Quantized Compressed Sensing with Score-Based Generative Models

Xiangming Meng and Yoshiyuki Kabashima

Institute for Physics of Intelligence and Department of Physics
The University of Tokyo, Tokyo, Japan

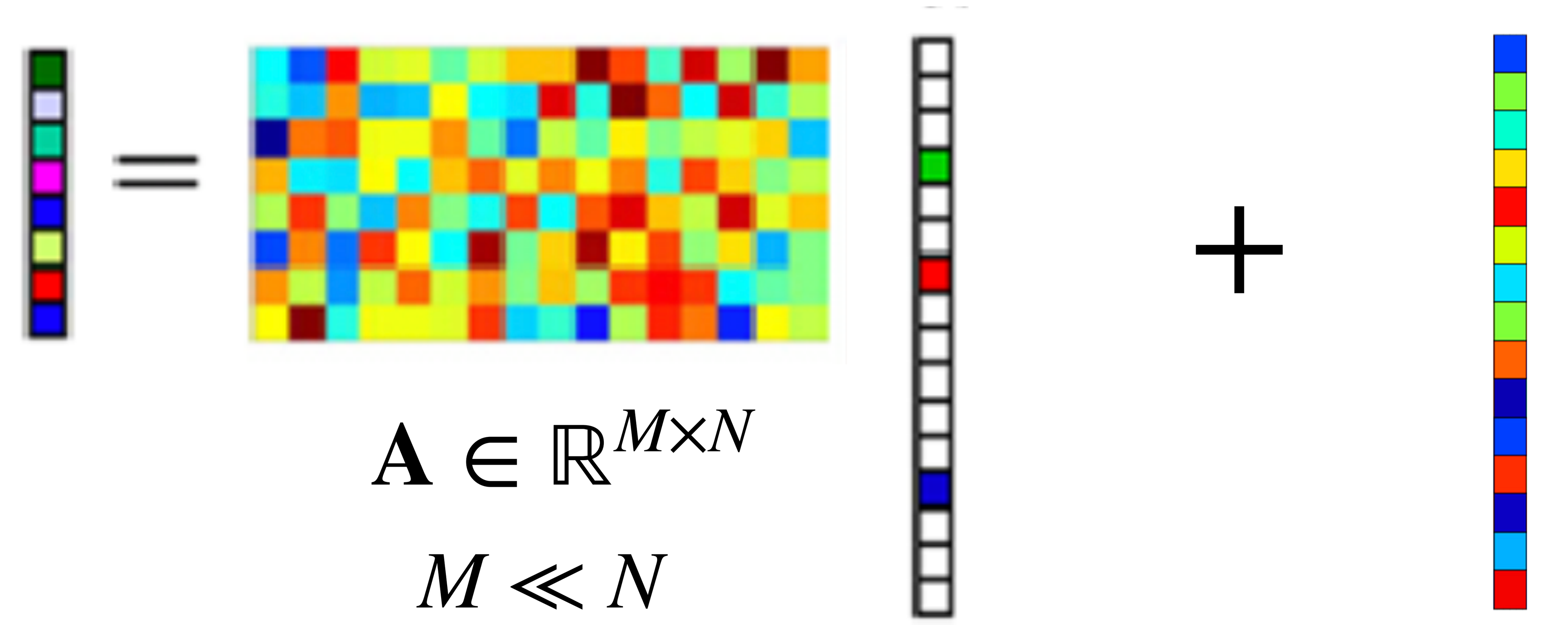
April 9th, 2023

Background and Problem Setup

- **Standard Compressed Sensing (CS)**

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

σ^2



$\mathbf{A} \in \mathbb{R}^{M\times N}$

$M \ll N$

$\mathbf{y} \in \mathbb{R}^{M\times 1}$
Observations

$\mathbf{x} \in \mathbb{R}^{N\times 1}$
High-dimensional
Sparse signal

$\mathbf{n} \in \mathbb{R}^{M\times 1}$
Noise

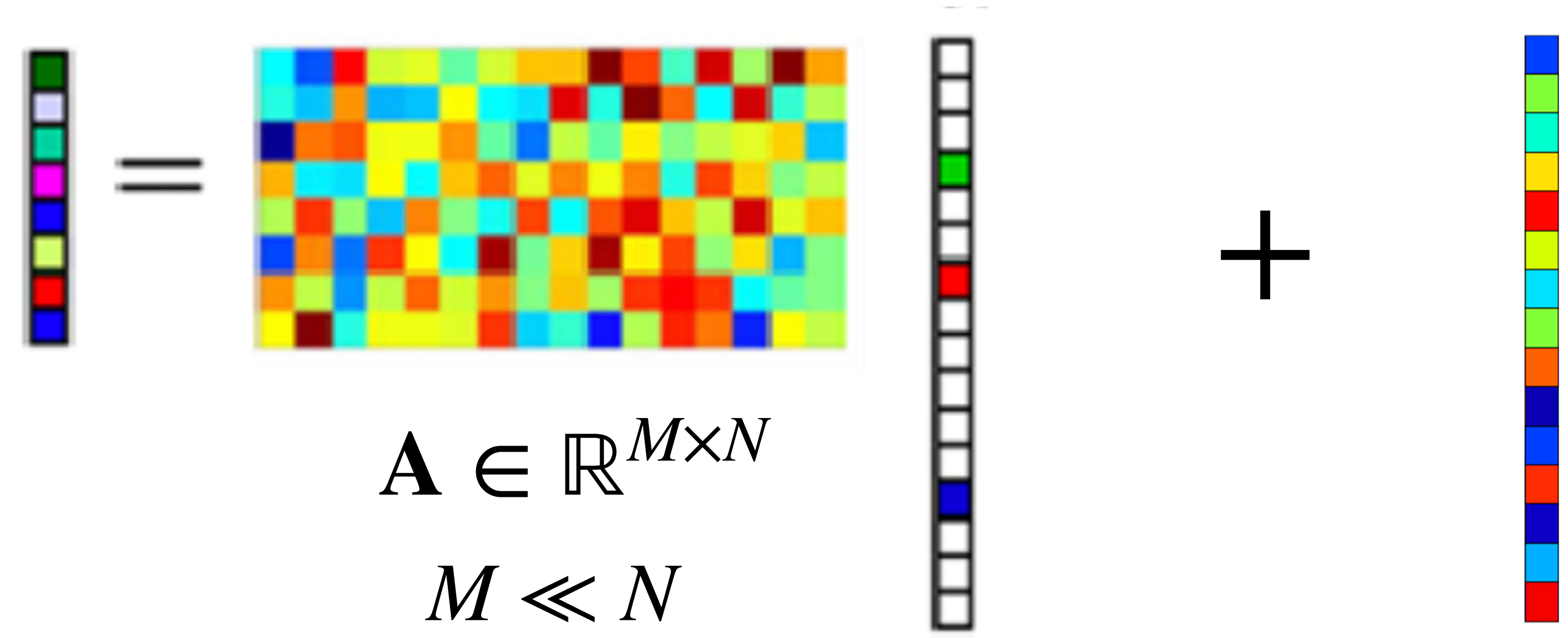
Goal: Recover the high-dimensional signal x from as few observations y as possible

