

RADIO-INTERFEROMETRIC IMAGE RECONSTRUCTION WITH DENOISING DIFFUSION RESTORATION MODELS

Emma Tolley, EPFL
Swiss SKA Days 2026
31 August 2026





INVERSE PROBLEMS IN RADIO ASTRONOMY

We correlate the voltages between different **pairs** of antennas to construct **visibilities**

INVERSE PROBLEMS IN RADIO ASTRONOMY

We correlate the voltages between different **pairs** of antennas to construct **visibilities**

Visibilities are related to the radio sky image $I(l, m)$:

$$V(u, v, w) = \iint I(l, m) e^{-2\pi i [ul + vm + wn]} \frac{dl \, dm}{n + 1} + z(u, v, w)$$

Reduces to 2D Fourier transform for narrow FoV, ie as $n \rightarrow 0$

INVERSE PROBLEMS IN RADIO ASTRONOMY

We correlate the voltages between different **pairs** of antennas to construct **visibilities**

Visibilities are related to the radio sky image $I(l, m)$:

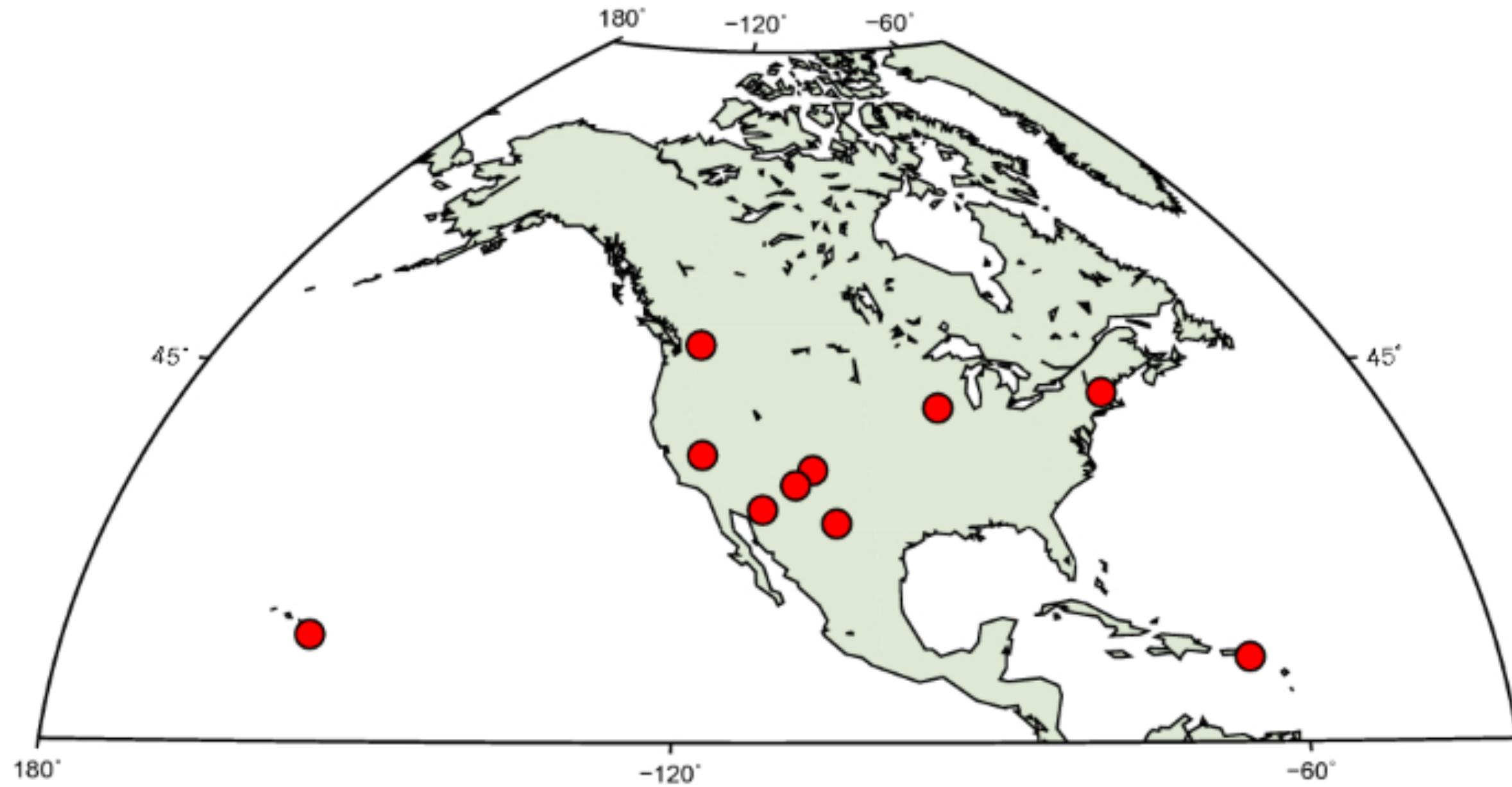
$$\underline{V}(u, v, w) = \iint \underline{I}(l, m) e^{-2\pi i [ul + vm + wn]} \frac{dl dm}{n + 1} + \underline{z}(u, v, w)$$

Reduces to 2D Fourier transform for narrow FoV, ie as $n \rightarrow 0$

Can write this as a linear inverse problem:

$$\underline{y} = \mathbf{S} \mathcal{F} \underline{x} + \underline{z}$$

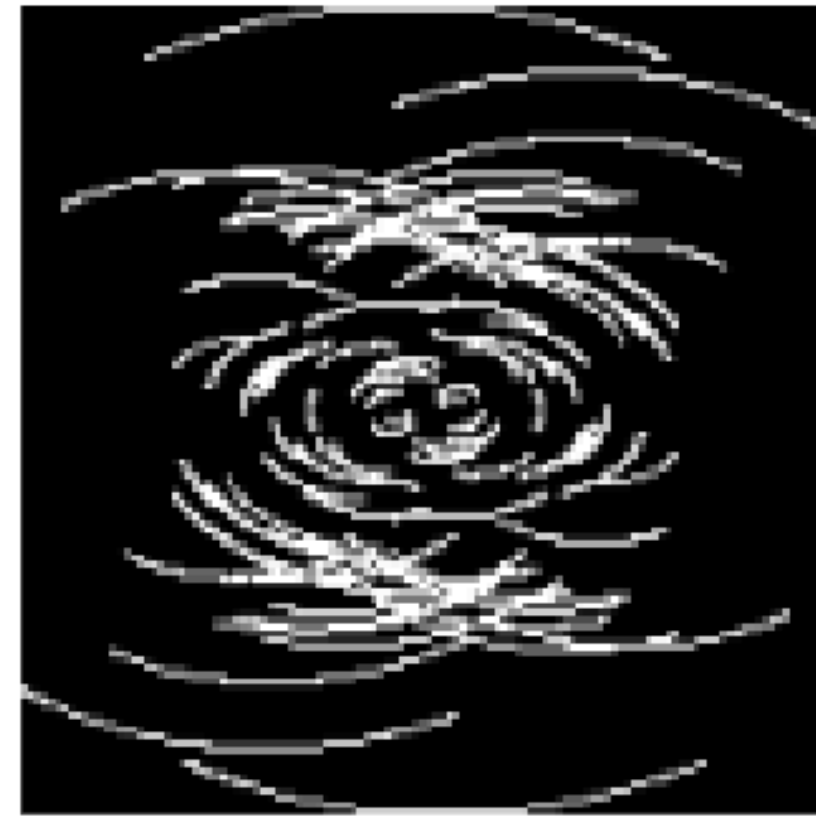
INVERSE PROBLEMS IN RADIO ASTRONOMY



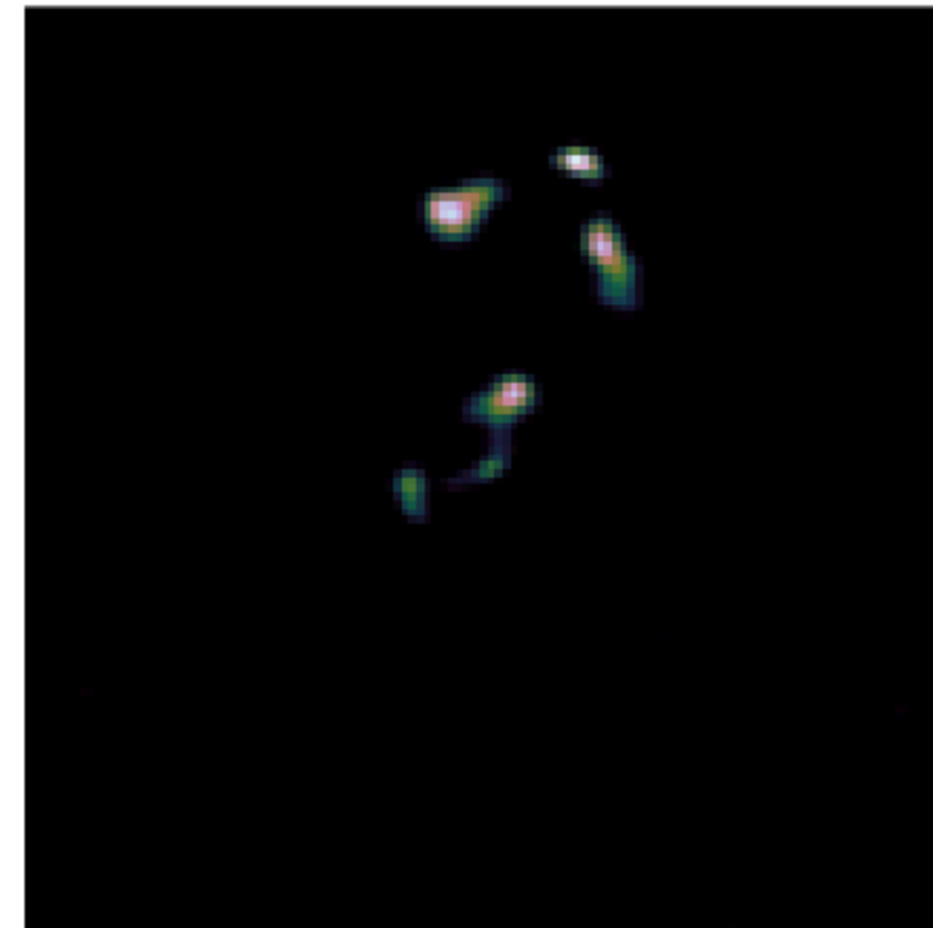
$$\mathbf{y} = \mathbf{S} \mathcal{F} \mathbf{x} + \mathbf{z}$$

$$\begin{aligned} \mathcal{F}^{-1}[\mathbf{y}] &= \mathcal{F}^{-1}[\mathbf{S} \mathcal{F} \mathbf{x} + \mathbf{z}] \\ &= \mathcal{F}^{-1}[\mathbf{S}] * \mathbf{x} + \mathcal{F}^{-1} \mathbf{z} \end{aligned}$$

