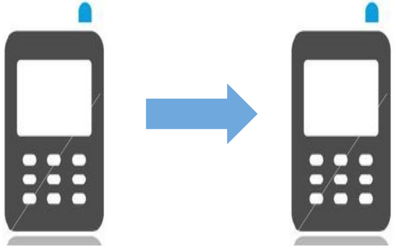
Neural Joint Source-Channel Coding

Kristy Choi, Kedar Tatwawadi, Aditya Grover,
Tsachy Weissman, Stefano Ermon

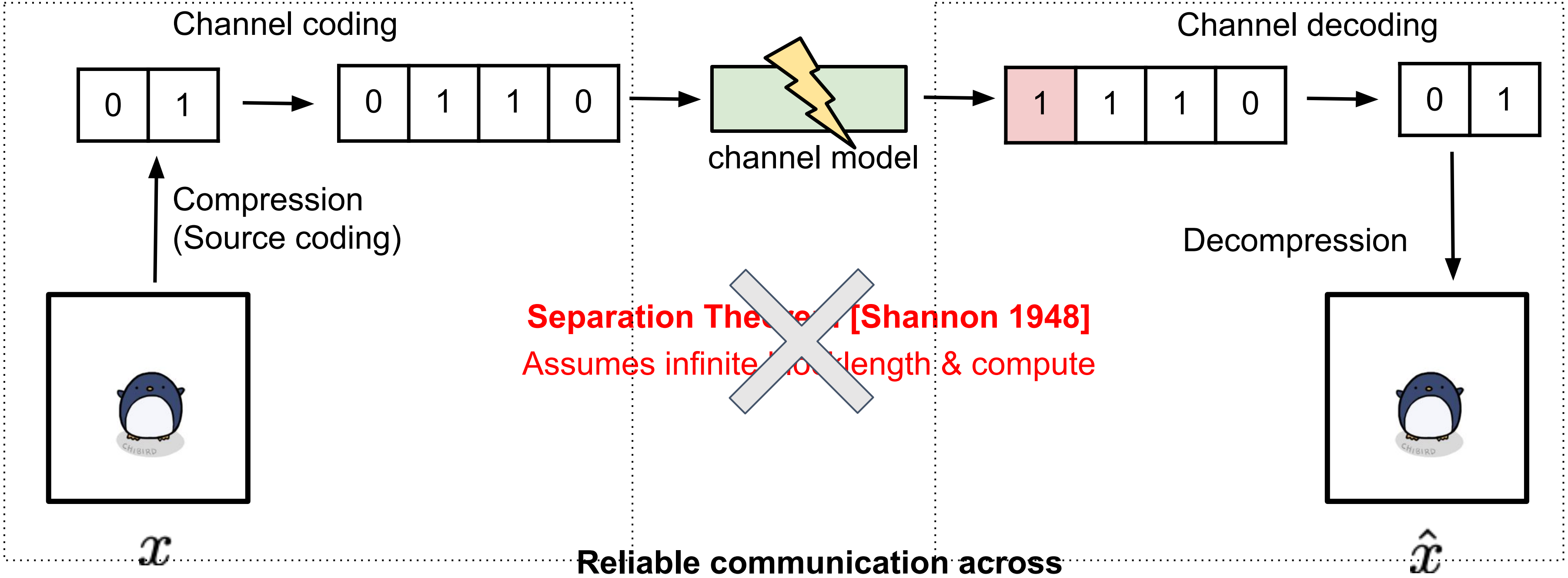
Computer Science Department, Stanford University

Motivation

Reliable, robust, and efficient information transmission is key for everyday communication