Background and Problem Setup

- **Standard Compressed Sensing (CS)**

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

σ^2



$\mathbf{A} \in \mathbb{R}^{M\times N}$

$M \ll N$

$\mathbf{y} \in \mathbb{R}^{M\times 1}$
Observations

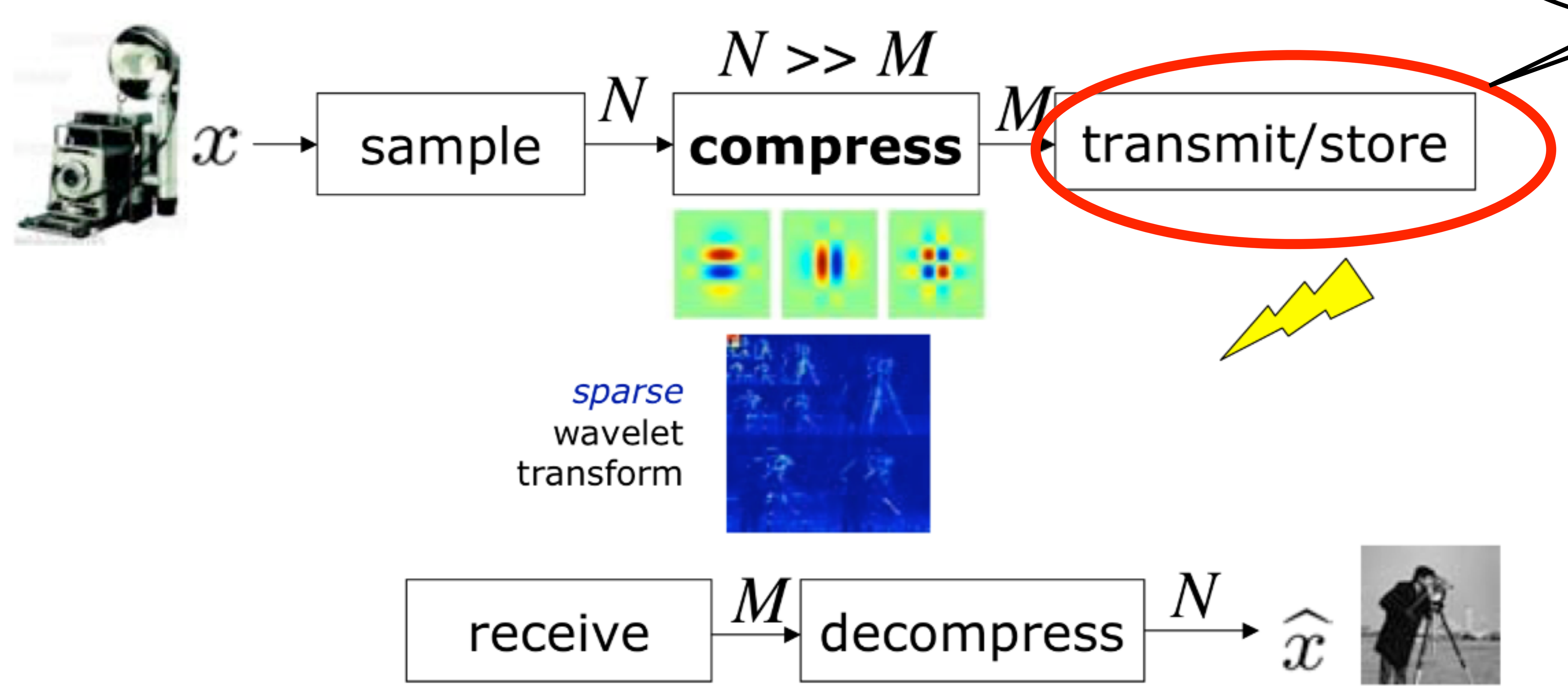
$\mathbf{x} \in \mathbb{R}^{N\times 1}$
High-dimensional
Sparse signal

$\mathbf{n} \in \mathbb{R}^{M\times 1}$
Noise

Goal: Recover the high-dimensional signal x from as few observations y as possible

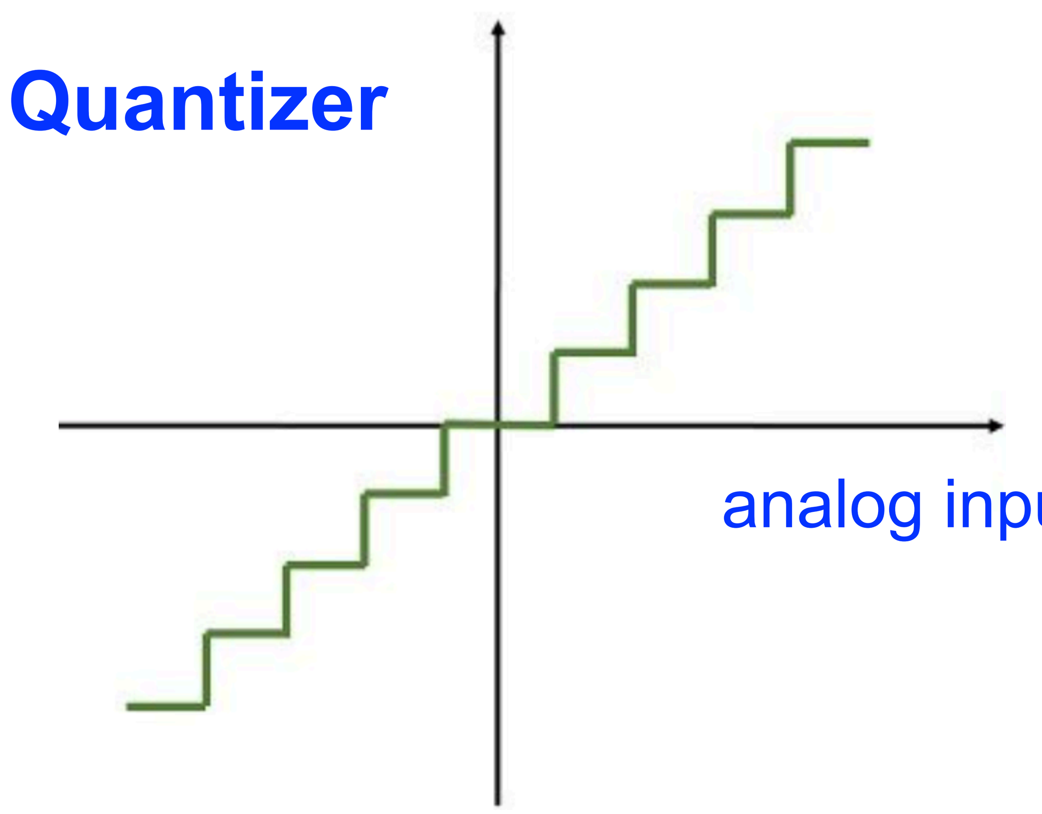
- **Quantized Compressed Sensing (QCS)**

