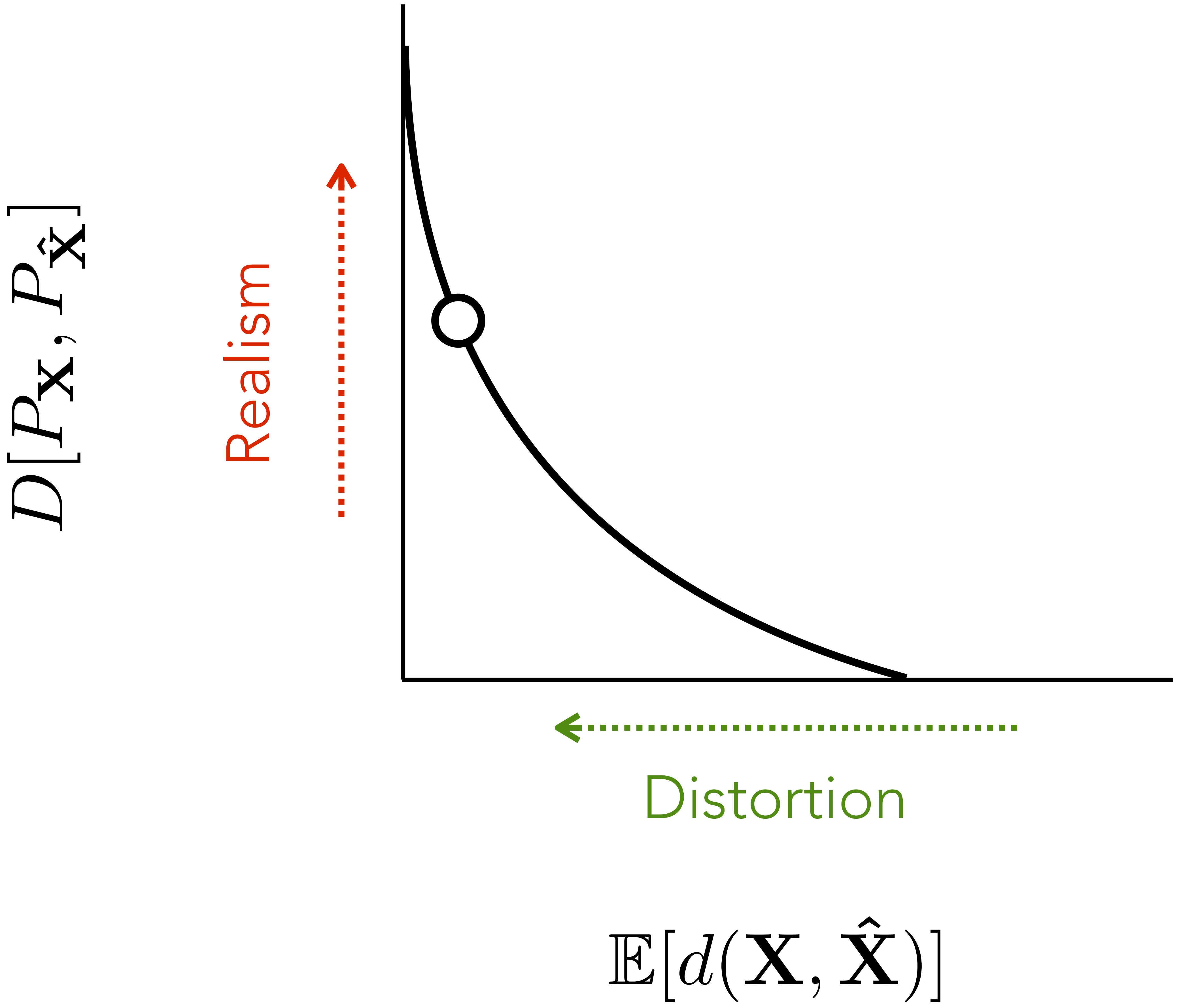
Blau & Michaeli (2018)

Realism-distortion trade-off



Blau & Michaeli (2018)

Realism-distortion trade-off

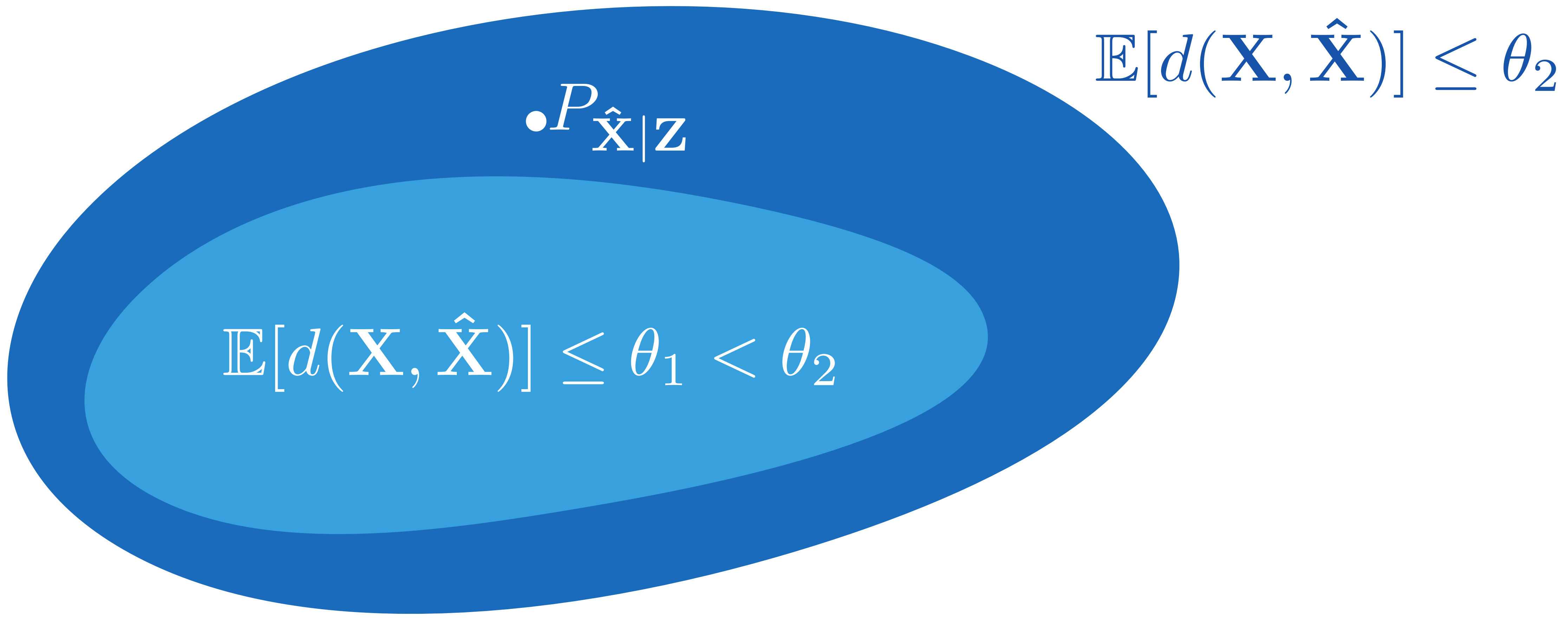


Blau & Michaeli (2018)

Realism-distortion trade-off

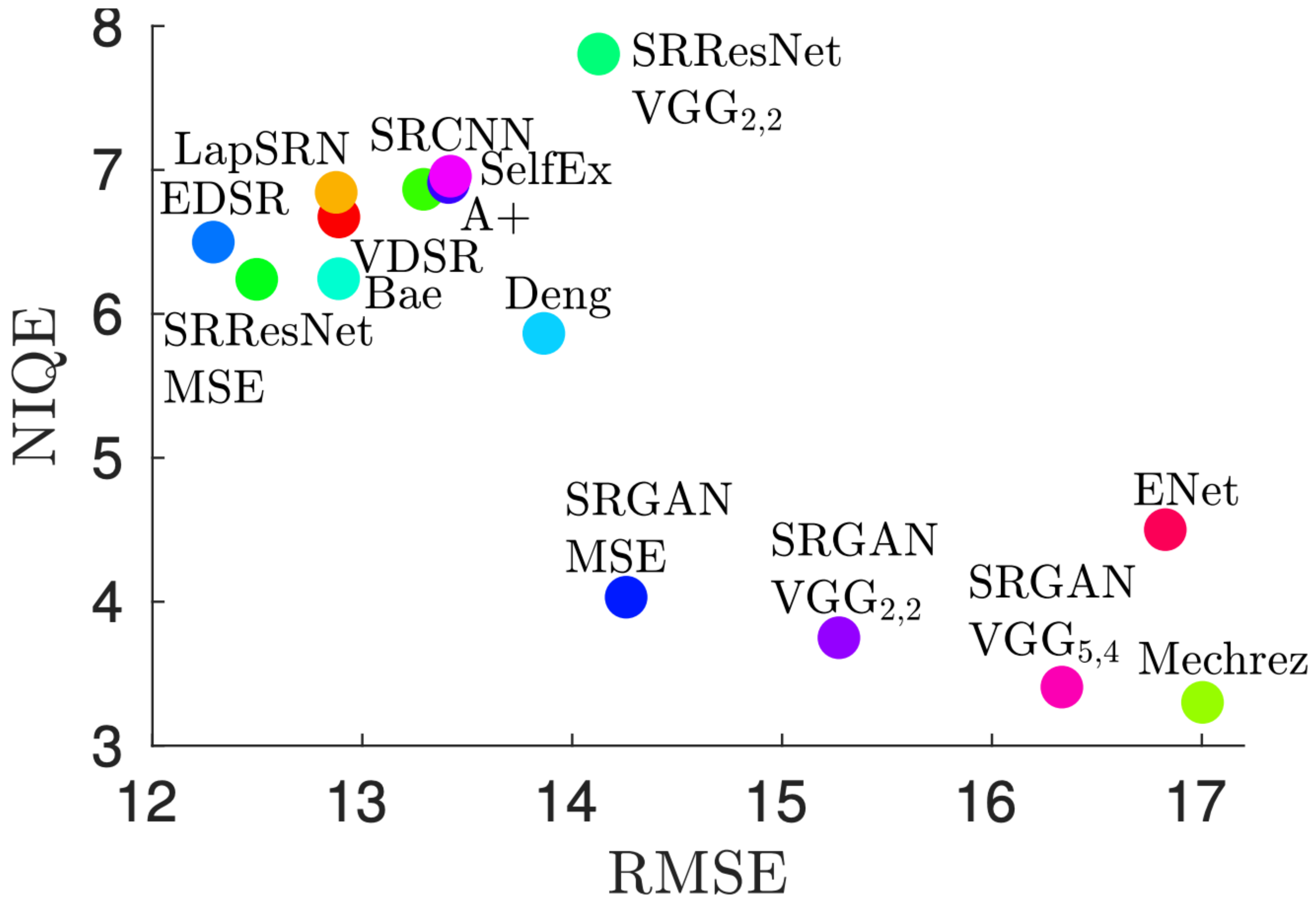
$$D(\theta_d) = \inf_{P_{\hat{\mathbf{X}}|\mathbf{Z}}} D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}] \quad \text{s.t.} \quad \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \leq \theta_d$$

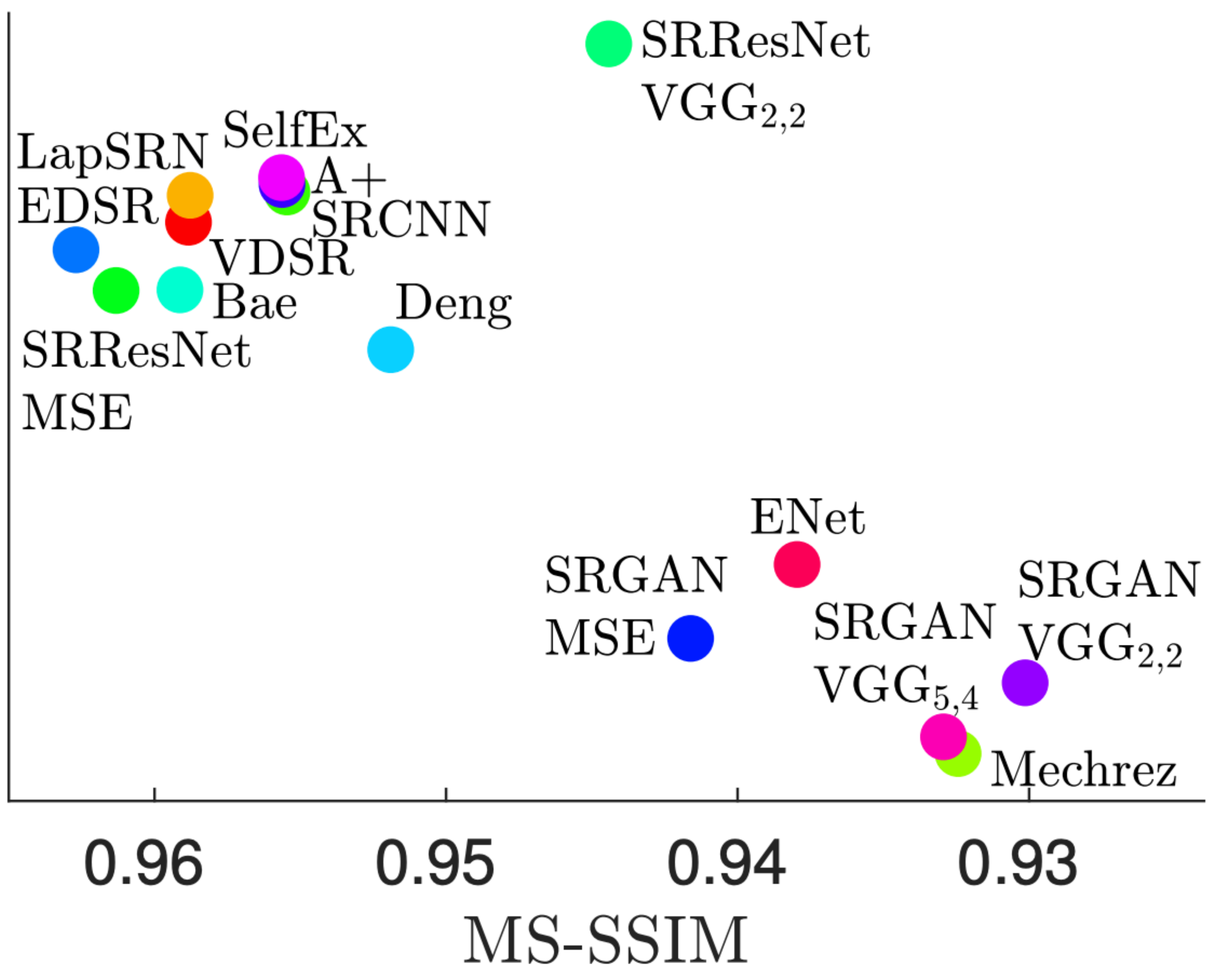
Theorem. If $D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}]$ is a divergence which is convex in its second argument, then the *perception-distortion function* $D(\theta_d)$ is (1) monotonically non-increasing and (2) convex.



Blau & Michaeli (2018)

Realism-distortion trade-off





Blau & Michaeli (2018)

(Information) rate-distortion function

$$R(\theta_d) = \inf_{P_{\hat{\mathbf{x}}|\mathbf{x}}} I[\mathbf{X}; \hat{\mathbf{X}}]$$

$$\text{s.t.} \quad \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \leq \theta_d$$

!

(Information) rate-distortion-perception function (RDPF)

$$R(\theta_d, \theta_D) = \inf_{P_{\hat{\mathbf{x}}|\mathbf{x}}} I[\mathbf{X}; \hat{\mathbf{X}}]$$

$$\text{s.t.} \quad \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \leq \theta_d \quad \text{and} \quad D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}] \leq \theta_D$$

Blau & Michaeli (2019)

E.g., $D_1[P_{\mathbf{X},\hat{\mathbf{X}}}] = \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})]$

General rate functions

Vector

$$R(\boldsymbol{\theta}) = \inf_{P_{\hat{\mathbf{X}}|\mathbf{X}}} I[\mathbf{X}; \hat{\mathbf{X}}]$$

$$\text{s.t.} \quad \forall i : D_i[P_{\mathbf{X},\hat{\mathbf{X}}}] \leq \theta_i$$

Shannon (1948)

equal fidelity. This means that a criterion of fidelity can be represented by a numerically valued function:

$$v\big(P(x,y)\big)$$

whose argument ranges over possible probability functions $P(x,y)$.

We will now show that under very general and reasonable assumptions the function $v\big(P(x,y)\big)$ can be written in a seemingly much more specialized form, namely as an average of a function $\rho(x,y)$ over the set of possible values of x and y:

$$v\big(P(x,y)\big) = \iint P(x,y)\rho(x,y)\,dx\,dy.$$

To obtain this we need only assume (1) that the source and system are ergodic so that a very long sample will be, with probability nearly 1, typical of the ensemble, and (2) that the evaluation is "reasonable" in the sense that it is possible, by observing a typical input and output x_1 and y_1, to form a tentative evaluation on the basis of these samples; and if these samples are increased in duration the tentative evaluation will, with probability 1, approach the exact evaluation based on a full knowledge of $P(x,y)$. Let the tentative

Perfect realism

$$\mathbf{X}^* \sim P_{\mathbf{X}|\hat{\mathbf{x}}}$$

Arbitrary reconstruction

Blau & Michaeli (2019)

Perfect realism

$$\mathbf{X}^* \sim P_{\mathbf{X}|\hat{\mathbf{x}}}$$

Corrupted data, e.g., reconstruction with minimal distortion

$$P_{\mathbf{X}^*}(\mathbf{x}^*) = \sum_{\hat{\mathbf{x}}} P_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}) P_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}^* \mid \hat{\mathbf{x}})$$

$$= \sum_{\hat{\mathbf{x}}} P_{\hat{\mathbf{X}},\mathbf{X}}(\hat{\mathbf{x}}, \mathbf{x}^*) = P_{\mathbf{X}}(\mathbf{x}^*)$$

Blau & Michaeli (2019)

Perfect realism

$$\mathbf{X}^* \sim P_{\mathbf{X}|\hat{\mathbf{x}}}$$

Corrupted data, e.g., reconstruction with minimal distortion

$$P_{\mathbf{X}^*}(\mathbf{x}^*) = \sum_{\hat{\mathbf{x}}} P_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}) P_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}^* \mid \hat{\mathbf{x}})$$

$$= \sum_{\hat{\mathbf{x}}} P_{\hat{\mathbf{X}},\mathbf{X}}(\hat{\mathbf{x}}, \mathbf{x}^*) = P_{\mathbf{X}}(\mathbf{x}^*)$$

$$D[P_{\mathbf{X}}, P_{\mathbf{X}^*}] = 0$$

$$\mathbb{E}[\|\mathbf{X} - \mathbf{X}^*\|^2] \leq 2\,\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$$

At most **2x** increase

Blau & Michaeli (2019)

Perfect realism

Proof:

$$\mathbb{E}[\|\mathbf{X} - \mathbf{X}^*\|^2] = \mathbb{E}[\|\mathbf{X}^* - \mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}] + \mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbf{X}\|^2]$$

$$= \mathbb{E}[\|\mathbf{X}^* - \mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}]\|^2 + \|\mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbf{X}\|^2]$$
$$+ \mathbb{E}[2(\mathbf{X}^* - \mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}])^\top (\mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbf{X})]]$$

$(\mathbf{X}^*, \hat{\mathbf{X}}) \sim (\mathbf{X}, \hat{\mathbf{X}})$

$$= \mathbb{E}[\|\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}]\|^2 + \|\mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbf{X}\|^2]$$
$$+ \mathbb{E}_{\hat{\mathbf{X}}}[2\,\mathbb{E}[\mathbf{X}^* - \mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}] \mid \hat{\mathbf{X}}]^\top \mathbb{E}[\mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbf{X} \mid \hat{\mathbf{X}}]]$$

0

$\mathbf{X}^* \perp\!\!\!\perp \mathbf{X} \mid \hat{\mathbf{X}}$

$$= 2\mathbb{E}[\|\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \hat{\mathbf{X}}]\|^2]$$

$$\leq 2\,\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$$

Blau & Michaeli (2019)

Bounds on the RDPF

$$\mathbb{E}[\|\mathbf{X} - \mathbf{X}^*\|^2] \leq 2\,\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$$

$$R(2\theta_d, 0) \leq R(\theta_d, \infty) = R(\theta_d)$$

Perfect realism

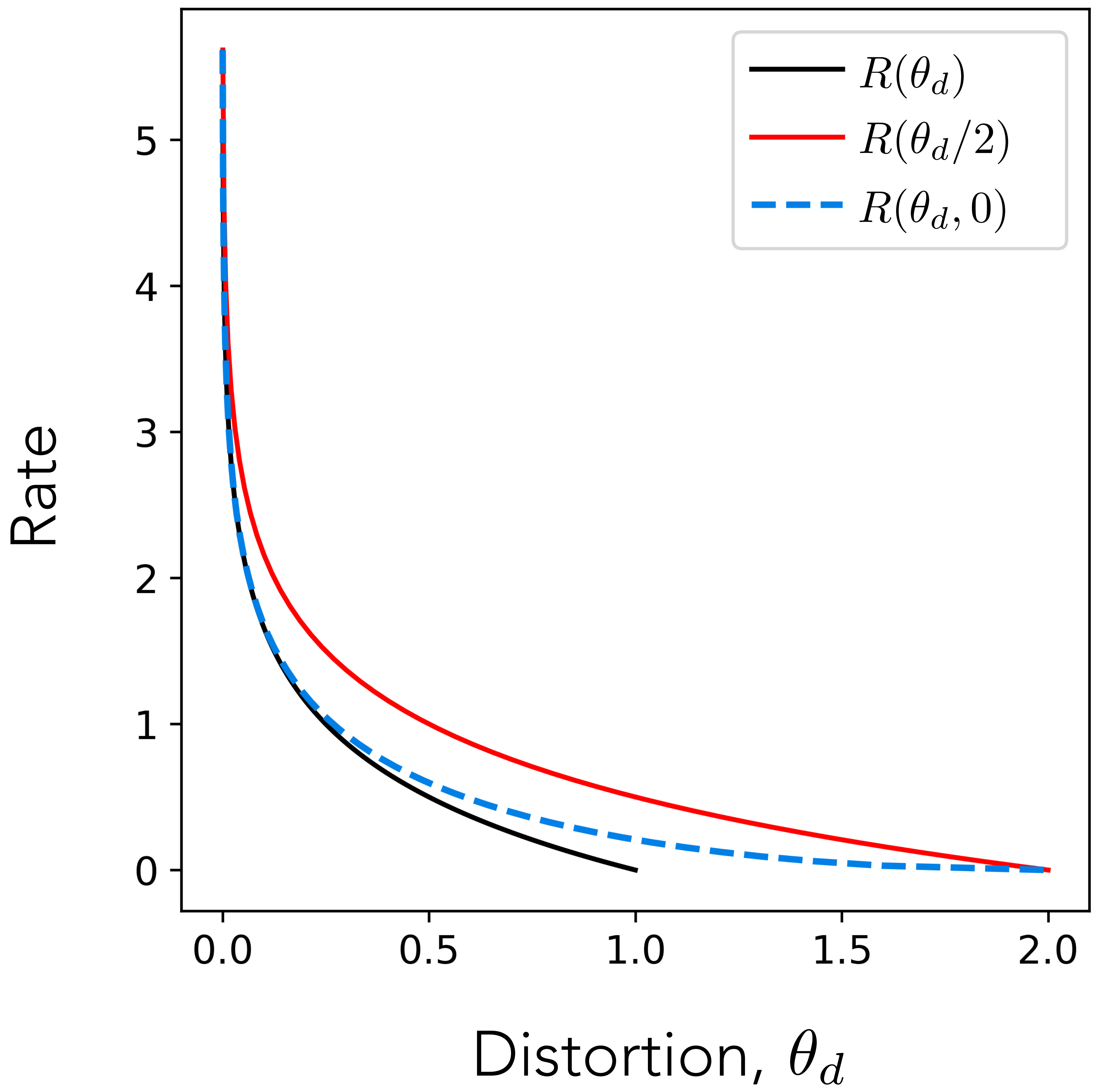
2x distortion

Bounds on the RDPF

$$\mathbb{E}[\|\mathbf{X} - \mathbf{X}^*\|^2] \leq 2\,\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$$

$$R(\theta_d, 0) \leq R(\theta_d/2, \infty) = R(\theta_d/2)$$

Bounds on the RDPF



How meaningful is $R(\theta_d, \theta_D)$?

E.g., $D_1[P_{\mathbf{X},\hat{\mathbf{X}}}] = \mathbb{E}[d(\mathbf{X},\hat{\mathbf{X}})]$

General rate functions

Vector

$$R(\boldsymbol{\theta}) = \inf_{P_{\hat{\mathbf{X}}|\mathbf{X}}} I[\mathbf{X};\hat{\mathbf{X}}]$$

$$\text{s.t.} \quad \forall i : D_i[P_{\mathbf{X},\hat{\mathbf{X}}}] \leq \theta_i$$

One-shot achievability

Definition. For a source $\mathbf{X} \sim P_{\mathbf{X}}$ and a given set of constraints, we say that a rate R is *one-shot achievable* if an encoder $f : \mathcal{X} \times \mathbb{R} \to \mathbb{N}_0$, a decoder $g : \mathbb{N}_0 \times \mathbb{R} \to \mathcal{X}$, and a random variable U (indep. of $\mathbf{X}$) exist with

$$K = f(\mathbf{X}, U) \quad \text{and} \quad \hat{\mathbf{X}} = g(K, U)$$

such that the joint distribution $P_{\mathbf{X},\hat{\mathbf{X}}}$ satisfies the constraints and the conditional entropy of K is not more than R, $H[K \mid U] \leq R$.

