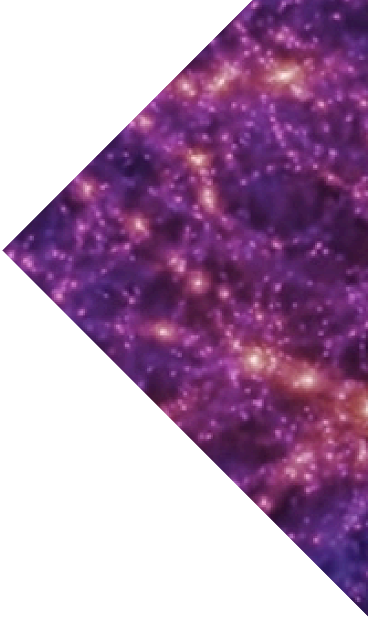
High-dimensional uncertainty quantification with deep data-driven priors

Jason McEwen
Fundamental Research, Alan Turing Institute
SciAI, Mullard Space Science Laboratory (MSSL), UCL

MINOAS workshop, FORTH, Heraklion, September 2025



Inverse imaging problems

Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x))$$

for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

Inverse imaging problems

Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \, ,$$

for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

Inverse imaging problems

Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \, ,$$

for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

Inverse imaging problems

Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \, ,$$

for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

$=$ 

Inverse imaging problems

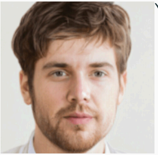
Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\mathbf{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \mathbf{\Phi} x + n \,,$$

for **image** x, deterministic **measurement model** $\mathbf{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

$=$ 

$\Big($  $\Big) + n$

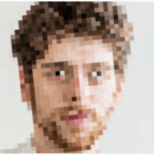
Inverse imaging problems

Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\mathbf{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \mathbf{\Phi} x + n \,,$$

for **image** x, deterministic **measurement model** $\mathbf{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.



$=$

$\Bigg($

$\Bigg) + n$

Inverse imaging problems

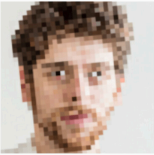
Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \,,$$

for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

$y \xleftarrow{\text{forward model}} x$



$=$

$+\, n$

Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \, ,$$

for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

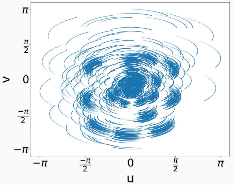
Inverse problem model

Consider **observations**

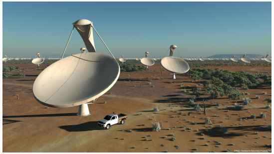
$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \, ,$$

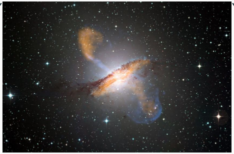
for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

$$y \xleftarrow{\text{forward model}} x$$



$=$



$\Big($  $\Big) + n$

$$\xrightarrow[\text{inverse problem}]{}$$

Inverse imaging problems

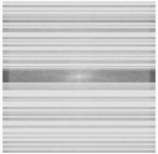
Inverse problem model

Consider **observations**

$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \,,$$

for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

y ← forward model — x



$=$

$\Big($  $\Big) + n$

→ inverse problem

Inverse imaging problems

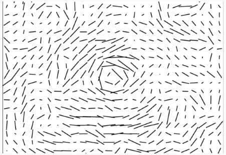
Inverse problem model

Consider **observations**

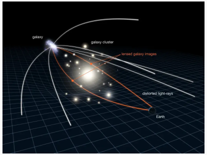
$$y \sim \mathbb{P}(\boldsymbol{\Phi}(x)) \quad \xrightarrow{\text{linear case}} \quad y = \boldsymbol{\Phi} x + n \, ,$$

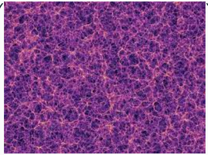
for **image** x, deterministic **measurement model** $\boldsymbol{\Phi}$, and stochastic aspects of data acquisition encoded by **statistical process** $\mathbb{P}$.

y ⟵ forward model ⟵ x



$=$



$\Bigg($  $\Bigg) + n$

⟶ inverse problem ⟶

Ill-conditioned and ill-posed problems

Inverse problems often **ill-conditioned** and **ill-posed** (in the sense of Hadamard):

1. Solution may not exist.
2. Solution may not be unique.
3. Solution may not be stable.

Ill-conditioned and ill-posed problems

Inverse problems often **ill-conditioned** and **ill-posed** (in the sense of Hadamard):

1. Solution may not exist.
2. Solution may not be unique.
3. Solution may not be stable.

- ▷ Inject **regularising prior information**
- ▷ **Quantify uncertainty**

$\Big\} \Rightarrow$ Bayesian inference

Bayesian inference

Bayes' theorem

$$\underbrace{p(\mathbf{x} \mid \mathbf{y}, M)}_{\text{posterior}} = \frac{\overbrace{p(\mathbf{y} \mid \mathbf{x}, M)}^{\text{likelihood}} \; \overbrace{p(\mathbf{x} \mid M)}^{\text{prior}}}{\underbrace{p(\mathbf{y} \mid M)}_{\text{marginal likelihood}}} = \frac{\overbrace{\mathcal{L}(\mathbf{x})}^{\text{likelihood}} \; \overbrace{\pi(\mathbf{x})}^{\text{prior}}}{\underbrace{Z}_{\text{marginal likelihood}}},$$

for parameters $\mathbf{x}$, model M and observed data $\mathbf{y}$.