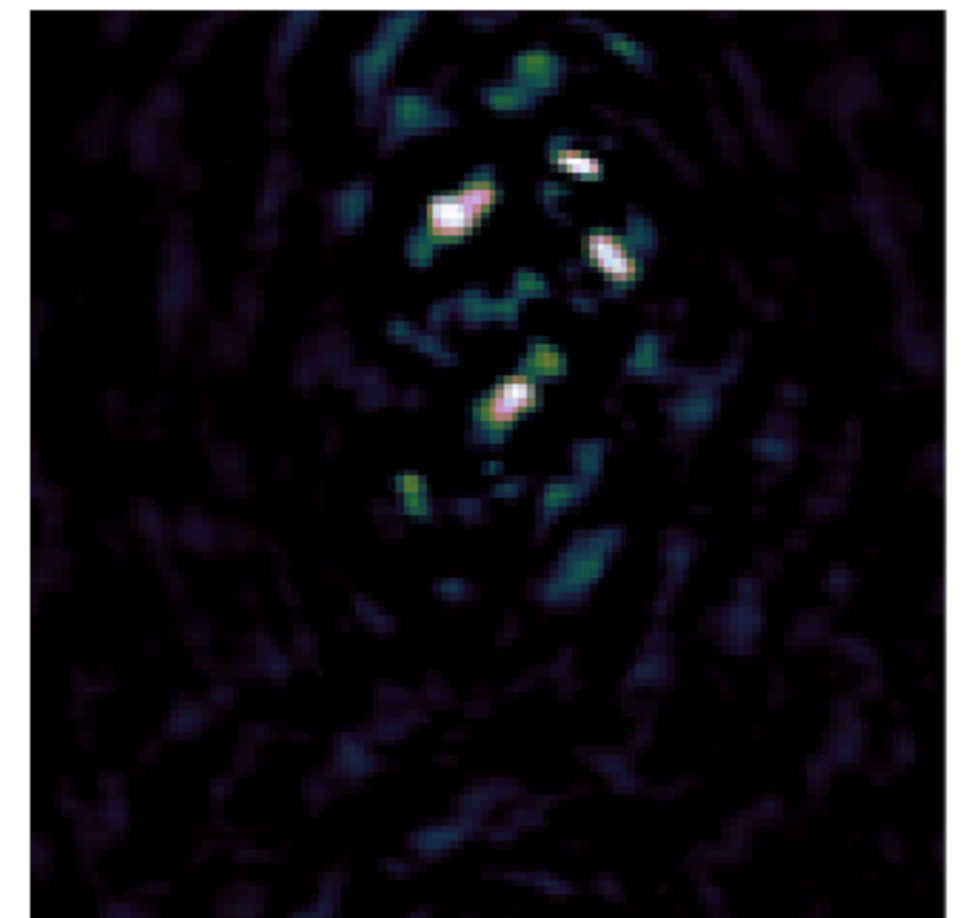
Sampling matrix



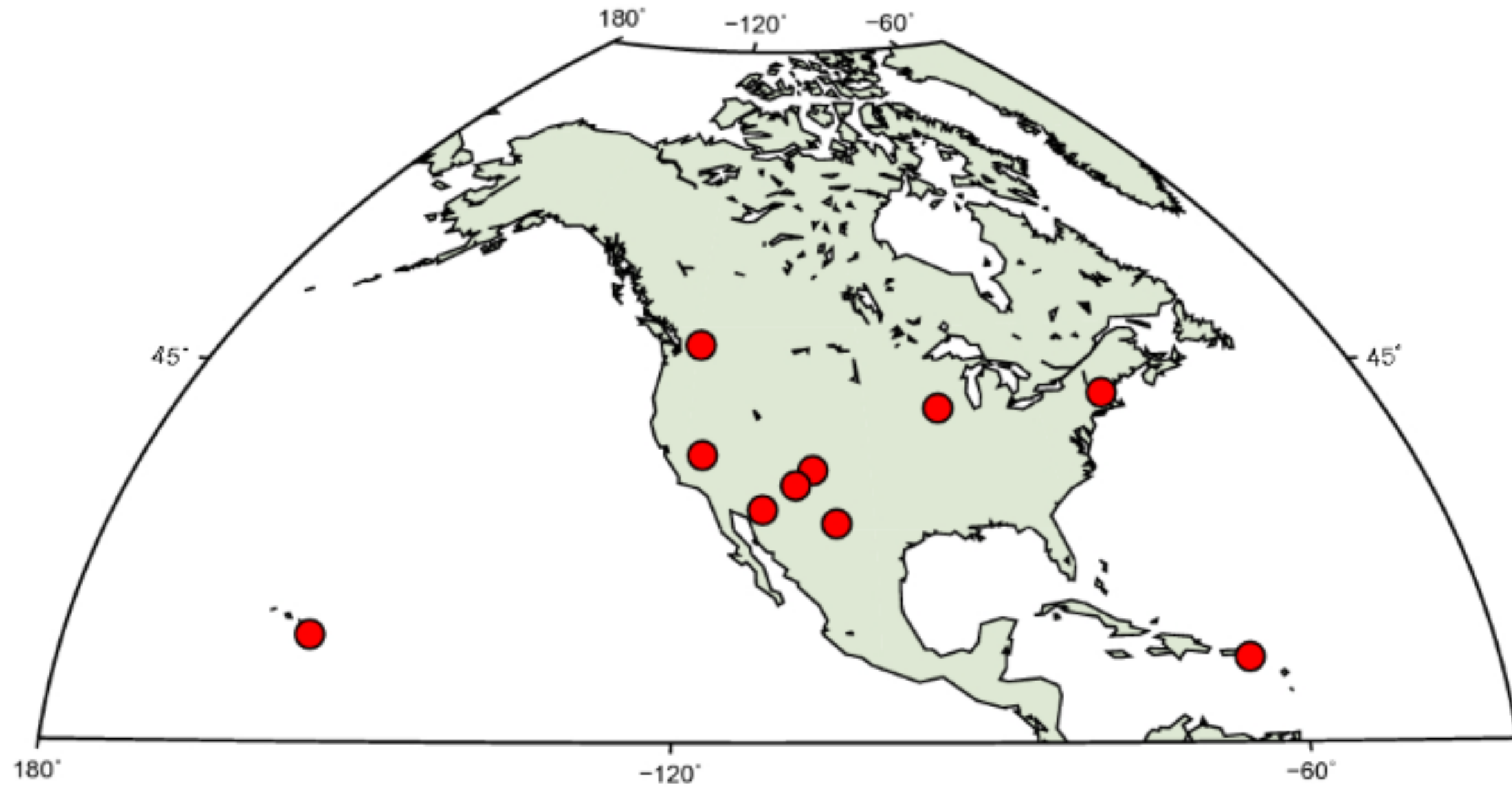
True image



Dirty image



INVERSE PROBLEMS IN RADIO ASTRONOMY

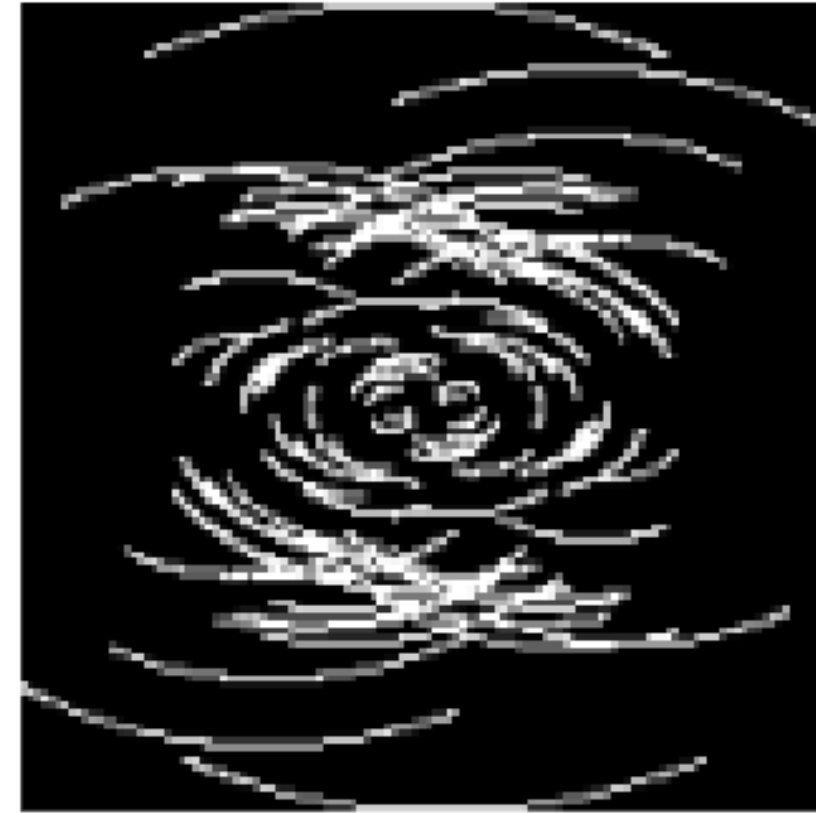


$$\mathbf{y} = \mathbf{S} \mathcal{F} \mathbf{x} + \mathbf{z}$$

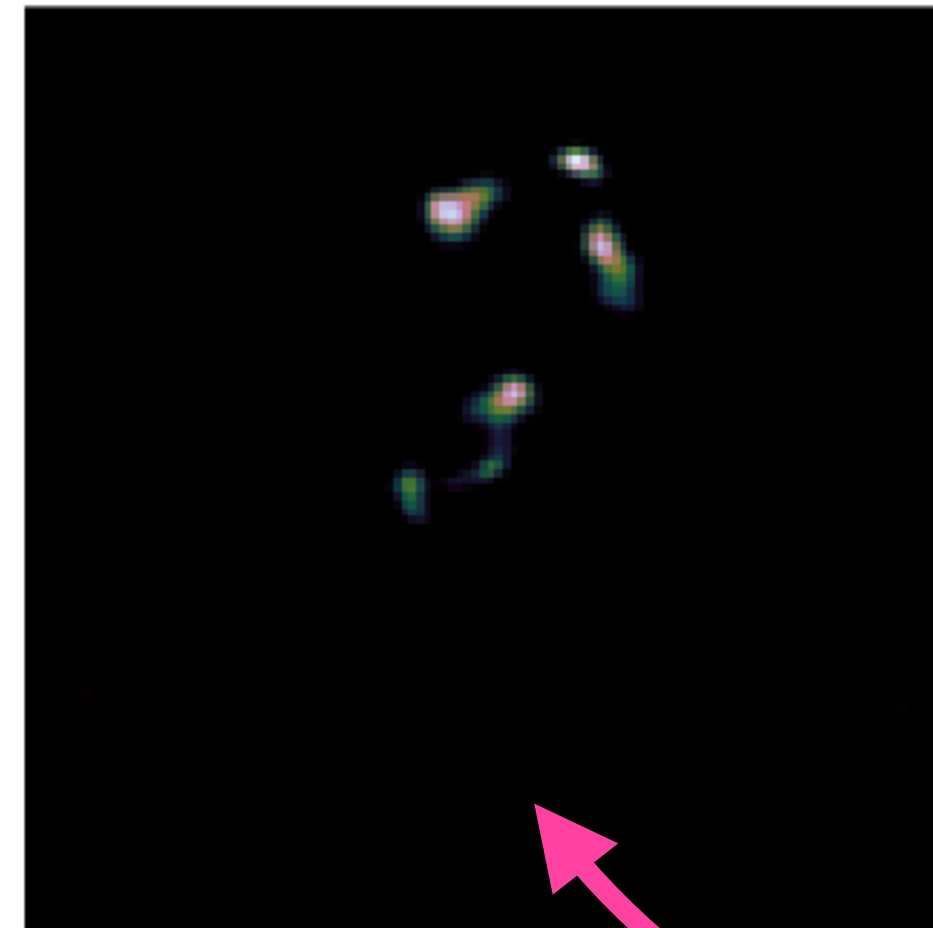
$$\begin{aligned} \mathcal{F}^{-1}[\mathbf{y}] &= \mathcal{F}^{-1}[\mathbf{S} \mathcal{F} \mathbf{x} + \mathbf{z}] \\ &= \mathcal{F}^{-1}[\mathbf{S}] * \mathbf{x} + \mathcal{F}^{-1} \mathbf{z} \end{aligned}$$

A pink curved arrow points from the $\mathcal{F}^{-1}[\mathbf{S}]$ term in the second equation to the 'Dirty image' panel on the right.

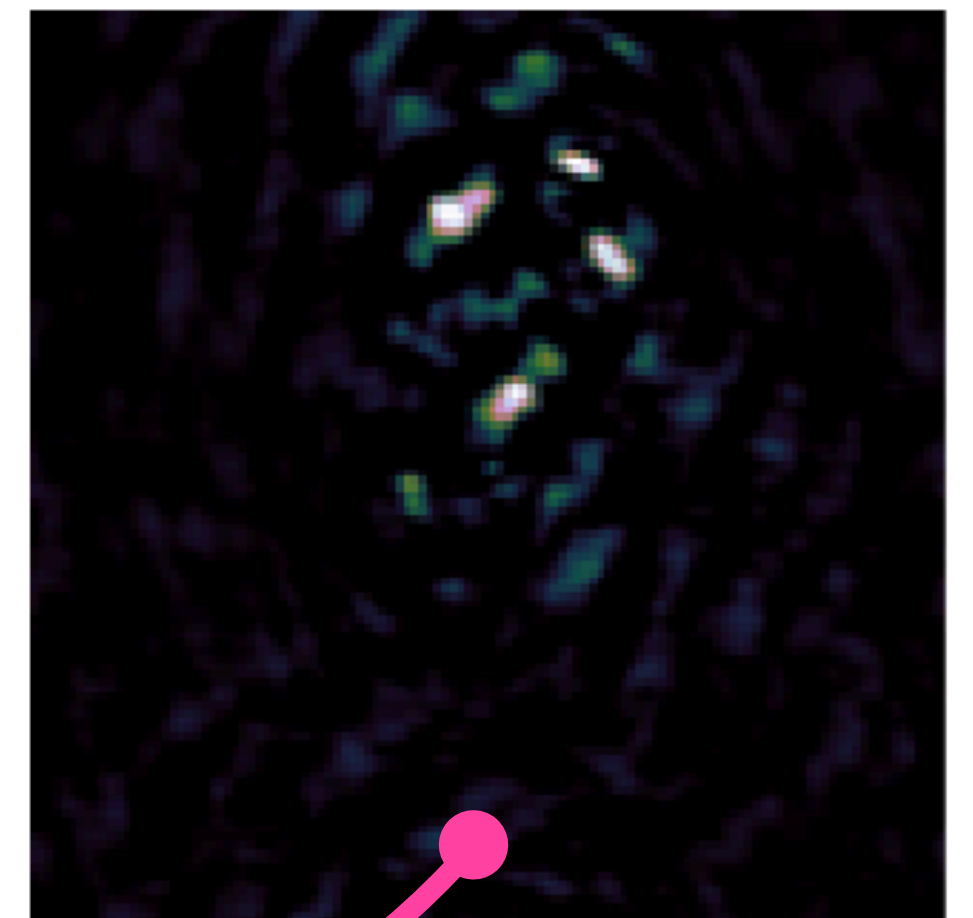
Sampling matrix



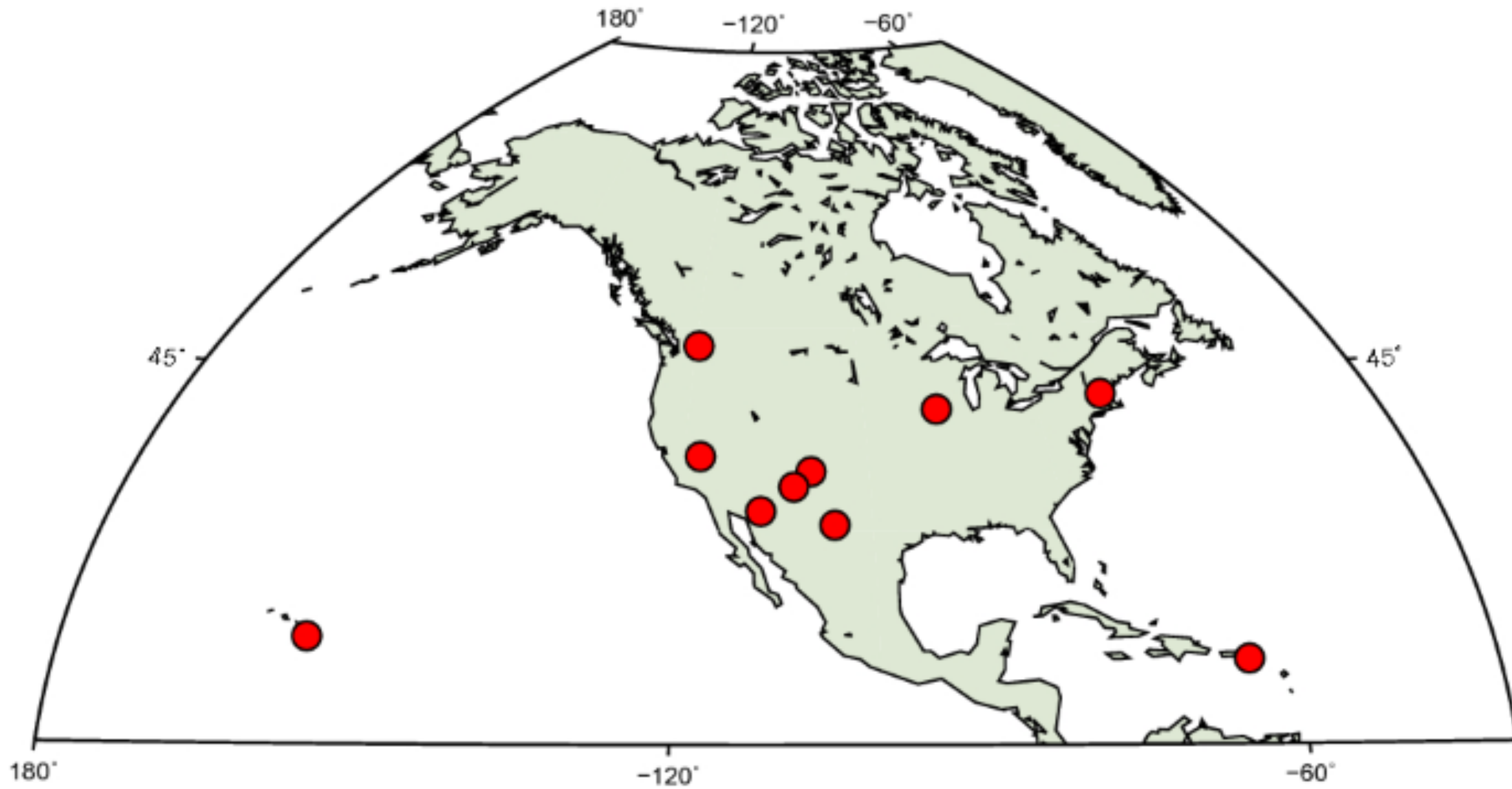
True image



Dirty image



INVERSE PROBLEMS IN RADIO ASTRONOMY

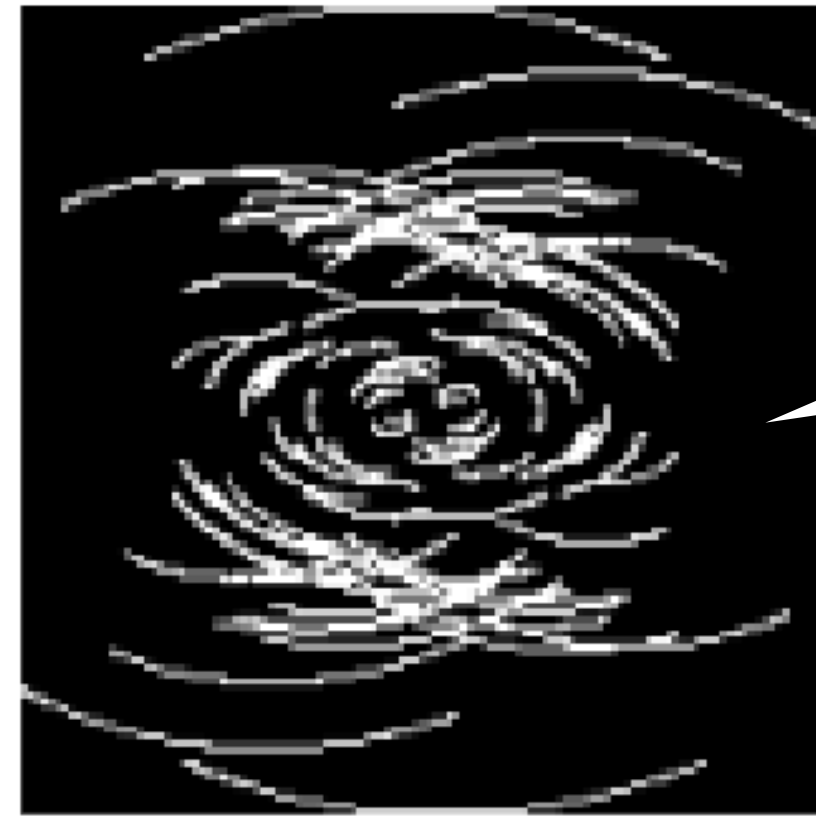


$$\mathbf{y} = \mathbf{S} \mathcal{F} \mathbf{x} + \mathbf{z}$$

$$\begin{aligned} \mathcal{F}^{-1}[\mathbf{y}] &= \mathcal{F}^{-1}[\mathbf{S} \mathcal{F} \mathbf{x} + \mathbf{z}] \\ &= \mathcal{F}^{-1}[\mathbf{S}] * \mathbf{x} + \mathcal{F}^{-1} \mathbf{z} \end{aligned}$$

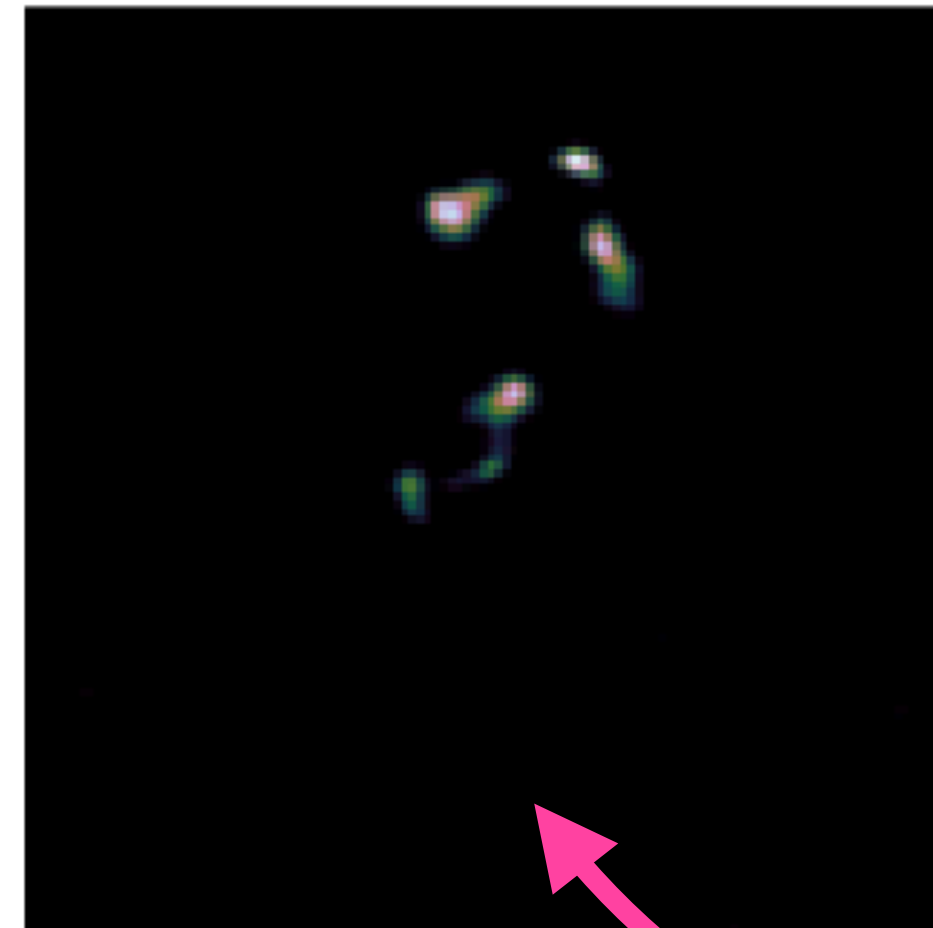
Equivalent to deconvolving the PSF

Sampling matrix

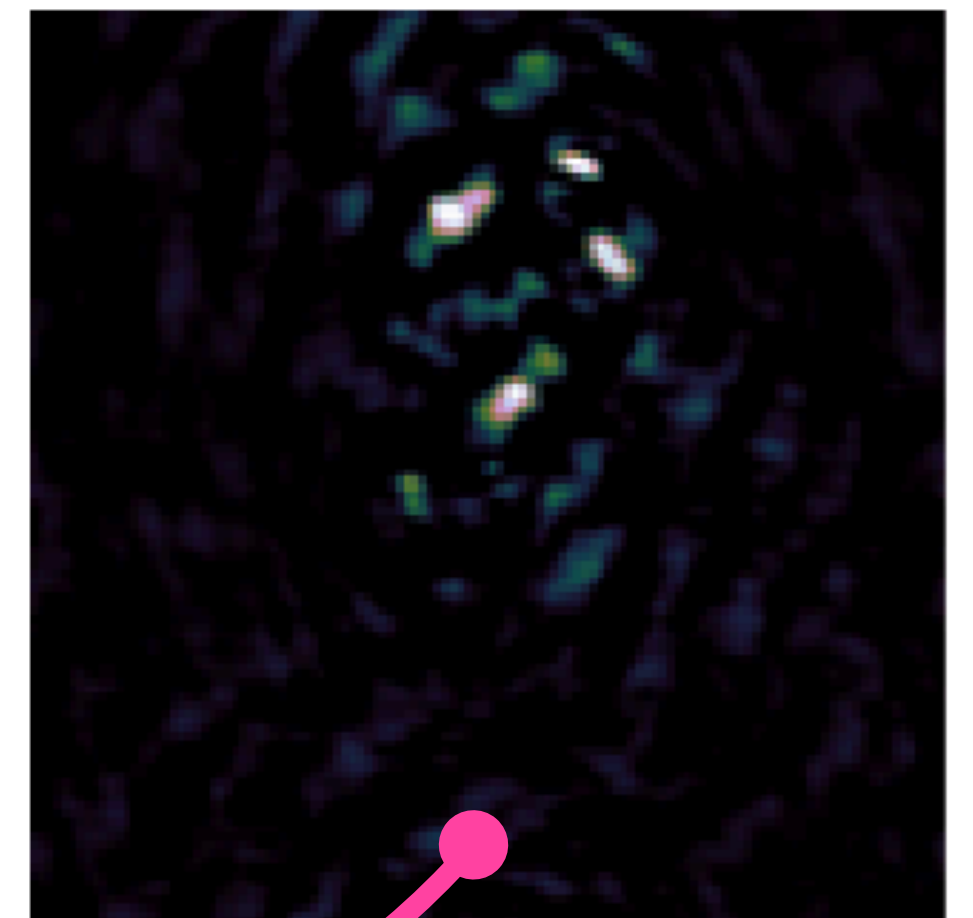


Inpaint missing visibilities

True image

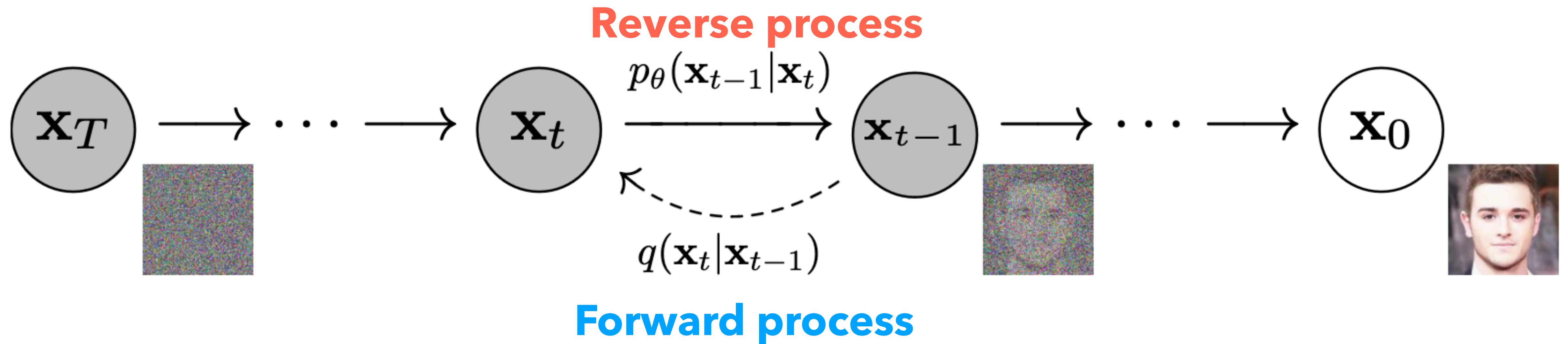


Dirty image



DENOISING DIFFUSION PROBABLISTIC MODELS (DDPM)