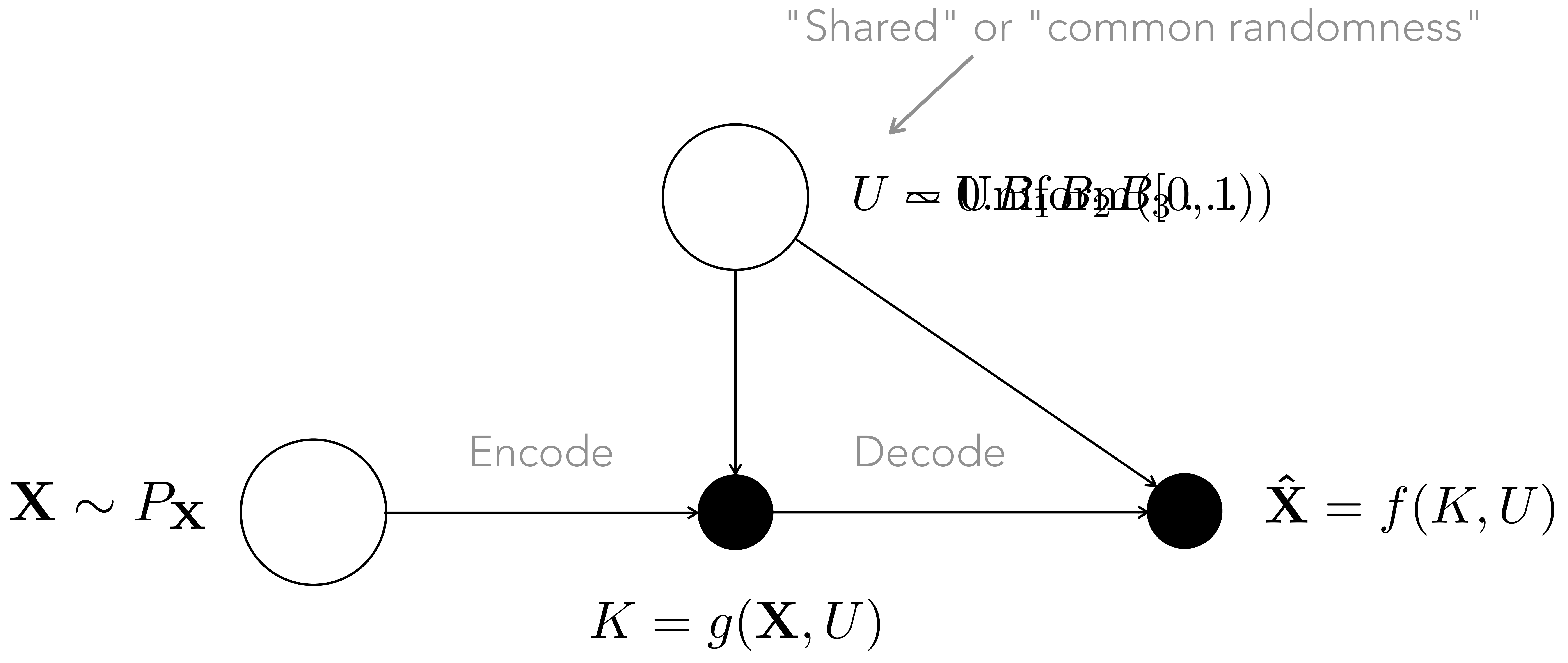
One-shot achievability

Definition. For a source $\mathbf{X} \sim P_{\mathbf{X}}$ and a given set of constraints, we say that a rate R is *one-shot achievable* if an encoder $f : \mathcal{X} \times \mathbb{R} \to \mathbb{N}_0$, a decoder $g : \mathbb{N}_0 \times \mathbb{R} \to \mathcal{X}$, and a random variable U (indep. of $\mathbf{X}$) exist with

$$K = f(\mathbf{X}, U) \quad \text{and} \quad \hat{\mathbf{X}} = g(K, U)$$

such that the joint distribution $P_{\mathbf{X},\hat{\mathbf{X}}}$ satisfies the constraints and the conditional entropy of K is not more than R, $H[K \mid U] \leq R$.

Shared randomness



Achievability

Definition. For a source $\mathbf{X} \sim P_{\mathbf{X}}$ and a given set of constraints, we say that a rate R is *(asymptotically) achievable* if there exists a random variable U independent of $\mathbf{X}$ and a sequence of encoders $f_N : \mathcal{X}^N \times \mathbb{R} \to \mathbb{N}_0$ and decoders $g_N : \mathbb{N}_0 \times \mathbb{R} \to \mathcal{X}^N$ with

$$K_N = f(\mathbf{X}^N, U) \quad \text{and} \quad \hat{\mathbf{X}}^N = g(K_N, U)$$

such that each joint distribution $P_{\mathbf{X}_n, \hat{\mathbf{X}}_n}$ $(n = 1, \ldots, N)$ satisfies the constraints and

$$\lim_{N \to \infty} H[K_N \mid U]/N \leq R.$$

Achievability

Definition. For a source $\mathbf{X} \sim P_{\mathbf{X}}$ and a given set of constraints, we say that a rate R is *(asymptotically) achievable* if there exists a random variable U independent of $\mathbf{X}$ and a sequence of encoders $f_N : \mathcal{X}^N \times \mathbb{R} \to \mathbb{N}_0$ and decoders $g_N : \mathbb{N}_0 \times \mathbb{R} \to \mathcal{X}^N$ with

$$K_N = f(\mathbf{X}^N, U) \quad \text{and} \quad \hat{\mathbf{X}}^N = g(K_N, U)$$

such that each joint distribution $P_{\mathbf{X}_n, \hat{\mathbf{X}}_n}$ $(n = 1, \dots, N)$ satisfies the constraints and

$$\lim_{N \to \infty} H[K_N \mid U]/N \leq R.$$

One-shot achievability

Theorem. Let an arbitrary source $\mathbf{X} \sim P_{\mathbf{X}}$ and constraints $D_i[P_{\mathbf{X},\hat{\mathbf{X}}}] \leq \theta_i$ be given. If

$$R > R(\boldsymbol{\theta}) + \log(R(\boldsymbol{\theta}) + 1) + 4,$$

then R is one-shot achievable.

Vector

Li & El Gamal (2018); Theis & Wagner (2021)

Achievability

Theorem. Let an arbitrary source $\mathbf{X} \sim P_{\mathbf{X}}$ and constraints $D_i[P_{\mathbf{X},\hat{\mathbf{X}}}] \leq \theta_i$ be given. Then $R < \infty$ is achievable if and only if $R \geq R(\boldsymbol{\theta})$.

Vector

Li & El Gamal (2018); Theis & Wagner (2021)

LEARNED COMPRESSION II:

Adversarial losses and diffusion

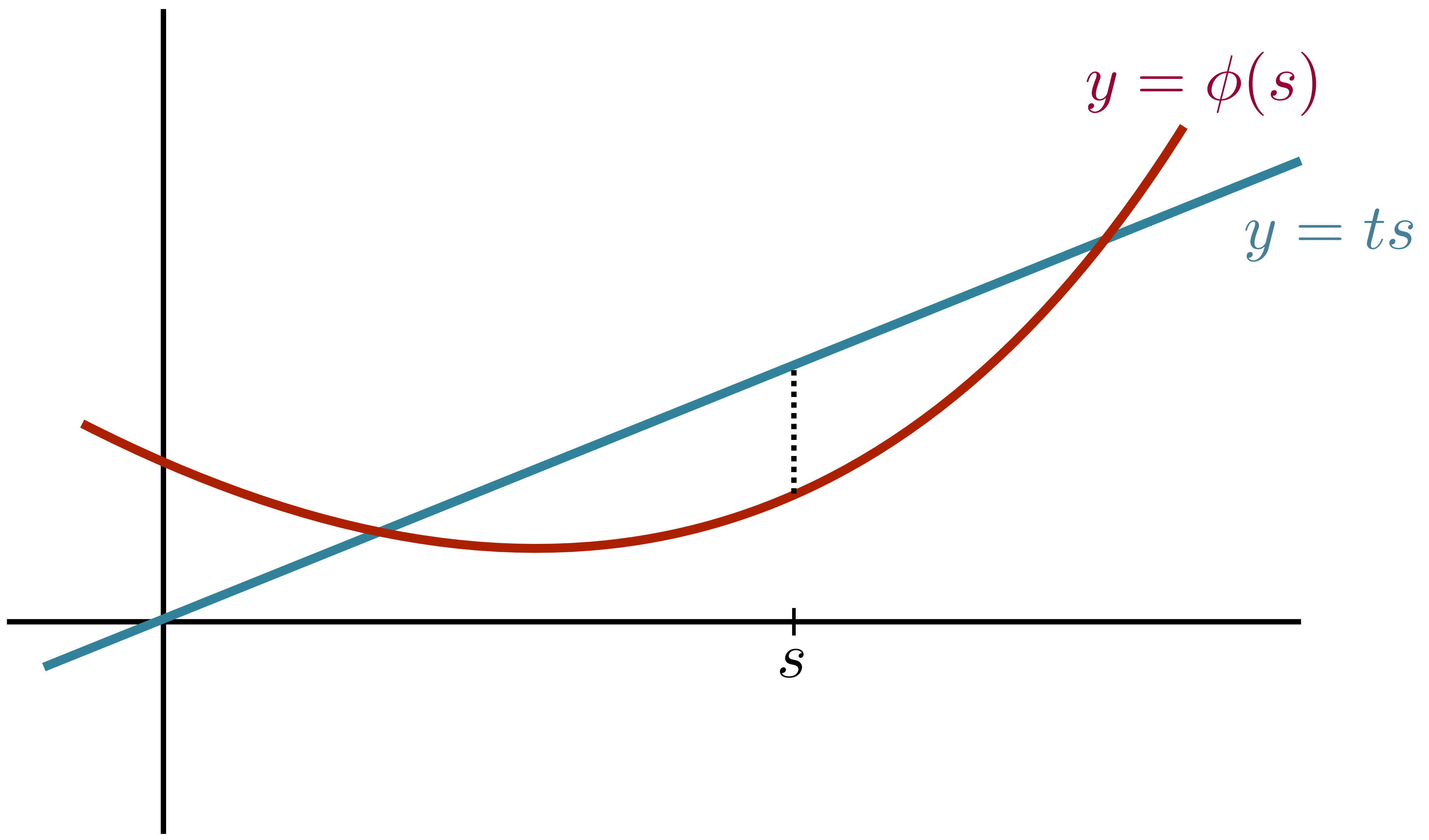
Legendre transform (convex conjugate)

$$\phi : I \to \mathbb{R}$$

$$\phi^* : I^* \to \mathbb{R},$$

$$\phi^*(t) = \sup_{s \in I} \{st - \phi(s)\}$$

$$\phi(s) = \sup_{t \in I^*} \{st - \phi^*(t)\}$$



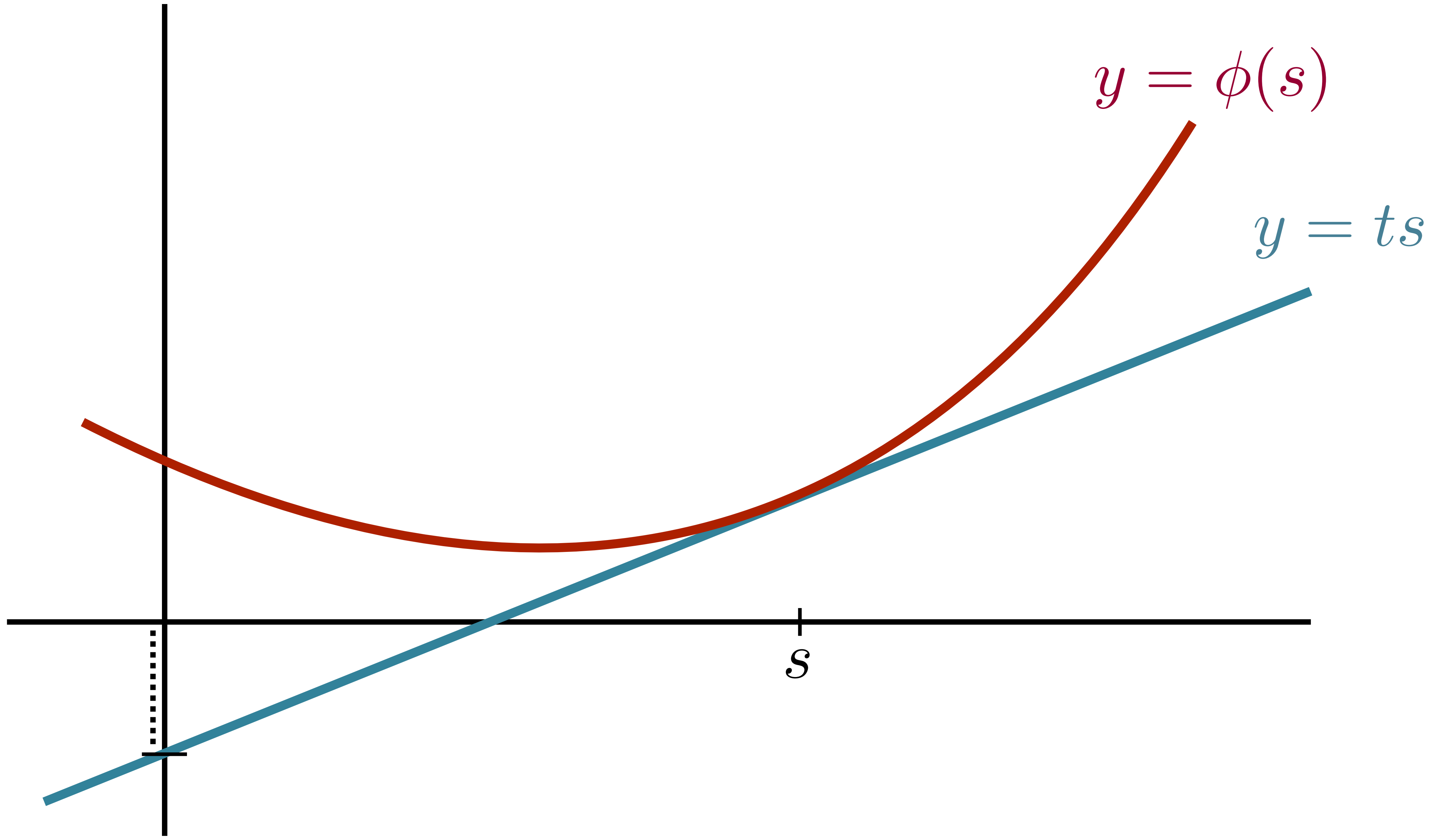
Legendre transform (convex conjugate)

$$\phi : I \to \mathbb{R}$$

$$\phi^* : I^* \to \mathbb{R},$$

$$\phi^*(t) = \sup_{s \in I} \{st - \phi(s)\}$$

$$\phi(s) = \sup_{t \in I^*} \{st - \phi^*(t)\}$$



Legendre transform (convex conjugate)

$$\phi(s) = s\log(s)$$

$$\phi^*(t) = \exp(t-1)$$

$$\phi(s) = \sup_{t\in\mathbb{R}}(st - \exp(t-1))$$

f-divergences

Convex function

$$D_\phi[P, Q] = \int \phi\left(\frac{dP}{dQ}\right) dQ$$

Rényi (1961)

f-divergences

$$D_\phi[p, q] = \int q(\mathbf{x})\phi\left(\frac{p(\mathbf{x})}{q(\mathbf{x})}\right) d\mathbf{x}$$

E.g., KL divergences: $\phi(s) = -\log s$ or $\phi(s) = s \log s$

Rényi (1961)

Adversarial losses

$$D_\phi[p, q] = \int q(\mathbf{x}) \phi\left(\frac{p(\mathbf{x})}{q(\mathbf{x})}\right)$$

$$= \int q(\mathbf{x}) \sup_t \left\{ t \frac{p(\mathbf{x})}{q(\mathbf{x})} - \phi^*(t) \right\} d\mathbf{x}$$

Goodfellow et al. (2014); Nowozin et al. (2016)

Adversarial losses

$$\begin{aligned}
D_\phi[p,q] &= \int q(\mathbf{x})\phi\left(\frac{p(\mathbf{x})}{q(\mathbf{x})}\right) \\
&= \int q(\mathbf{x}) \sup_t \left\{ t\frac{p(\mathbf{x})}{q(\mathbf{x})} - \phi^*(t) \right\} d\mathbf{x} \\
&\geq \sup_{\boldsymbol{\eta}} \int q(\mathbf{x}) \left(T_{\boldsymbol{\eta}}(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} - \phi^*(T_{\boldsymbol{\eta}}(\mathbf{x})) \right) d\mathbf{x} \\
&= \sup_{\boldsymbol{\eta}} \int p(\mathbf{x}) T_{\boldsymbol{\eta}}(\mathbf{x})\, d\mathbf{x} - \int q(\mathbf{x}) \phi^*(T_{\boldsymbol{\eta}}(\mathbf{x}))\, d\mathbf{x}
\end{aligned}$$

Goodfellow et al. (2014); Nowozin et al. (2016)

Adversarial losses

$$\sup_{\boldsymbol{\eta}} \int p(\mathbf{x}) T_{\boldsymbol{\eta}}(\mathbf{x}) \, d\mathbf{x} - \int q(\mathbf{x}) \phi^*(T_{\boldsymbol{\eta}}(\mathbf{x})) \, d\mathbf{x}$$

Goodfellow et al. (2014); Nowozin et al. (2016)

Adversarial losses

"Adversary" or "critic"

$$\sup_{\boldsymbol{\eta}} \mathbb{E}[T_{\boldsymbol{\eta}}(\mathbf{X})] - \mathbb{E}[\phi^*(T_{\boldsymbol{\eta}}(\hat{\mathbf{X}}))]$$

Adversarial losses

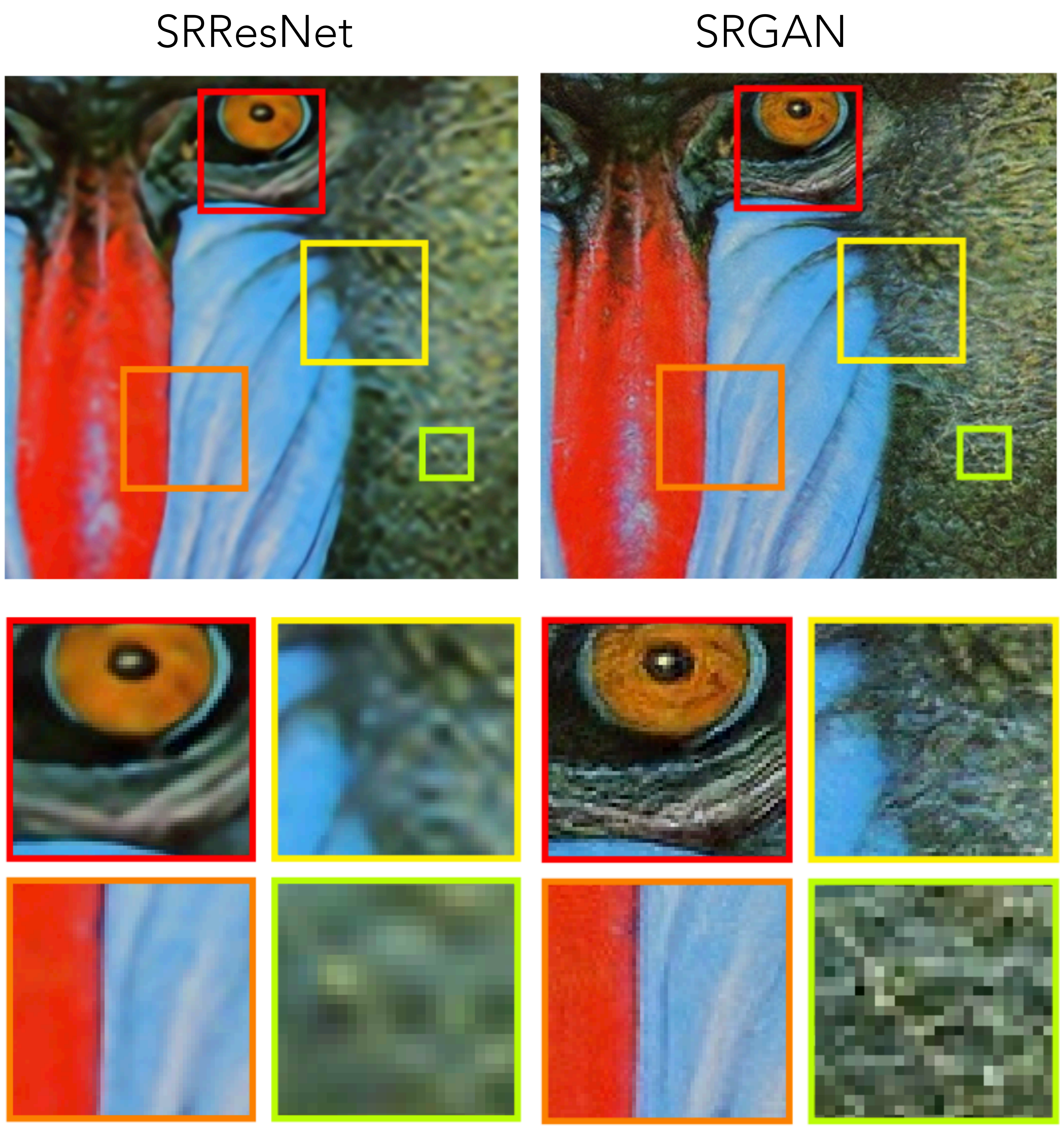
$$\mathbf{Z} = f_{\boldsymbol{\theta}}(\mathbf{X}) + \mathbf{U}$$

$$\hat{\mathbf{X}} = g_{\boldsymbol{\theta}}(\mathbf{Z})$$

Repeat:

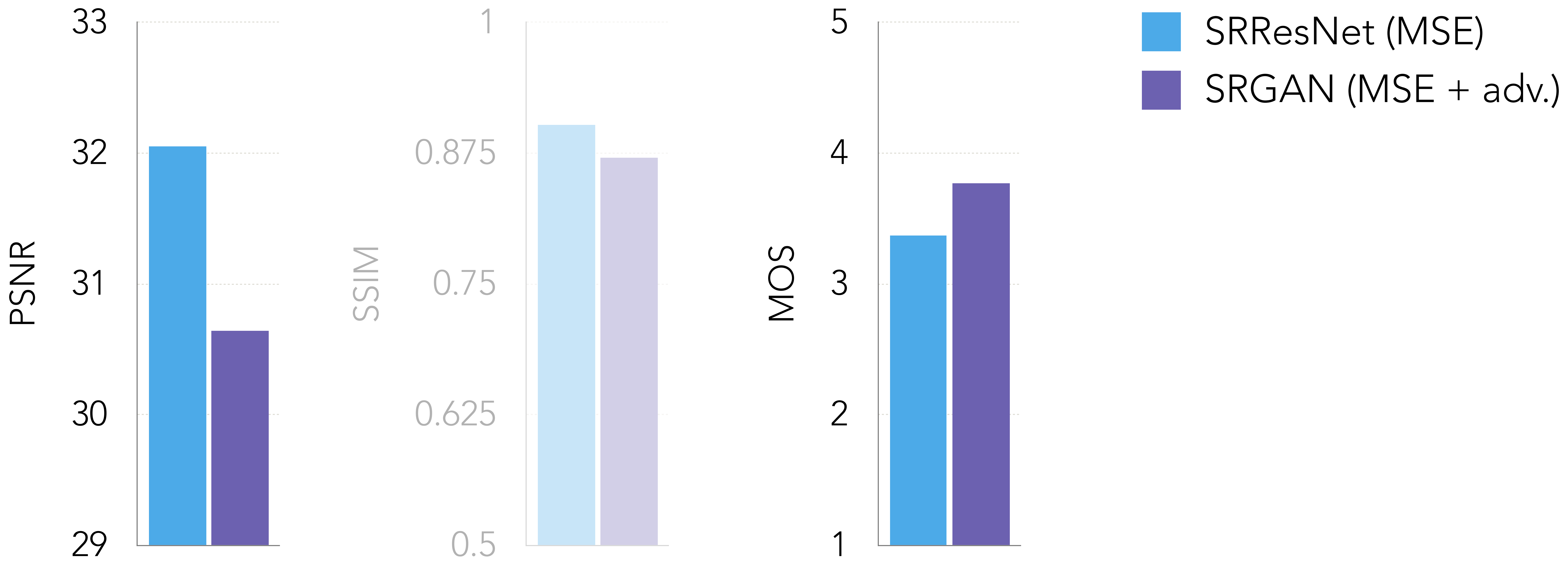
$$\boldsymbol{\eta} \leftarrow \operatorname*{argmin}_{\eta} \left(\mathbb{E}[T_{\boldsymbol{\eta}}(\mathbf{X})] - \mathbb{E}[\phi^*(T_{\boldsymbol{\eta}}(\hat{\mathbf{X}}))] \right)$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \varepsilon \nabla_{\boldsymbol{\theta}} \left(\underbrace{\mathbb{E}[-\log p(\mathbf{Z})]}_{\text{Rate}} + \lambda \underbrace{\mathbb{E}[-\log \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})]}_{\text{Distortion}} + \beta \underbrace{\mathbb{E}[-\phi^*(T_{\boldsymbol{\eta}}(\hat{\mathbf{X}}))]}_{\text{Realism}} \right)$$



Ledig et al. (2017)

Example: SRGAN



Ledig et al. (2017)

A first approach using *diffusion*

Perfect realism

Powerful (conditional) generative model

Rate-constrained representation

$$\hat{\mathbf{X}} \sim P_{\mathbf{X}|\mathbf{Z}}$$ (2)

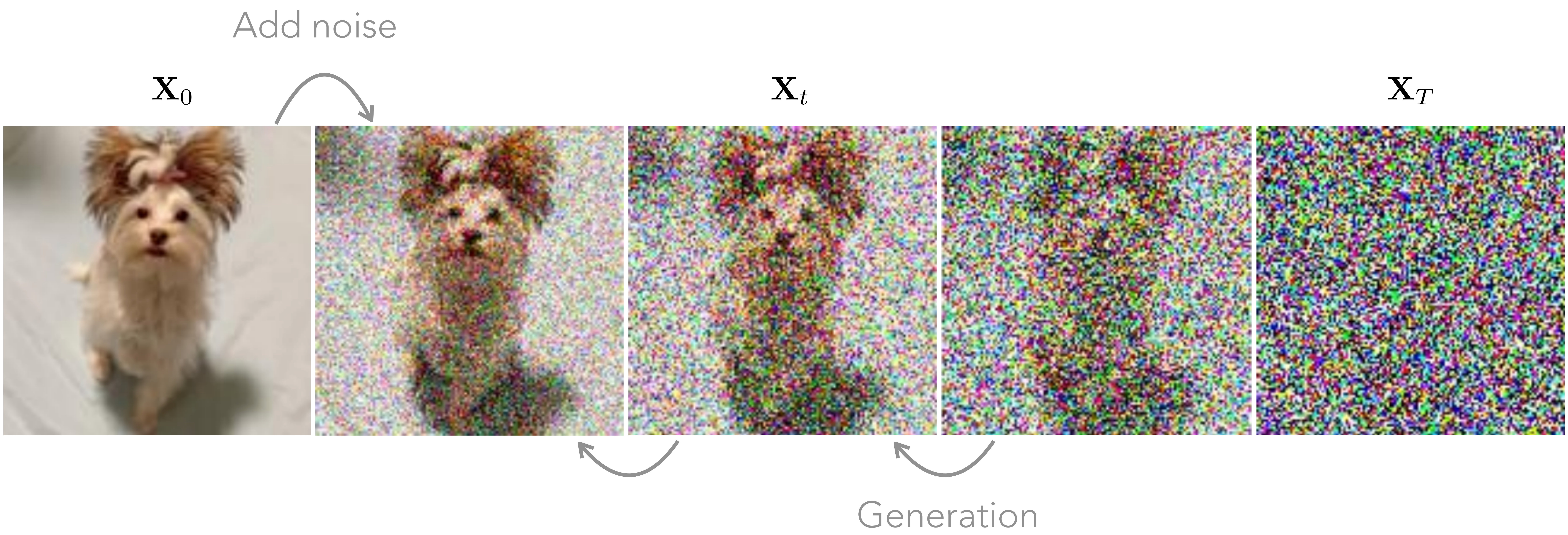
$$D[P_{\mathbf{X}}, P_{\hat{\mathbf{X}}}] = 0$$

$$\mathbb{E}[\|\hat{\mathbf{X}} - \mathbf{X}\|^2] = 2\,\mathbb{E}[\|\mathbb{E}[\mathbf{X} \mid \mathbf{Z}] - \mathbf{X}\|^2]$$ (1)

Difficult to optimize

Easy to optimize

Diffusion



Sohl-Dickstein et al. (2015); Song et al. (2021)

Diffusion

💡 $P_{\mathbf{X}_{t-\delta}|\mathbf{X}_t}$ is approximately Gaussian (e.g., Feller, 1949; Anderson, 1982).