Quantization is essential



$$\mathbf{y} = \mathrm{Q}(\mathbf{A}\mathbf{x} + \mathbf{n})$$

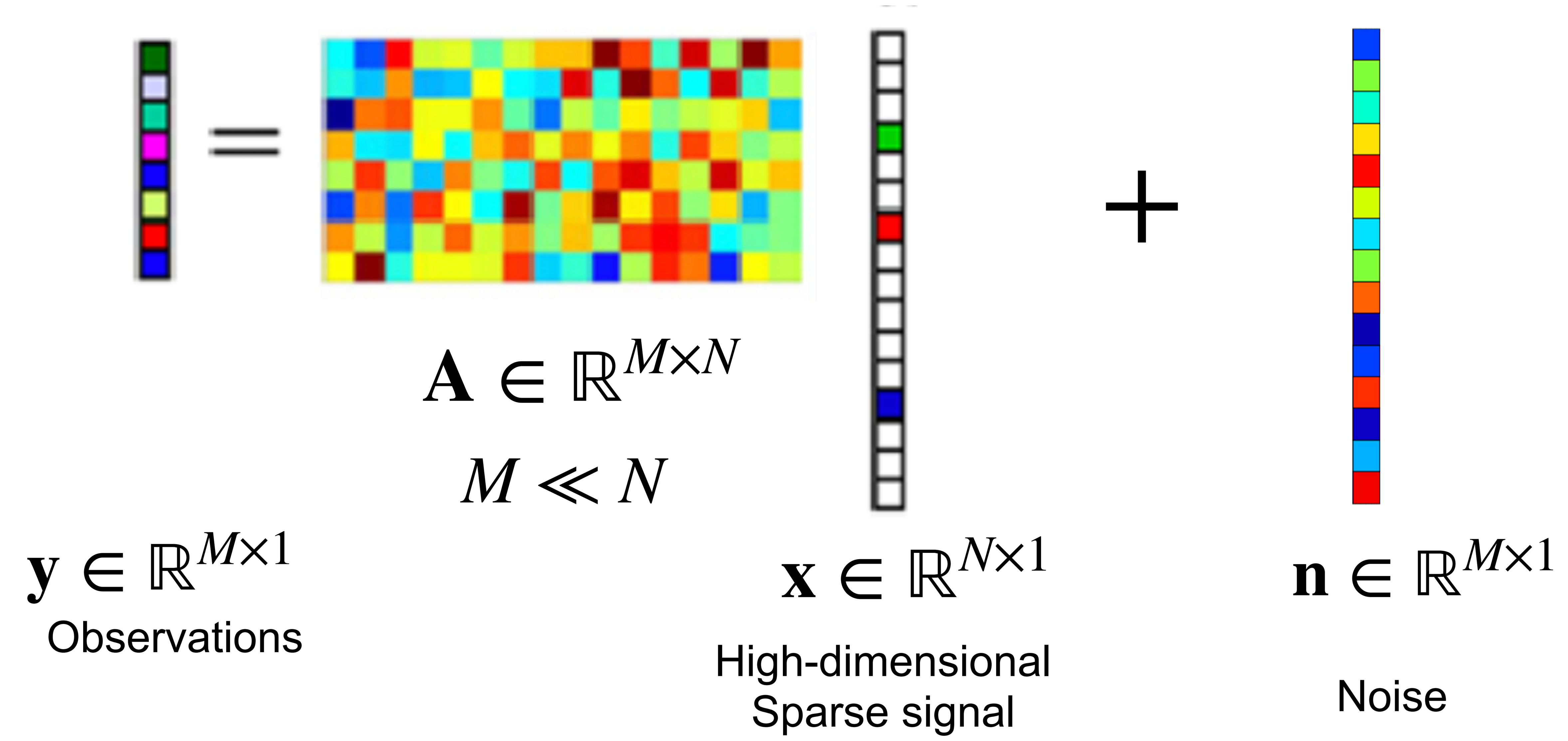
Quantizer



Background and Problem Setup

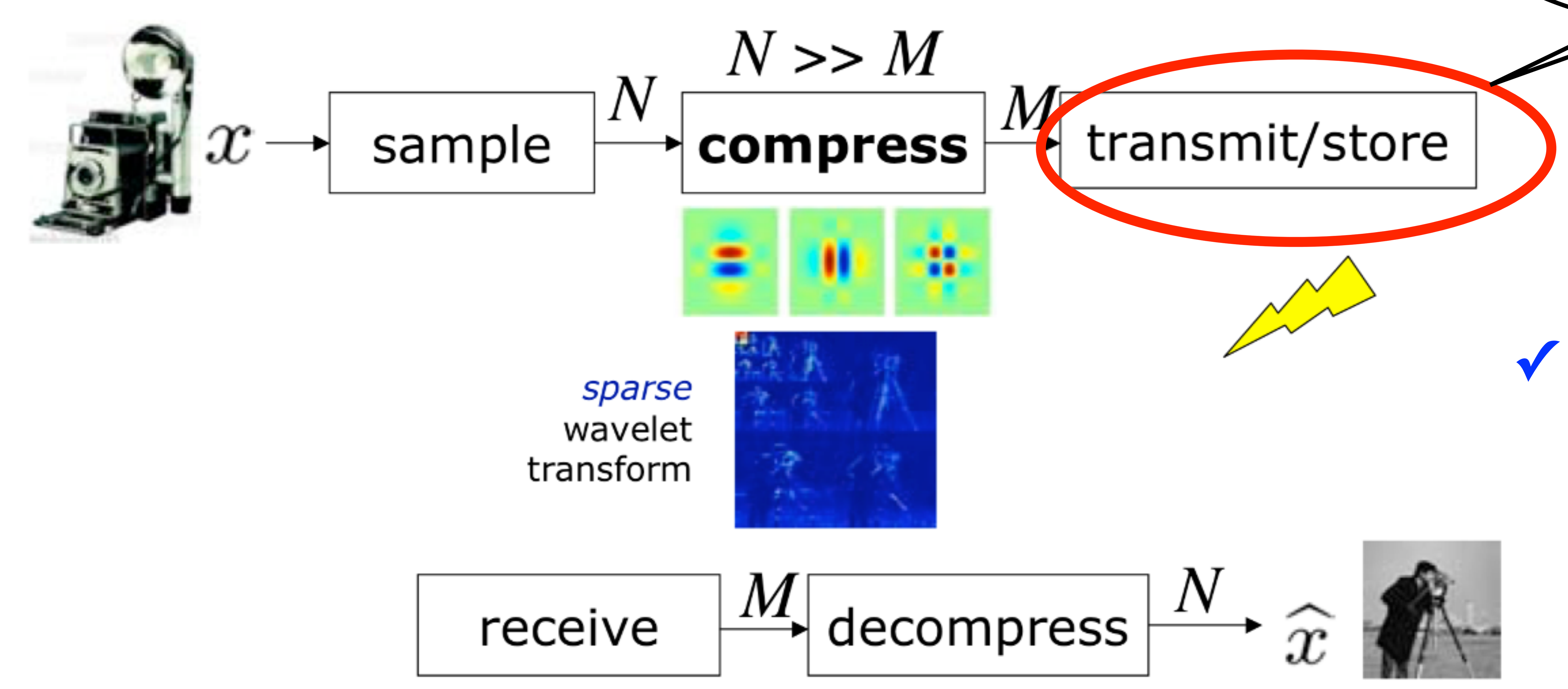
- **Standard Compressed Sensing (CS)**

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

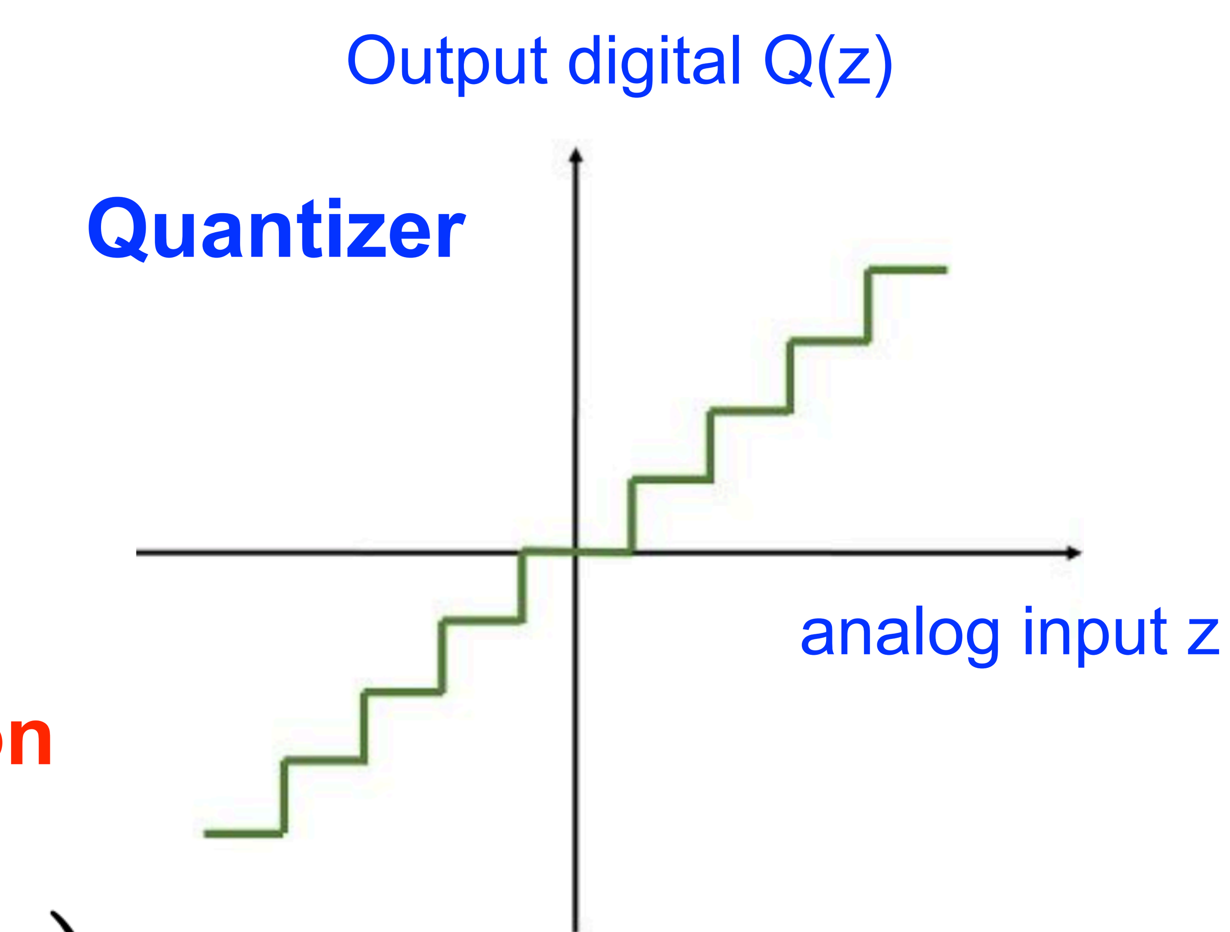


Goal: Recover the high-dimensional signal x from as few observations y as possible