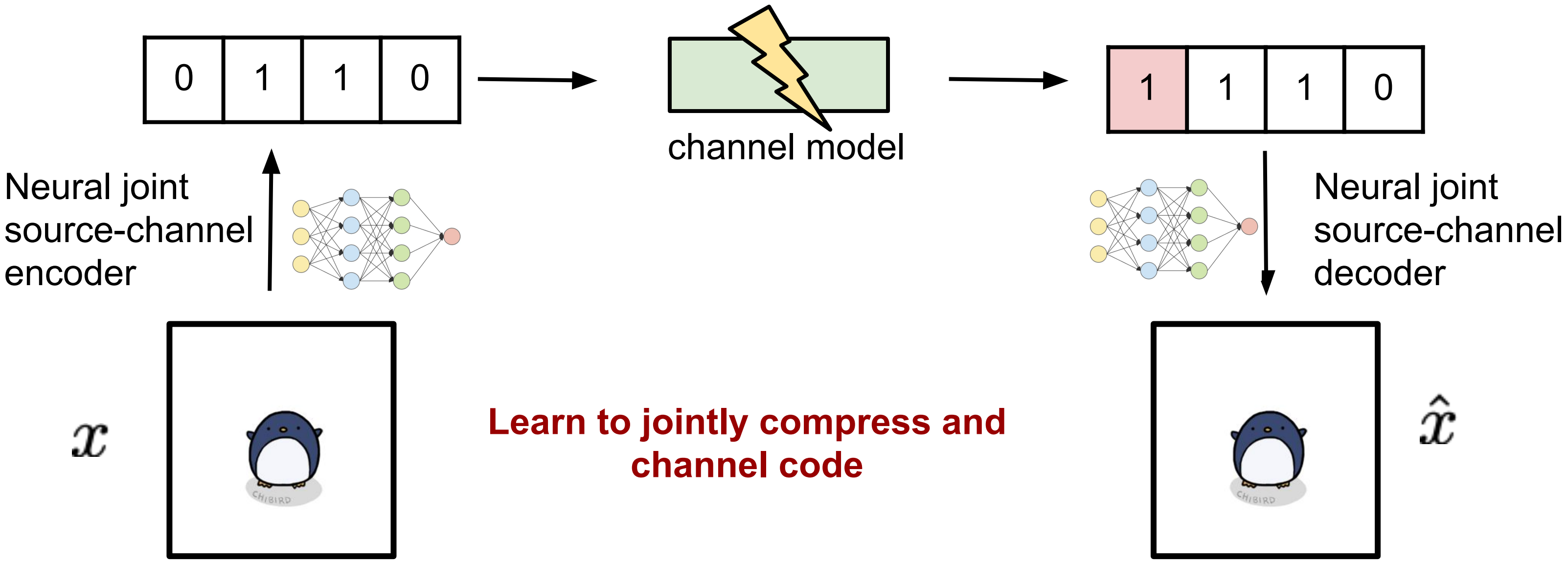


Problem Statement

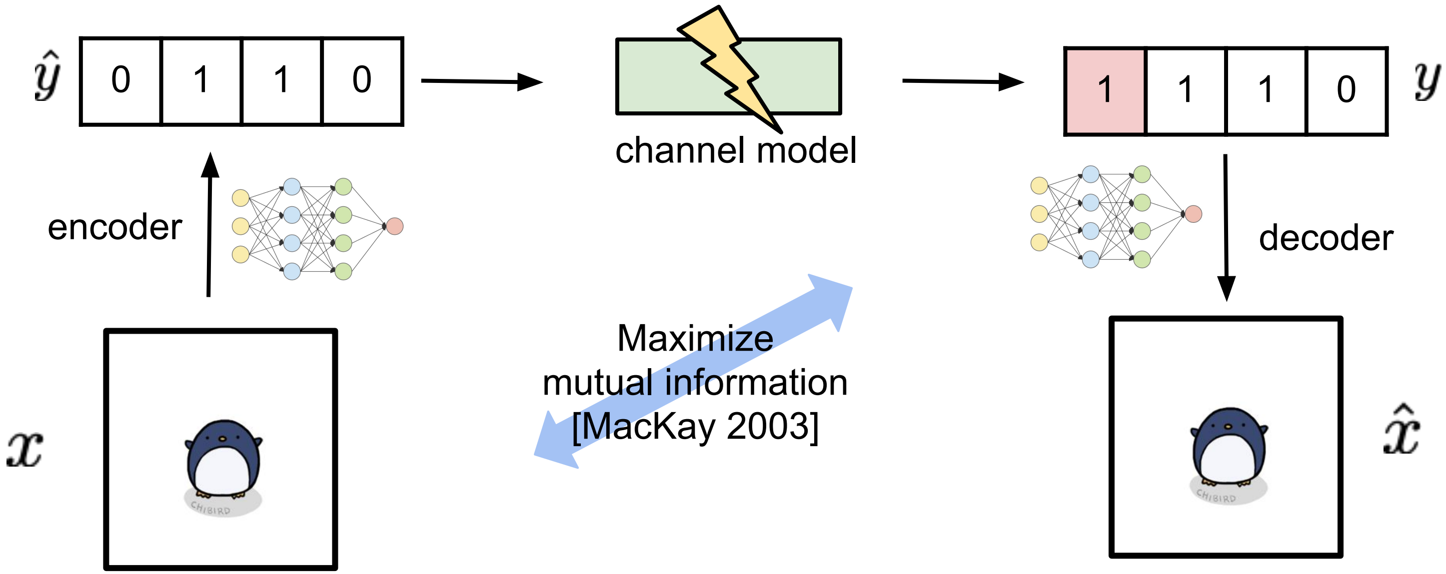


Reliable communication across noisy channel

Neural Joint Source-Channel Coding