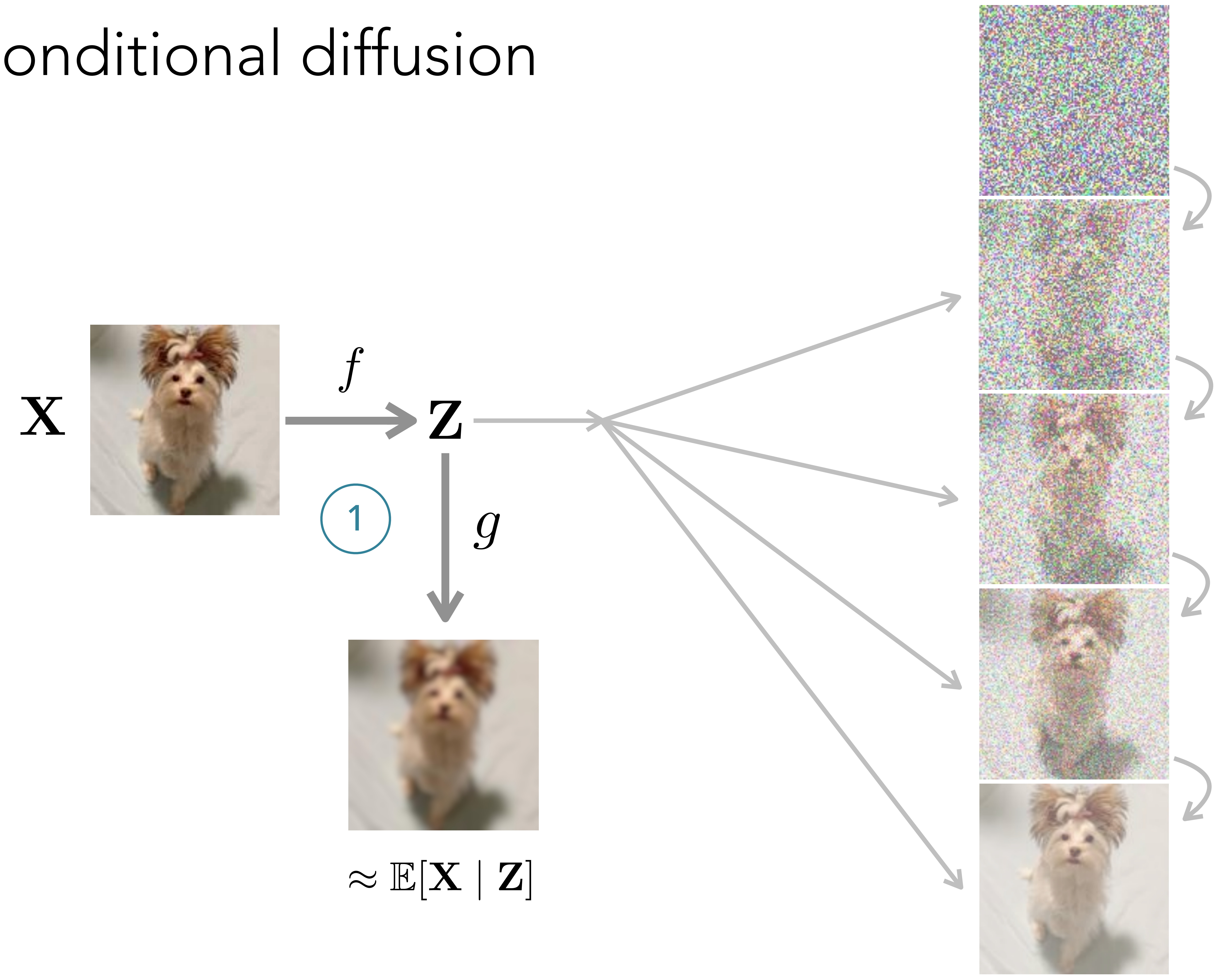
💡 Optimize $\mathbb{E}[\|\mathbf{X}_{t-\delta} - m_{\boldsymbol{\theta}}(\mathbf{X}_t)\|^2]$ so that $m_{\boldsymbol{\theta}}(\mathbf{X}_t) \approx \mathbb{E}[\mathbf{X}_{t-\delta} \mid \mathbf{X}_t]$.

Diffusion

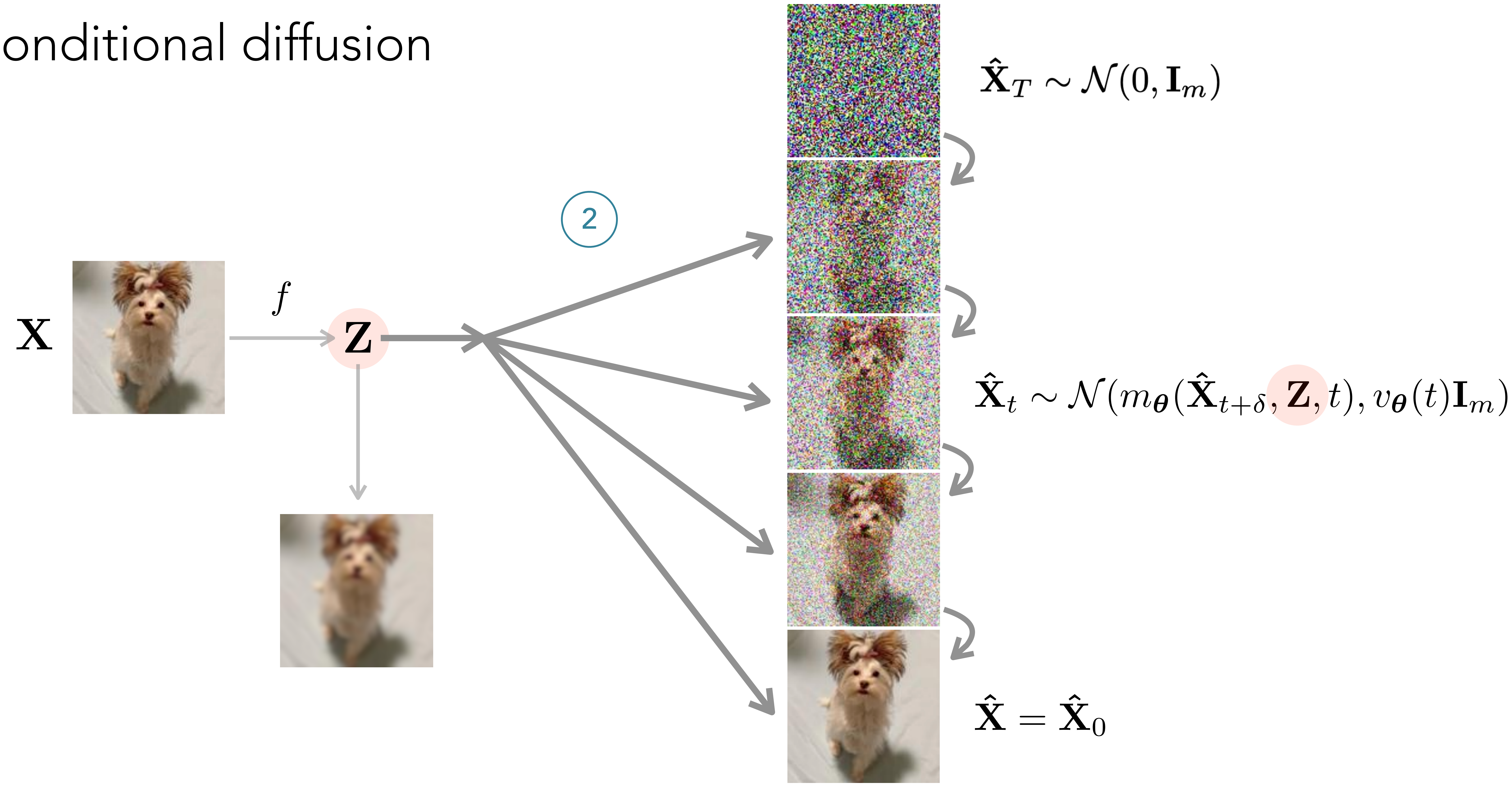
💡 $P_{\mathbf{X}_{t-\delta}|\mathbf{X}_t}$ is approximately Gaussian (e.g., Feller, 1949; Anderson, 1982).

💡 Optimize $\sum_t w_t \mathbb{E}[\|\mathbf{X}_{t-\delta} - m_{\boldsymbol{\theta}}(\mathbf{X}_t, t)\|^2]$ so that $m_{\boldsymbol{\theta}}(\mathbf{X}_t, t) \approx \mathbb{E}[\mathbf{X}_{t-\delta} \mid \mathbf{X}_t]$.

Conditional diffusion



Conditional diffusion

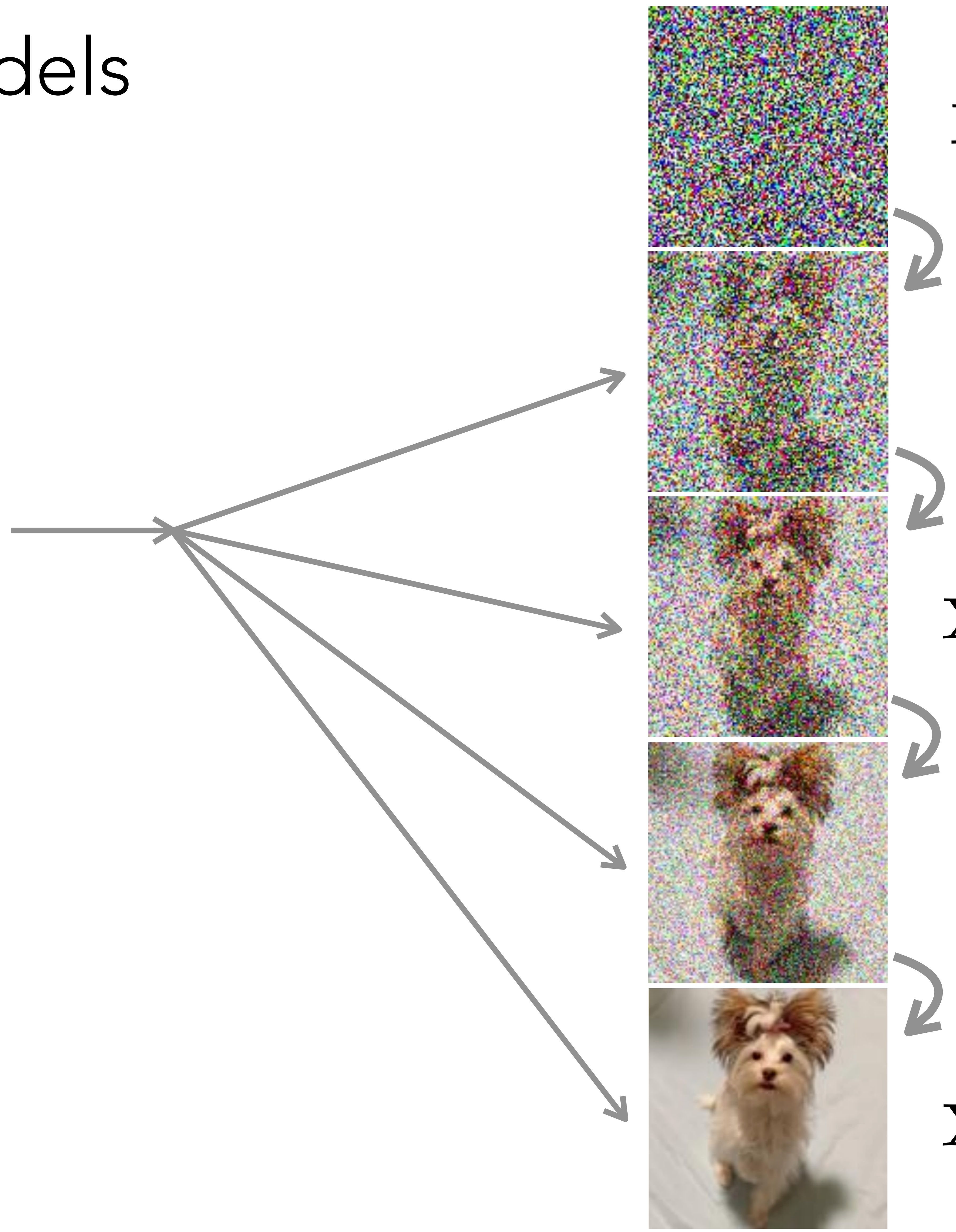



MSE
0.0562 bpp


MSE + diffusion
0.0562 bpp

Text-to-image models

"A cute dog." → $\mathbf{Z}$



$$\hat{\mathbf{X}}_T \sim \mathcal{N}(0, \mathbf{I}_m)$$

$$\hat{\mathbf{X}}_t \sim \mathcal{N}(m_{\boldsymbol{\theta}}(\hat{\mathbf{X}}_{t+\delta}, \mathbf{Z}, t), v_{\boldsymbol{\theta}}(t)\mathbf{I}_m)$$

$$\hat{\mathbf{X}} = \hat{\mathbf{X}}_0$$

Text-to-image models

kodim05.png
(Kodak, 1993)

→ "seven motocross bikes facing right at the starting line" → $\mathbf{z}$ →

Imagen
(Saharia et al., 2022)

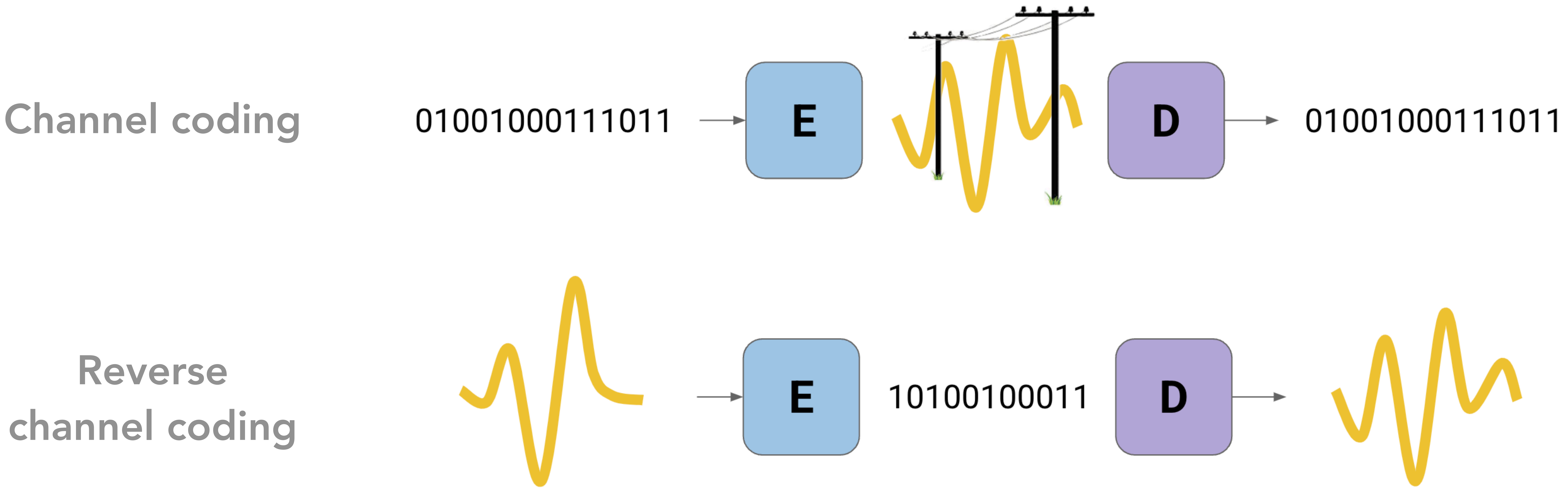
Hoogeboom et al. (2023)

Rombach et al. (2022)

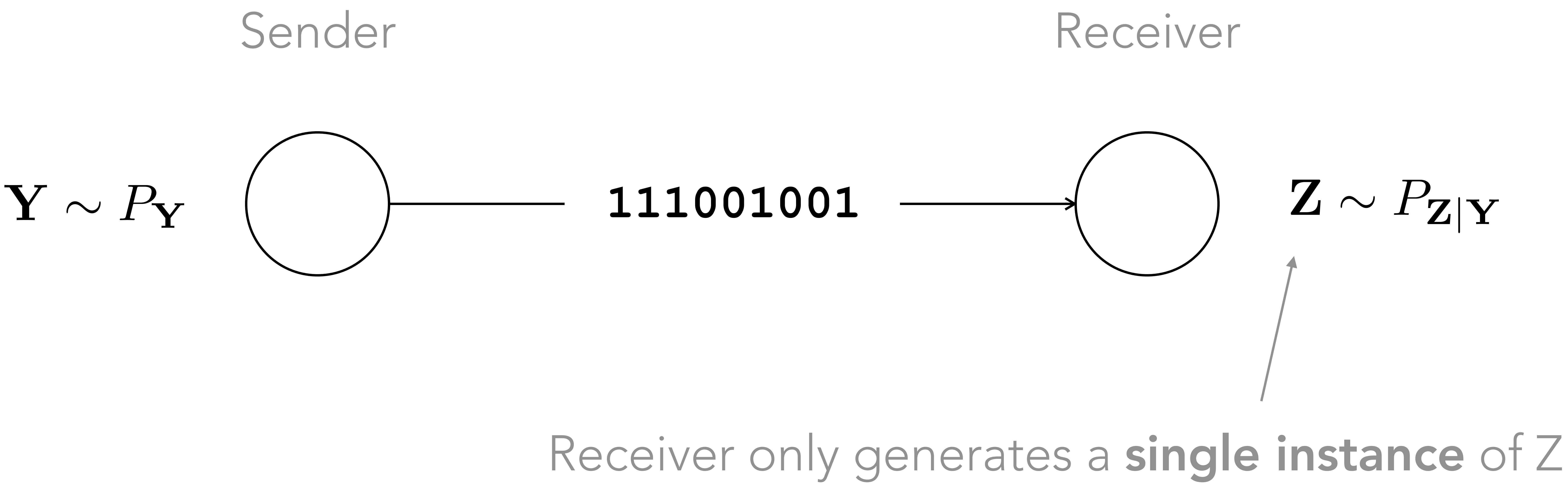
REALISM II:

Channel simulation

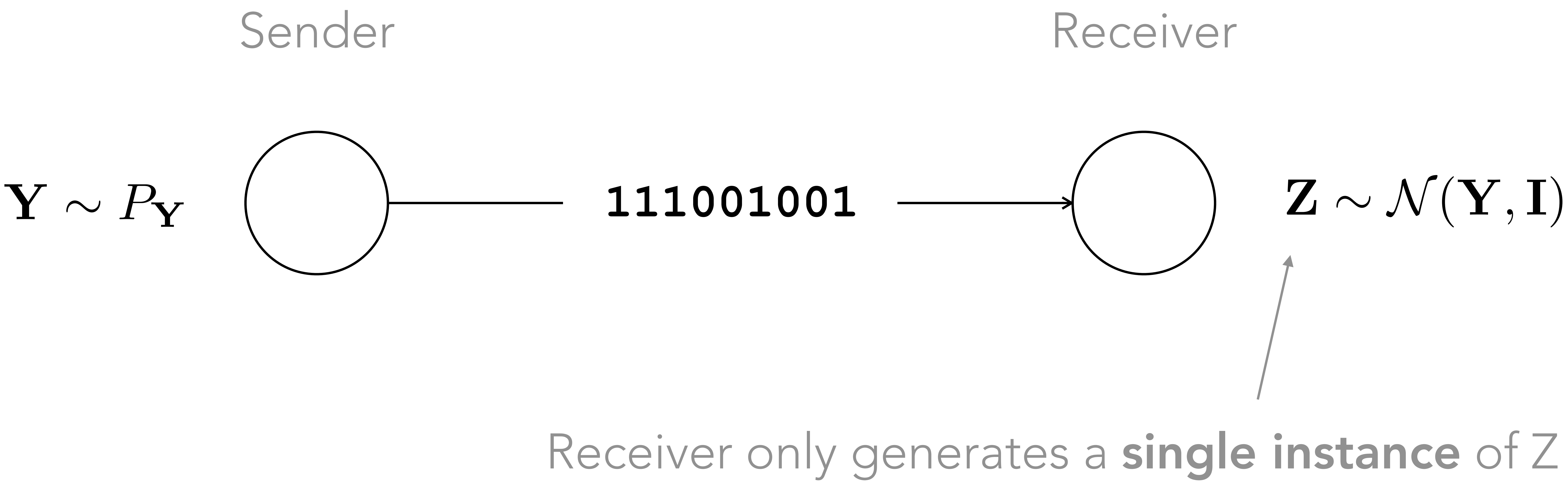
Channel simulation (or *reverse* channel coding)



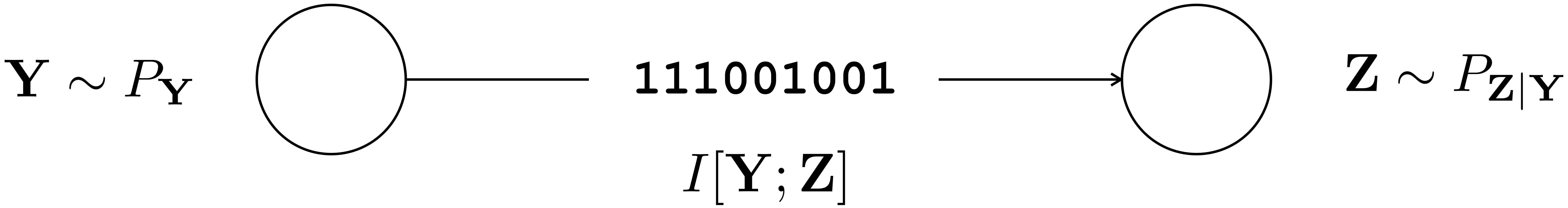
Channel simulation (or *reverse* channel coding)



Channel simulation (or *reverse* channel coding)