- **Quantized Compressed Sensing (QCS)**



$$\mathbf{y} = \mathrm{Q}(\mathbf{A}\mathbf{x} + \mathbf{n})$$



✓ **An extreme case: 1-bit quantization**

$$\mathbf{y} = \mathrm{sign}(\mathbf{A}\mathbf{x} + \mathbf{n})$$

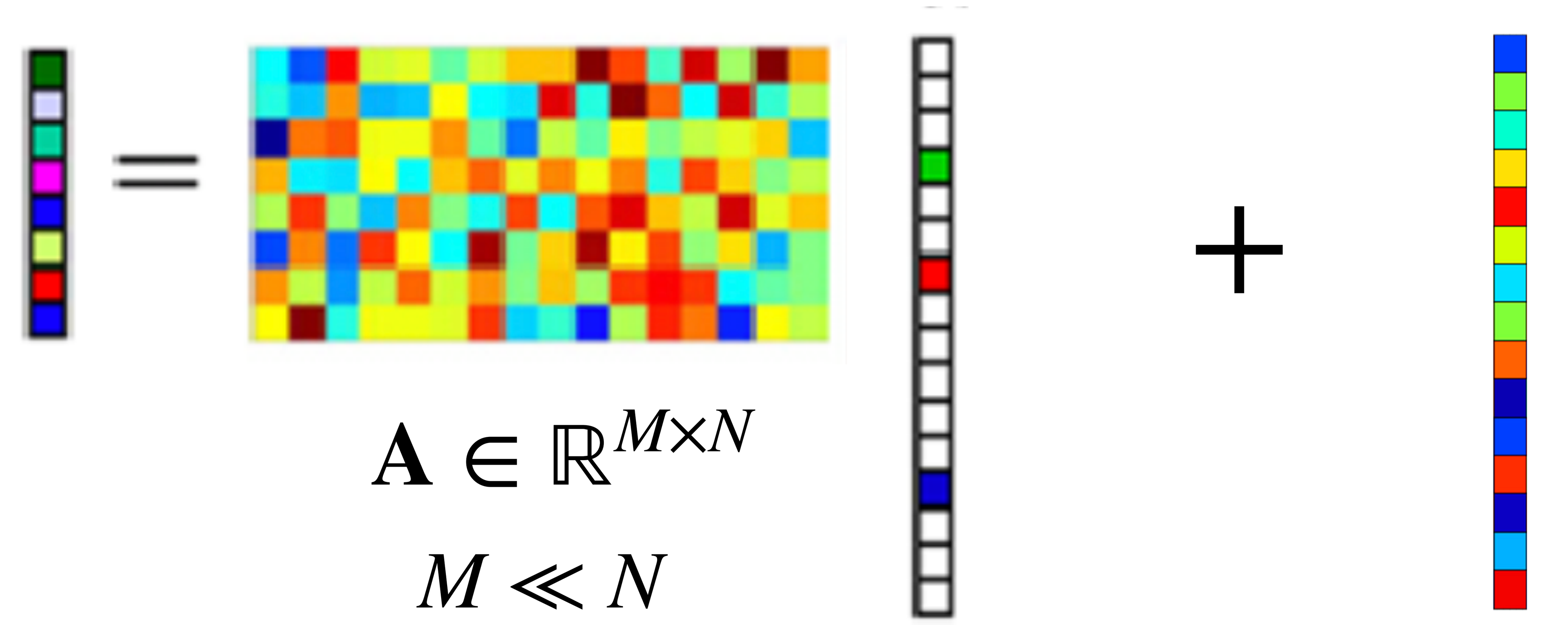
1-bit CS: Hardware friendly

Background and Problem Setup

- **Standard Compressed Sensing (CS)**

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

σ^2



$\mathbf{A} \in \mathbb{R}^{M\times N}$

$M \ll N$

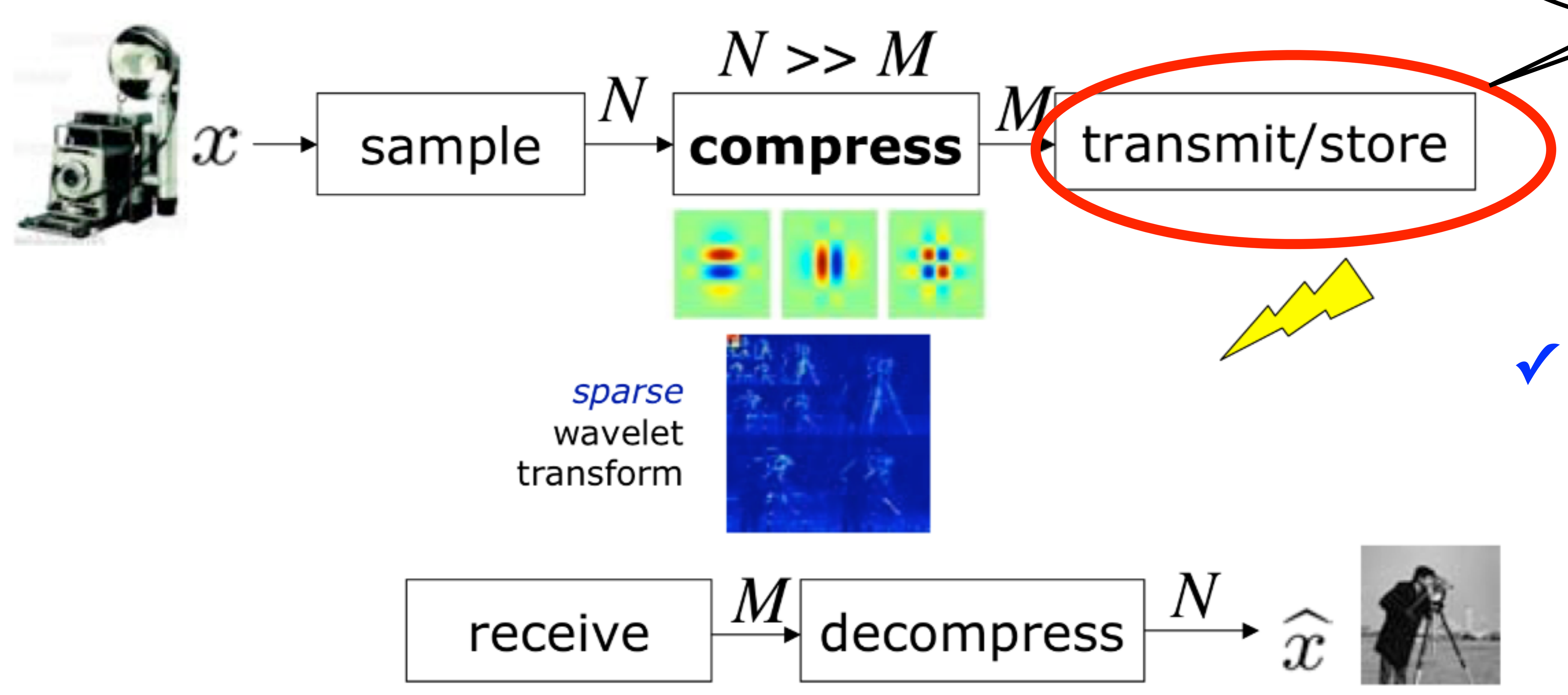
$\mathbf{y} \in \mathbb{R}^{M\times 1}$ Observations

$\mathbf{x} \in \mathbb{R}^{N\times 1}$ High-dimensional Sparse signal

$\mathbf{n} \in \mathbb{R}^{M\times 1}$ Noise

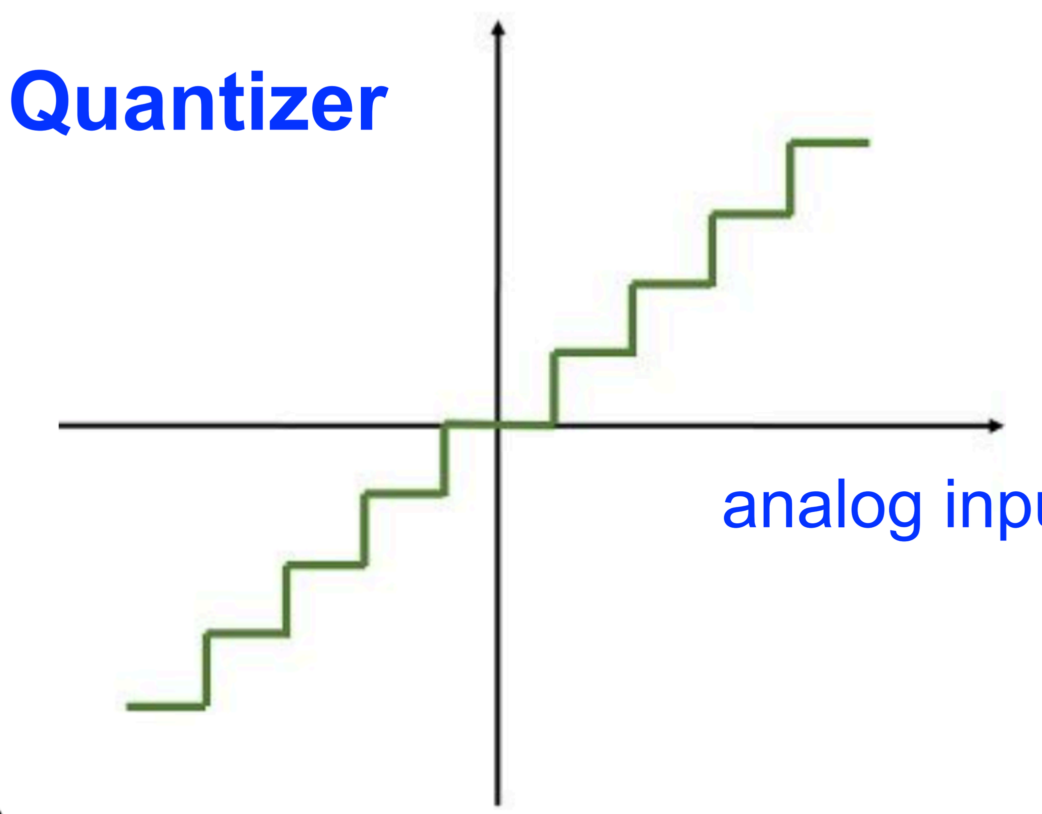
Goal: Recover the high-dimensional signal x from as few observations y as possible

- **Quantized Compressed Sensing (QCS)**



Quantization is essential

$$\mathbf{y} = \mathrm{Q}(\mathbf{A}\mathbf{x} + \mathbf{n})$$