Train a neural network to learn this denoising step



Denoising Diffusion Probabilistic Models

arXiv:2006.11239

Jonathan Ho
UC Berkeley

jonathanho@berkeley.edu

Ajay Jain
UC Berkeley

ajayj@berkeley.edu

Pieter Abbeel
UC Berkeley

pabbeel@cs.berkeley.edu

DENOISING DIFFUSION PROBABLISTIC MODELS (DDPM)

Why Diffusion Models Don't Memorize: The Role of Implicit Dynamical Regularization in Training

Tony Bonnaire[†]
LPENS

Université PSL, Paris
tony.bonnaire@phys.ens.fr

Raphaël Urfin[†]
LPENS

Université PSL, Paris
raphael.urfin@phys.ens.fr

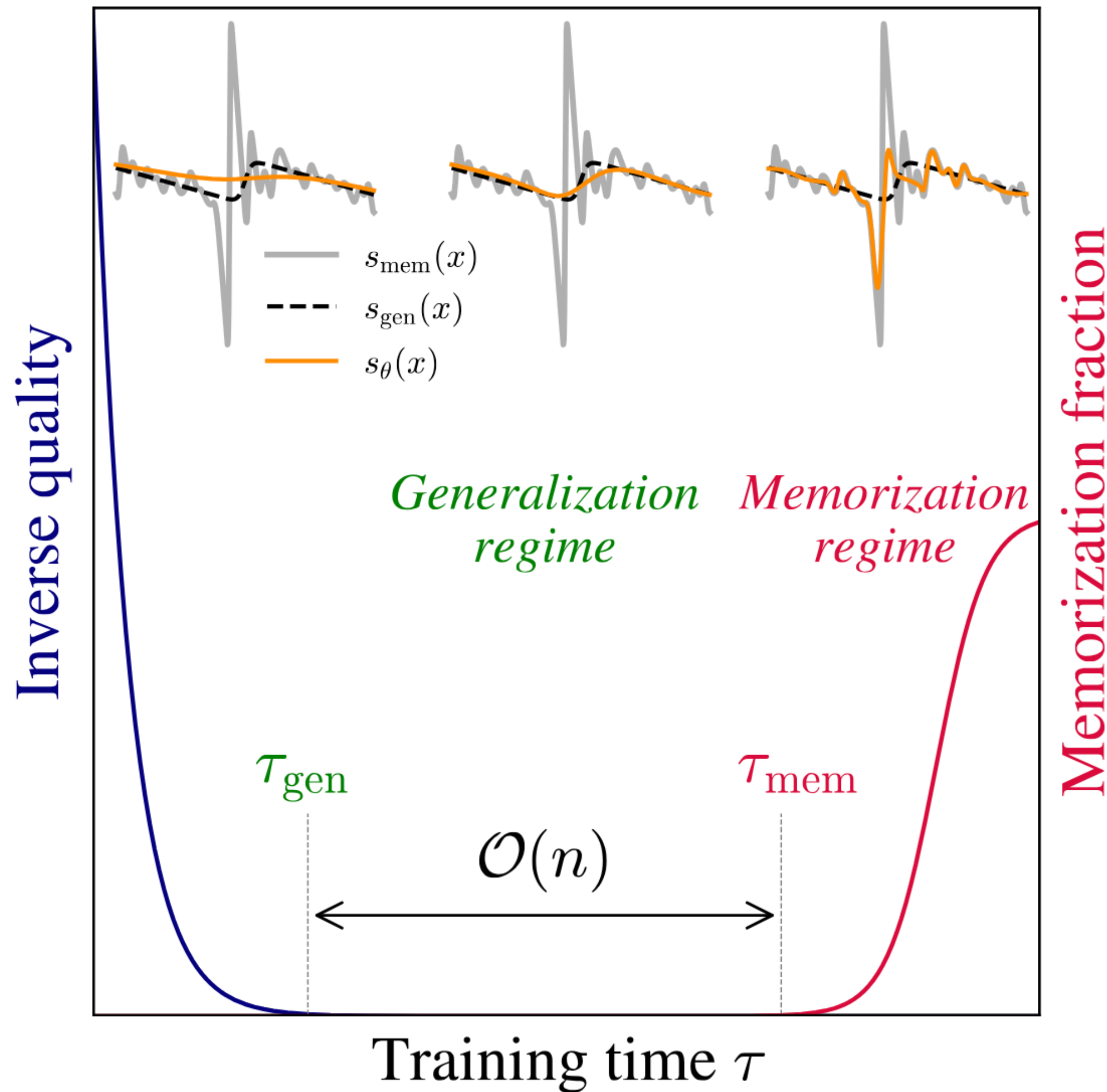
Giulio Biroli
LPENS

Université PSL, Paris
giulio.biroli@phys.ens.fr

Marc Mézard

Department of Computing Sciences
Bocconi University, Milano
marc.mezard@unibocconi.it

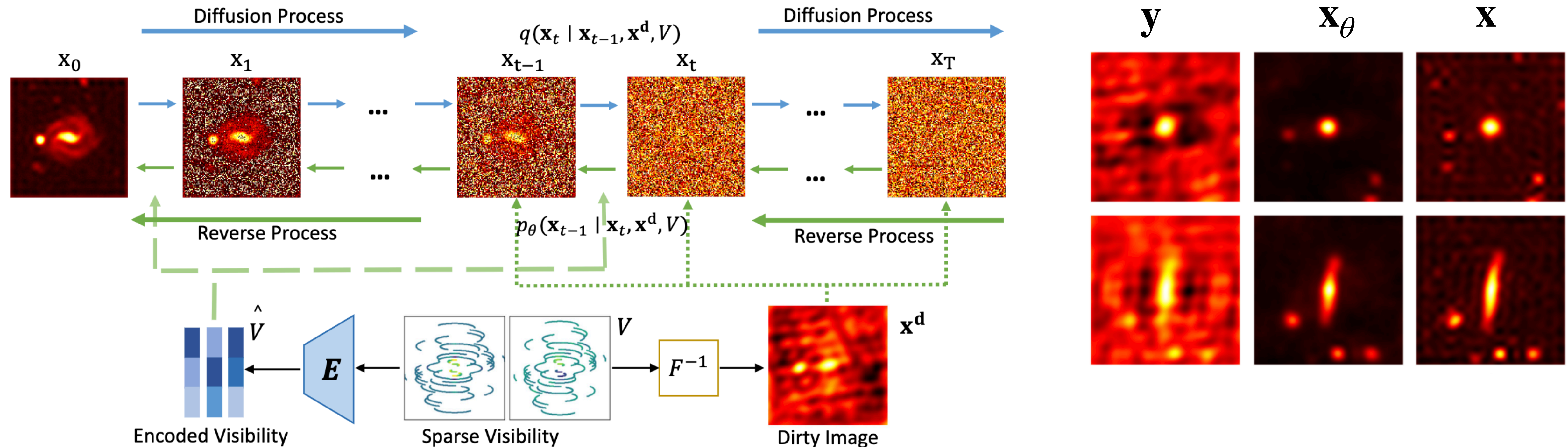
“[Diffusion models have a] form of implicit dynamical regularization in the training dynamics, which allow to avoid memorization even in highly overparameterized settings”



DENOISING DIFFUSION PROBABLISTIC MODELS (DDPM)

A Conditional Denoising Diffusion Probabilistic Model for Radio Interferometric Image Reconstruction

Ruoqi Wang^a, Zhuoyang Chen^a, Qiong Luo^{a,*} and Feng Wang^b

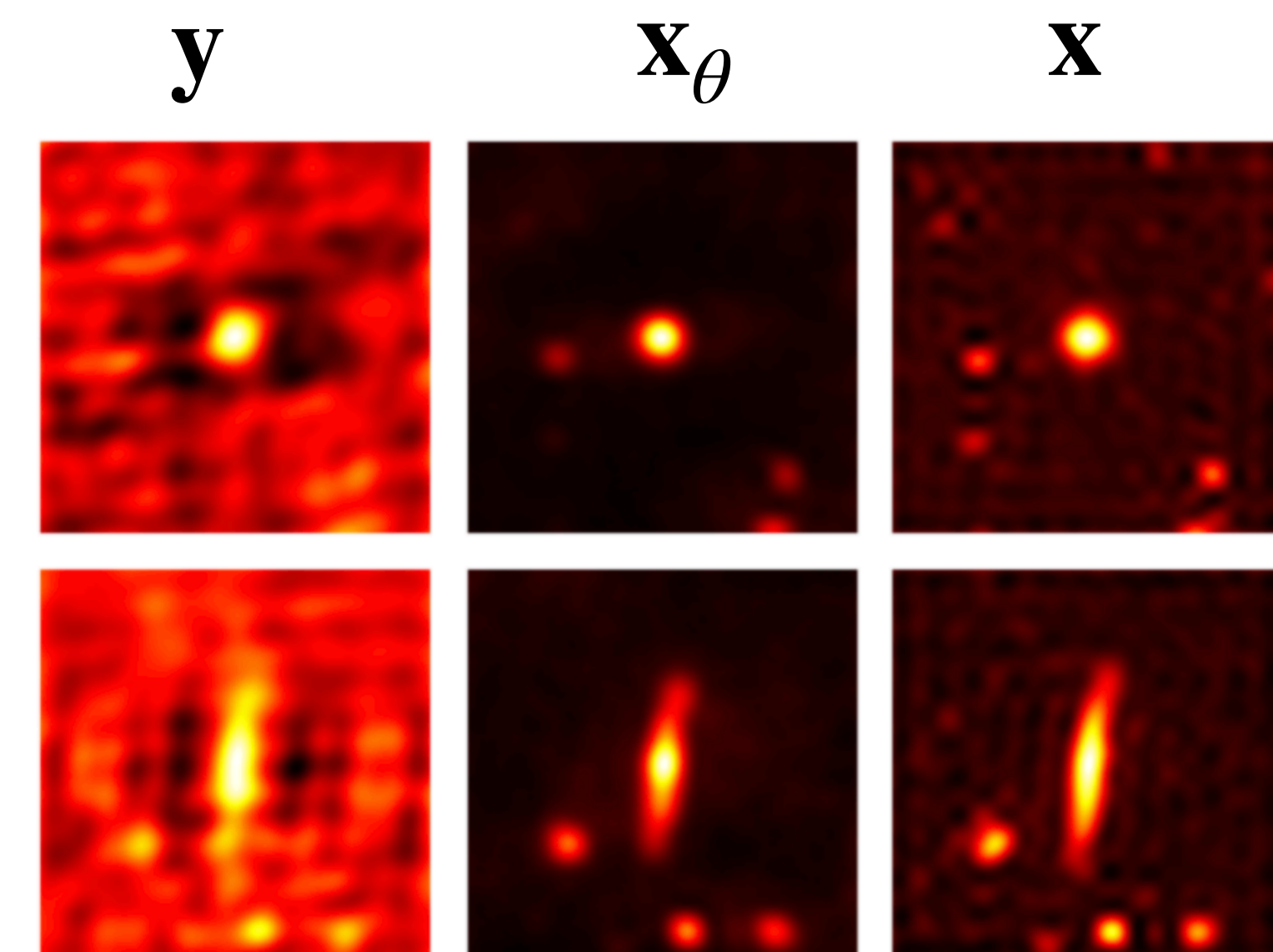
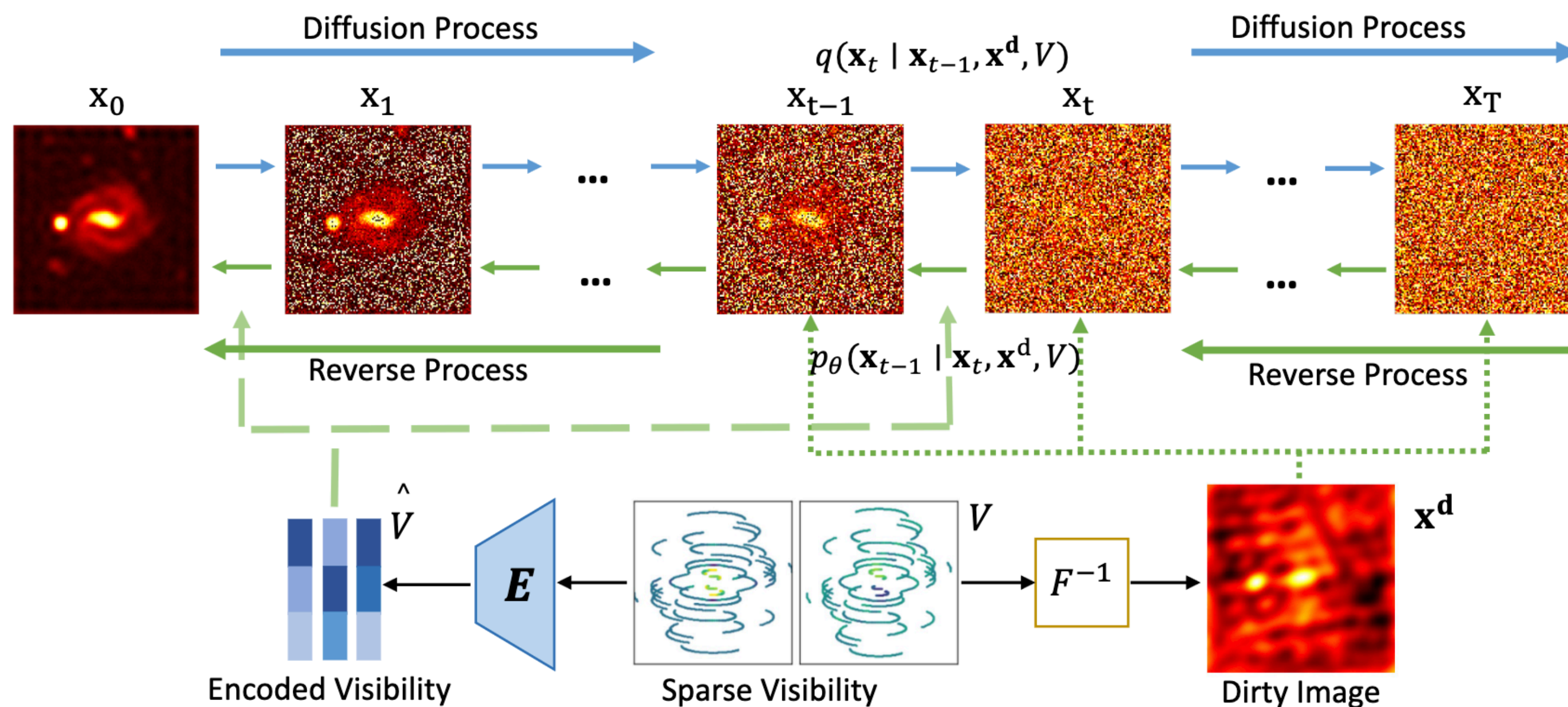


Radio-astronomical image reconstruction with a conditional denoising diffusion model[★]

See also:

M. Drozdova¹, V. Kinakh¹, O. Bait², O. Taran¹, E. Lastufka¹, M. Dessauges-Zavadsky², T. Holotyak¹, D. Schaerer², and S. Voloshynovskiy¹

DENOISING DIFFUSION PROBABLISTIC MODELS (DDPM)



Problem: Network learns $\mathbf{S}$, but $\mathbf{S}$ depends on instrument, observation time, pointing direction, etc etc => retrain for every new instrument & observation configuration

DENOISING DIFFUSION RESTORATION MODELS (DDRM)

Can we have the best of both worlds?

Goal: Use DDPM to learn a data-driven prior of the sky, and reconstruct images without needing to learn the instrument configuration

Denoising Diffusion Restoration Models

arXiv: 2201.11793

Bahjat Kawar

Department of Computer Science
Technion, Haifa, Israel
bahjat.kawar@cs.technion.ac.il

Michael Elad

Department of Computer Science
Technion, Haifa, Israel
elad@cs.technion.ac.il

Stefano Ermon

Department of Computer Science
Stanford, California, USA
ermon@cs.stanford.edu

Jiaming Song

NVIDIA
Santa Clara, California, USA
jiamings@nvidia.com



Radio-Interferometric Image Reconstruction with Denoising Diffusion Restoration Models

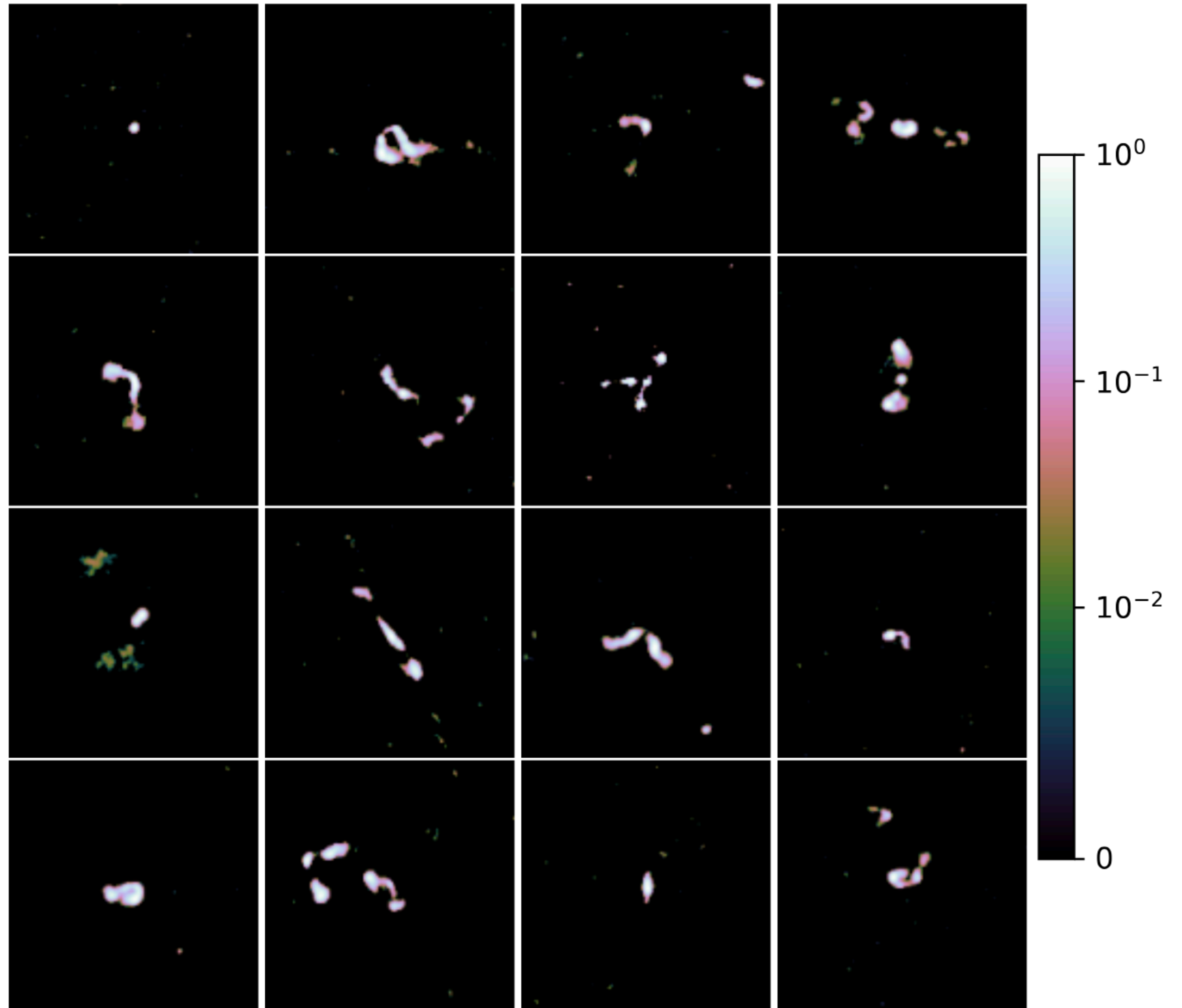
M. Morales,¹ E. Tolley,¹★ R. Poitevineau¹