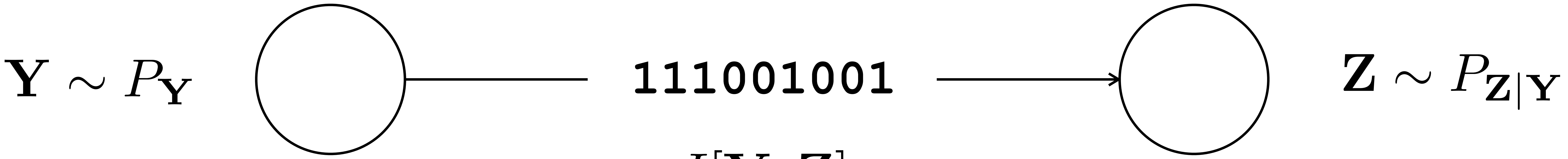


Channel simulation (or *reverse* channel coding)



Bennett & Shor (2002); Harsha et al. (2007)

Channel simulation (or *reverse* channel coding)

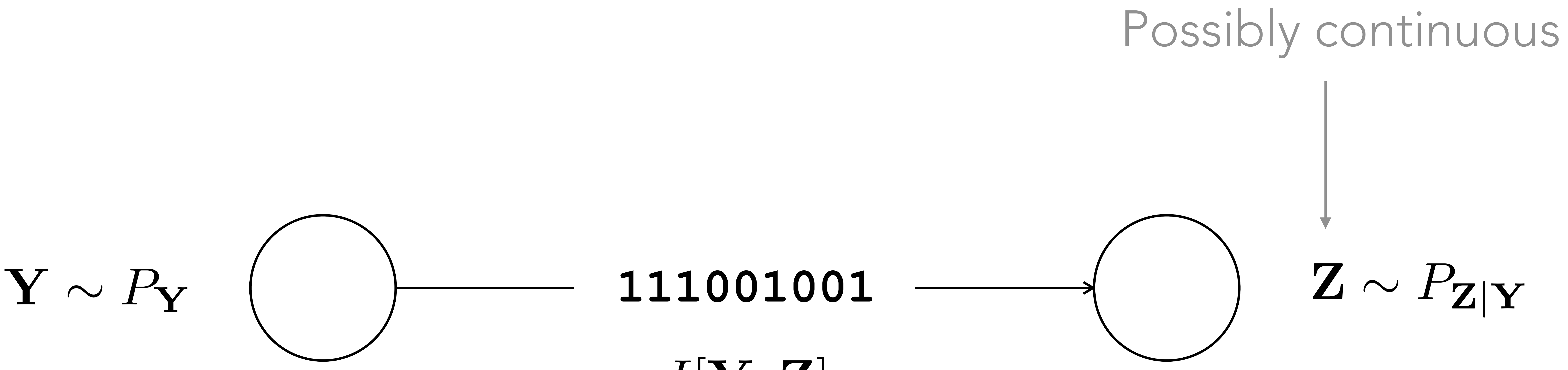


$$I[\mathbf{Y};\mathbf{Z}]$$

$$= H[\mathbf{Z}] - H[\mathbf{Z} \mid \mathbf{Y}]$$

Bennett & Shor (2002); Harsha et al. (2007)

Channel simulation (or *reverse* channel coding)



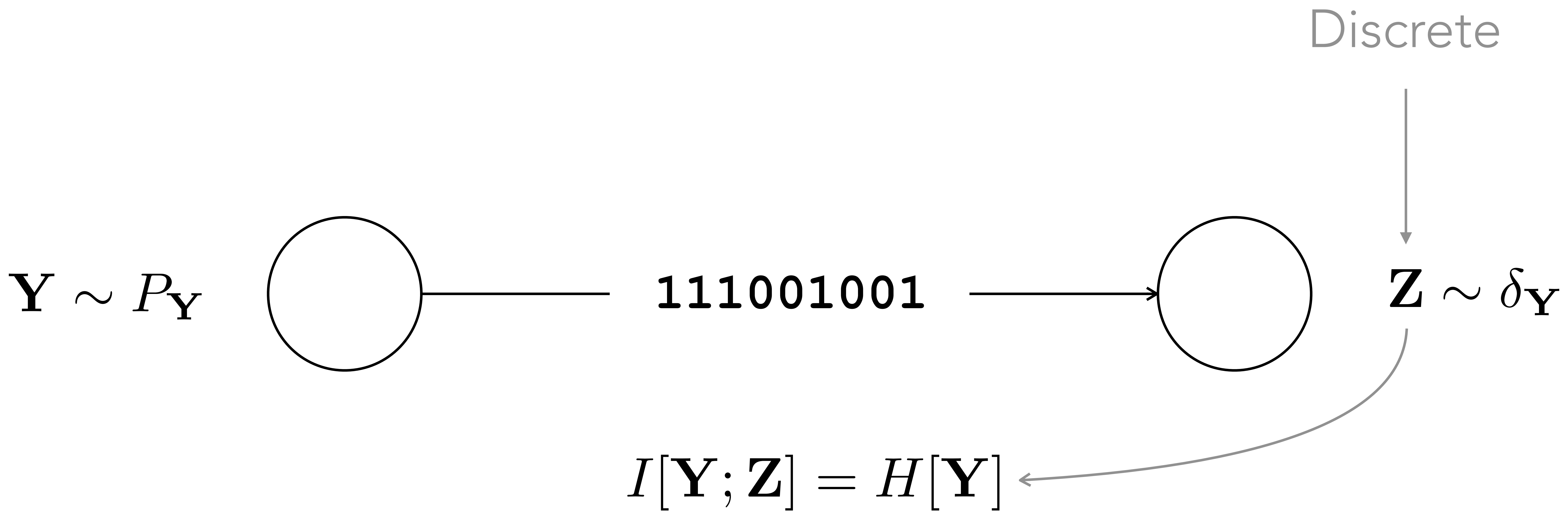
$$I[\mathbf{Y}; \mathbf{Z}]$$

$$= h[\mathbf{Z}] - h[\mathbf{Z} \mid \mathbf{Y}]$$

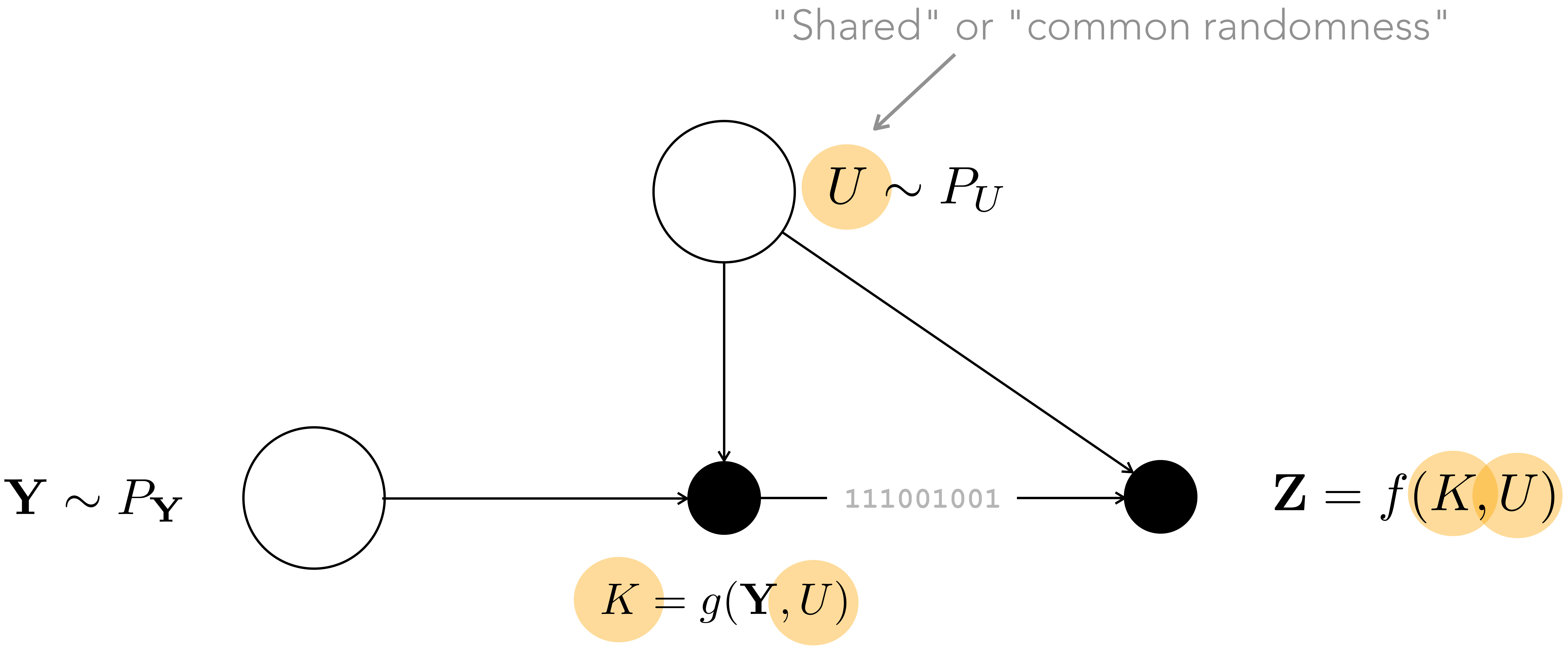
Bennett & Shor (2002); Harsha et al. (2007)

Channel simulation...

... generalizes lossless source coding



Channel simulation



Dithered quantization

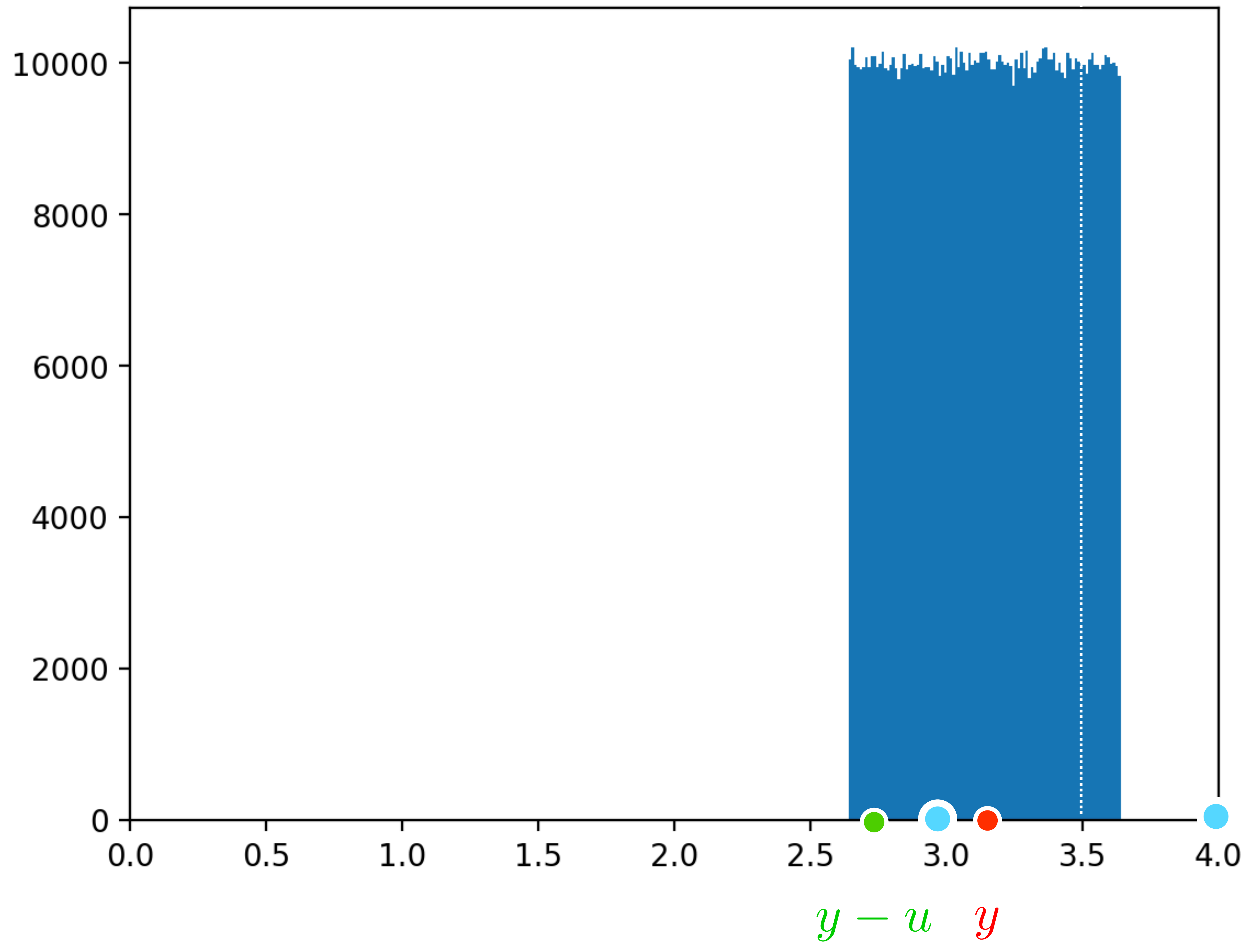
Encoder Decoder

$$Z = \underbrace{\lfloor Y - U \rceil}_{K} + U \sim Y + U' \qquad U, U' \sim \mathrm{Uniform}([-1/2, 1/2))$$

Roberts (1962)

```
y = 3.14
U = np.random.rand(1000000) - 0.5
K = np.round(y - U)
Z = K + U

plt.hist(Z, 100)
plt.xlim([0, 4]);
```



Dithered quantization

$Z = K + U$

$$P(K = k \mid U = u) = p_Z(k + u)$$

$$\begin{aligned} H[K \mid U] &= \mathbb{E}[-\log p_Z(K + U)] \\ &= \mathbb{E}[-\log p_Z(Z)] \\ &= h[Z] \\ &= h[Z] - h[Z \mid Y] \\ &= I[Z, Y] \end{aligned}$$

0

Zamir & Feder (1992)

Dithered quantization

$$P(K = k \mid U = u) = P(Y - u \in [k - 0.5, k + 0.5))$$

$$= P(Y \in [k + u - 0.5, k + u + 0.5))$$

$$= \int p_Y(k + u - u')[\![u' \in [-0.5, 0.5)]\!]\, du'$$

$$= p_{Y+U'}(k + u)$$

$$= p_Z(k + u)$$

Zamir & Feder (1992)

Dithered quantization

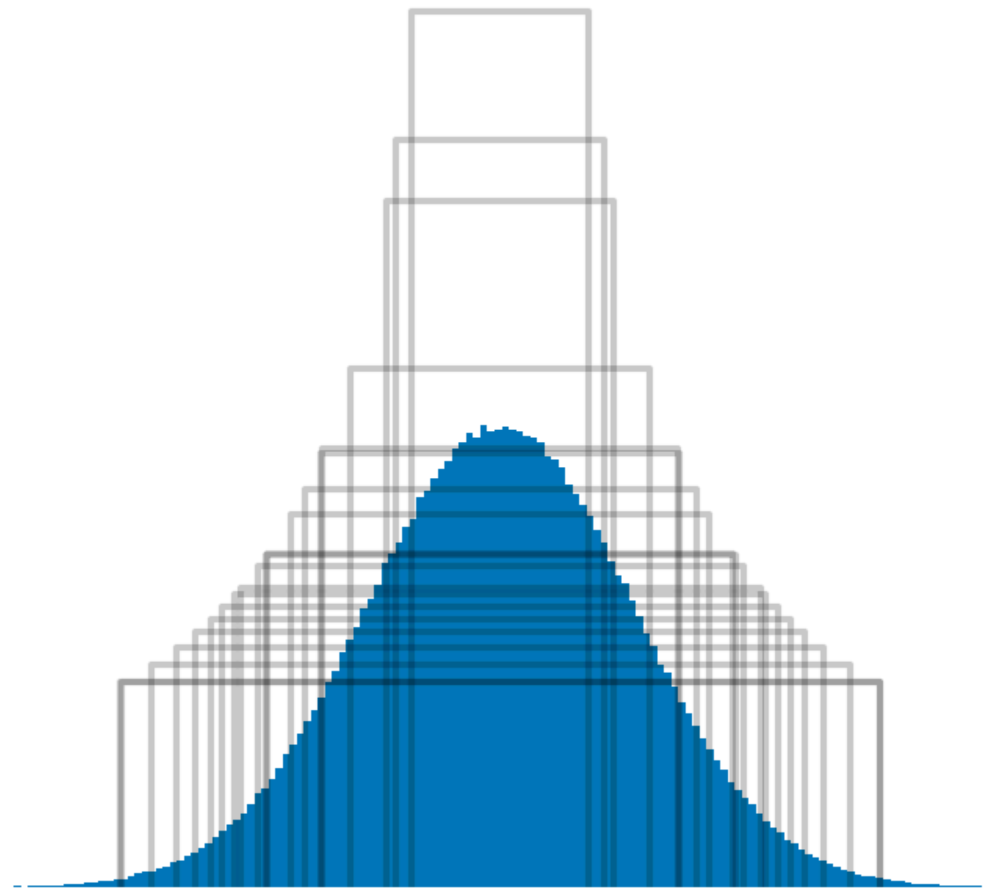
$$Z = \lfloor Y - (U + b) \rceil + (U + b)$$

$$U' \sim \mathrm{Uniform}([-0.5, 0.5))$$

Dithered quantization

$$Z = \left(\lfloor Y / s - U \rceil + U \right) s$$

Gaussian channel



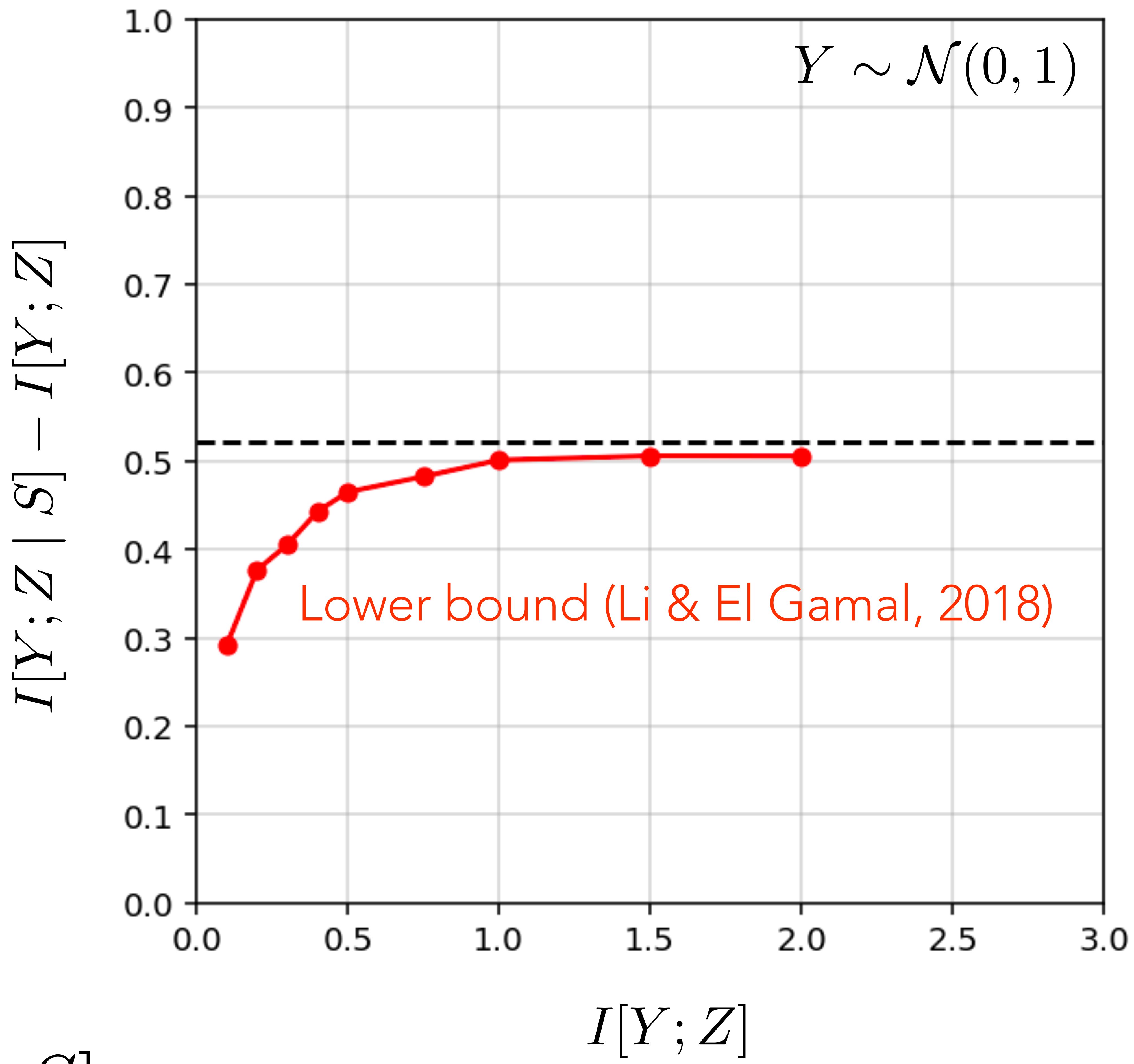
$$Z = (\lfloor Y/S - U \rceil + U)\, S$$
$$\sim Y + SU'$$

Qin et al. (2003); Agustsson & Theis (2020)

Gaussian channel

$$K = \lfloor Y/S - U \rceil$$

$$\begin{aligned}
H[K \mid U, S] &= I[Y; Z \mid S] \\
&= h[Z \mid S] - h[Z \mid Y, S] \\
&= h[Z \mid S] - \mathbb{E}[\log_2 S] \\
&\leq h[Z] - \mathbb{E}[\log_2 S] \\
&= I[Y; Z] + h[Z \mid Y] - \mathbb{E}[\log_2 S] \\
&= I[Y; Z] + \frac{1}{2}\log_2(\pi) + \frac{1 - \psi(3/2)}{2 \ln 2} \\
&\leq I[Y; Z] + 0.521
\end{aligned}$$



Dithered quantization and VAEs

Encoder Decoder

$$Z = \underbrace{\lfloor Y - U \rceil + U}_{\text{Inference}} \sim \underbrace{Y + U'}_{\text{Training}} \qquad U, U' \sim \mathrm{Uniform}([-{}^{1}\!/\!{}_{2}, {}^{1}\!/\!{}_{2}))$$

Inference Training

Roberts (1962)

VAEs revisited

Inference:

$$\mathbf{Z} = \lfloor f_{\boldsymbol{\theta}}(\mathbf{X}) \rceil$$

$$\hat{\mathbf{X}} = g_{\boldsymbol{\theta}}(\mathbf{Z})$$

Mismatch

Training:

$$\mathbf{Z} = f_{\boldsymbol{\theta}}(\mathbf{X}) + \mathbf{U}$$

$$\hat{\mathbf{X}} = g_{\boldsymbol{\theta}}(\mathbf{Z})$$

VAEs revisited

Encoder

$$\mathbf{Y} = f_{\boldsymbol{\theta}}(\mathbf{X})$$

$$K_i = \lfloor Y_i - U_i \rceil$$

Decoder

$$Z_i = K_i + U_i$$

$$\hat{\mathbf{X}} = g_{\boldsymbol{\theta}}(\mathbf{Z})$$

Rate-distortion trade-off

$$\ell(\boldsymbol{\theta}) = \lambda \mathbb{E}[d(\mathbf{X}, g_{\boldsymbol{\theta}}(\lfloor \mathbf{Y} - \mathbf{U} \rceil + \mathbf{U}))] + \mathbb{E}[-\log P(\mathbf{K} \mid \mathbf{U})]$$

VAEs revisited

Encoder

$$\mathbf{Y} = f_{\boldsymbol{\theta}}(\mathbf{X})$$

$$K_i = \lfloor Y_i - U_i \rceil$$

Decoder

$$Z_i = K_i + U_i$$

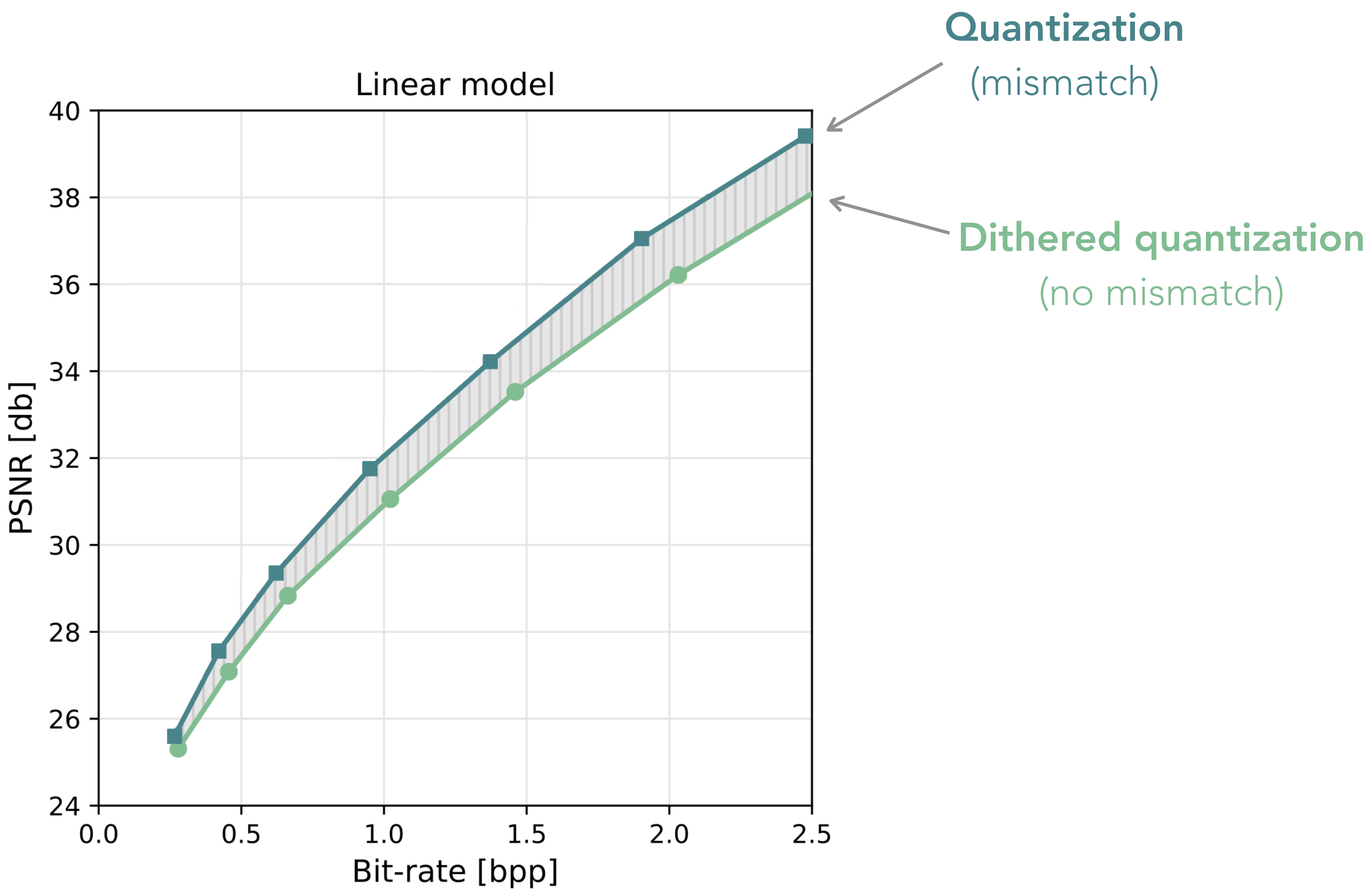
$$\hat{\mathbf{X}} = g_{\boldsymbol{\theta}}(\mathbf{Z})$$