✓ An extreme case: 1-bit quantization

$$\mathbf{y} = \mathrm{sign}(\mathbf{A}\mathbf{x} + \mathbf{n})$$

1-bit CS: Hardware friendly

Goal: How to accurately recover signal x from minimal quantized observations y?

A Bayesian Perspective

- **A Bayesian Perspective**

$$\mathbf{y} = \mathrm{Q}(\mathbf{A}\mathbf{x} + \mathbf{n})$$

Posterior — Prior — Likelihood

Bayes' rule

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{x})\,p(\mathbf{y} \mid \mathbf{x})}{p(\mathbf{y})}$$

Key idea: The more you know *a priori*, the less you need

Thomas Bayes (1702-1761)

A Bayesian Perspective

- **A Bayesian Perspective**

$$\mathbf{y} = \mathrm{Q}(\mathbf{A}\mathbf{x} + \mathbf{n})$$

Bayes' rule

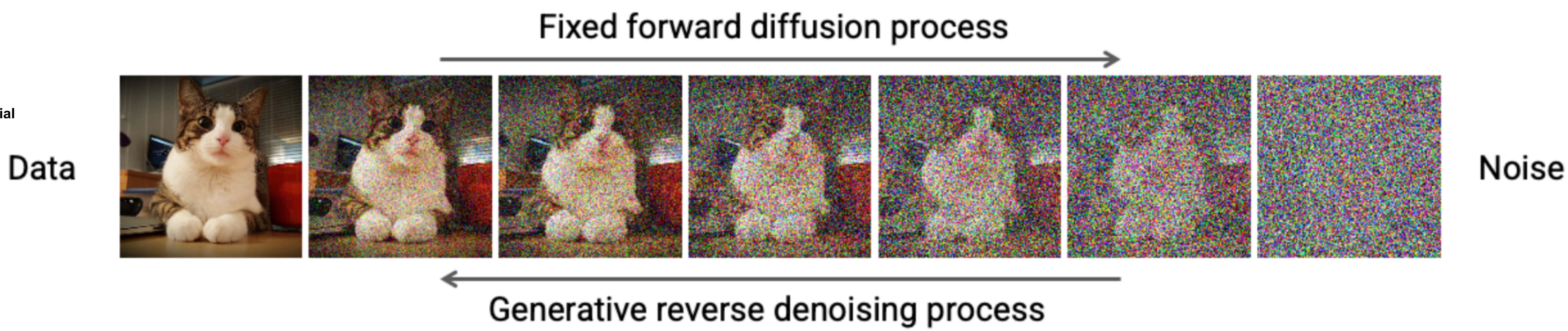
$$\underbrace{p(\mathbf{x}\,|\,\mathbf{y})}_{\text{Posterior}} = \frac{\underbrace{p(\mathbf{x})}_{\text{Prior}}\,\underbrace{p(\mathbf{y}\,|\,\mathbf{x})}_{\text{Likelihood}}}{p(\mathbf{y})}$$