¹ *Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), 1290 Sauverny, Switzerland*

DDRM FOR RADIO ASTRONOMY

First, train a **DDPM** on ~20k
150x150 radio galaxy image
cutouts from radio galaxy zoo DR1
(arXiv: 2412.14502)

Reserve 200 for validation
& 200 for test



DDRM FOR RADIO ASTRONOMY

The radio astronomy forward model: $\mathbf{y} = \mathbf{S}\mathcal{F}\mathbf{x} + \mathbf{z}$

Start by calculating SVD of $\mathbf{S}$:

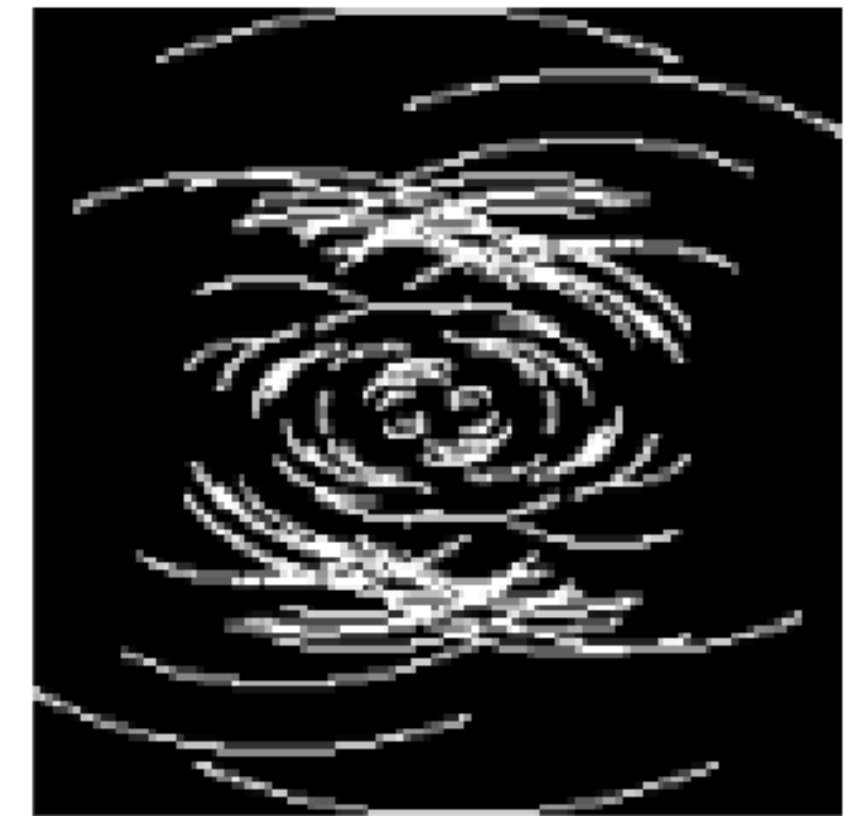
$$\mathbf{S} = \mathbf{I}\mathbf{\Sigma}\mathbf{P}$$

identity

Permutation

Rectangular diagonal matrix of
the **singular values** s_i

Sampling matrix

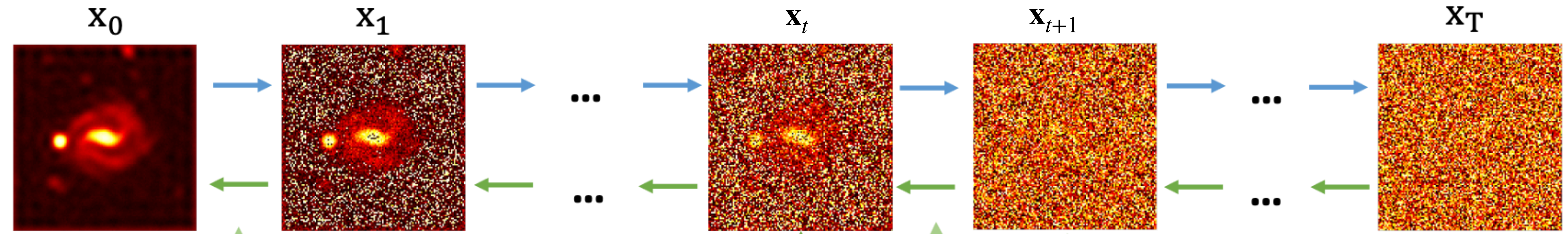


Transform the measurement and and outputs:

$$\bar{\mathbf{x}} \equiv \mathbf{P}\mathcal{F}\mathbf{x} \quad \bar{\mathbf{y}} \equiv \mathbf{\Sigma}^+\mathbf{y}$$

Implement $\mathcal{F}$ using PyTorch FFT instead of as a matrix

DDRM FOR RADIO ASTRONOMY



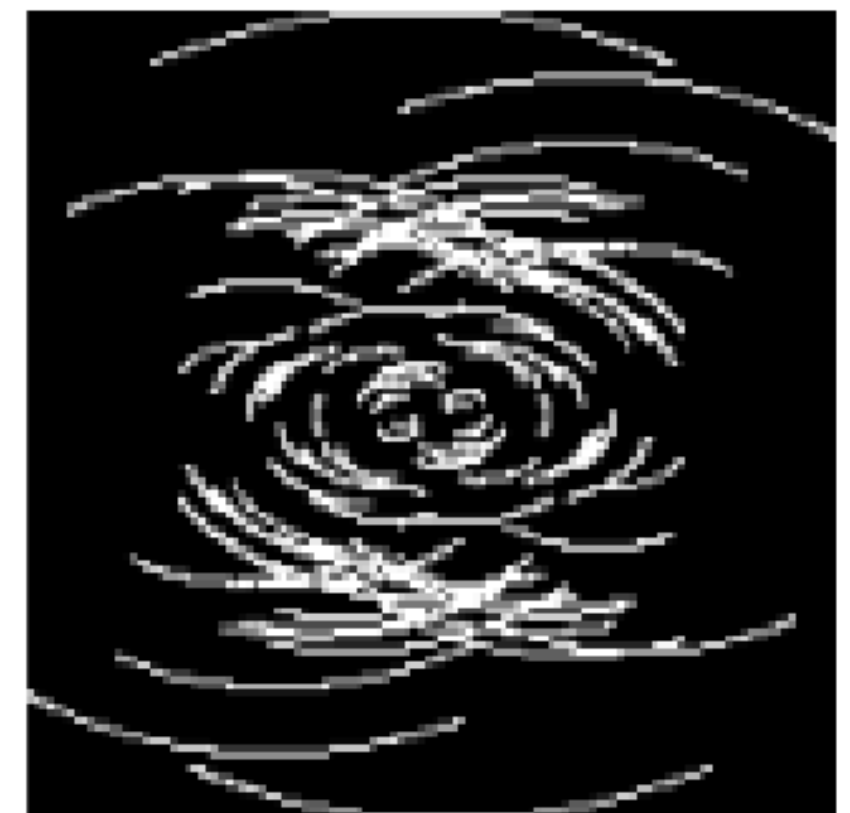
We have our pre-trained DDPM: $f_{\theta}(\mathbf{x}_{t+1}) = \mathbf{x}_{\theta,t}$

Transform: $\bar{\mathbf{x}}_{\theta,t} = \mathbf{P}\mathcal{F}\mathbf{x}_{\theta,t}$

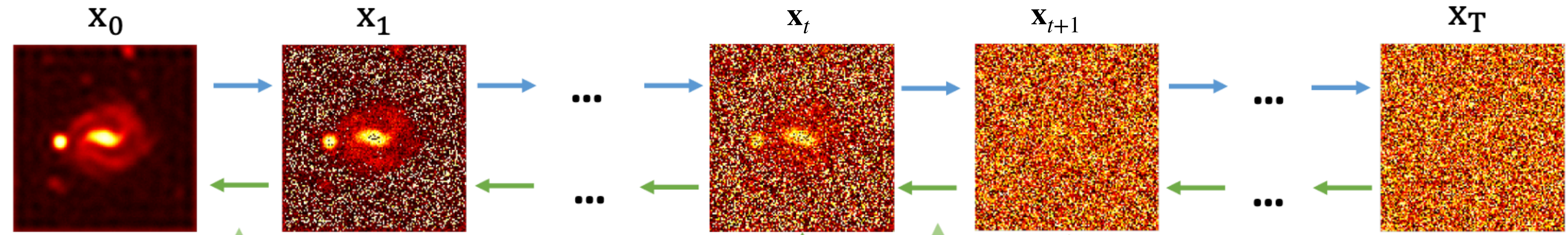
Update using the observation $\mathbf{y}$:

$$\bar{\mathbf{x}}_{\theta,t}^{(i)} = \left\{ \begin{array}{l} \end{array} \right.$$

Sampling matrix



DDRM FOR RADIO ASTRONOMY



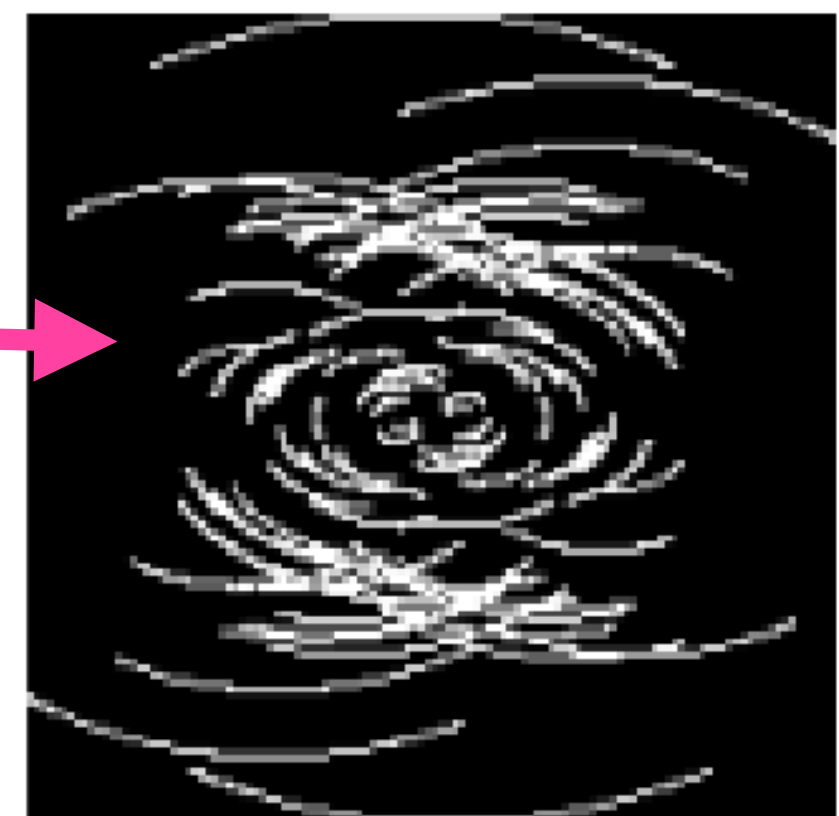
We have our pre-trained DDPM: $f_{\theta}(\mathbf{x}_{t+1}) = \mathbf{x}_{\theta,t}$

Transform: $\bar{\mathbf{x}}_{\theta,t} = \mathbf{P}\mathcal{F}\mathbf{x}_{\theta,t}$

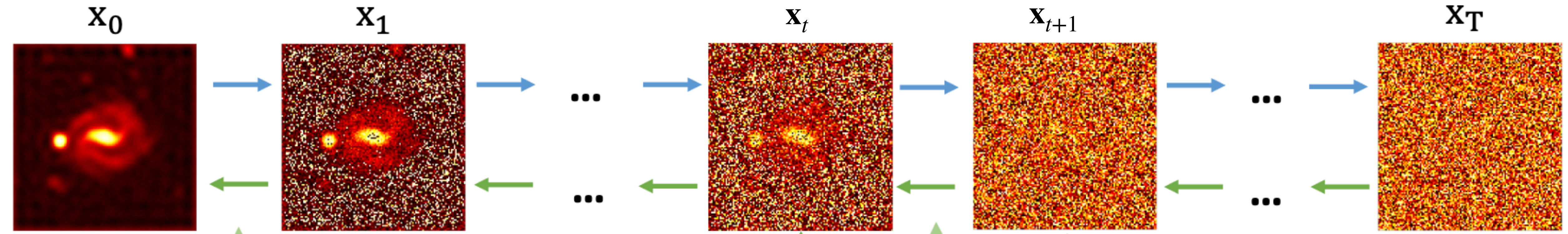
Update using the observation $\mathbf{y}$:

$$\bar{\mathbf{x}}_{\theta,t}^{(i)} = \begin{cases} \bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}} & \text{if } s_i = 0 \end{cases}$$

Sampling matrix



DDRM FOR RADIO ASTRONOMY



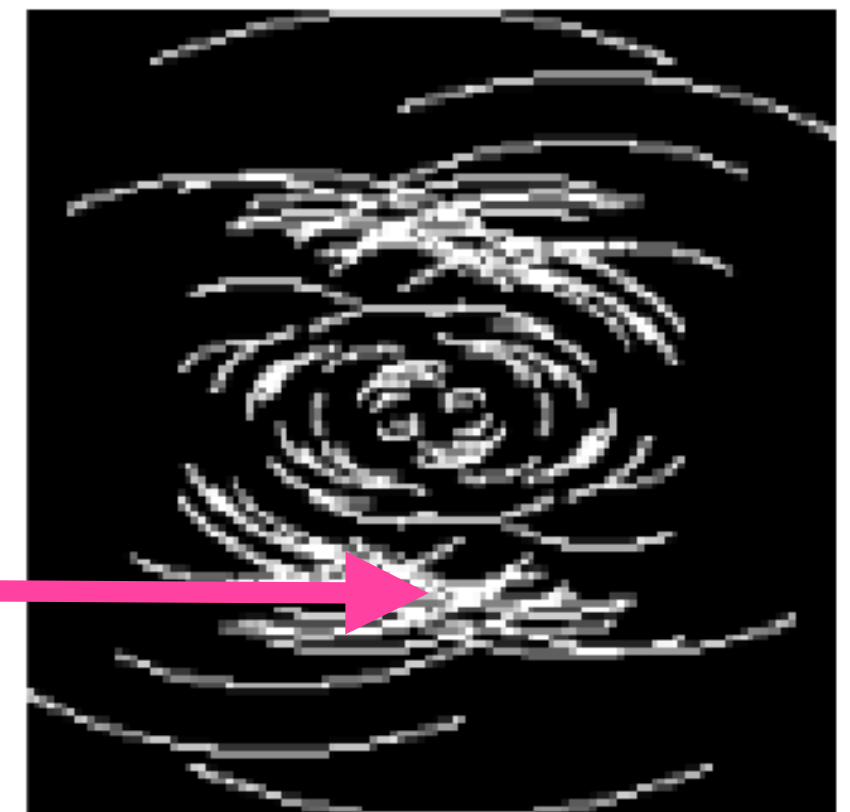
We have our pre-trained DDPM: $f_{\theta}(\mathbf{x}_{t+1}) = \mathbf{x}_{\theta,t}$

Transform: $\bar{\mathbf{x}}_{\theta,t} = \mathbf{P}\mathcal{F}\mathbf{x}_{\theta,t}$

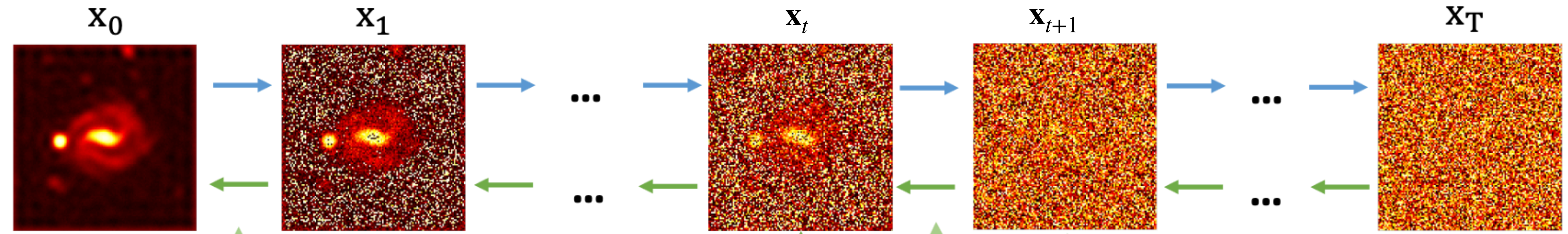
Update using the observation $\mathbf{y}$:

$$\bar{\mathbf{x}}_{\theta,t}^{(i)} = \begin{cases} \bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}} & \text{if } s_i = 0 \\ \bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{y}}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{\mathbf{y}}/s_i} & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_i} \end{cases}$$

Sampling matrix



DDRM FOR RADIO ASTRONOMY



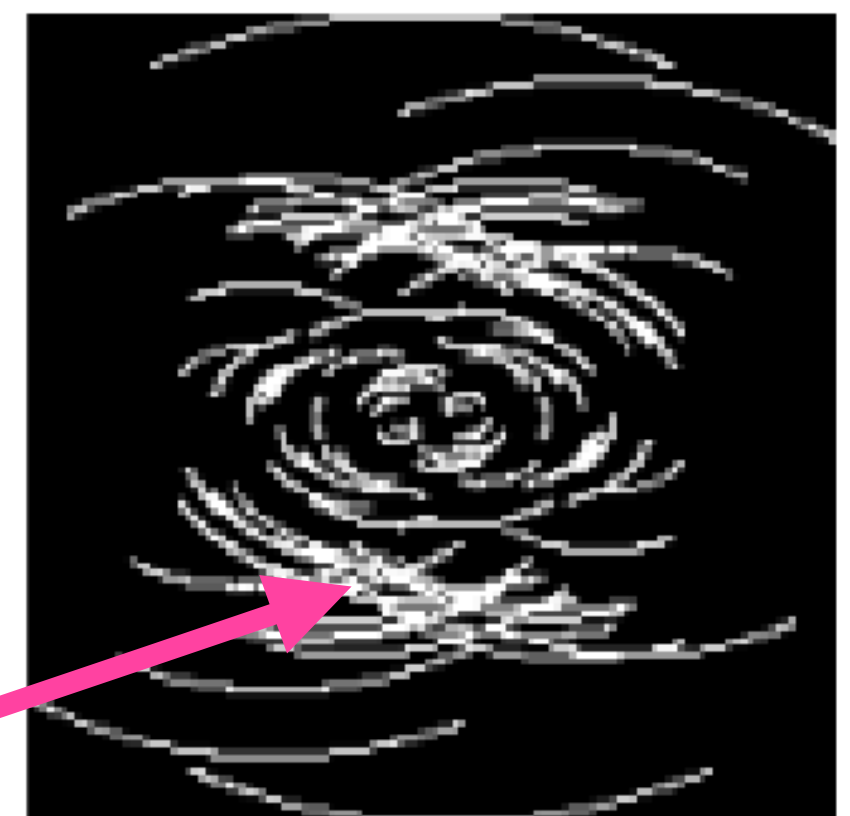
We have our pre-trained DDPM: $f_{\theta}(\mathbf{x}_{t+1}) = \mathbf{x}_{\theta,t}$

Transform: $\bar{\mathbf{x}}_{\theta,t} = \mathbf{P}\mathcal{F}\mathbf{x}_{\theta,t}$

Update using the observation $\mathbf{y}$:

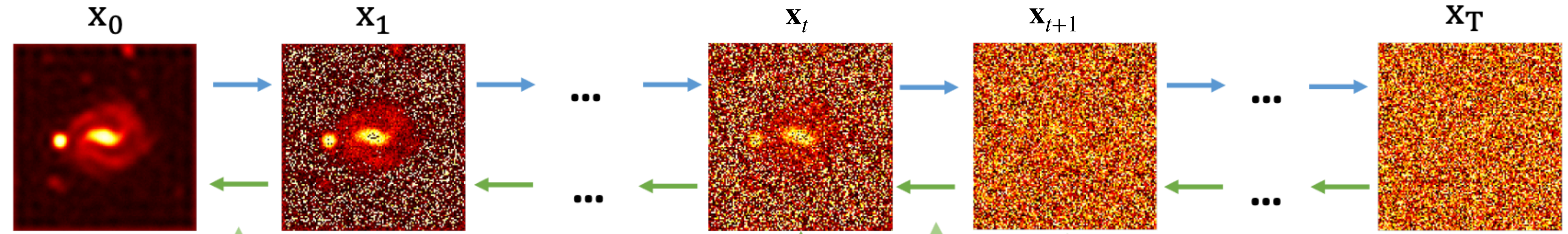
$$\bar{\mathbf{x}}_{\theta,t}^{(i)} = \begin{cases} \bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}} & \text{if } s_i = 0 \\ \bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{y}}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{\mathbf{y}}/s_i} & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_i} \\ (1 - \eta_b) \bar{\mathbf{x}}_{\theta,t}^{(i)} + \eta_b \bar{\mathbf{y}}^{(i)} & \text{if } \sigma_t \geq \frac{\sigma_{\mathbf{y}}}{s_i} \end{cases}$$

Sampling matrix



We use $\eta_b = 1$ to prefer the observation

DDRM FOR RADIO ASTRONOMY



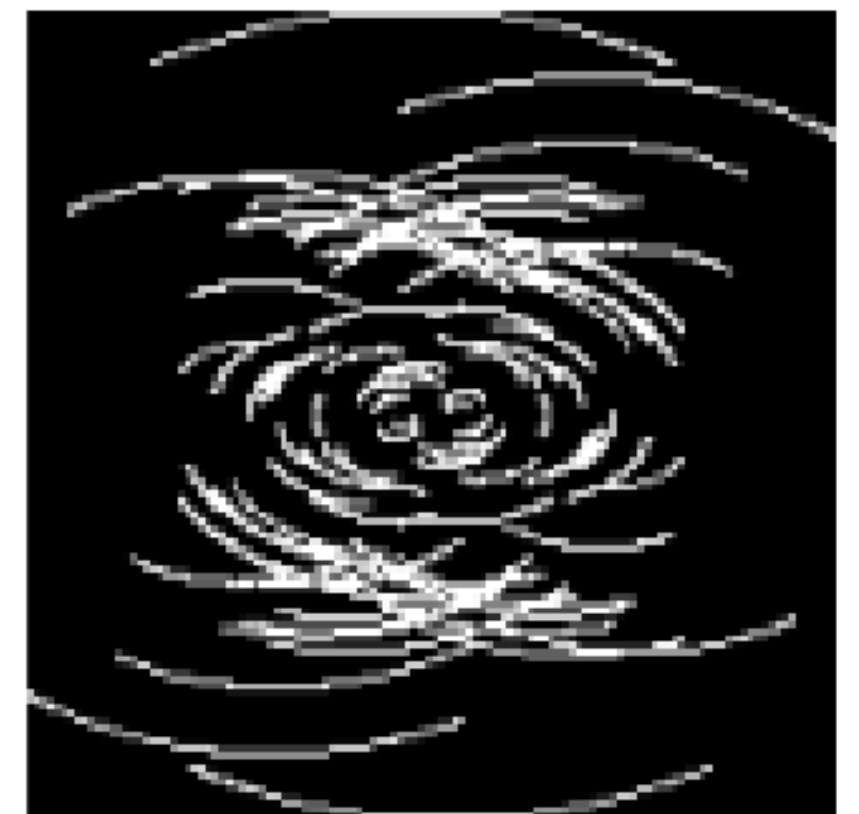
We have our pre-trained DDPM: $f_{\theta}(\mathbf{x}_{t+1}) = \mathbf{x}_{\theta,t}$

Transform: $\bar{\mathbf{x}}_{\theta,t} = \mathbf{P}\mathcal{F}\mathbf{x}_{\theta,t}$

Update using the observation $\mathbf{y}$:

$$\bar{\mathbf{x}}_{\theta,t}^{(i)} = \begin{cases} \bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}} & \text{if } s_i = 0 \\ \bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{y}}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{\mathbf{y}}/s_i} & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_i} \\ (1 - \eta_b) \bar{\mathbf{x}}_{\theta,t}^{(i)} + \eta_b \bar{\mathbf{y}}^{(i)} & \text{if } \sigma_t \geq \frac{\sigma_{\mathbf{y}}}{s_i} \end{cases}$$

Sampling matrix



Transform back: $\mathbf{x}_{\theta,t} = \mathcal{F}^{-1}\mathbf{P}^{\top}\bar{\mathbf{x}}_{\theta,t}$

DDRM FOR RADIO ASTRONOMY

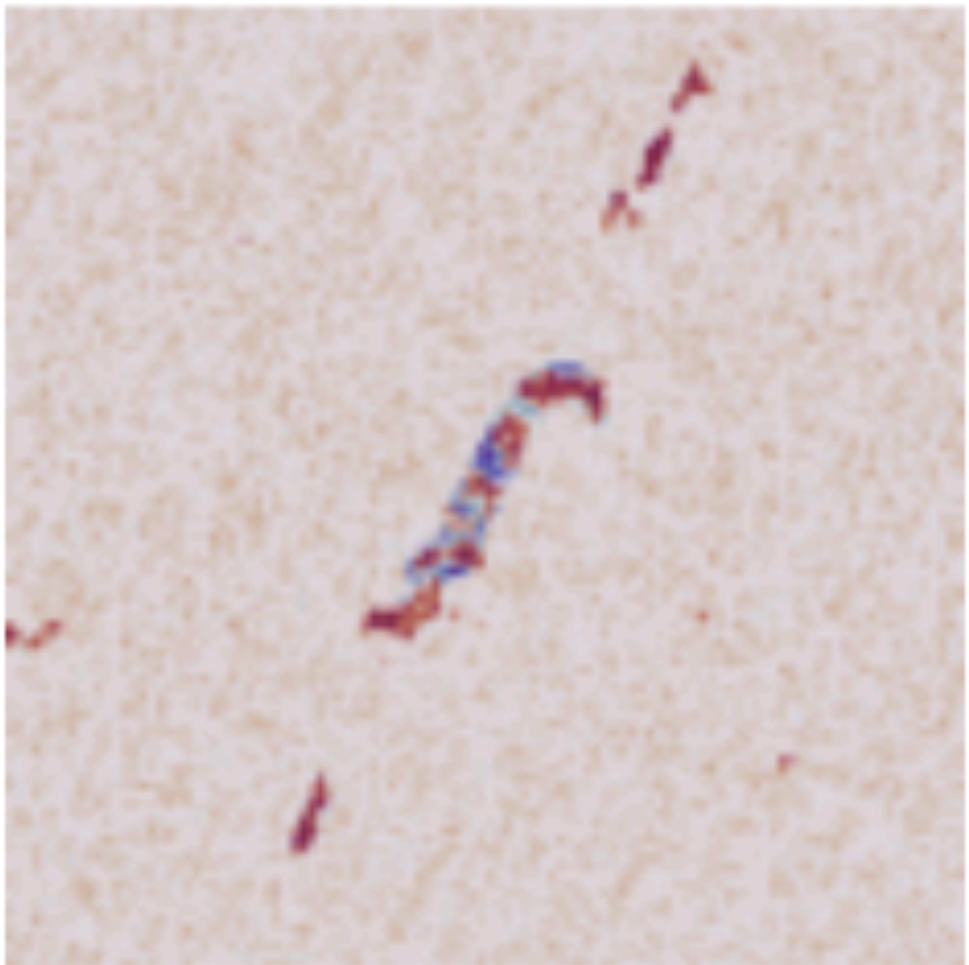
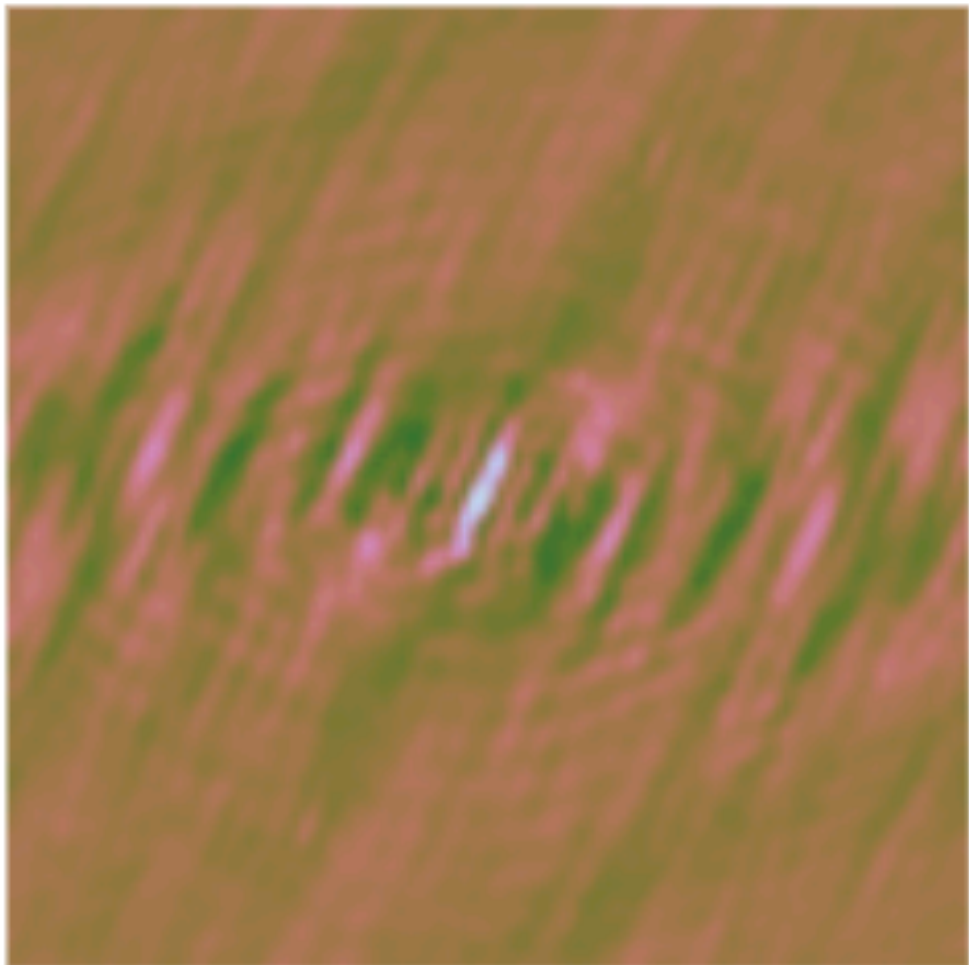
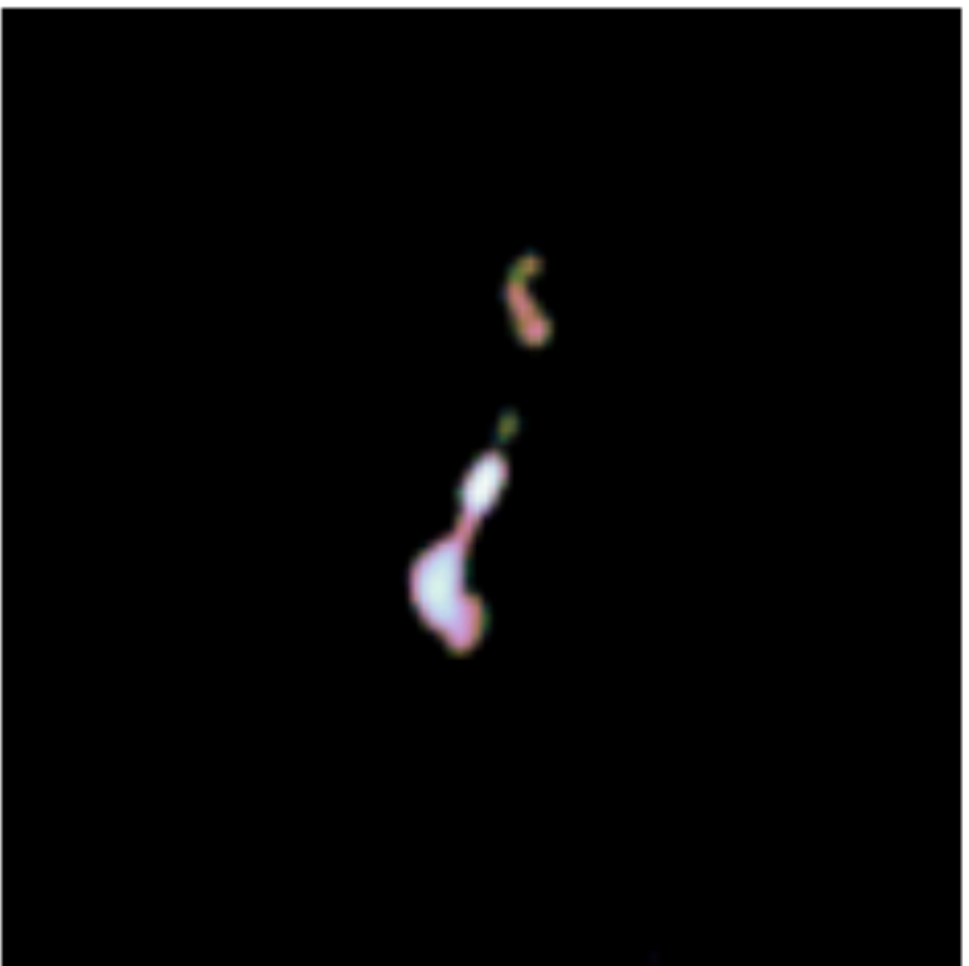
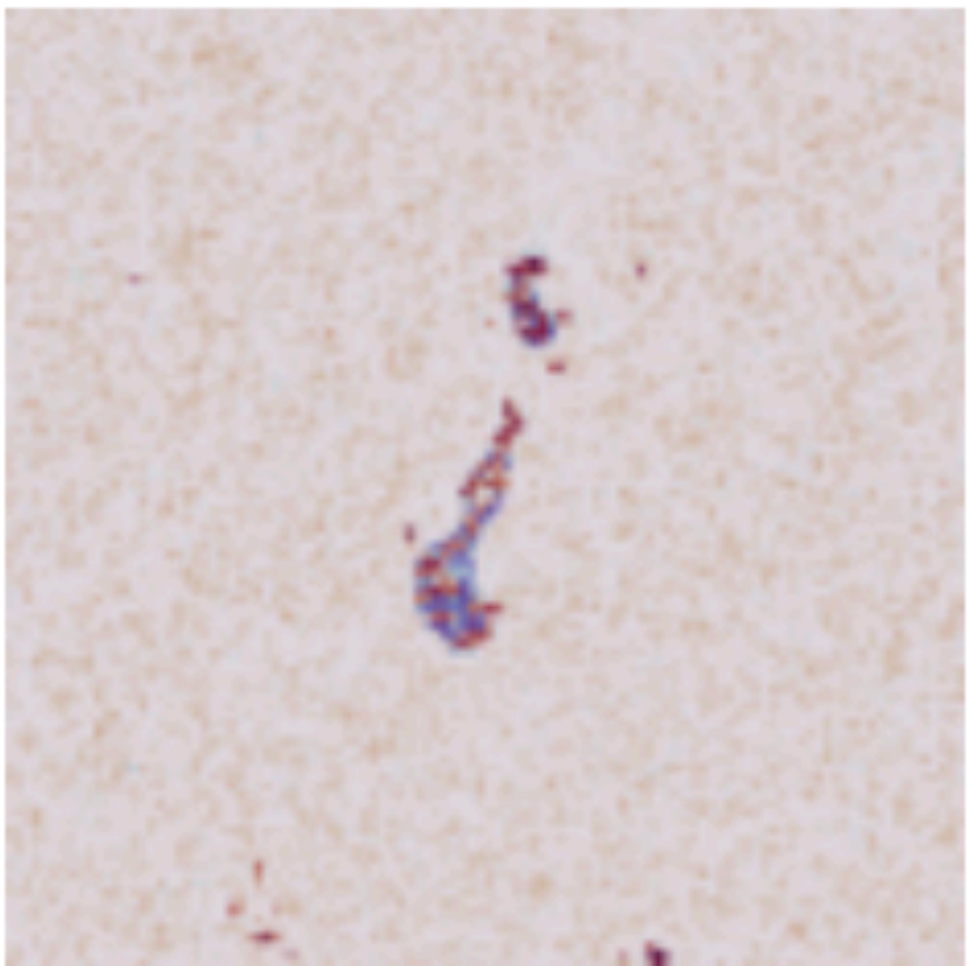
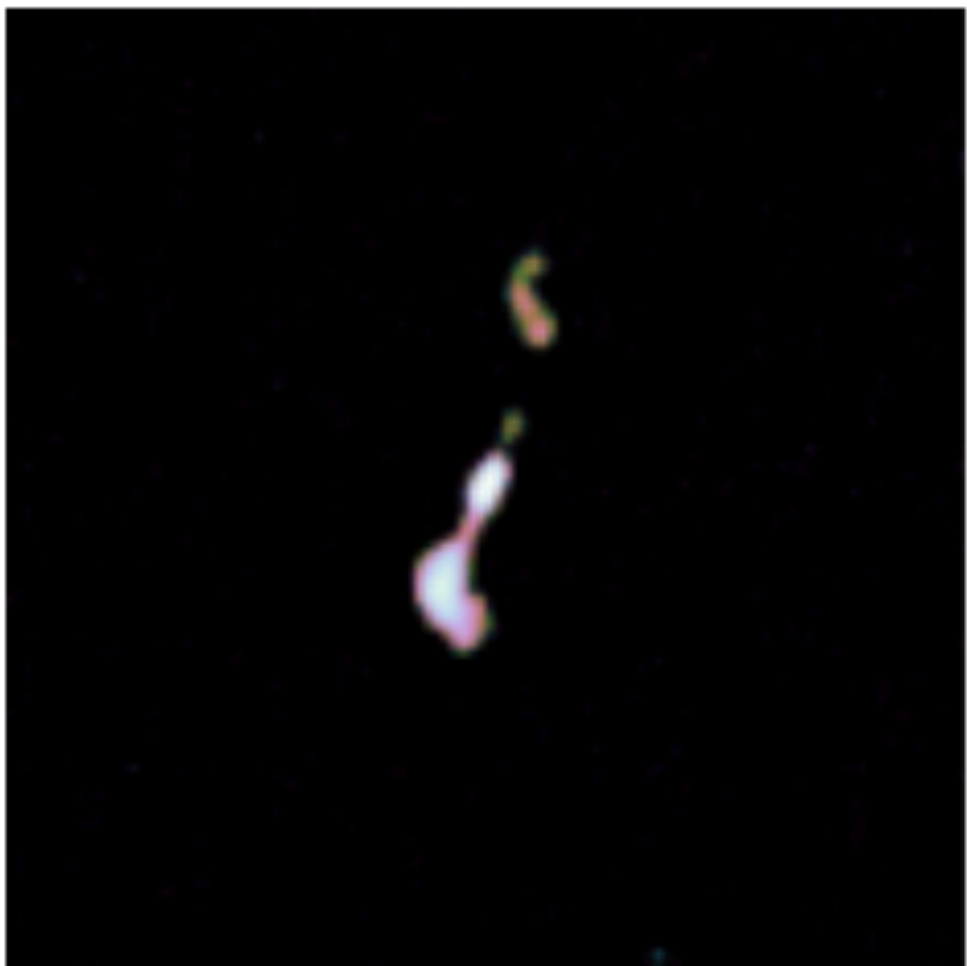
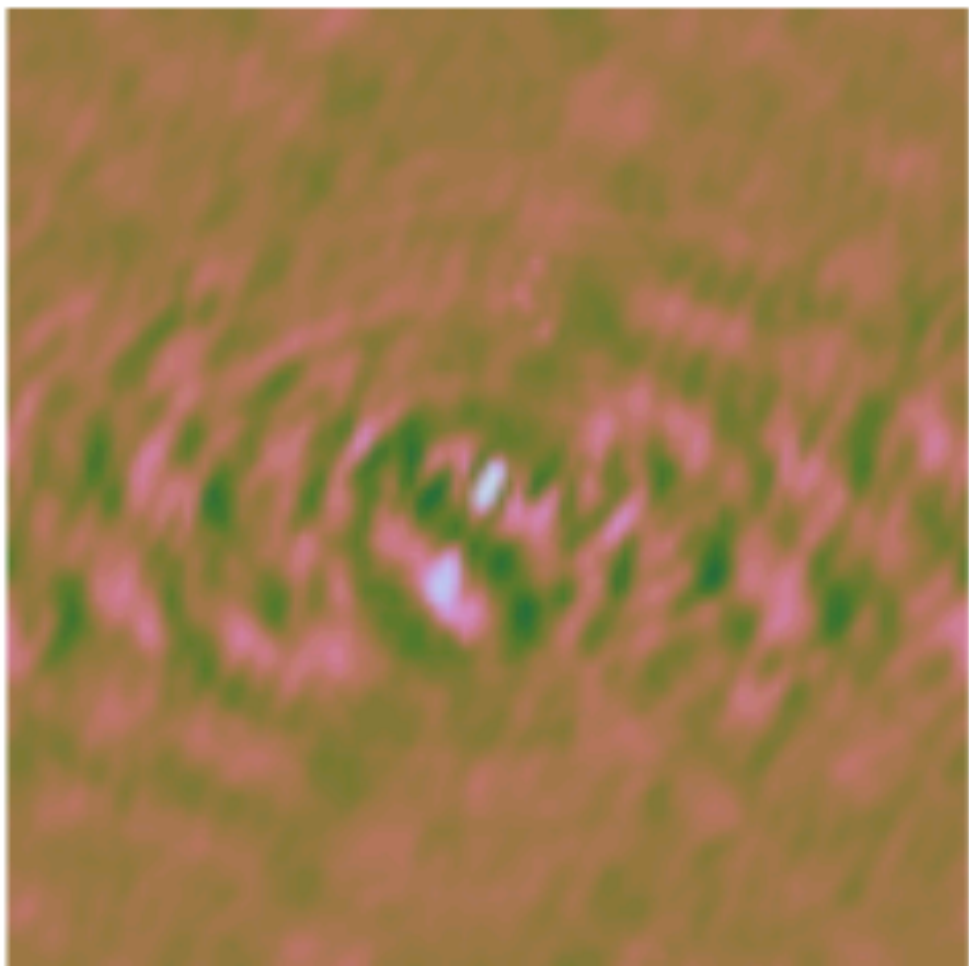
True image