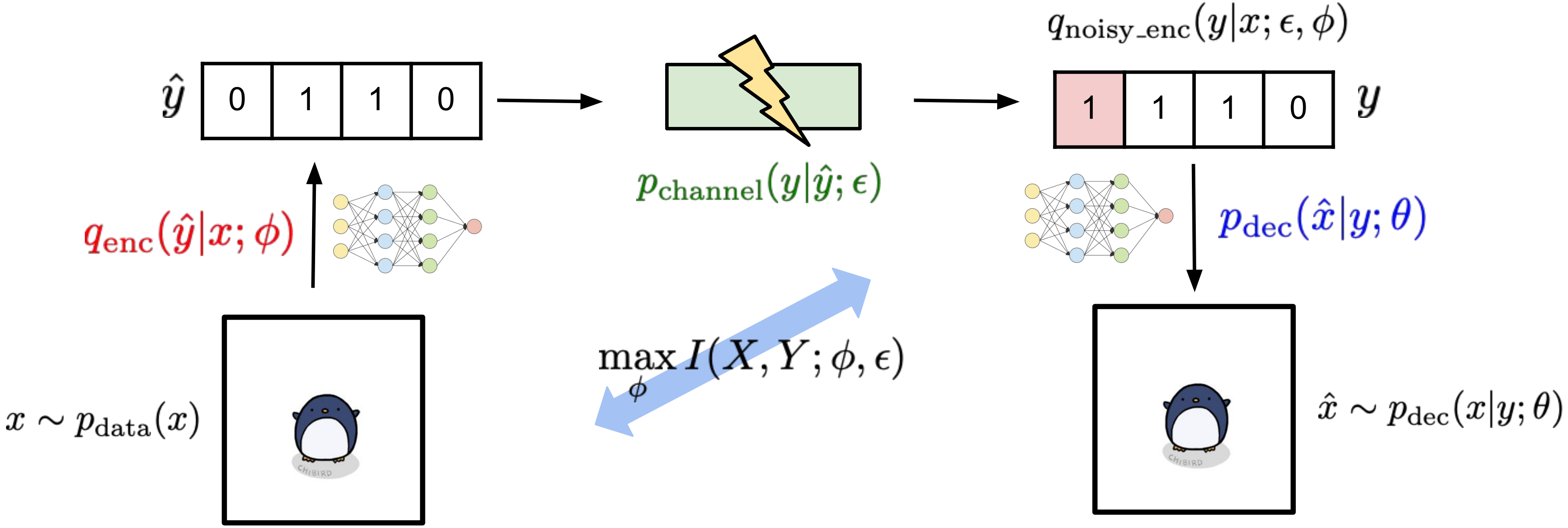


NECST Model



$\hat{y}$
0
1
1
0
channel model
1
1
1
0
y
encoder
decoder
Maximize
mutual information
[MacKay 2003]
x
CHIBIRD
CHIBIRD
$\hat{x}$