Thomas Bayes (1702-1761)

Key idea: The more you know *a priori*, the less you need

- **Score-based Generative Models (SGM, also known as diffusion models)**

SGM
credited to
CVPR 2022 Tutorial



Reverse denoising process

Annealed Langevin dynamics (ALD)

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \alpha_t[\underbrace{\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t)}_{\text{Score Function}}] + \underbrace{\sqrt{2\alpha_t}\,\mathbf{z}_t}_{\text{Gaussian Noise}}$$

Score Function **Gaussian Noise**

Score-based Models (SGM) as an Implicit Prior

- **Posterior Score**

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

$$\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}) = \nabla_{\mathbf{x}} \log p(\mathbf{x}) + \nabla_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x})$$

Posterior score | Prior score | Likelihood score

- **Posterior Sampling via Annealed Langevin dynamics (ALD)**

intractable!

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \alpha_t[\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)] + \sqrt{2\alpha_t}\mathbf{z}_t$$

noise-perturbed prior
(from pre-trained SGM)

noise-perturbed likelihood score
(from quantized measurements)

Key Problem: How to Compute the Noise-perturbed Likelihood Score?

Two Assumptions

- **Assumption 1**

The prior $p(\mathbf{x})$ is non-informative w.r.t. $p(\mathbf{x}_t \mid \mathbf{x})$

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

Asymptotically accurate when the perturbed noise is negligible

- **Assumption 2**

The sensing matrix **A** is row-orthogonal, i.e.,

$$\mathbf{A}\mathbf{A}^T = \text{Diagonal matrix}$$

(Approximately) satisfied by many popular CS matrices
e.g., DFT, DCT, Hadamard, and random Gaussian matrices, etc.

Results of Pseudo-likelihood Score

- **Theorem 1: Under assumptions 1 and 2, we obtain a closed-form solution to the likelihood score**

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) = \mathbf{A}^T \mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t)$$

where

$$\mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t) = [g_1, g_2, \ldots, g_M]^T \in \mathbb{R}^{M \times 1}$$

$$g_m = \frac{\exp\left(-\frac{\tilde{u}_{y_m}^2}{2}\right) - \exp\left(-\frac{\tilde{l}_{y_m}^2}{2}\right)}{\sqrt{\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2} \int_{\tilde{l}_{y_m}}^{\tilde{u}_{y_m}} \exp\left(-\frac{t^2}{2}\right) dt} \qquad \tilde{u}_{y_m} = \frac{\mathbf{a}_m^T \mathbf{x}_t - u_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2}} \qquad \tilde{l}_{y_m} = \frac{\mathbf{a}_m^T \mathbf{x}_t - l_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2}}$$

Results of Pseudo-likelihood Score

- **Theorem 1: Under assumptions 1 and 2, we obtain a closed-form solution to the likelihood score**

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) = \mathbf{A}^T \mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t)$$

where

$$\mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t) = [g_1, g_2, \ldots, g_M]^T \in \mathbb{R}^{M \times 1}$$

$$g_m = \frac{\exp\left(-\frac{\tilde{u}_{y_m}^2}{2}\right) - \exp\left(-\frac{\tilde{l}_{y_m}^2}{2}\right)}{\sqrt{\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2} \int_{\tilde{l}_{y_m}}^{\tilde{u}_{y_m}} \exp\left(-\frac{t^2}{2}\right) dt} \qquad \tilde{u}_{y_m} = \frac{\mathbf{a}_m^T \mathbf{x}_t - u_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2}} \qquad \tilde{l}_{y_m} = \frac{\mathbf{a}_m^T \mathbf{x}_t - l_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2}}$$

- **Corollary 1.1: In the special case of 1-bit CS, results can be further simplified**

$$g_m = \left[\frac{1 + y_m}{2\Phi(\tilde{z}_m)} - \frac{1 - y_m}{2\left(1 - \Phi(\tilde{z}_m)\right)}\right] \frac{\exp\left(-\frac{\tilde{z}_m^2}{2}\right)}{\sqrt{2\pi(\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2)}} \qquad \tilde{z}_m = \frac{\mathbf{a}_m^T \mathbf{x}_t}{\sqrt{\sigma^2 + \beta_t^2 \parallel \mathbf{a}_m^T \parallel_2^2}} \qquad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

Results of Pseudo-likelihood Score

- **Theorem 1**: **Under assumptions 1 and 2, we obtain a closed-form solution to the likelihood score**

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) = \mathbf{A}^T \mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t)$$

where

$$\mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t) = [g_1, g_2, \ldots, g_M]^T \in \mathbb{R}^{M \times 1}$$

$$g_m = \frac{\exp\left(-\frac{\tilde{u}_{y_m}^2}{2}\right) - \exp\left(-\frac{\tilde{l}_{y_m}^2}{2}\right)}{\sqrt{\sigma^2 + \beta_t^2 \left\| \mathbf{a}_m^T \right\|_2^2} \int_{\tilde{l}_{y_m}}^{\tilde{u}_{y_m}} \exp\left(-\frac{t^2}{2}\right) dt} \qquad \tilde{u}_{y_m} = \frac{\mathbf{a}_m^T \mathbf{x}_t - u_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \left\| \mathbf{a}_m^T \right\|_2^2}} \qquad \tilde{l}_{y_m} = \frac{\mathbf{a}_m^T \mathbf{x}_t - l_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \left\| \mathbf{a}_m^T \right\|_2^2}}$$

- **Corollary 1.1**: **In the special case of 1-bit CS, results can be further simplified**

$$g_m = \left[\frac{1 + y_m}{2\Phi(\tilde{z}_m)} - \frac{1 - y_m}{2\left(1 - \Phi(\tilde{z}_m)\right)} \right] \frac{\exp\left(-\frac{\tilde{z}_m^2}{2}\right)}{\sqrt{2\pi(\sigma^2 + \beta_t^2 \left\| \mathbf{a}_m^T \right\|_2^2)}} \qquad \tilde{z}_m = \frac{\mathbf{a}_m^T \mathbf{x}_t}{\sqrt{\sigma^2 + \beta_t^2 \left\| \mathbf{a}_m^T \right\|_2^2}} \qquad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

- **Corollary 1.2**: **In the special case of standard CS**

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) = \mathbf{A}^T \left(\sigma^2 \mathbf{I} + \beta_t^2 \mathbf{A}\mathbf{A}^T\right)^{-1} \left(\mathbf{y} - \mathbf{A}\mathbf{x}_\mathbf{t}\right)$$

✓Explain the necessity of annealing term in Jalal et al. (2021a)
✓Extend and improve Jalal et al. (2021a) in the general case

The Proposed QCS-SGM Algorithm

- QCS-SGM: Quantized Compressed Sensing with SGM

Algorithm 1: Quantized Compressed Sensing with SGM (QCS-SGM)