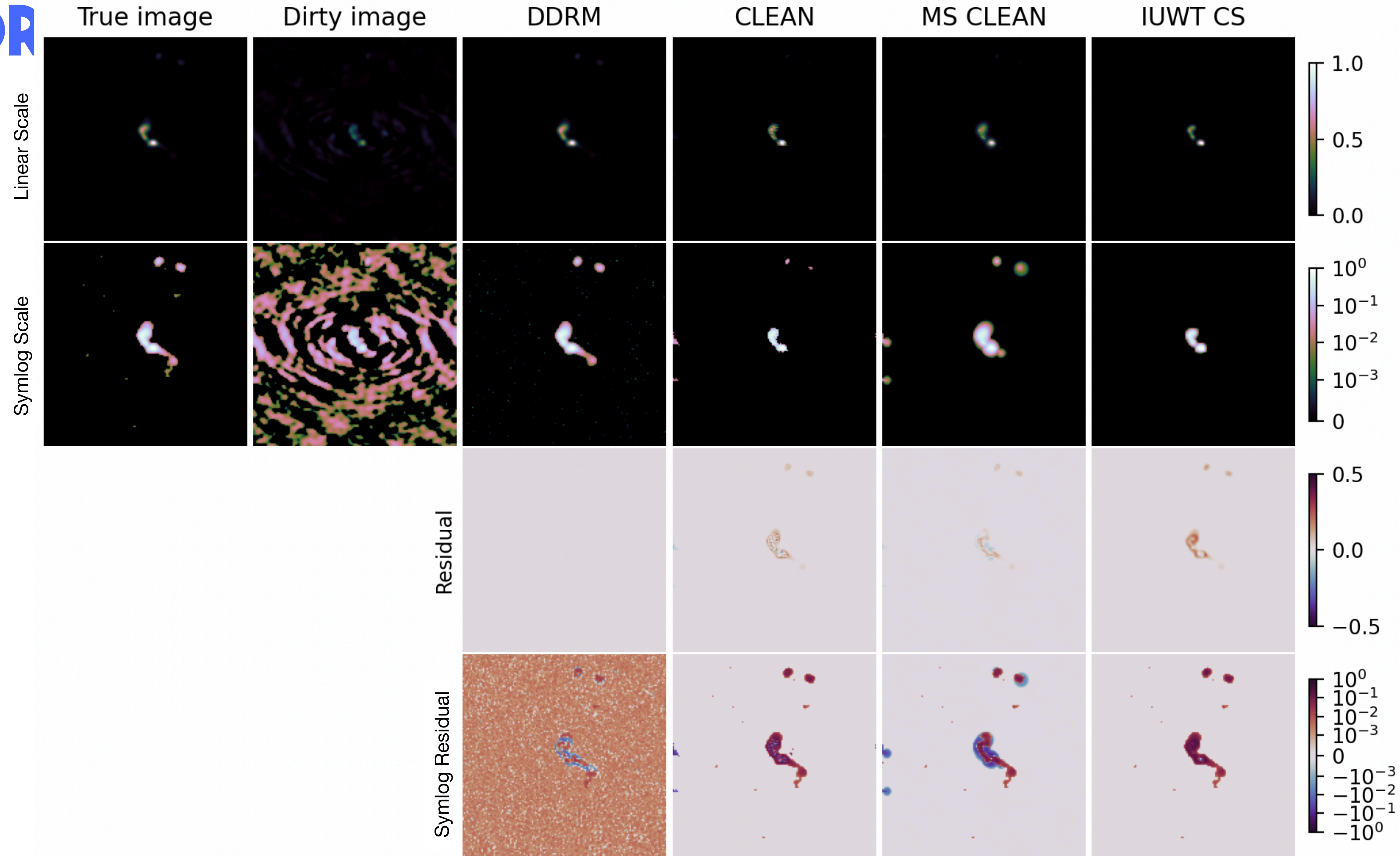
Dirty image

DDRM

True - DDRM

DDRM mean





True image

Dirty image

DDRM

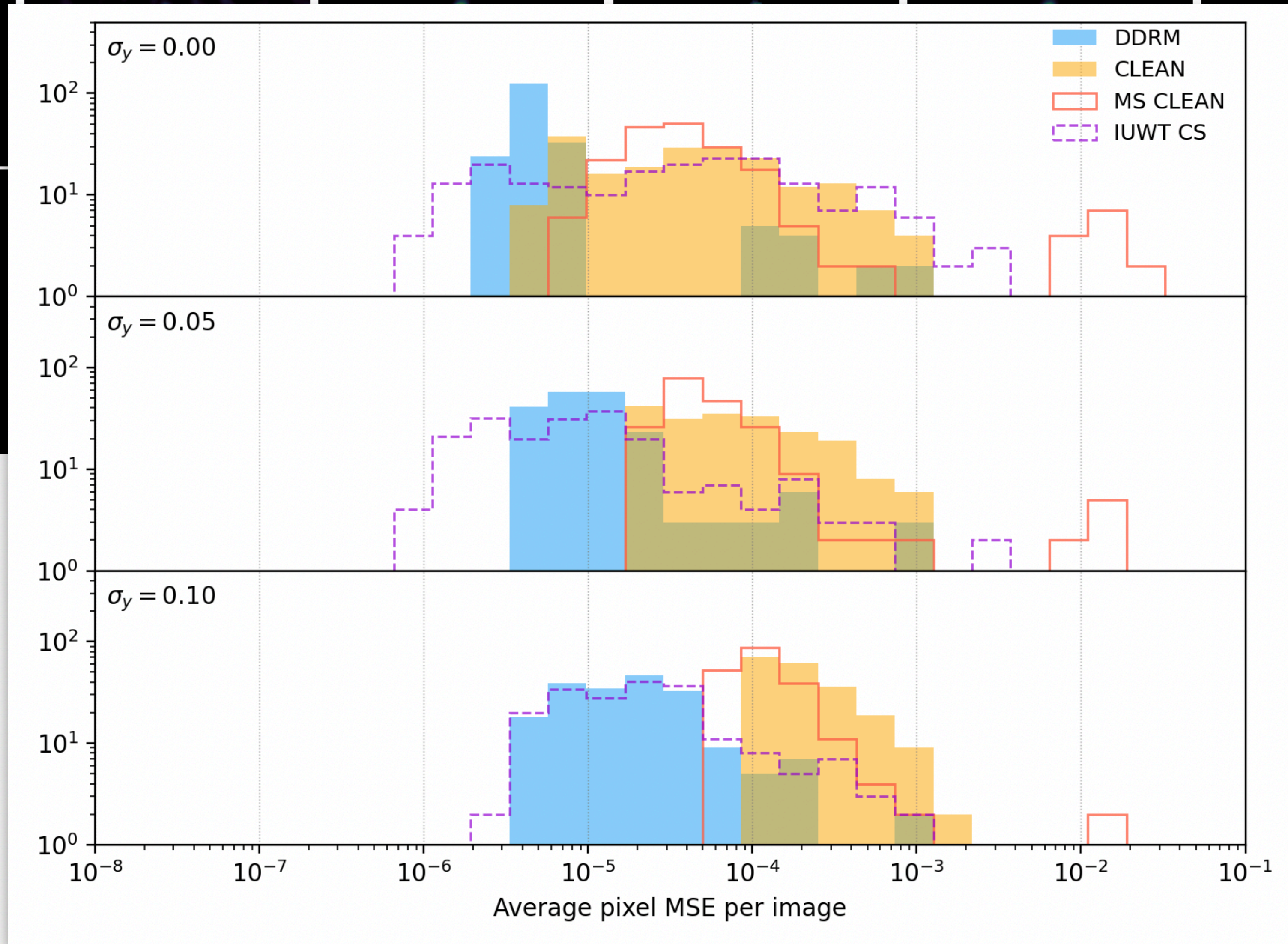
CLEAN

MS CLEAN

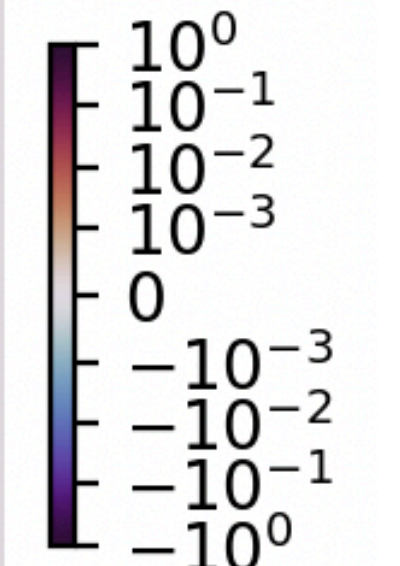
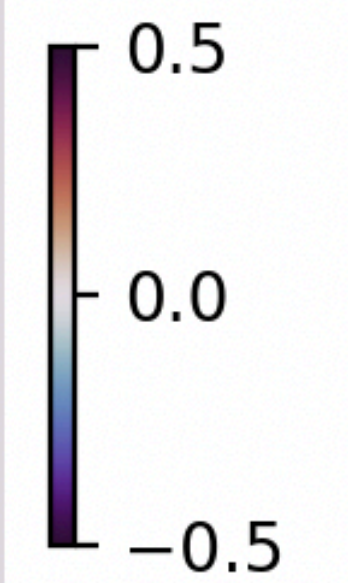
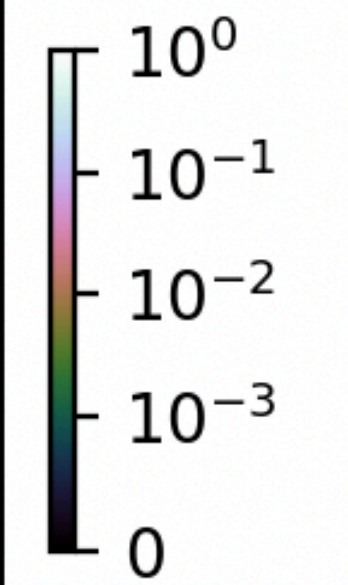
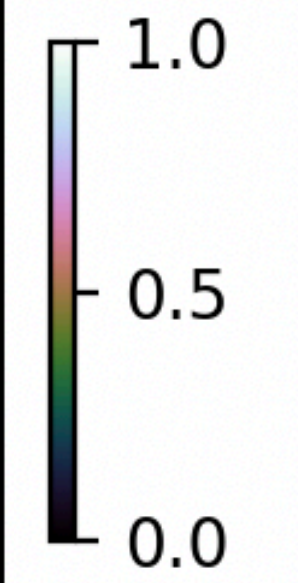
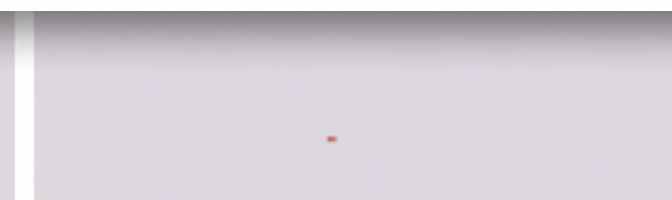
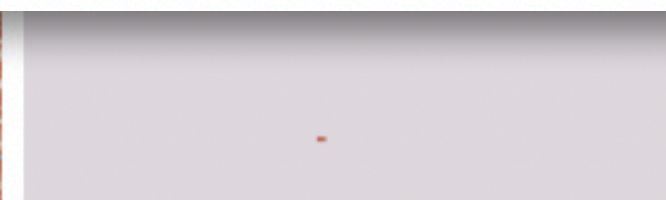
IUWT CS