Rate-distortion trade-off = ELBO

$$\ell(\boldsymbol{\theta}) = \lambda \mathbb{E}[d(\mathbf{X}, g_{\boldsymbol{\theta}}(\mathbf{Y} + \mathbf{U}'))] + \mathbb{E}[-\log p_{\mathbf{Z}}(\mathbf{Y} + \mathbf{U}')]$$

Agustsson & Theis (2020)

Variational autoencoders revisited



Agustsson & Theis (2020)

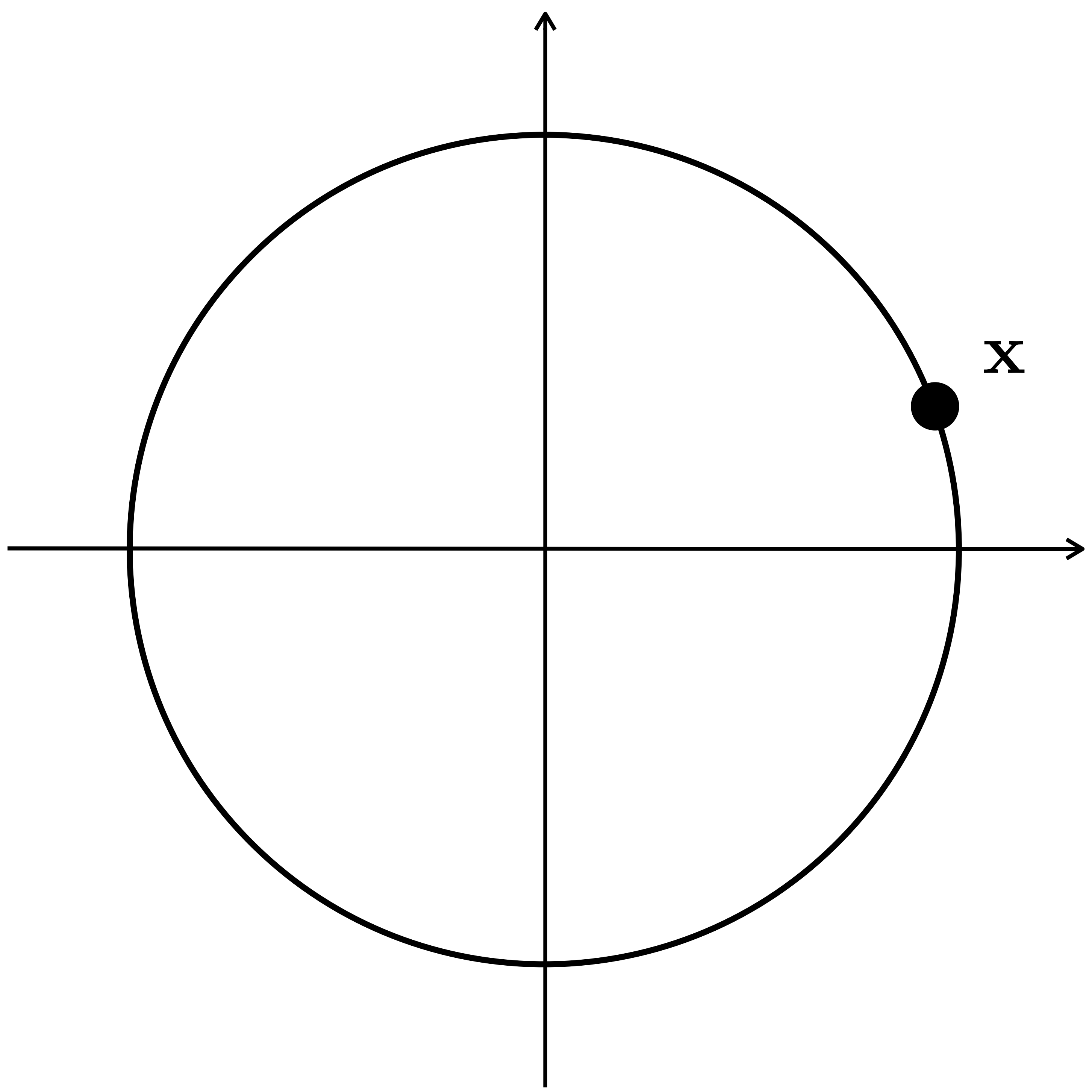
Randomness and rate-distortion trade-offs

$$\mathbb{E}[\ell_\lambda(\mathbf{X},\mathbf{U})] = \mathbb{E}_{\mathbf{U}}[\mathbb{E}_{\mathbf{X}}[\ell_\lambda(\mathbf{X},\mathbf{U})]]$$
$$\geq \min_{\mathbf{U}} \mathbb{E}_{\mathbf{X}}[\ell_\lambda(\mathbf{X},\mathbf{U})]$$
$$= \mathbb{E}_{\mathbf{X}}[\ell_\lambda(\mathbf{X},\mathbf{u}^*)]$$

Fix randomness

Ballé et al. (2021); Theis & Agustsson (2021)

Randomness and realism



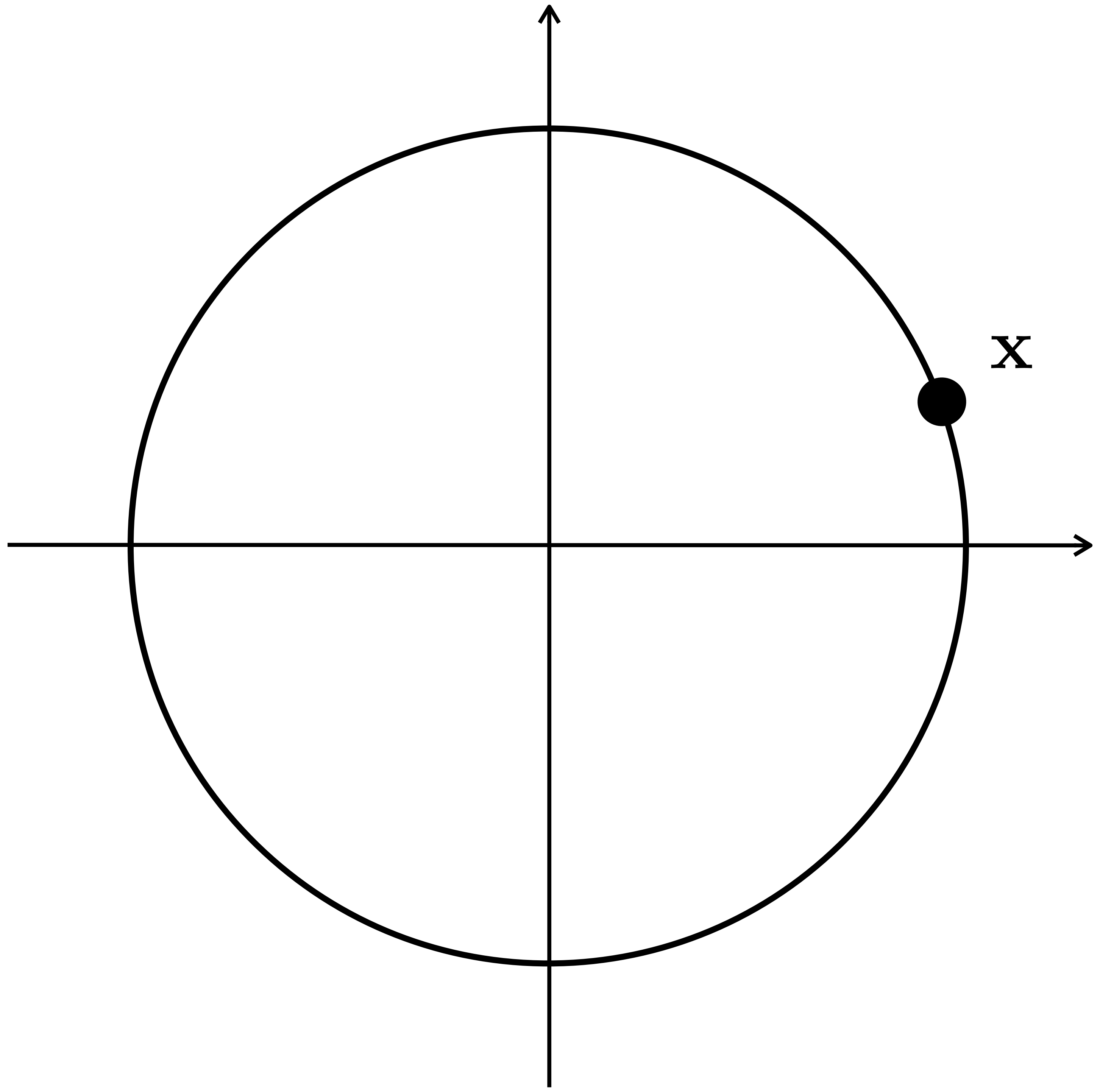
$$\Theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$\mathbf{X} = \begin{pmatrix} \cos\Theta \\ \sin\Theta \end{pmatrix}$$

Theis & Agustsson (2021)

Randomness and realism

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{2}\|\mathbf{x} - \hat{\mathbf{x}}\|^2 = 1 - \cos(\theta - \hat{\theta})$$



$$\Theta \sim \text{Uniform}(0, 2\pi)$$

$$\mathbf{X} = \begin{pmatrix} \cos\Theta \\ \sin\Theta \end{pmatrix}$$

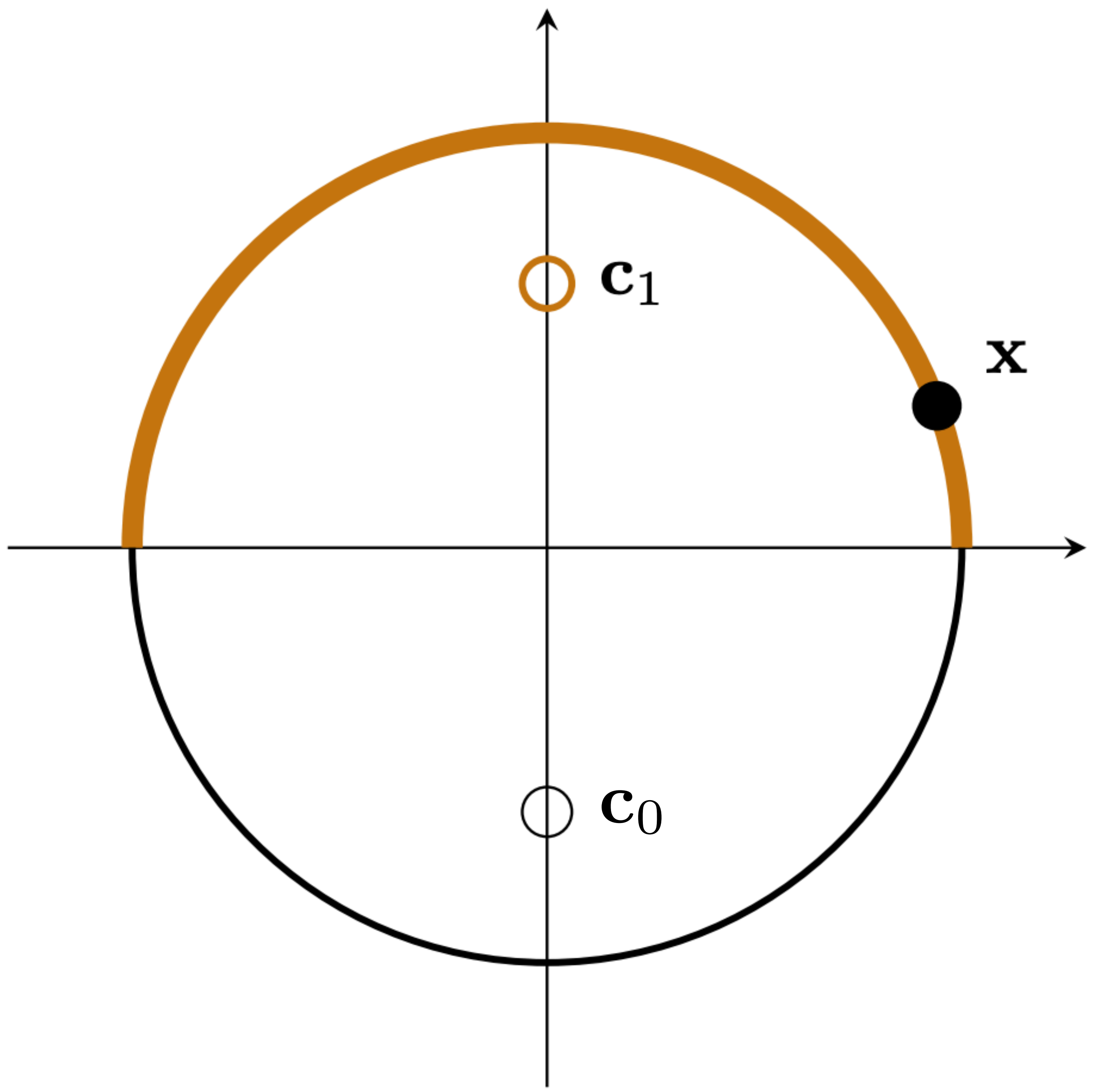
$$U \sim \text{Uniform}(0, 1) \qquad f : \mathbb{R}^2 \times \mathbb{R} \to \{0, 1\}$$

$$\hat{\mathbf{X}} = g(f(\mathbf{X}, U), U) \qquad g : \{0, 1\} \times \mathbb{R} \to \mathbb{R}^2$$

$$\underset{f,g}{\text{minimize}} \quad \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \quad \text{s.t.} \quad \hat{\mathbf{X}} \sim \mathbf{X}$$

Theis & Agustsson (2021)

Randomness and realism



$$\Theta \sim \mathrm{Uniform}(0, 2\pi)$$

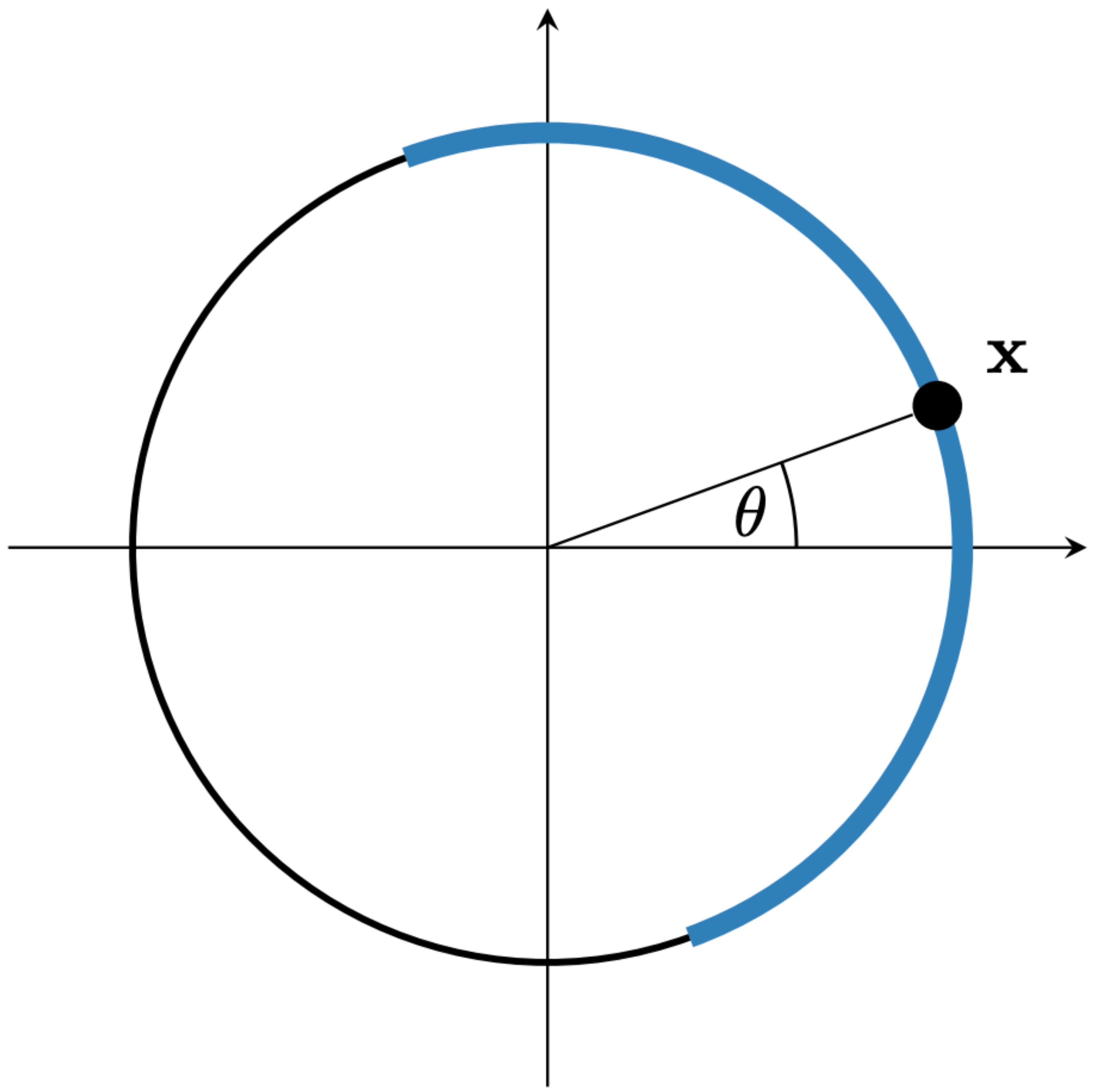
$$\mathbf{X} = \begin{pmatrix} \cos\Theta \\ \sin\Theta \end{pmatrix}$$

$$U \sim \mathrm{Uniform}(0, 1)$$

$$\hat{\mathbf{X}} = g(f(\mathbf{X}), U) \sim P_{\mathbf{X} \mid f(\mathbf{X})}$$

Theis & Agustsson (2021)

Randomness and realism



$$\Theta \sim \text{Uniform}(0, 2\pi)$$

$$\mathbf{X} = \begin{pmatrix} \cos\Theta \\ \sin\Theta \end{pmatrix}$$

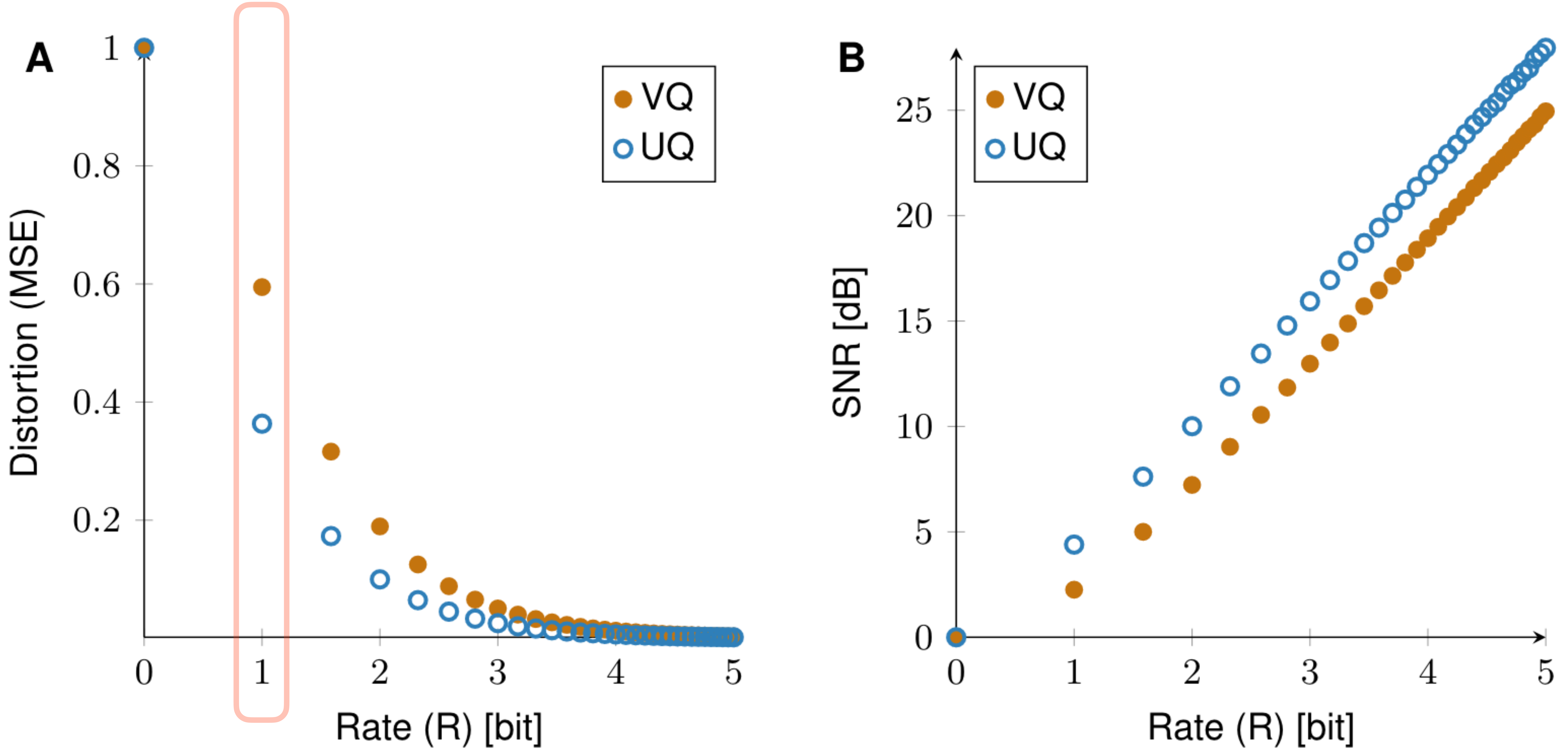
$$U \sim \text{Uniform}(0, 1)$$

$$K = f(\mathbf{X}, U) = \left\lfloor \frac{\Theta}{\pi} - U \right\rceil \mod 2$$

$$\hat{\Theta} = \pi(K + U)$$

Theis & Agustsson (2021)

Randomness and realism



Theis & Agustsson (2021)