Input: $\{\beta_t\}_{t=1}^{T}$, ϵ, K, $\mathbf{y}$, $\mathbf{A}$, σ^2, quantization codewords $\mathcal{Q}$ and thresholds $\{[l_q, u_q)|q \in \mathcal{Q}\}$

Initialization: $\mathbf{x}_1^0 \sim \mathcal{U}(0, 1)$

for $t = 1$ ***to*** T **do**

  $\alpha_t \leftarrow \epsilon\beta_t^2/\beta_T^2$

  for $k = 1$ ***to*** K **do**

    Draw $\mathbf{z}_t^k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

    Compute $\mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t^{k-1})$ as (12) (or (15) for 1-bit)

    $\mathbf{x}_t^k = \mathbf{x}_t^{k-1} + \alpha_t \left[\mathrm{s}_{\boldsymbol{\theta}}(\mathbf{x}_t^{k-1}, \beta_t) + \mathbf{A}^T \mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \mathbf{x}_t^{k-1}) \right] + \sqrt{2\alpha_t}\mathbf{z}_t^k$

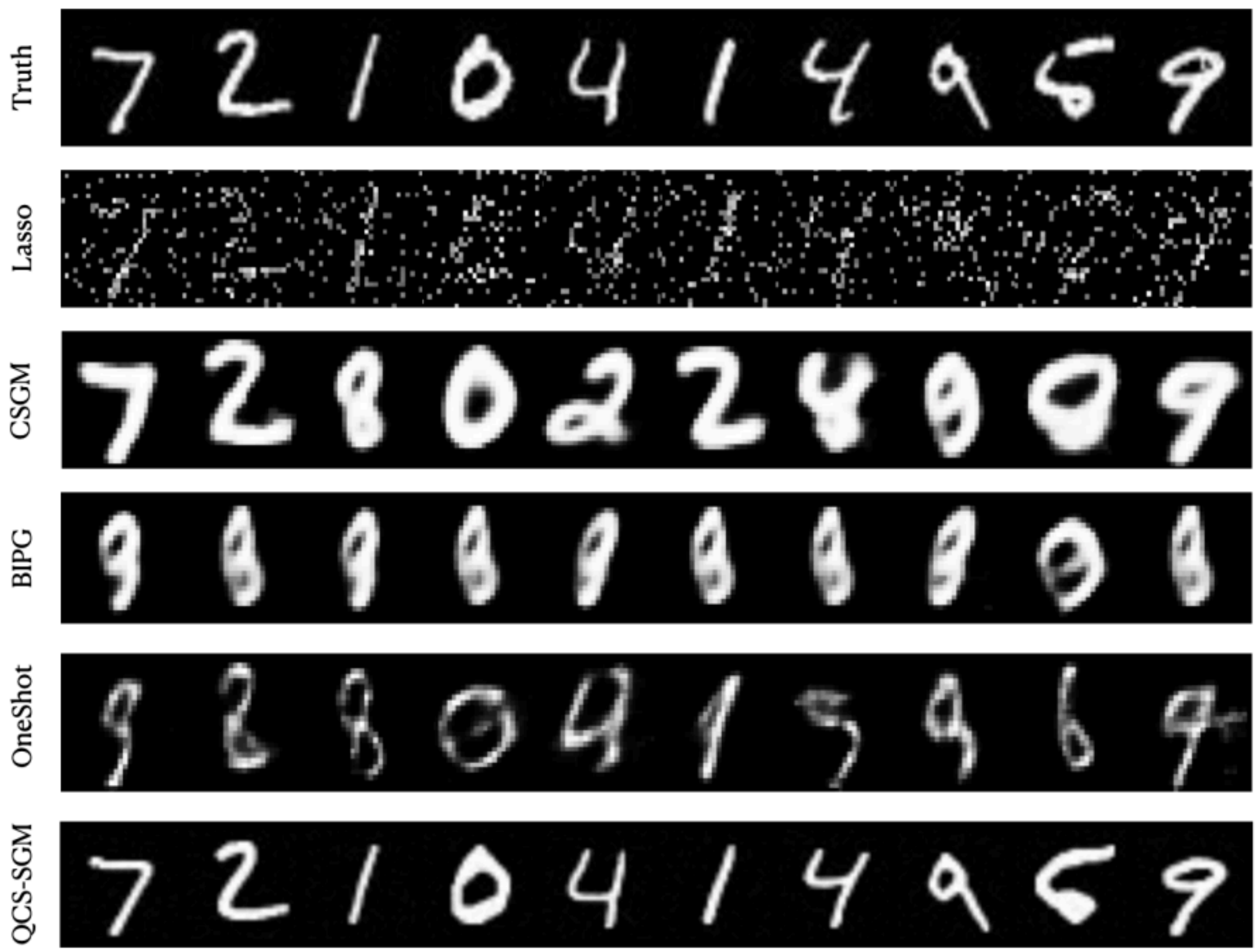
  $\mathbf{x}_{t+1}^0 \leftarrow \mathbf{x}_t^K$

Output: $\hat{\mathbf{x}} = \mathbf{x}_T^K$

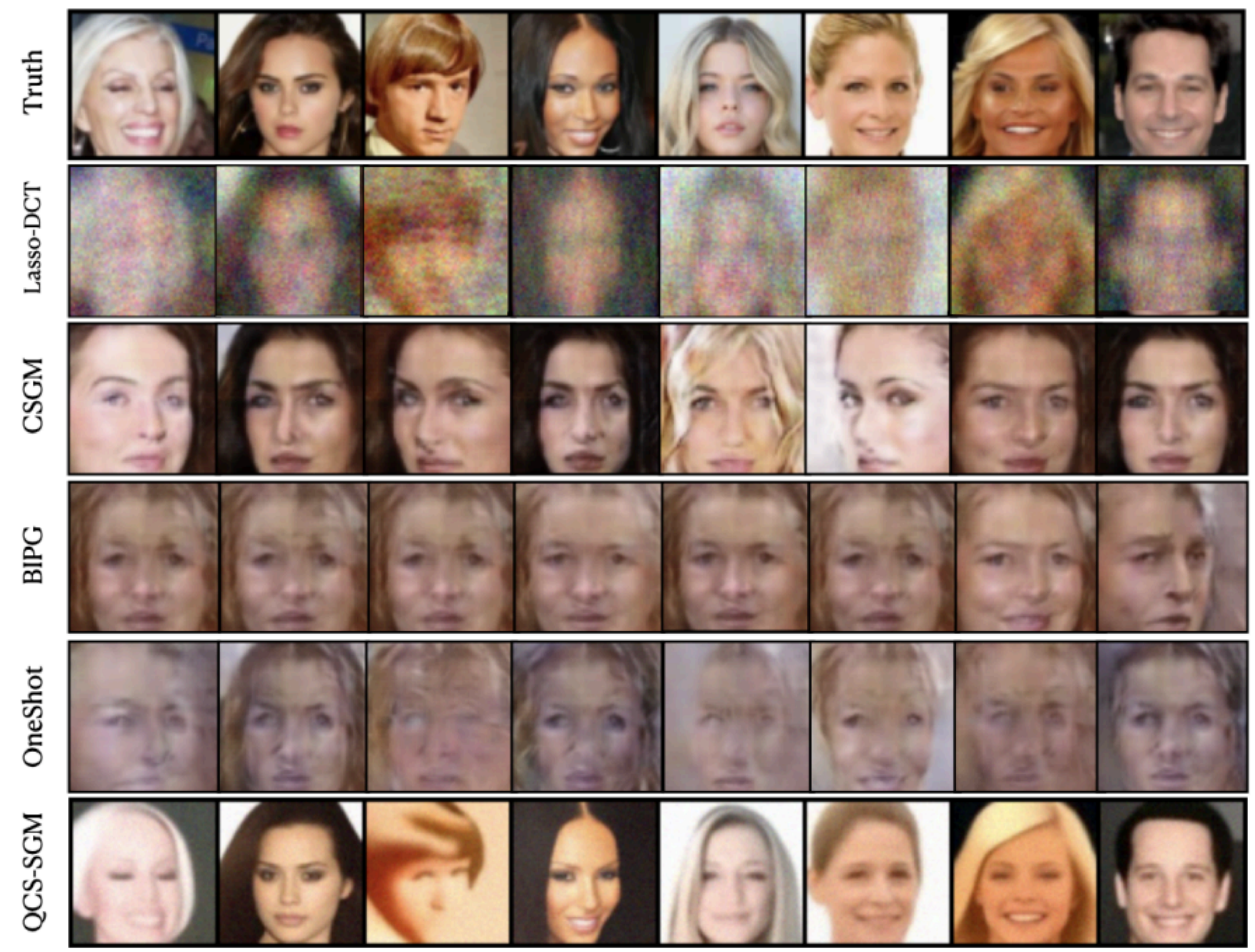
Only this term is different from SGM!