Coding Process



$$p(x,\hat{y},y,\hat{x}) = p_{\text{data}}(x)q_{\text{enc}}(\hat{y}|x;\phi)p_{\text{channel}}(y|\hat{y};\epsilon)p_{\text{dec}}(\hat{x}|y;\theta)$$

Learning Objective

- Mutual information maximization
  - *Y* should capture as much information about *X* as possible, *even after corruption!*
  - Estimation is hard ☹️ [Barber & Agakov 2004]
- Variational lower bound is nicer:

$$\max_{\phi} I(X,Y;\phi,\epsilon) = \max_{\phi} H(X) - H(X|Y;\phi,\epsilon)$$

$$\geq \max_{\theta,\phi} \boxed{\mathbb{E}_{x\sim p_{\mathrm{data}}(x)}\mathbb{E}_{y\sim q_{\mathrm{noisy\_enc}}(y|x;\epsilon,\phi)}[\log p_{\mathrm{dec}}(x|y;\theta)]}$$

[Kingma & Welling 2014]
[Vincent 2008]

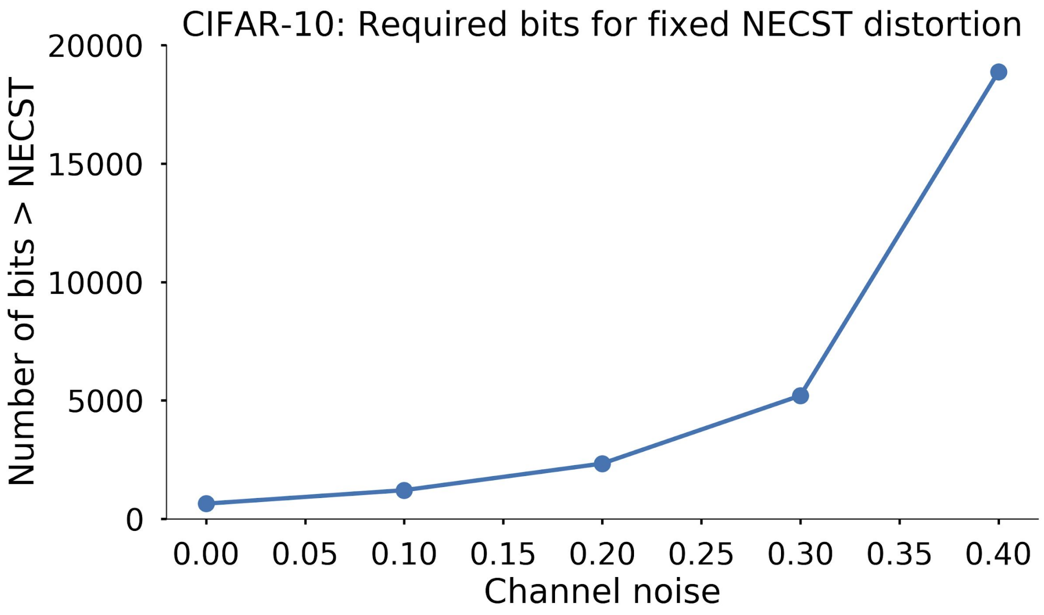
Reconstruction loss!

Optimization Procedure

- Our latent variables *y* are discrete ☹️
- Use VIMCO: [Mnih and Rezende 2016]
  - Draw multiple (*K*) samples from inference network, get tighter lower bound