Linear Scale

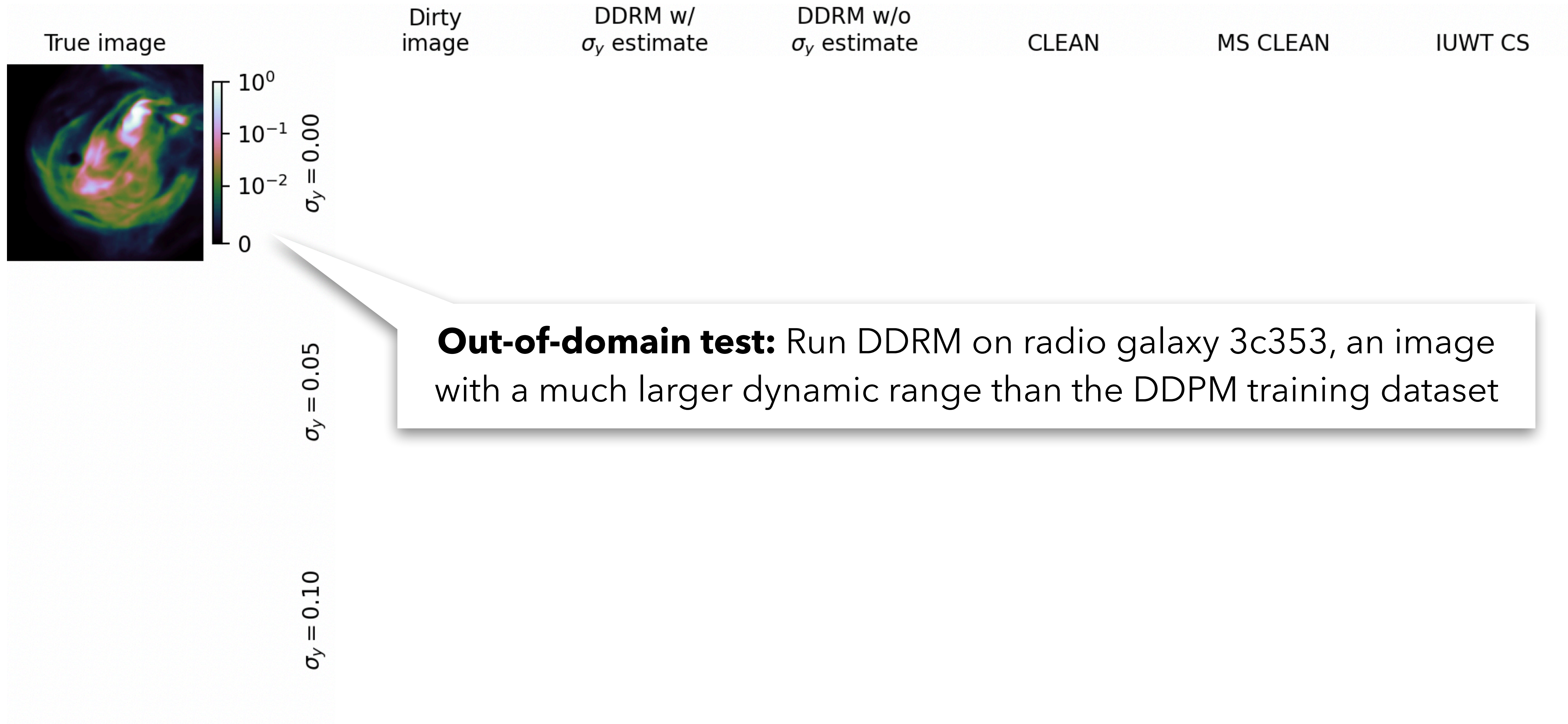
Symlog Scale



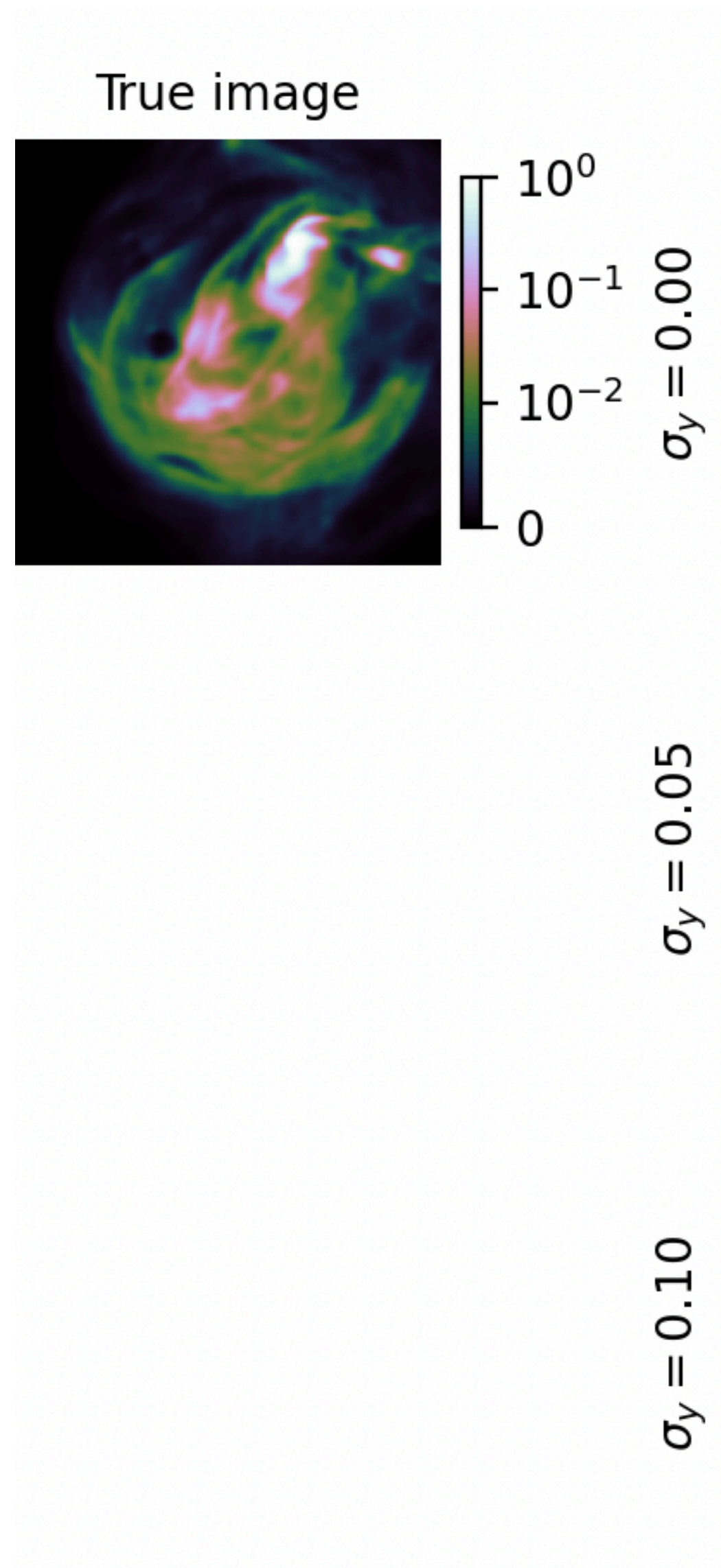
Sym



DDRM FOR RADIO ASTRONOMY



DDRM FOR RADIO ASTRONOMY



Dirty
image

DDRM w/
 σ_y estimate

DDRM w/o
 σ_y estimate

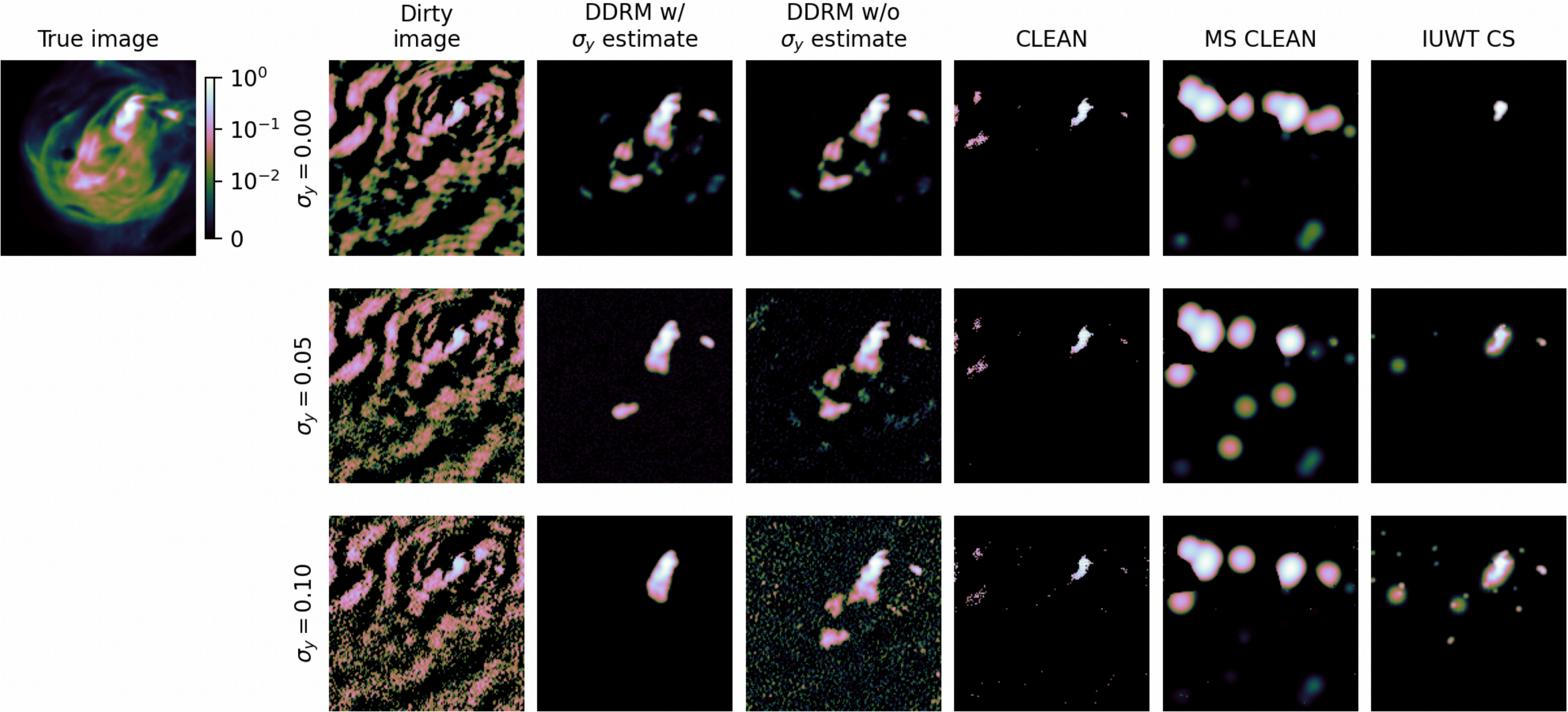
CLEAN

MS CLEAN

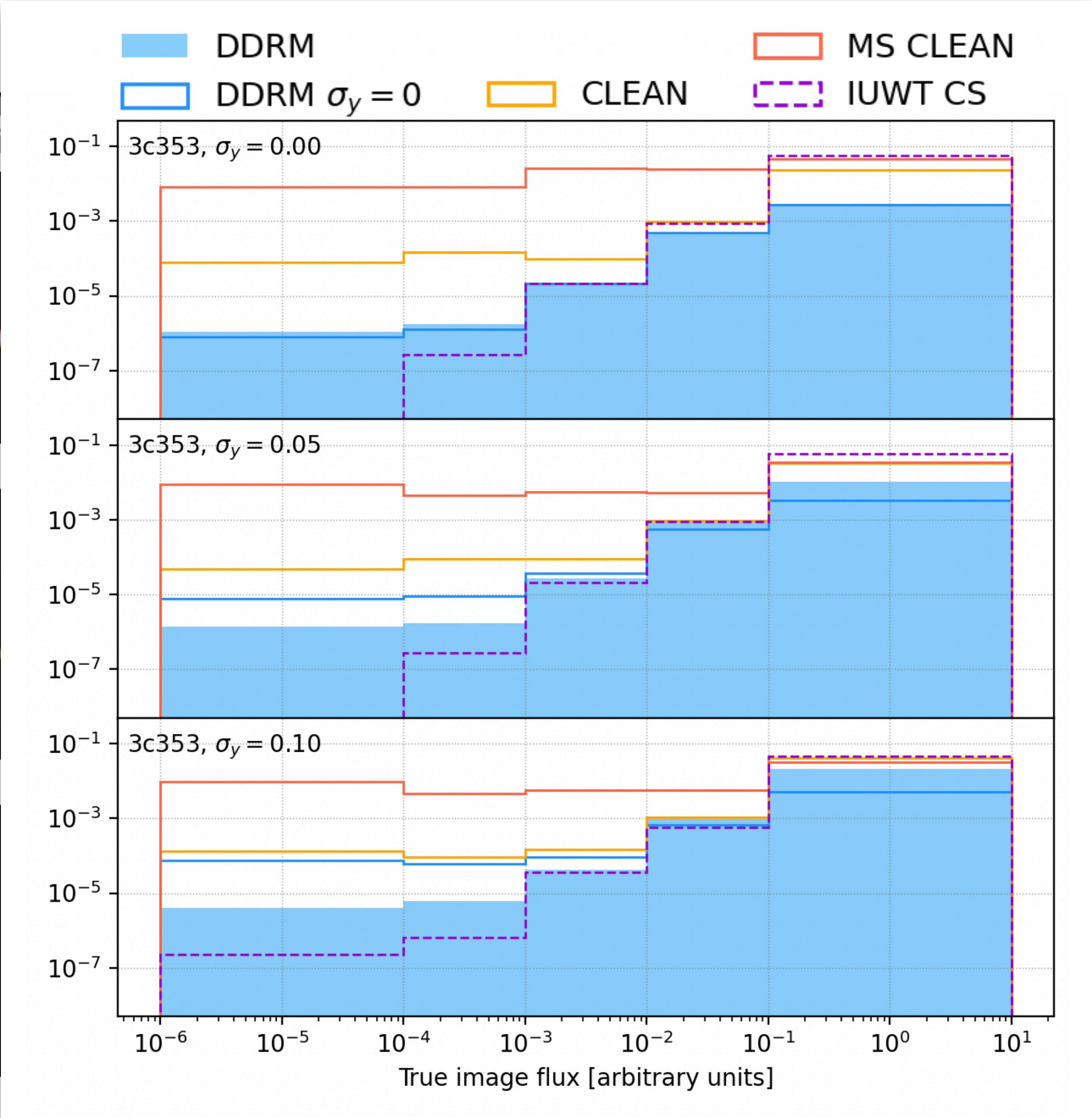
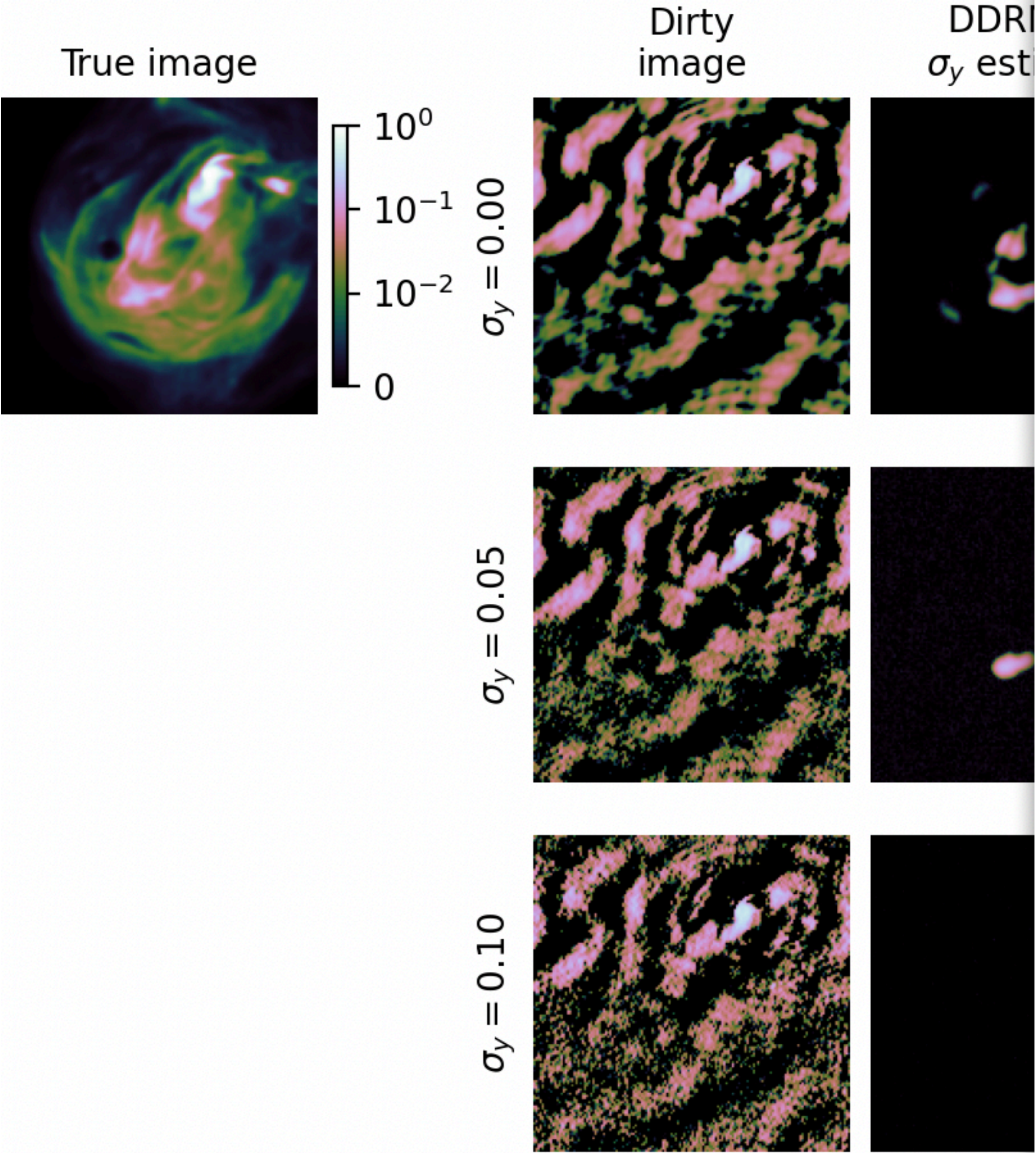
IUWT CS

Incorrect parameter test: Run DDRM with the incorrect observation noise parameter σ_y

DDRM FOR RADIO ASTRONOMY

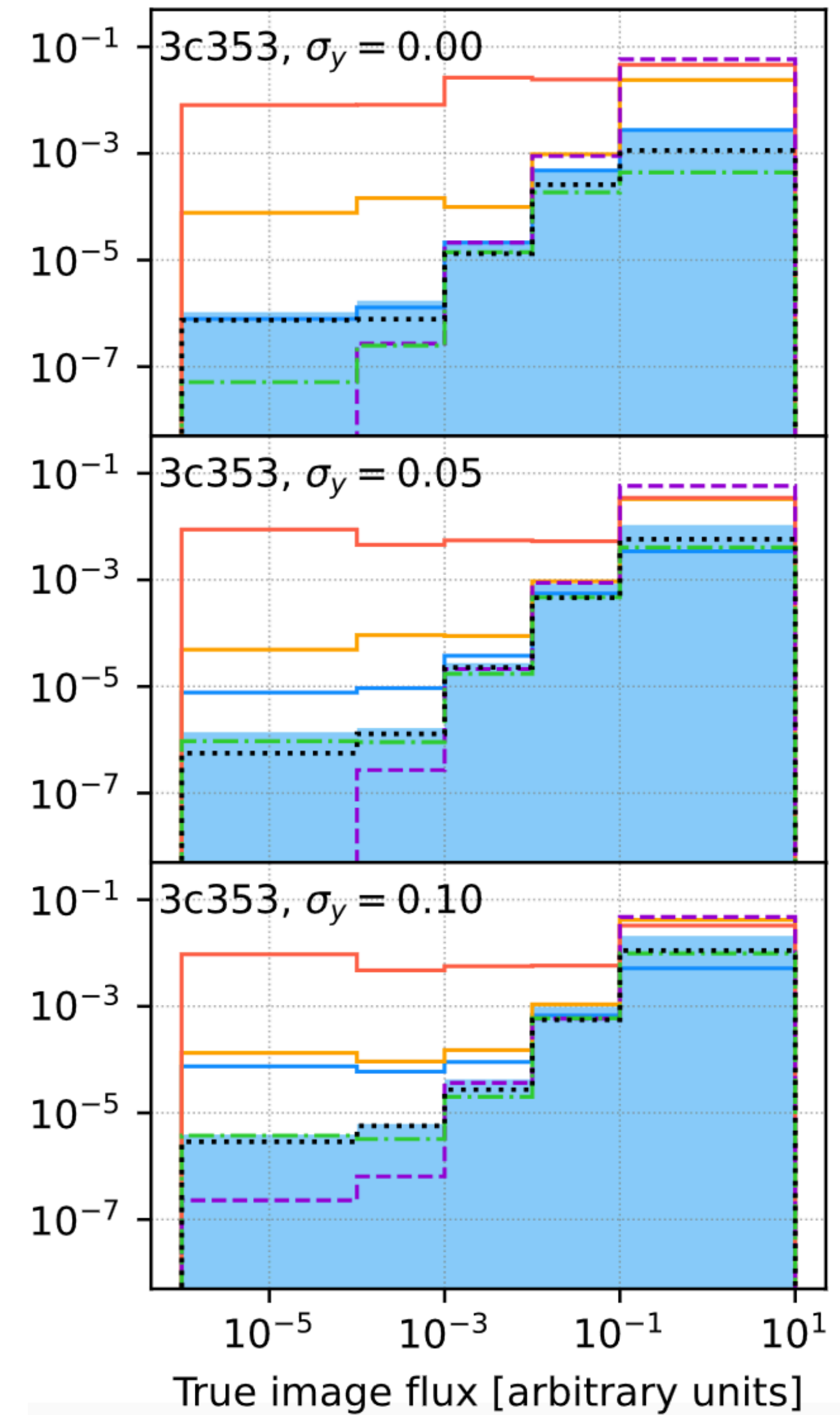
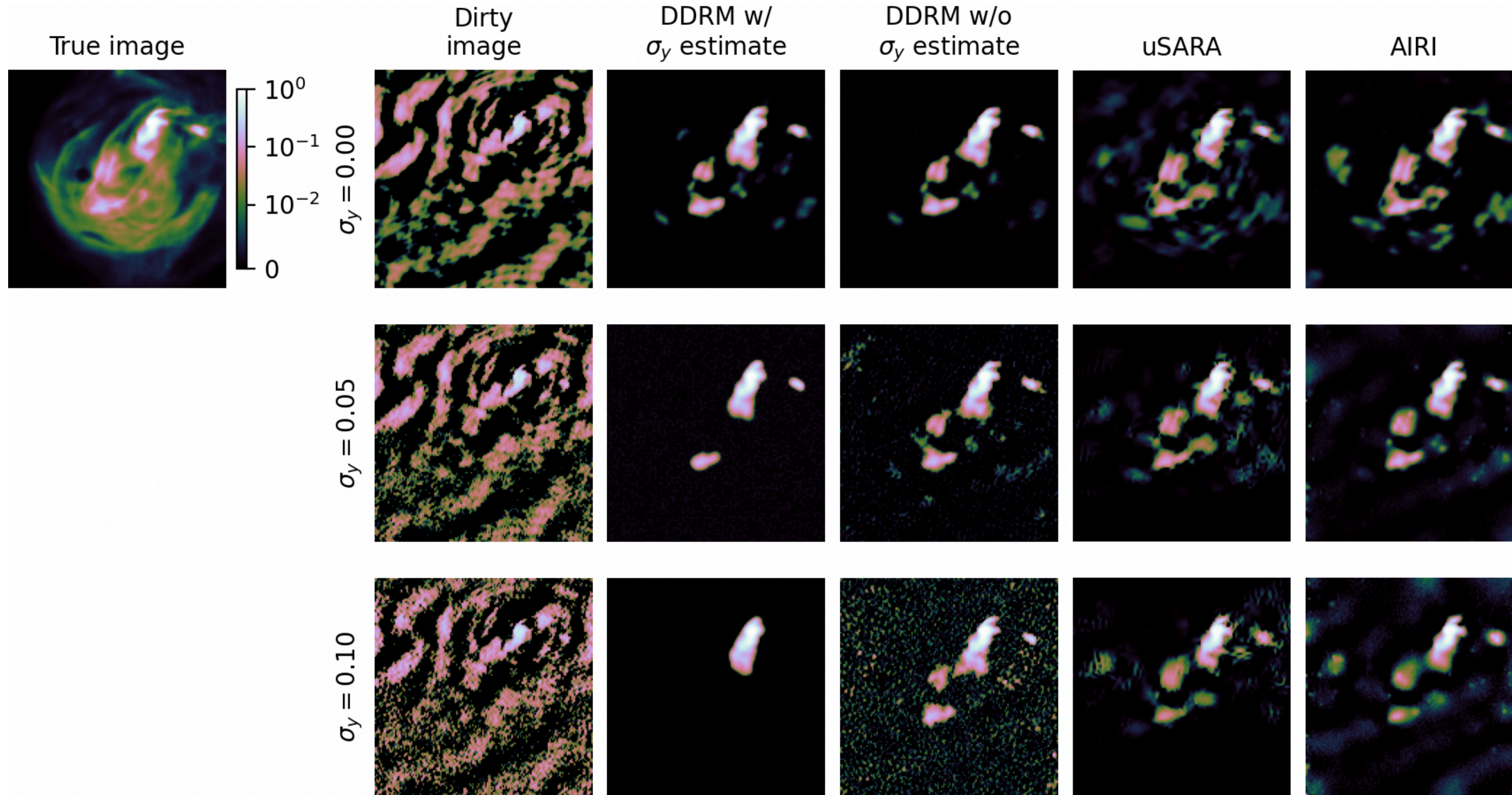