Experimental Results

- **In-Distribution Results on MNIST and CelebA**



(a) MNIST, $M = 200, \sigma = 0.05$

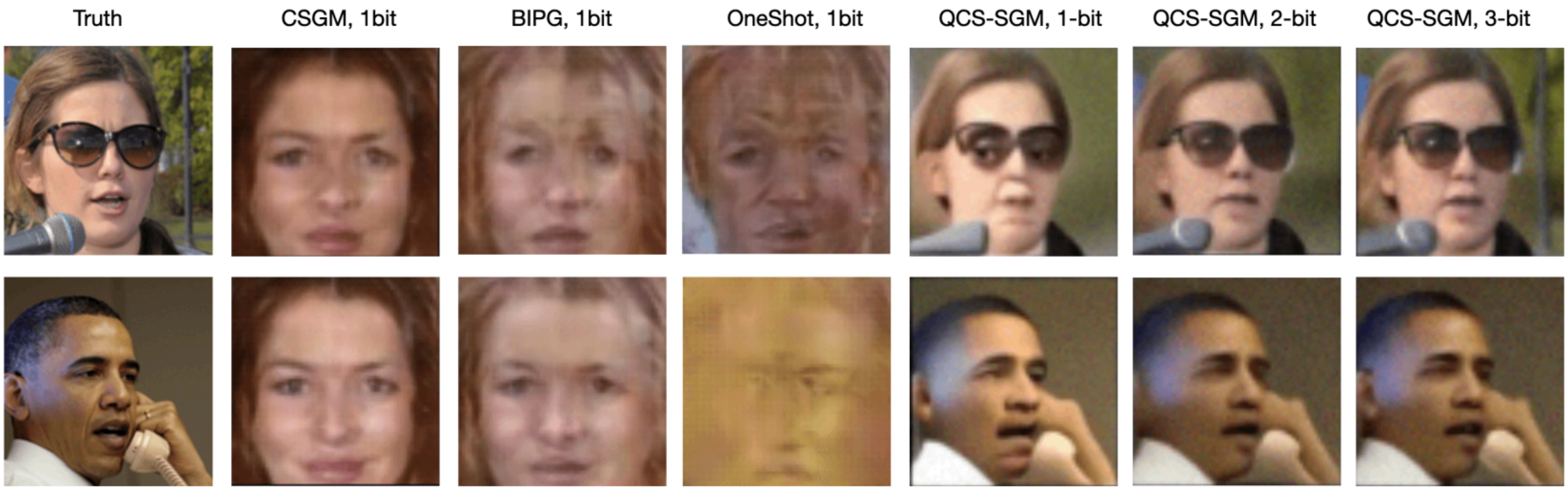


(b) CelebA, $M = 4000, \sigma = 0.001$

The proposed QCS-SGM significantly outperforms existing methods!

Experimental Results

- **Out-of-Distribution (OOD) Results on FFHQ**



✓SGM model trained on CelebA dataset
✓Images tested on FFHQ dataset

The proposed QCS-SGM significantly outperforms existing methods!

Experimental Results

- **High-Resolution (256*256) Image Results on FFHQ**

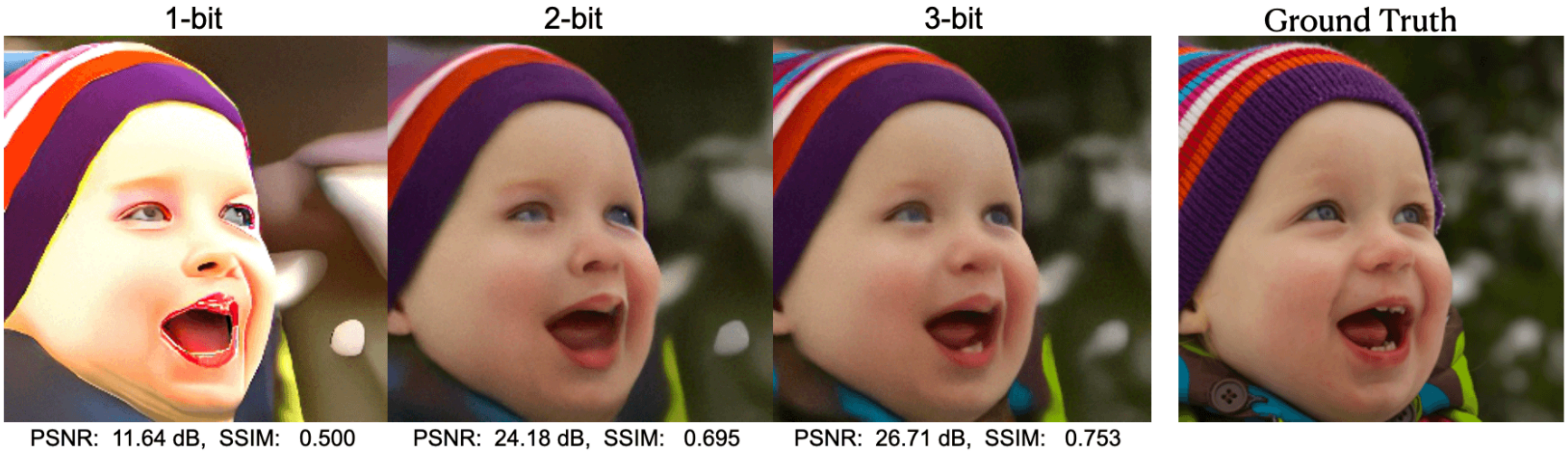


Figure 1: Reconstructed images of our QCS-SGM for one FFHQ 256×256 high-resolution RGB test image ($N = 256 \times 256 \times 3 = 196608$ pixels) from noisy heavily quantized (1bit, 2-bit and 3-bit) CS $8\times$ measurements $\mathbf{y} = \mathrm{Q}(\mathbf{A}\mathbf{x} + \mathbf{n})$, i.e., $M = 24576 \ll N$. The measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ is i.i.d. Gaussian, i.e., $A_{ij} \sim \mathcal{N}(0, \frac{1}{M})$, and a Gaussian noise $\mathbf{n}$ is added with standard deviation $\sigma = 10^{-3}$.

The proposed QCS-SGM can accurately recover high-resolution images from a small number of heavily quantized noisy measurements!

Thank you!
Q&A