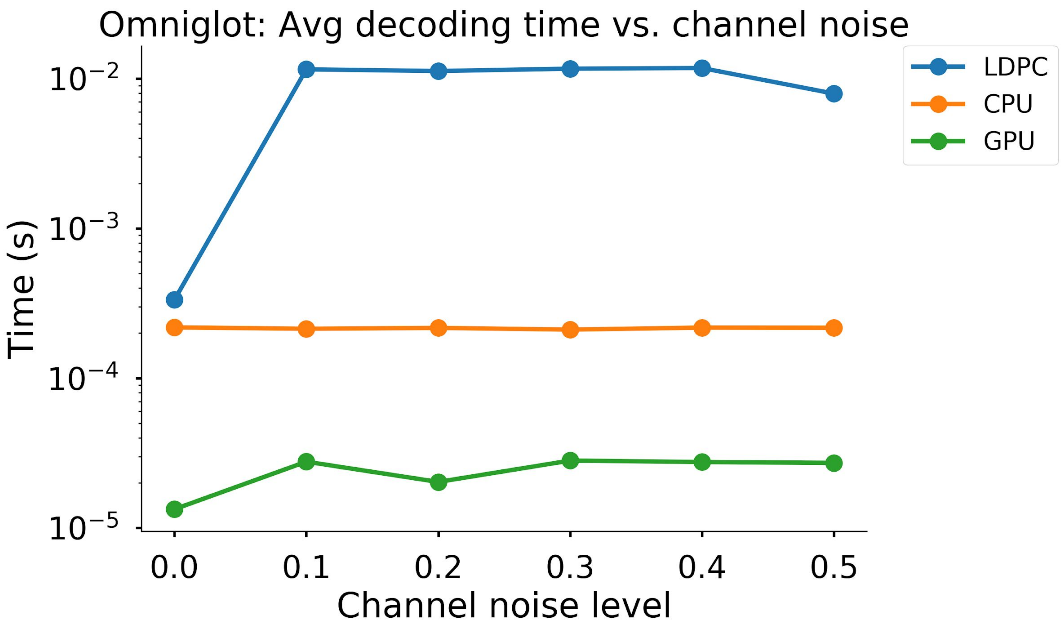
$$\mathcal{L}^K(\phi, \theta; x, \epsilon) = \sum_{x \in \mathcal{D}} \underbrace{\boxed{\mathbb{E}_{y^{1:K} \sim q_{\text{noisy\_enc}}(y|x;\epsilon,\phi)}}}_{\text{Multiple samples of } y} \underbrace{\boxed{\left[\frac{1}{K} \sum_{i=1}^{K} p_{\text{dec}}(x|y^i;\theta) \right]}}_{\text{Multiple reconstruction loss terms}}$$

Fixed Rate: Comparison vs. Ideal Codes



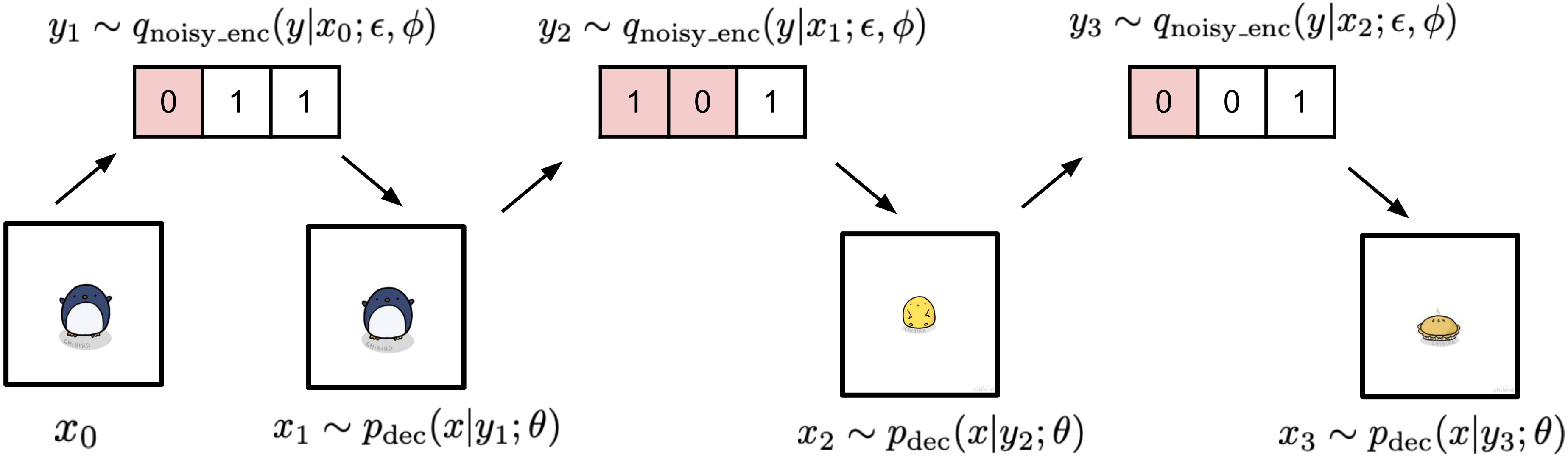
We need *a much smaller number of bits* to get the same level of distortion, even vs. WebP [Google 2010] + ideal channel code