Rejection sampling

$k \leftarrow 0$

$\mathbf{repeat}$

$\quad k \leftarrow k + 1$

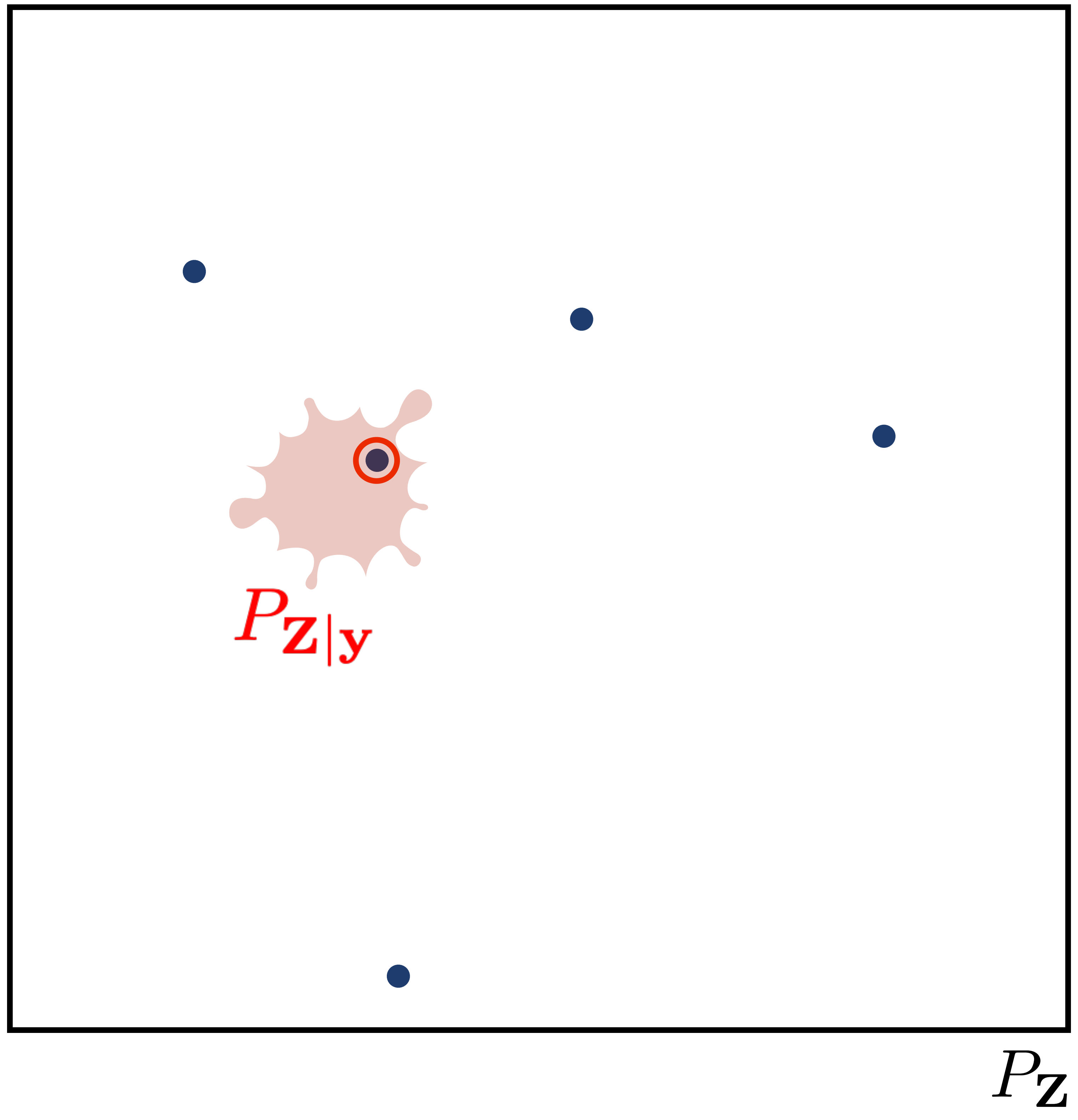
$\quad \mathbf{Z}_k \sim P_{\mathbf{Z}}$

$\quad U_k \sim \mathrm{Uniform}([0, 1))$

$$\mathbf{until}\ U_k < \frac{1}{M} \frac{dP_{\mathbf{Z}|\mathbf{y}}}{dP_{\mathbf{Z}}}(\mathbf{Z}_k)$$

$\mathbf{return}\ k$

Using shared source of randomness



Rejection sampling

$$k \leftarrow 0$$

repeat

$$k \leftarrow k + 1$$

$$\mathbf{Z}_k \sim P_{\mathbf{Z}}$$

Using shared source of randomness

$$U_k \sim \mathrm{Uniform}([0, 1))$$

until $U_k < \frac{1}{M} \frac{dP_{\mathbf{Z}|\mathbf{y}}}{dP_{\mathbf{Z}}}(\mathbf{Z}_k)$

return k

Poisson functional representation

$$S_k \sim \mathrm{Exp}(1)$$

$$T_k = \sum_{i=1}^{k} S_i$$

$$Z_k \sim P_{\mathbf{Z}}$$

Using shared source of randomness

$$K = \operatorname*{argmin}_{k \in \mathbb{N}} \left\{ T_k \frac{dP_{\mathbf{Z}}}{dP_{\mathbf{Z}|\mathbf{y}}}(\mathbf{Z}_k) \right\}$$

$$\mathbf{Z}_K \sim P_{\mathbf{Z}|\mathbf{y}}$$ ✓

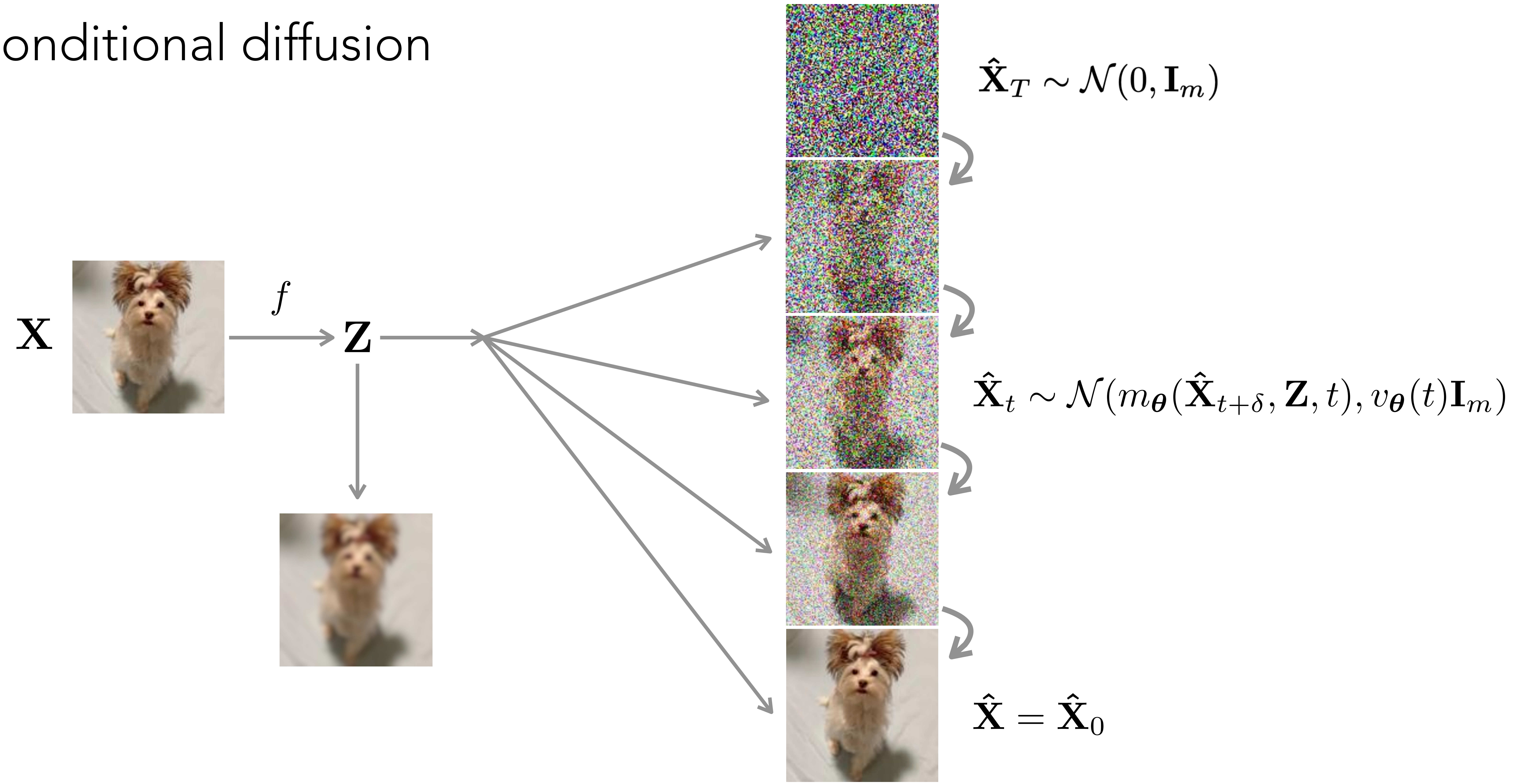
$$H[K] < I[\mathbf{Z};\mathbf{Y}] + \log_2(I[\mathbf{Z};\mathbf{Y}] + 1) + 4$$ ✓

Li & El Gamal (2018)

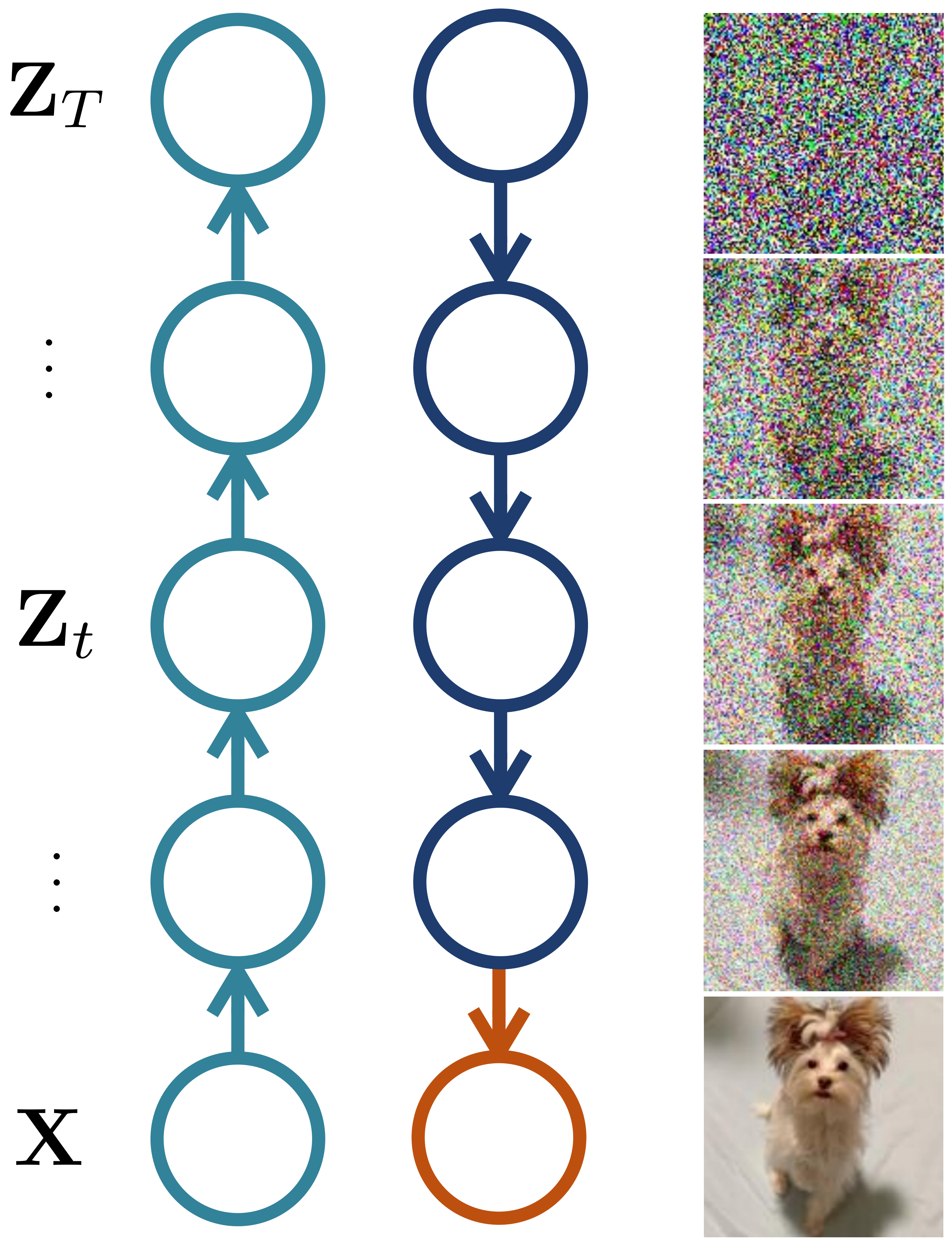
LEARNED COMPRESSION III:

Diffusion-based compression

Conditional diffusion

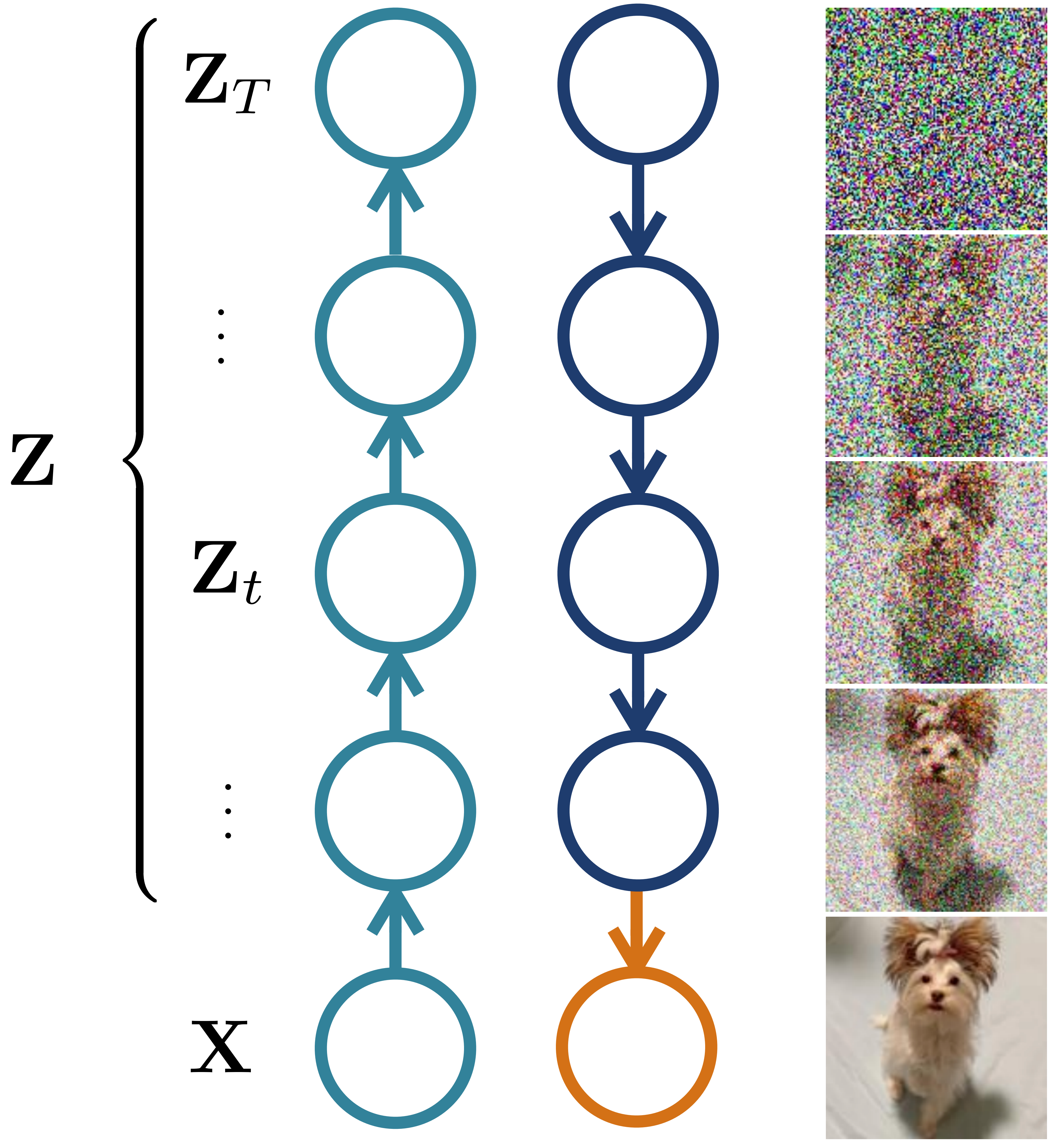


Diffusion as a VAE



Diffusion as a VAE

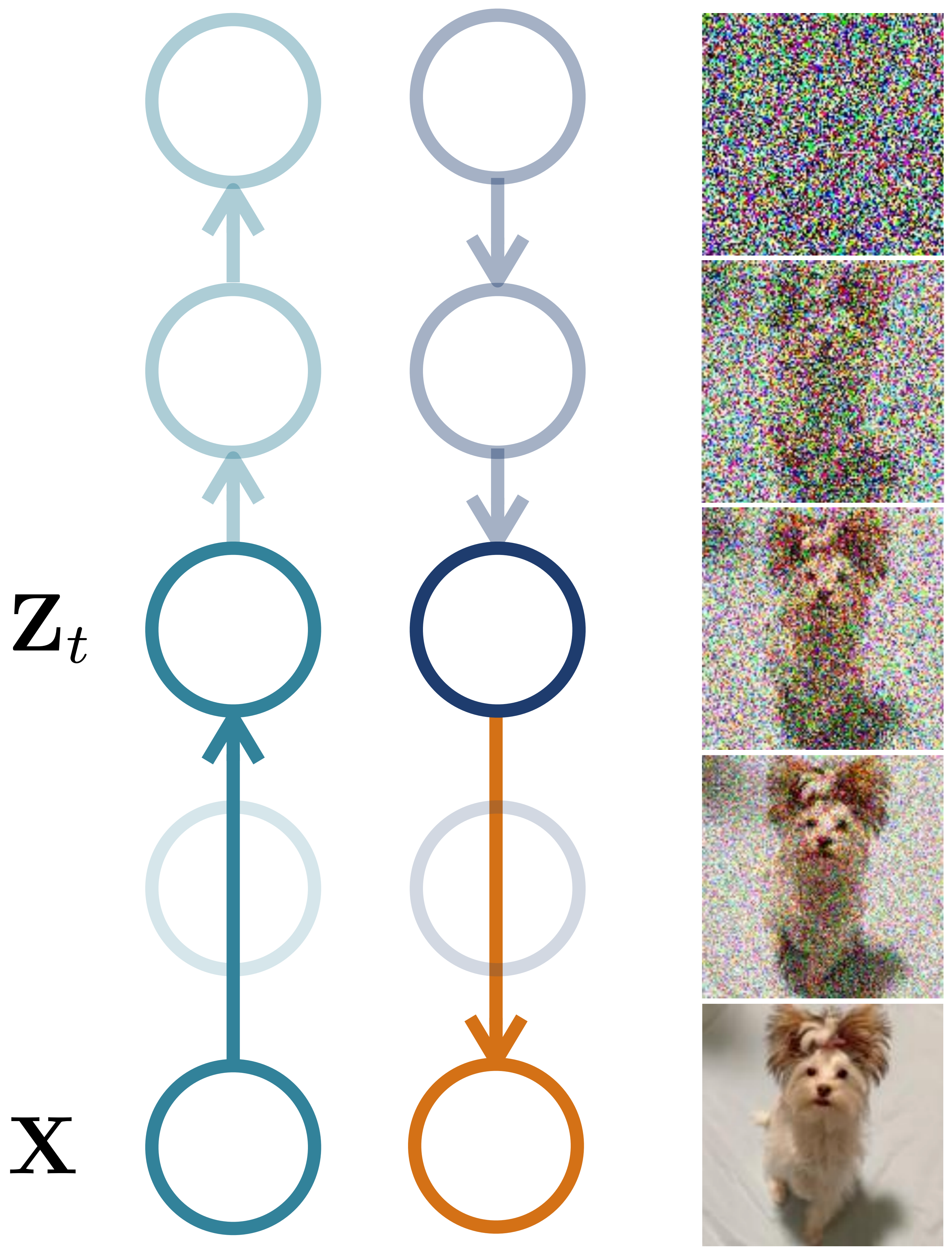
$q(\mathbf{z} \mid \mathbf{x})$



$p(\mathbf{z}) = p(\mathbf{z}_T)p(\mathbf{z}_{T-1} \mid \mathbf{z}_T)\cdots$

$p(\mathbf{x} \mid \mathbf{z})$

Diffusion as a VAE

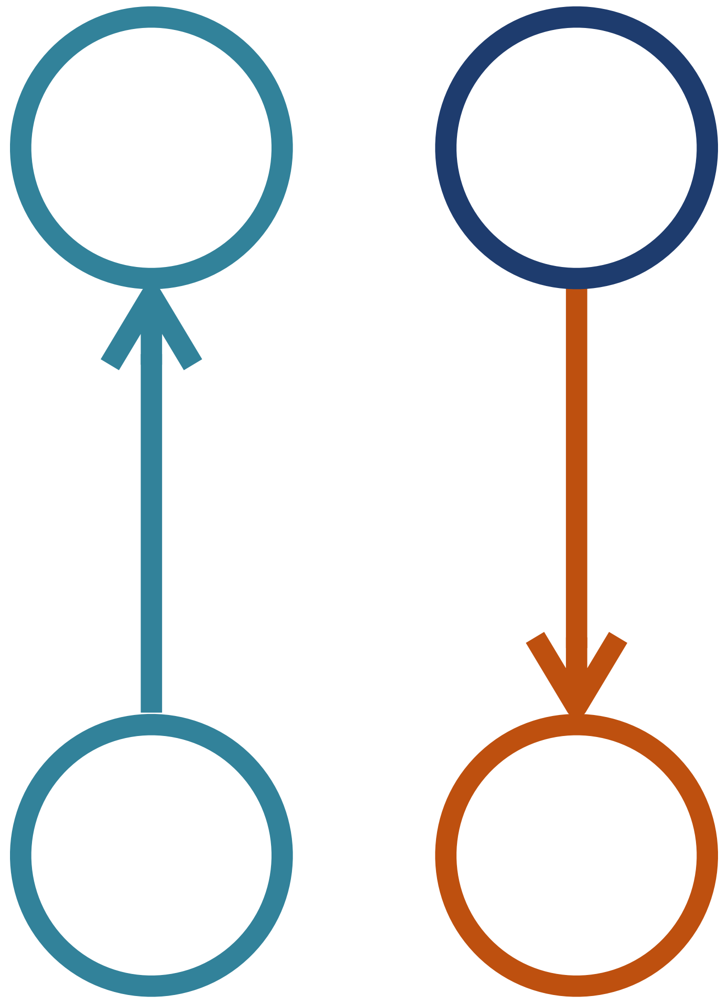


$p(\mathbf{x} \mid \mathbf{z}_t)$

Diffusion as a VAE

Simple

$$q(\mathbf{z}_t \mid \mathbf{x}) = \mathcal{N}(\mathbf{z}_t; \alpha_t \mathbf{x}, \sigma_t^2 \mathbf{I}_m)$$