DDRM FOR RADIO ASTRONOMY

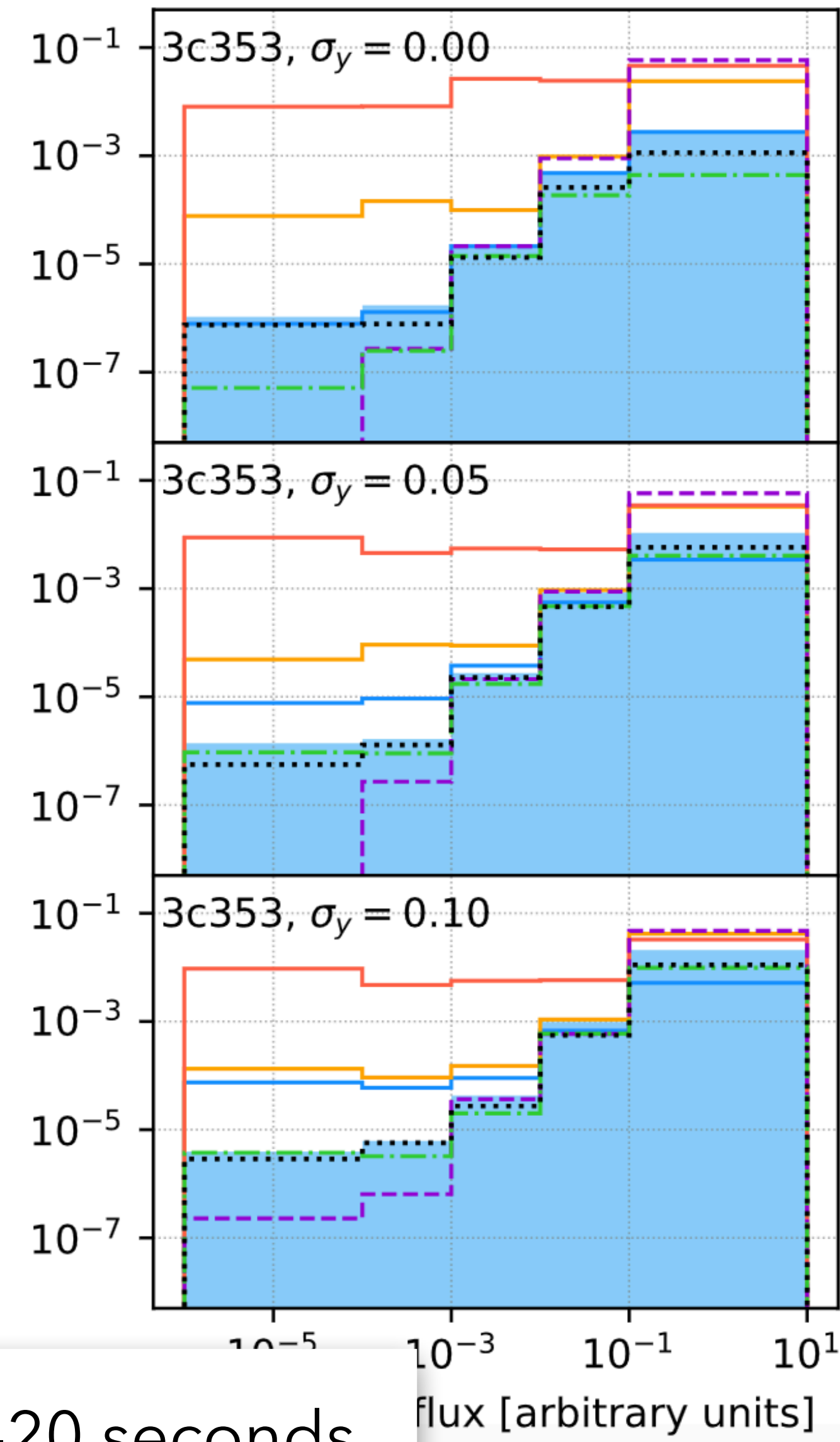
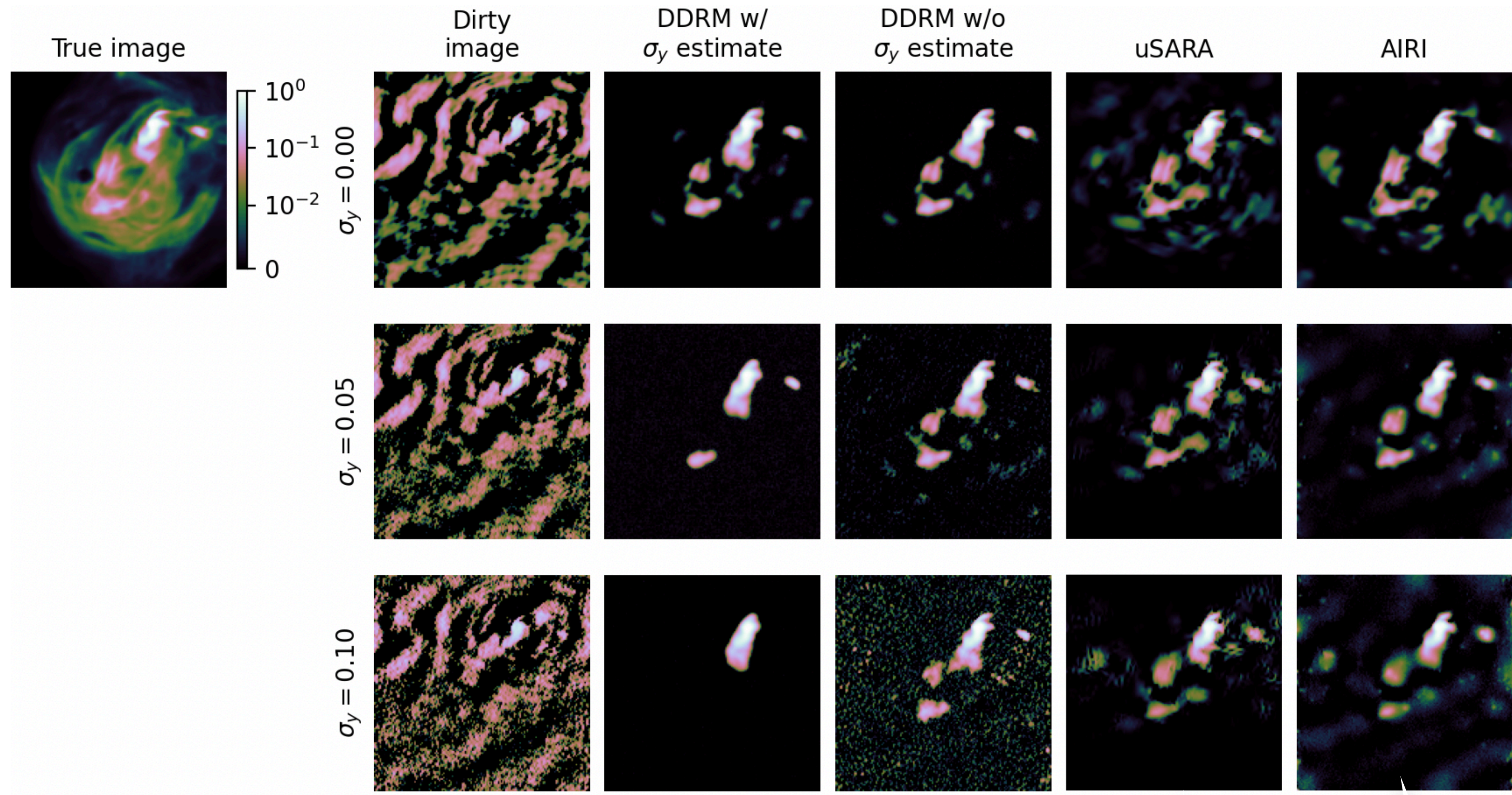
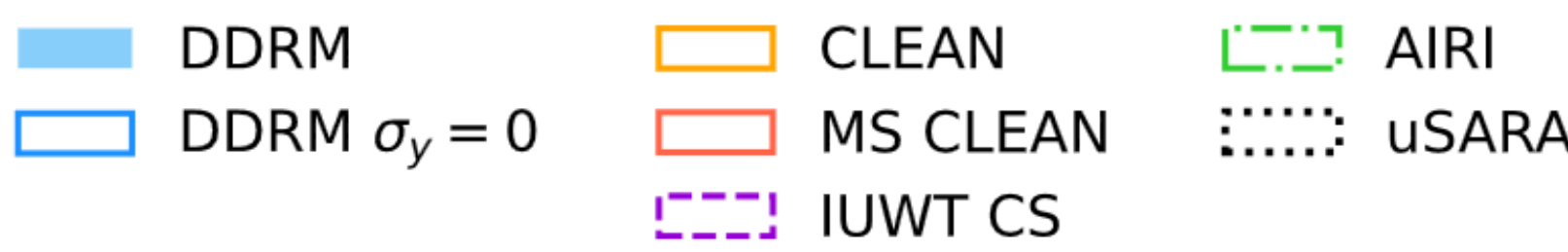


COMPARED TO AIRI & USARA

■ DDRM
□ DDRM $\sigma_y = 0$
■ CLEAN
■ MS CLEAN
- - - AIRI
⋯ uSARA
- - - IUWT CS



COMPARED TO AIRI & USARA



2-3 seconds

1-2 minutes

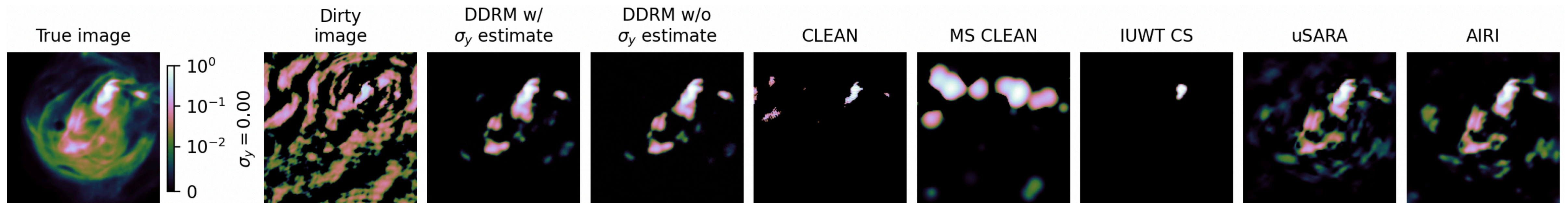
15-20 seconds

IN CONCLUSION

DDRM offers **excellent reconstruction performance** with a small execution time, outperforming CLEAN by an order of magnitude.

However...

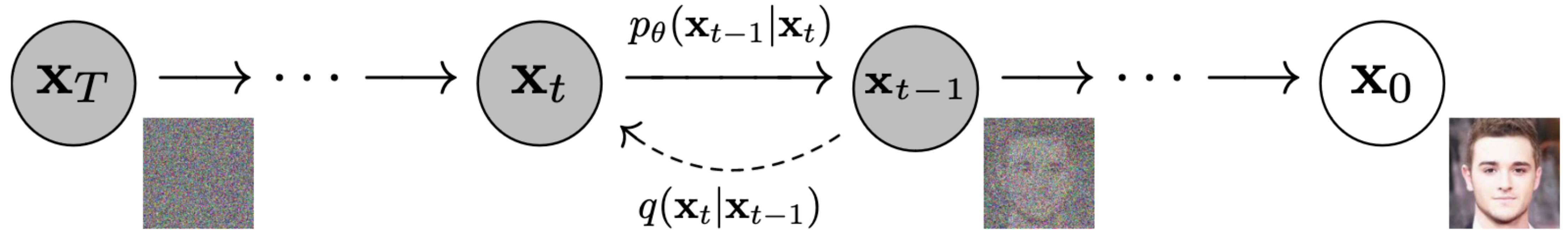
- **fixed image size** (150x150) set by available training data.
- Does **not** provide a useful **uncertainty estimate**, but could possibly fine-tune DDPM to better calibrate the uncertainty (arXiv: 2302.10688)
- Limited to **linear** forward operators, though could extend this with Diffusion Posterior Sampling (arXiv: 2209.14687)



A photograph of a radio telescope array, likely the Very Large Array, featuring several large parabolic dish antennas mounted on complex metal structures. The scene is set against a clear blue sky at dusk or dawn, with the sun low on the horizon. The foreground shows the intricate lattice of the largest dish on the left, while other smaller dishes are visible in the background and to the right. The overall tone is serene and technological.

BACKUP SLIDES

DENOISING DIFFUSION PROBABLISTIC MODELS (DDPM)



Forward process:

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Reverse process:

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \quad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Training is performed by optimizing the usual variational bound on negative log likelihood:

$$\mathbb{E}[-\log p_\theta(\mathbf{x}_0)] \leq \mathbb{E}_q \left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \right] = \mathbb{E}_q \left[-\log p(\mathbf{x}_T) - \sum_{t \geq 1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \right] =: L$$

Algorithm 1 Training

```
1: repeat  
2:    $\mathbf{x}_0 \sim q(\mathbf{x}_0)$   
3:    $t \sim \text{Uniform}(\{1, \dots, T\})$   
4:    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$   
5:   Take gradient descent step on  
        $\nabla_{\theta} \|\epsilon - \epsilon_{\theta}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2$   
6: until converged
```

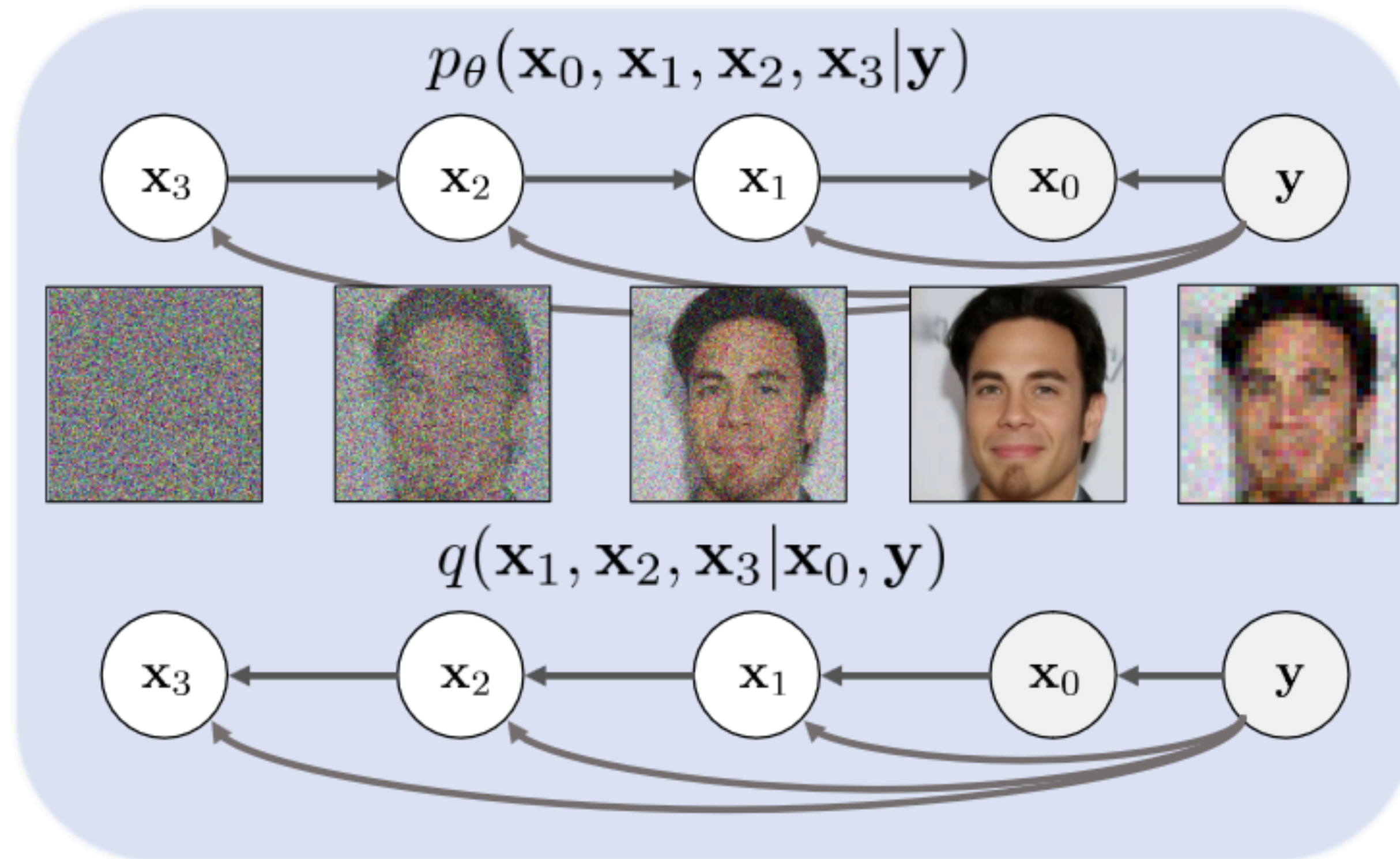
$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\boldsymbol{\beta}}_t \mathbf{I})$$

$$p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t), \boldsymbol{\Sigma}_{\theta}(\mathbf{x}_t, t))$$

The loss term L_t is parameterized to minimize the difference from $\tilde{\boldsymbol{\mu}}$:

$$\begin{aligned} L_t &= \mathbb{E}_{\mathbf{x}_0, \epsilon} \left[\frac{1}{2\|\boldsymbol{\Sigma}_{\theta}(\mathbf{x}_t, t)\|_2^2} \|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) - \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t)\|^2 \right] \\ &= \mathbb{E}_{\mathbf{x}_0, \epsilon} \left[\frac{1}{2\|\boldsymbol{\Sigma}_{\theta}\|_2^2} \left\| \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_t \right) - \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta}(\mathbf{x}_t, t) \right) \right\|^2 \right] \\ &= \mathbb{E}_{\mathbf{x}_0, \epsilon} \left[\frac{(1 - \alpha_t)^2}{2\alpha_t(1 - \bar{\alpha}_t)\|\boldsymbol{\Sigma}_{\theta}\|_2^2} \|\epsilon_t - \epsilon_{\theta}(\mathbf{x}_t, t)\|^2 \right] \\ &= \mathbb{E}_{\mathbf{x}_0, \epsilon} \left[\frac{(1 - \alpha_t)^2}{2\alpha_t(1 - \bar{\alpha}_t)\|\boldsymbol{\Sigma}_{\theta}\|_2^2} \|\epsilon_t - \epsilon_{\theta}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon_t, t)\|^2 \right] \end{aligned}$$

DENOISING DIFFUSION RESTORATION MODELS (DDRM)



Denoising Diffusion Restoration Models
(Dependent on inverse problem)

DDRM is a Markov chain conditioned on $\mathbf{y}$ which corresponds to the evidence lower bound:

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0), \mathbf{y} \sim q(\mathbf{y} | \mathbf{x}_0)} [\log p_{\theta}(\mathbf{x}_0 | \mathbf{y})] \\ & \geq -\mathbb{E} \left[\sum_{t=1}^{T-1} D_{\text{KL}}(q^{(t)}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{x}_0, \mathbf{y}) \| p_{\theta}^{(t)}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y})) \right] \\ & \quad + \mathbb{E} \left[\log p_{\theta}^{(0)}(\mathbf{x}_0 | \mathbf{x}_1, \mathbf{y}) \right] \\ & \quad - \mathbb{E} [D_{\text{KL}}(q^{(T)}(\mathbf{x}_T | \mathbf{x}_0, \mathbf{y}) \| p_{\theta}^{(T)}(\mathbf{x}_T | \mathbf{y}))] \end{aligned}$$

Uses a pre-trained DDPM as a prior on $\mathbf{x}$
 $\mathbf{x}$ and $\mathbf{y}$ are related by a linear forward model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$$