Extremely Fast Decoding



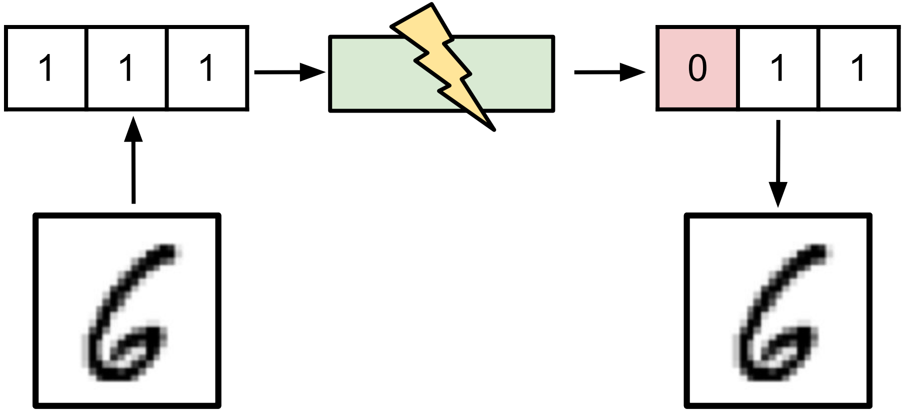
Up to 2x orders of magnitude in speedup on GPU vs. LDPC decoder [Gallager 1963]

Learning the Data Distribution

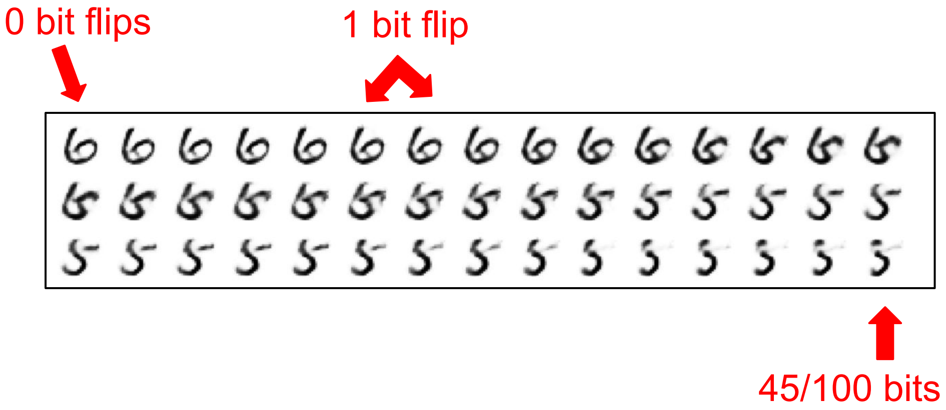


<u>Theorem (informal):</u> NECST learns an implicit model of $p_{\text{data}}(x)$

Robust Representation Learning



1) Encoded redundancies:
interpolation in latent space by bit-flip



2) Improved downstream classification: improves accuracy by as much as 29% across variety of classifiers when inputs are corrupted by noise!

Summary

- End-to-end deep generative modeling framework for the JSCC problem
- Better bitlength efficiency than separation scheme on CIFAR10, CelebA, SVHN
- Another way to learn robust latent representations
- Get an extremely fast decoder for free

Thanks!


Kedar Tatwawadi


Aditya Grover


Tsachy Weissman


Stefano Ermon

Contact: kechoi@stanford.edu

Code: https://github.com/ermongroup/necst

Poster #165: Tuesday, June 11th @ **Pacific Ballroom**