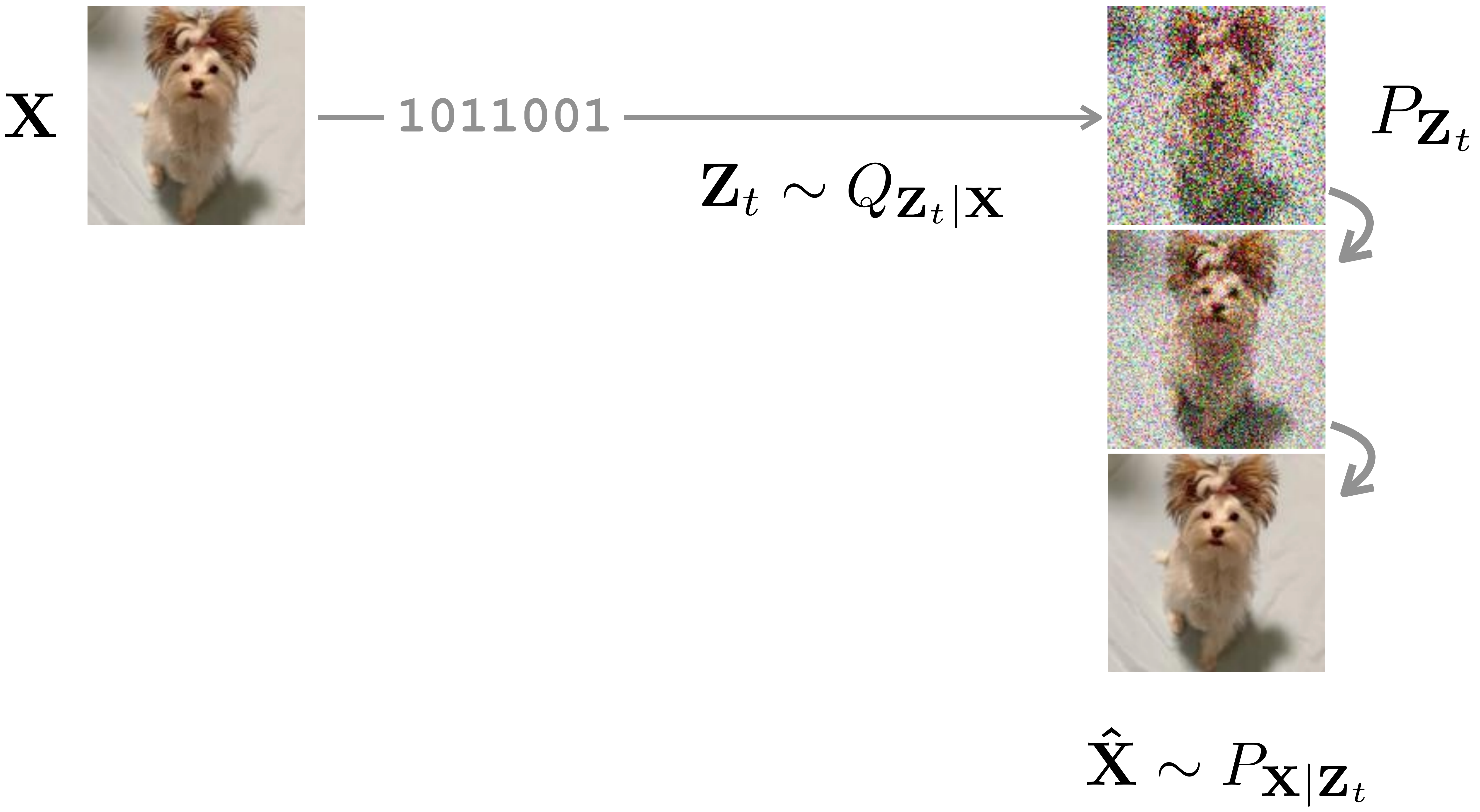


$p(\mathbf{z}_t)$

Complicated

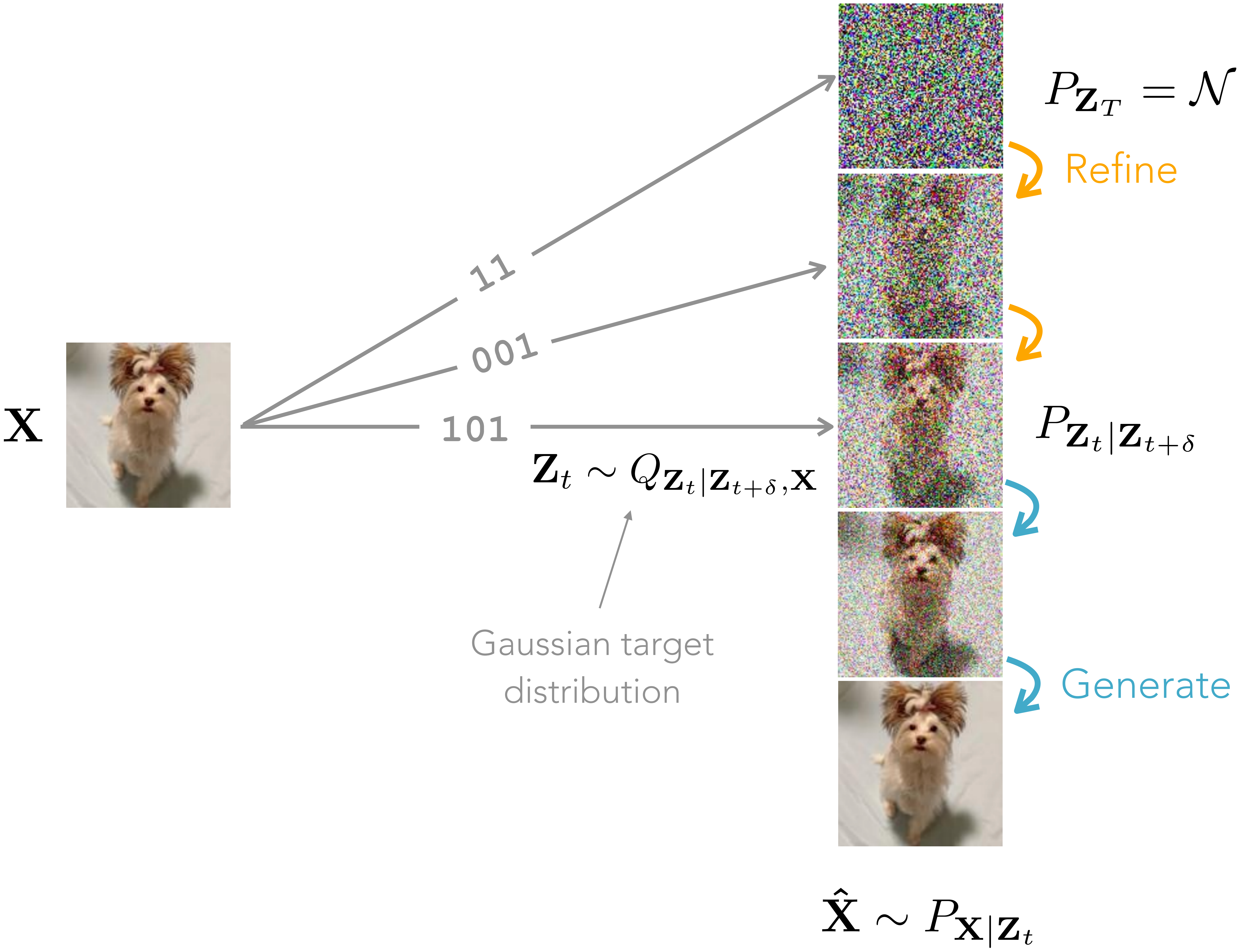
$p(\mathbf{x} \mid \mathbf{z}_t)$

Diffusion



$\mathbf{X}$
1011001
$\mathbf{Z}_t \sim Q_{\mathbf{Z}_t|\mathbf{X}}$
$P_{\mathbf{Z}_t}$
$\hat{\mathbf{X}} \sim P_{\mathbf{X}|\mathbf{Z}_t}$

Diffusion



$\mathbf{X}$
11
001
101
$\mathbf{Z}_t \sim Q_{\mathbf{Z}_t|\mathbf{Z}_{t+\delta},\mathbf{X}}$
Gaussian target distribution
$P_{\mathbf{Z}_T} = \mathcal{N}$
Refine
$P_{\mathbf{Z}_t|\mathbf{Z}_{t+\delta}}$
Generate
$\hat{\mathbf{X}} \sim P_{\mathbf{X}|\mathbf{Z}_t}$

Example: Normal

$$X \sim \mathcal{N}(0,1)$$

$$Z_t = \sqrt{1-\sigma^2}X + \sigma U \qquad U \sim \mathcal{N}(0,1) \qquad (0 < \sigma \leq 1)$$

$$\hat{X} \sim P_{X|Z_t}$$

Zhang et al. (2021); Wagner (2022); Theis et al. (2022)

Example: Normal

Already follows correct distribution

$$X \sim \mathcal{N}(0,1)$$

$$Z_t = \sqrt{1-\sigma^2}X + \sigma U \qquad U \sim \mathcal{N}(0,1) \qquad (0 < \sigma \leq 1)$$

$$\hat{X} = \sqrt{1-\sigma^2}Z_t + \sigma V \qquad V \sim \mathcal{N}(0,1)$$

Realism: $D[P_X, P_{\hat{X}}] = 0$

Distortion: $\mathbb{E}[(X - \hat{X})^2] = 2\sigma^2 = d$

RD function of standard normal

Rate: $I[X, Z_t] = -\log_2 \sigma = \frac{1}{2}\log\frac{2}{d} = R(d/2)$

Zhang et al. (2021); Wagner (2022); Theis et al. (2022)

Example: Normal

$$X \sim \mathcal{N}(0,1)$$

$$Z_t = \sqrt{1-\sigma^2} X + \sigma U \qquad U \sim \mathcal{N}(0,1) \qquad (0 < \sigma \leq 1)$$

$$\hat{X} = Z_t$$

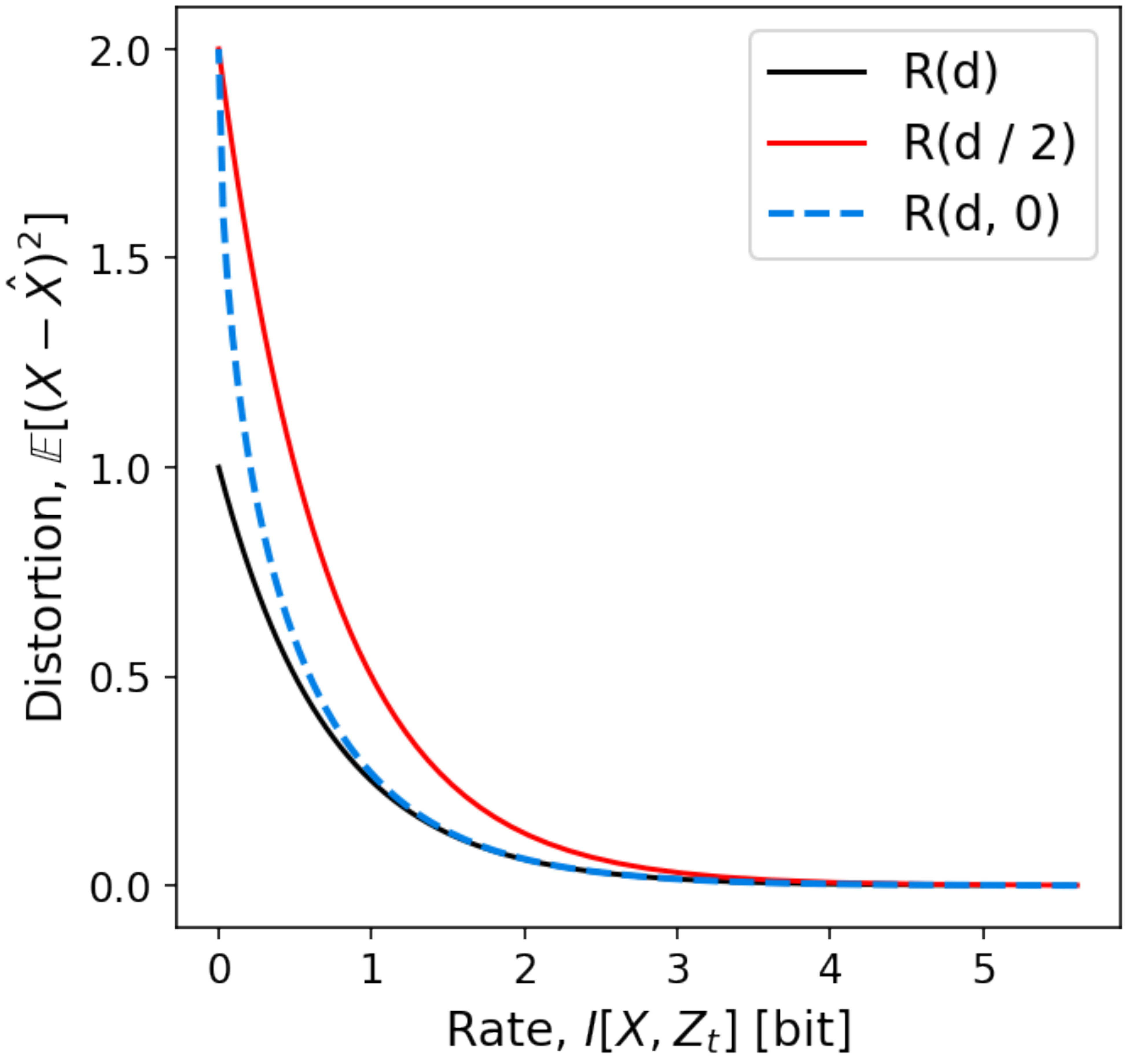
Realism: $D[P_X, P_{\hat{X}}] = 0$

Distortion: $\mathbb{E}[(X - \hat{X})^2] = 2 - 2\sqrt{1-\sigma^2} < 2\sigma^2$

Rate: $I[X, X_t] = -\log_2 \sigma = \frac{1}{2} \log \frac{2}{d}$

Zhang et al. (2021); Wagner (2022); Theis et al. (2022)

Example: Normal



Zhang et al. (2021); Wagner (2022); Theis et al. (2022)

SDE / ODE

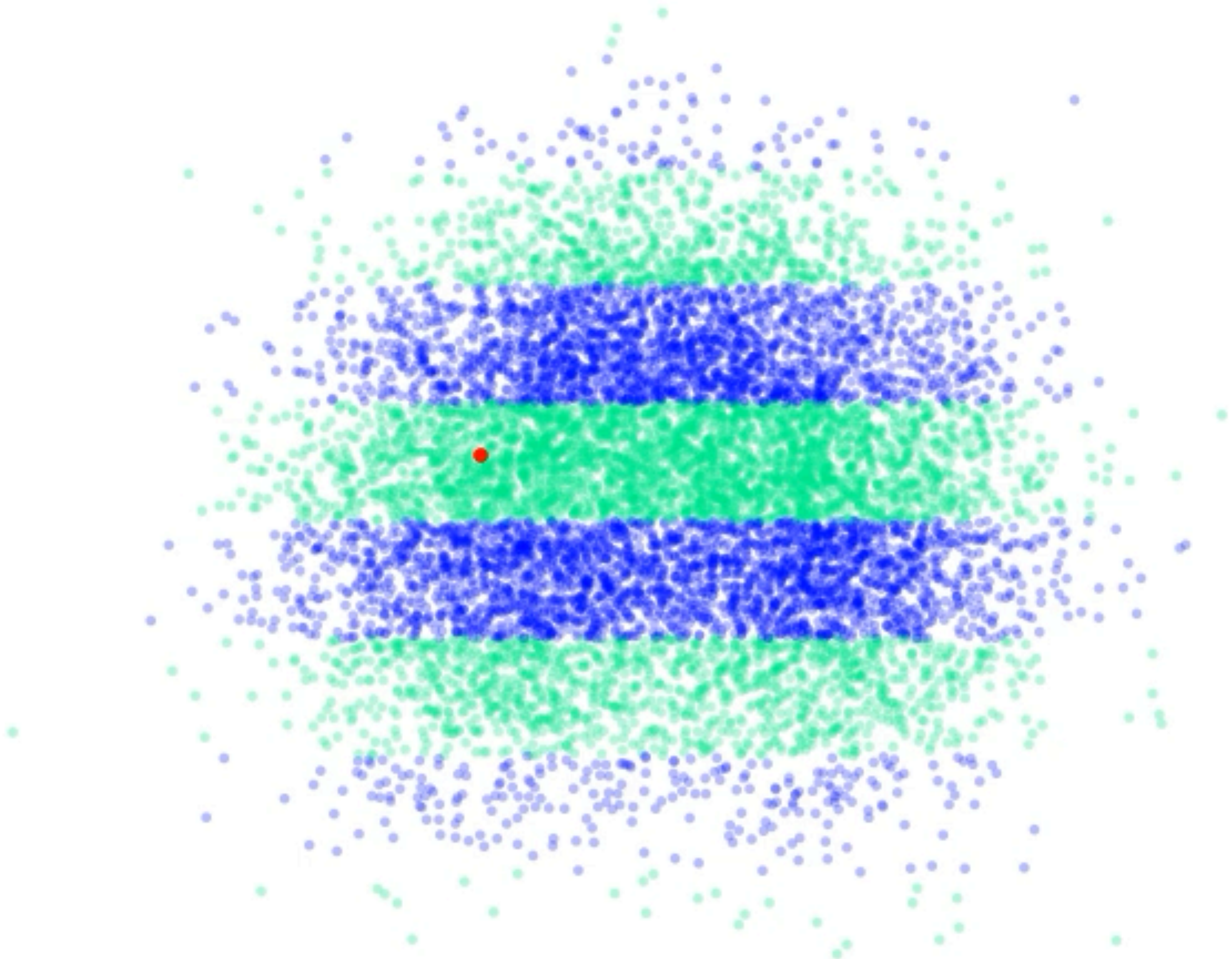
Neural network goes in here

SDE:

$$d\mathbf{Z}_t = \left(-\frac{1}{2}\beta_t \mathbf{Z}_t - \beta_t \nabla \ln p_t(\mathbf{Z}_t)\right) dt + \sqrt{\beta_t}\, d\bar{\mathbf{W}}_t$$

Anderson (1982); Maoutsa et al. (2020); Song et al. (2021)

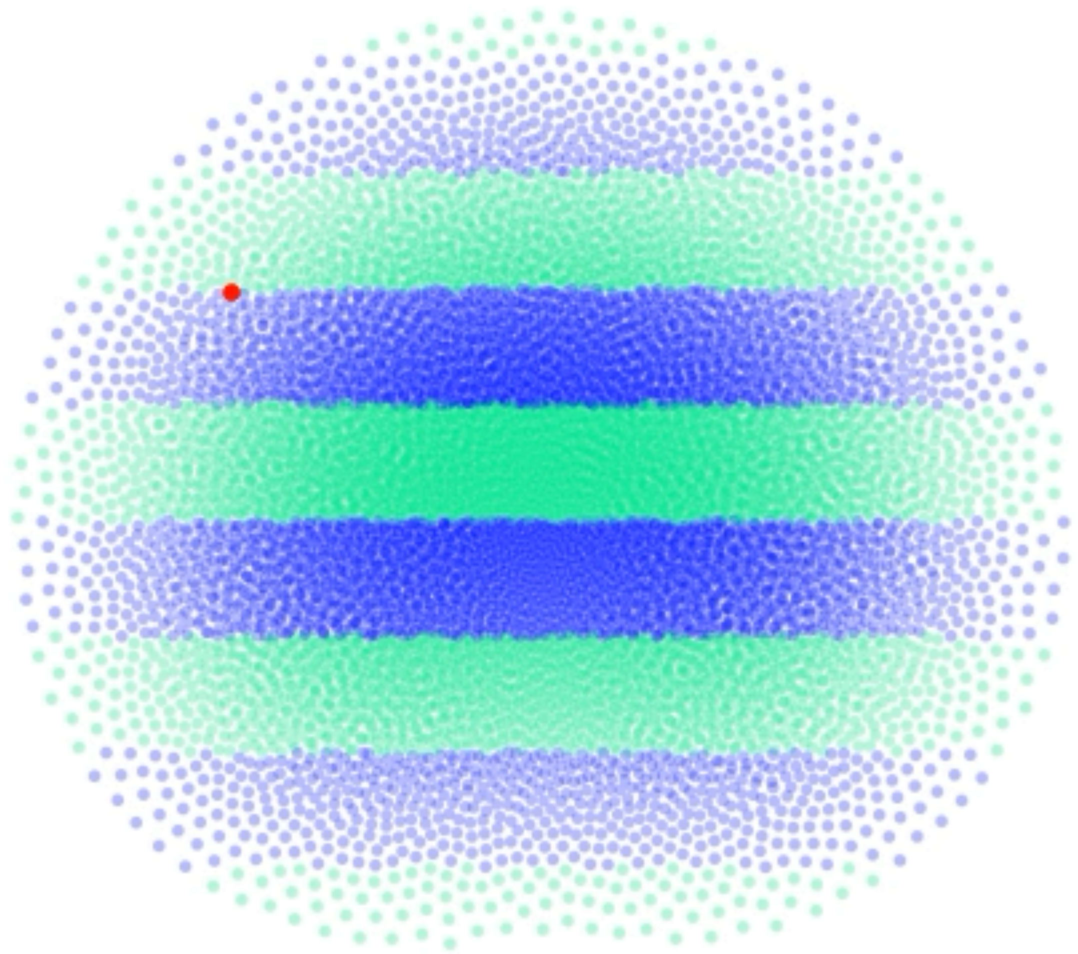
SDE



$$dX_t = \left(-\frac{1}{2}\beta_t X_t - \beta_t \nabla \log p_t(X_t)\right) dt + \sqrt{\beta_t} d\bar{W}_t$$

(simulated in reverse)

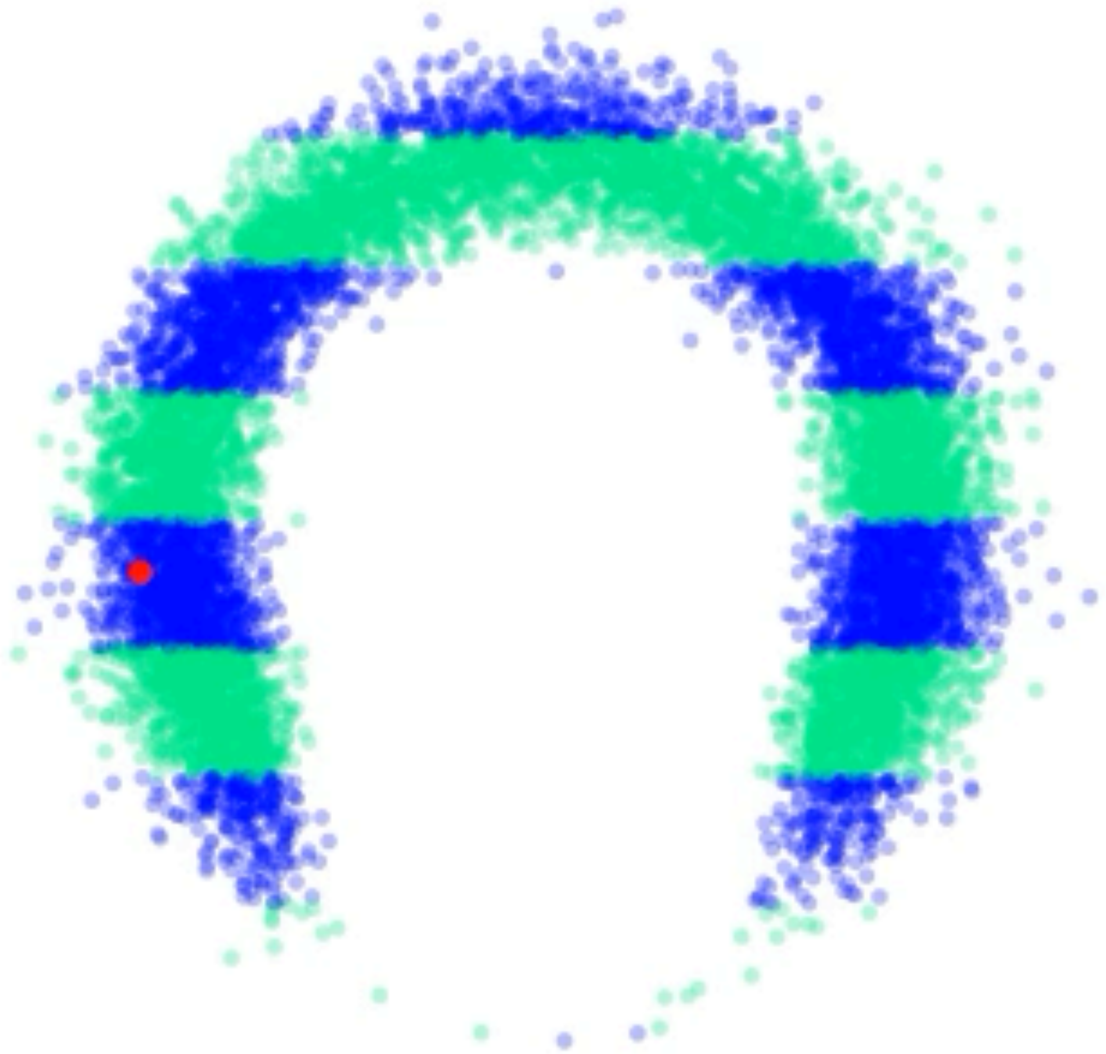
ODE



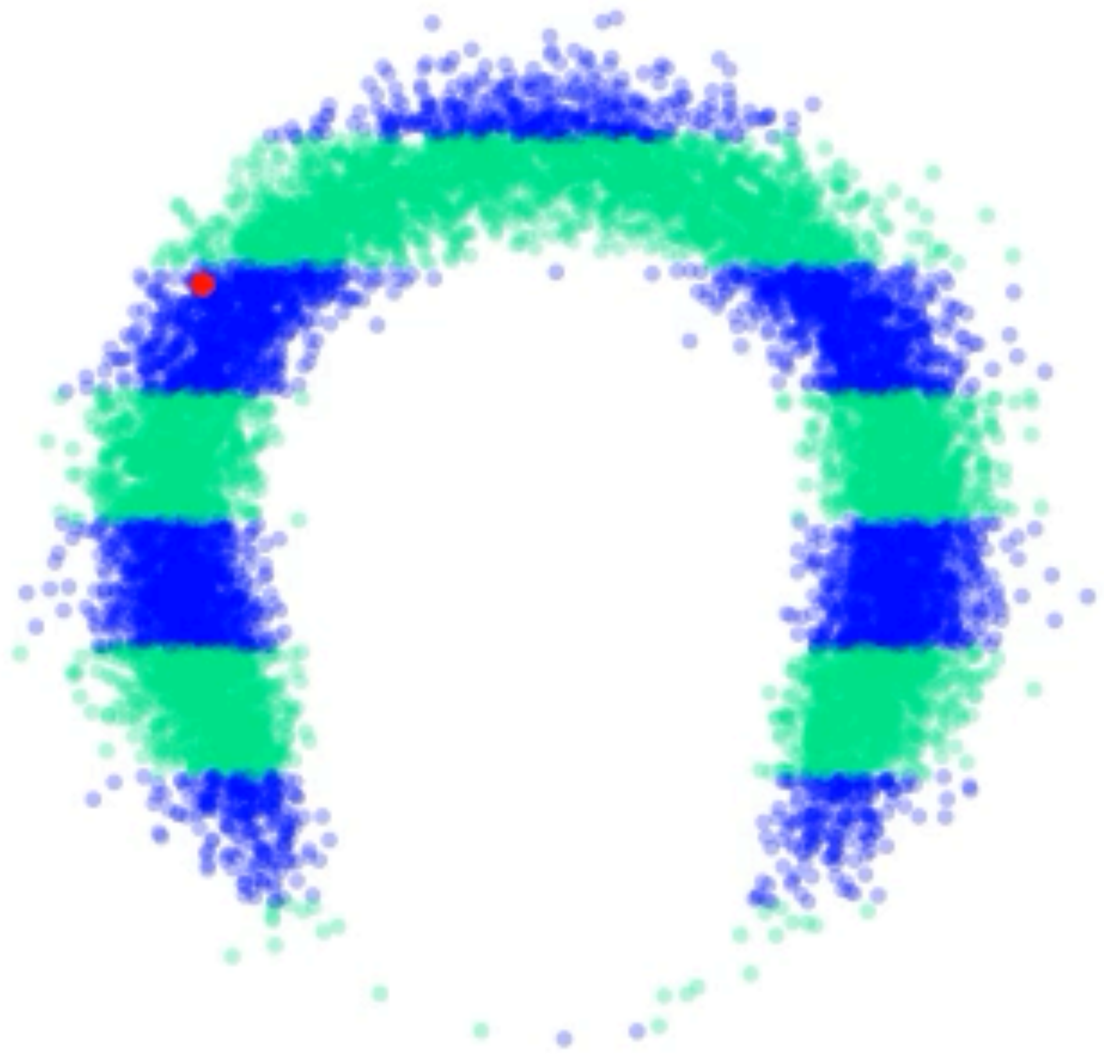
$$d\mathbf{x}_t = \left(-\frac{1}{2}\beta_t \mathbf{x}_t - \frac{1}{2}\beta_t \nabla \log p_t(\mathbf{x}_t)\right) dt$$

(simulated in reverse)

DiffC-A (SDE)



DiffC-F (ODE)



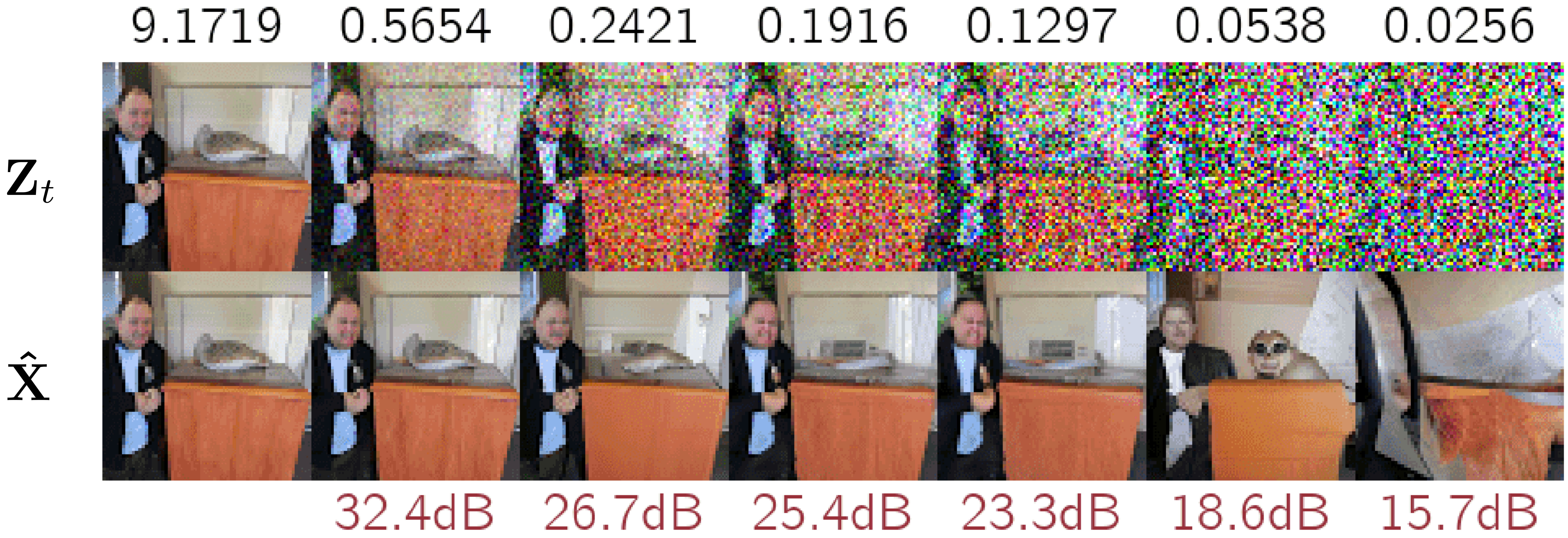
Tweedie's formula

$$\mathbf{Z} = \mathbf{X} + \sigma \mathbf{V} \qquad \mathbf{V} \sim \mathcal{N}(0, \mathbf{I}_m)$$

$$\mathbb{E}[\mathbf{X} \mid \mathbf{z}] = \mathbf{z} + \sigma^2 \nabla \ln p_{\mathbf{Z}}(\mathbf{z})$$

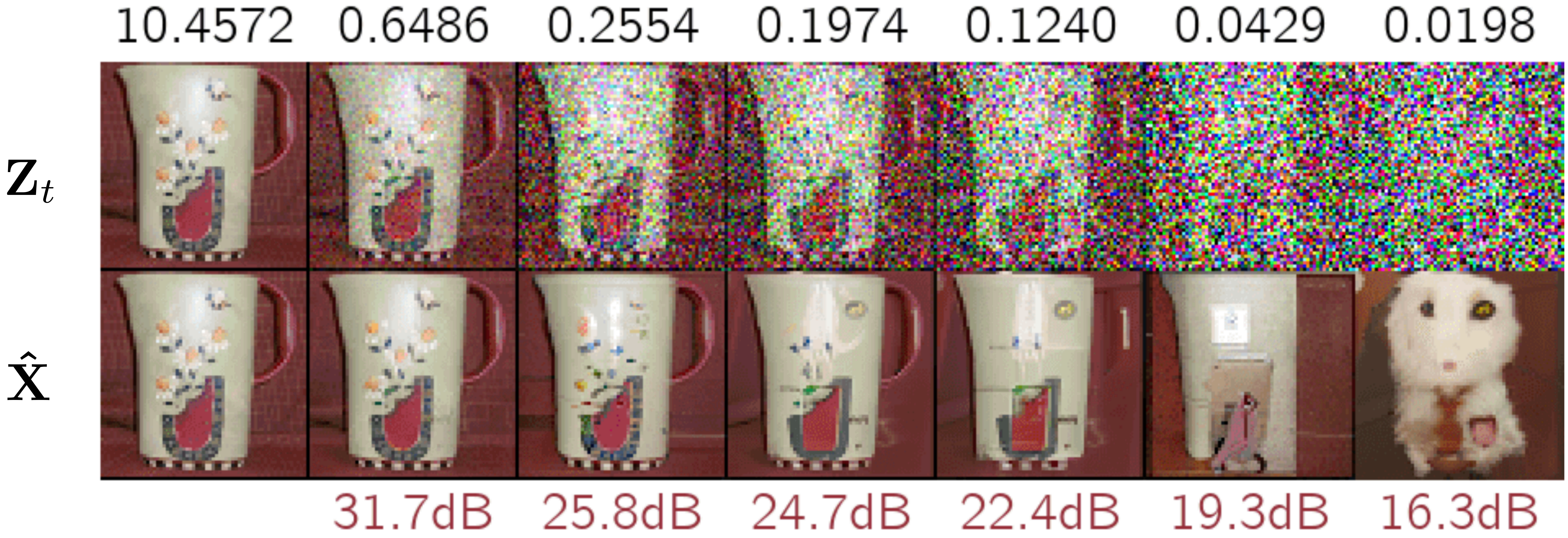
Robbins (1956)

DiffC-F (ODE)



Theis et al. (2022)

DiffC-F (ODE)



Theis et al. (2022)

DiffC-F (ODE) vs DiffC-A (SDE)

Theorem. Let $\mathbf{X} : \Omega \to \mathbb{R}^M$ have a smooth density p with finite

$$G = \mathbb{E}[\|\nabla \ln p(\mathbf{X})\|^2].$$

Let $\mathbf{Z}_t = \sqrt{1 - \sigma_t^2}\mathbf{X} + \sigma_t \mathbf{U}$ with $\mathbf{U} \sim \mathcal{N}(0, \mathbf{I})$. Let $\hat{\mathbf{X}}_A \sim P(\mathbf{X} \mid \mathbf{Z}_t)$ and let $\hat{\mathbf{X}}_F = \mathbf{Z}_0$ be the solution to the ODE with $\mathbf{Z}_t$ as initial condition. Then

$$\lim_{\sigma_t \to 0} \frac{\mathbb{E}[\|\hat{\mathbf{X}}_F - \mathbf{X}\|^2]}{\mathbb{E}[\|\hat{\mathbf{X}}_A - \mathbf{X}\|^2]} = \frac{1}{2}$$

Theis et al. (2022)

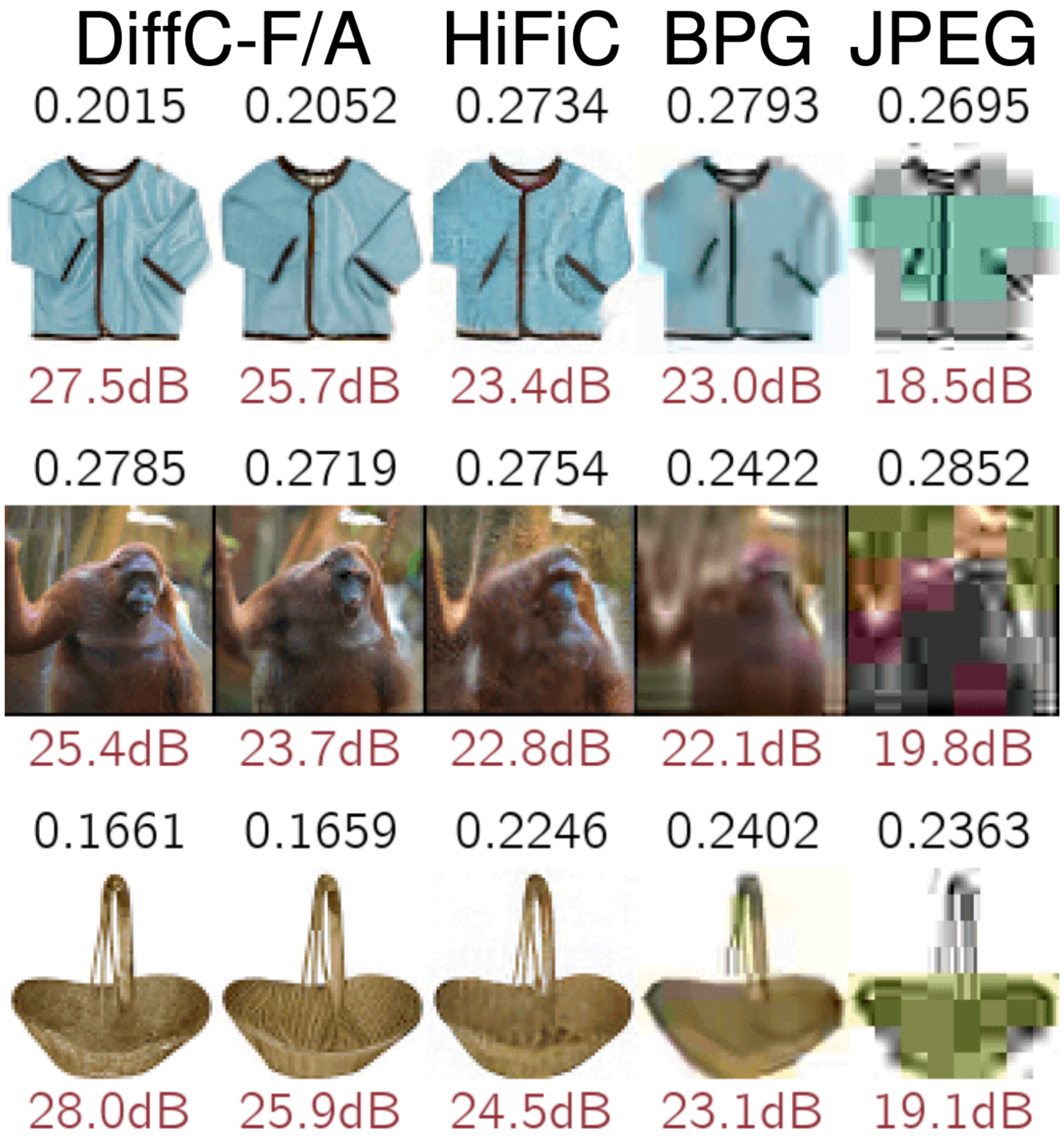
DiffC-F (ODE)

Theorem. Let $\mathbf{X} = \mathbf{QS}$ where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{S} : \Omega \to \mathbb{R}^M$ is a random vector with smooth density and $S_i \perp\!\!\!\perp S_j$ for all $i \neq j$. Define $\mathbf{Z}_t$ as before. If $\hat{\mathbf{X}}_F = \mathbf{Z}_0$ is the solution to the ODE given $\mathbf{Z}_t$ as initial condition, then

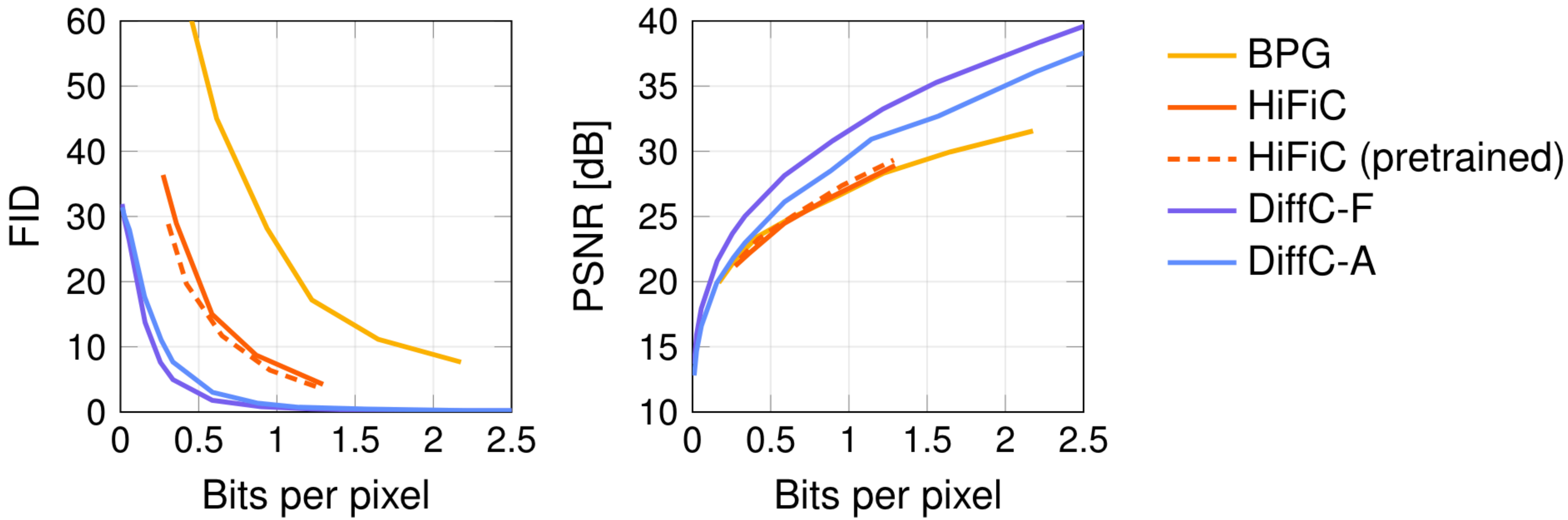
$$\mathbb{E}[\|\hat{\mathbf{X}}_F - \mathbf{X}\|^2] \leq \mathbb{E}[\|\hat{\mathbf{X}}' - \mathbf{X}\|^2]$$

for any $\hat{\mathbf{X}}'$ with $\hat{\mathbf{X}}' \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}_t$ which achieves perfect realism, $\hat{\mathbf{X}}' \sim \mathbf{X}$.

Theis et al. (2022)



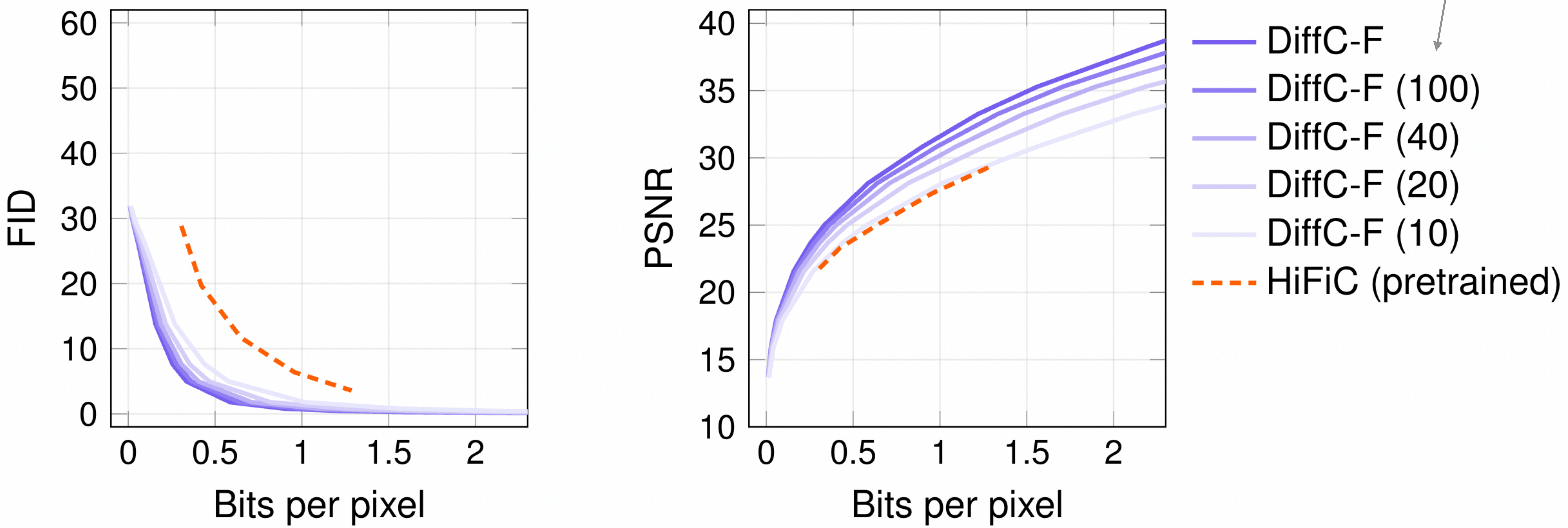
Theis et al. (2022)



Theis et al. (2022)

Refinement

Bits communicated per refinement



Theis et al. (2022)

Example: Multivariate Gaussian

$$\mathbf{X} \sim \mathcal{N}(0, \mathbf{\Sigma})$$

$$\mathbf{Z}_t = \sqrt{1 - \sigma^2}\mathbf{X} + \sigma \mathbf{V} \qquad \mathbf{V} \sim \mathcal{N}(0, \mathbf{I})$$

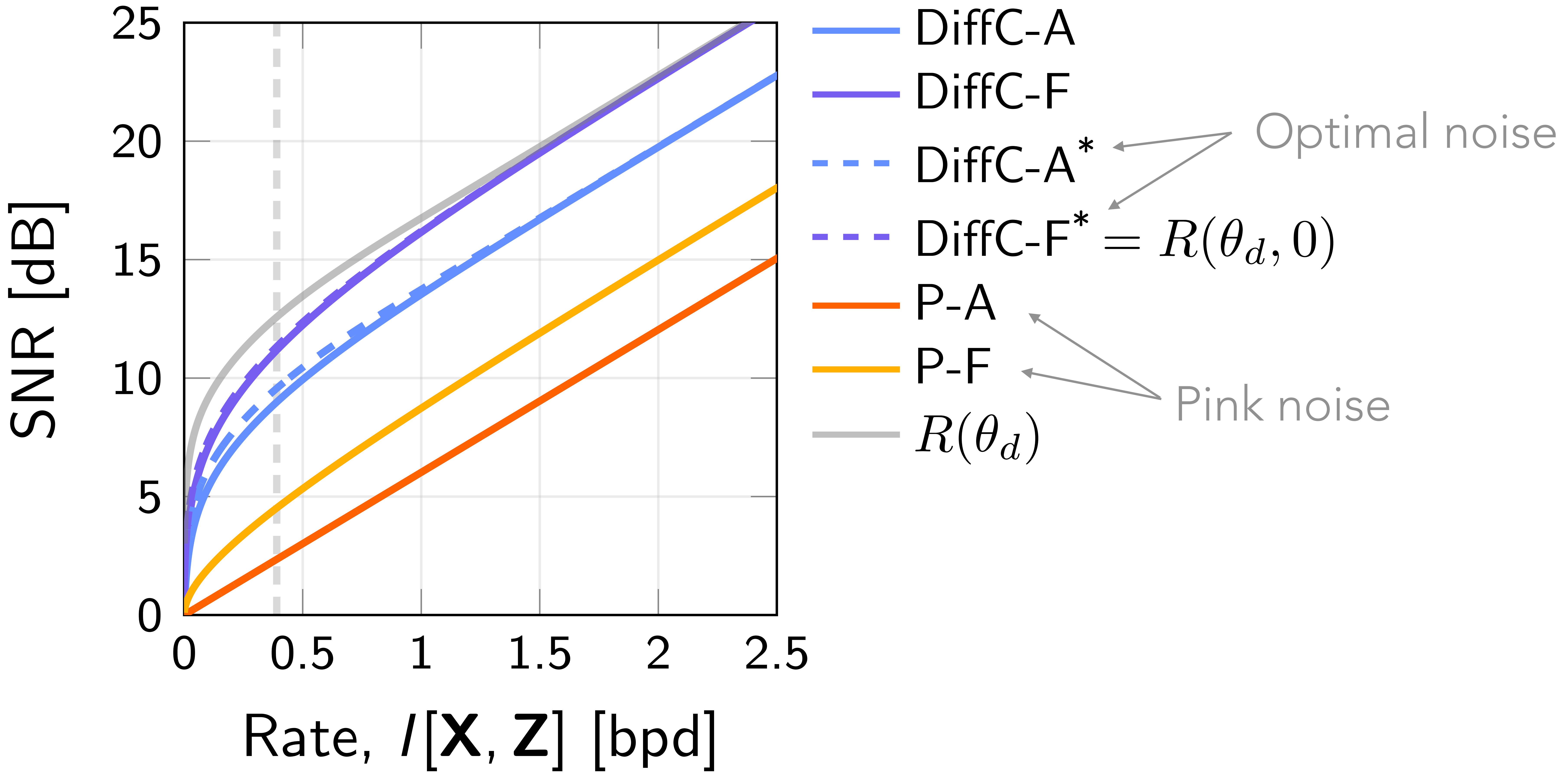
$$\hat{\mathbf{X}} \sim P_{\mathbf{X}|\mathbf{Z}_t}$$

$$I[\mathbf{X}, \mathbf{Z}_t] \geq R(d/2) \qquad d = \mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$$

Worse than 1D case

Theis et al. (2022)

Example: Multivariate Gaussian



Theis et al. (2022)

Open problems

Channel simulation

Computational complexity

- No general algorithm exists with runtime polynomial in D_{KL} (Agustsson & Theis, 2020).
- Existing algorithms' runtime is exponential in D_∞ (infinity divergence), can we find one that's exponential in D_{KL}?
- What is the computational complexity of simulating Gaussian channels? What is the computational complexity of the closest-vector problem (CVP) for optimal lattices?
- Practical implementation of entropy-coded lattice quantization (e.g., for A2 or E8 lattice)?

Coding cost

- Bounds for the one-shot coding cost of simulating multivariate Gaussian channels

Open problems

Learned compression

DiffC

- Can we bound the rate-distortion-perception function of DiffC-A? (For DiffC-A*, see Theis et al., 2022).
- How can we make DiffC computationally more efficient?
- What about other types of noise? Dithered quantization?

Generative compression

- Training in two stages *vs* end-to-end training
- Better approaches targeting $D(\theta_d, \theta_D)$ where $0 < \theta_D < \infty$

Open problems

Realism

Shared randomness

- The role of shared randomness (e.g., Wagner, 2022; Chen et al., 2022)

Universal critics

- Adversarial losses can be challenging to optimize. Can we construct a "universal critic" (or a distortion over sequences) which encourages realistic reconstructions when optimized?

  (*Yes* in theory, but how do we make it practical?)

files.theis.io/nasit2023_solutions.pdf

References

Y. Yang, S. Mandt, and L. Theis, *An Introduction to Neural Data Compression*, 2023

D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, *ELIC: Efficient Learned Image Compression with Unevenly Grouped Space-Channel Contextual Adaptive Coding,* 2022

E. Agustsson, D. Minnen, G. Toderici, and F. Mentzer, *Multi-Realism Image Compression with a Conditional Generator*, 2023

Y. Blau and T. Michaeli, *The Perception-Distortion Tradeoff*, 2018

Y. Blau and T. Michaeli, *Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff*, 2019

L. Theis and A. B. Wagner, *A coding theorem for the rate-distortion-perception function*, 2021

C. Shannon, *A Mathematical Theory of Communication*, 1948

S. Nowozin, B. Cseke, and R.Tomioka, *f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization,* 2016

References

C. Ledig *et al.*, *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*, 2017

J. Sohl-Dickstein *et al.*, *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*, 2015

Y. Song *et al.*, *Score-Based Generative Modeling through Stochastic Differential Equations*, 2021

H. Robbins, *An empirical Bayes approach to statistics*, 1956

W. Feller, *On the theory of stochastic processes, with particular reference to applications*, 1949

B. D. O. Anderson, *Reverse-time diffusion equation models*, 1982

L. G. Roberts, *Picture Coding Using Pseudo-Random Noise*, 1962

J. Ballé et al., *Nonlinear transform coding*, 2021

Z. Qin, P. Damien, and S. Walker, *Uniform Scale Mixture Models With Applications to Bayesian Inference*, 2003

C. T. Li and A. El Gamal, *Strong Functional Representation Lemma and Applications to Coding Theorem*, 2018

References

E. Hoogeboom et al., *High-Fidelity Image Compression with Score-based Generative Models*, 2023

D. Maoutsa et al., *Interacting particle solutions of Fokker-Planck equations through gradient-log-density estimation,* 2020

L. Theis et al., *Lossy Compression with Gaussian Diffusion*, 2022

G. Zhang et al., *Universal rate-distortion-perception representations for lossy compression*, 2021

A. Wagner, *The Rate-Distortion-Perception Tradeoff: The Role of Common Randomness*, 2022

R. Rombach et al., *High-resolution image synthesis with latent diffusion models*, 2022

P. Harsha et al., *The Communication Complexity of Correlation*, 2007

C. H. Bennett and P. W. Shor, *Entanglement-Assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem*, 2002

C. T. Li and A. El Gamal, *Strong Functional Representation Lemma and Applications to Coding Theorems*, 2018

References

D. Minnen and S. Singh, *Channel-wise autoregressive entropy models for learned image compression*, 2020

J. Ballé et al., *End-to-end Optimized Image Compression*, 2017

L. Theis et al., *Lossy Image Compression with Compressive Autoencoders*, 2017

B. K. Sriperumbudur, *On Integral Probability Metrics, phi-Divergences, and Binary Classification*, 2009

J. Chen et al., *On the Rate-Distortion-Perception Function*, 2022