For **parameter estimation**, typically draw samples from the posterior by *Markov chain Monte Carlo (MCMC)* sampling.

Computational challenge of MCMC sampling can be prohibitive

- Parameter space high dimensional, *i.e.* $\boldsymbol{x} \in \mathbb{R}^N$ with large N.
- Large data volume, *i.e.* $\boldsymbol{y} \in \mathbb{R}^M$ with large M.
- Computationally costly measurement operator $\boldsymbol{\Phi} : \mathbb{R}^N \rightarrow \mathbb{R}^M$.

Computational challenge of MCMC sampling can be prohibitive

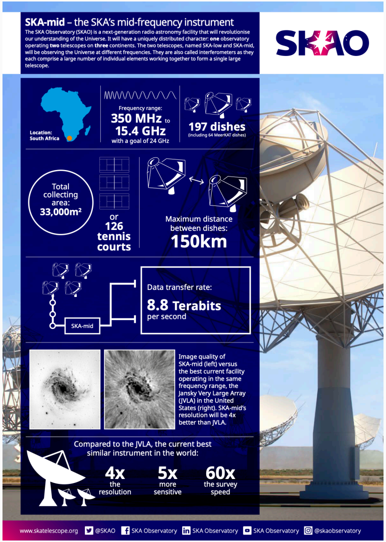
- **Parameter space high dimensional**, *i.e.* $\boldsymbol{x} \in \mathbb{R}^N$ with large N.
- **Large data volume**, *i.e.* $\boldsymbol{y} \in \mathbb{R}^M$ with large M.
- **Computationally costly measurement operator** $\boldsymbol{\Phi} : \mathbb{R}^N \rightarrow \mathbb{R}^M$.

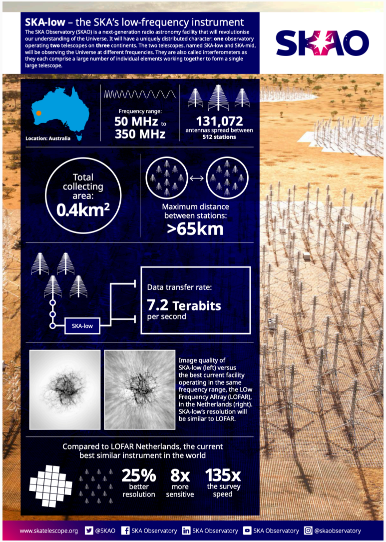
In many settings we have one of these challenges... in some we have all!

Square Kilometre Array (SKA)

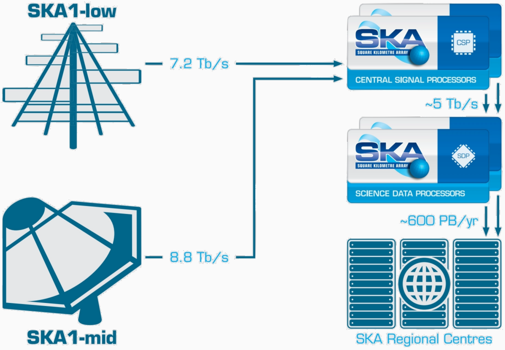
Artist impression of the Square Kilometer Array (SKA)

SKA sites


SKA-mid – the SKA's mid-frequency instrument
The SKA Observatory (SKAO) is a next-generation radio astronomy facility that will revolutionise our understanding of the Universe. It will have a uniquely distributed character: one observatory operating two telescopes on three continents. The two telescopes, named SKA-low and SKA-mid, will be observing the Universe at different frequencies. They are also called interferometers as they each comprise a large number of individual elements working together to form a single large telescope.
Location: South Africa
Frequency range:
350 MHz to 15.4 GHz
with a goal of 24 GHz
197 dishes
(including 64 MeerKAT dishes)
Total collecting area: 33,000m²
or 126 tennis courts
Maximum distance between dishes:
150km
SKA-mid
Data transfer rate:
8.8 Terabits
per second
Image quality of SKA-mid (left) versus the best current facility operating in the same frequency range, the Jansky Very Large Array (JVLA) in the United States (right). SKA-mid's resolution will be 4x better than JVLA.
Compared to the JVLA, the current best similar instrument in the world:
4x the resolution
5x more sensitive
60x the survey speed
www.skatelescope.org
@SKAO
SKA Observatory
SKA Observatory
SKA Observatory
@skaobservatory


SKA-low – the SKA's low-frequency instrument
The SKA Observatory (SKAO) is a next-generation radio astronomy facility that will revolutionise our understanding of the Universe. It will have a uniquely distributed character: one observatory operating two telescopes on three continents. The two telescopes, named SKA-low and SKA-mid, will be observing the Universe at different frequencies. They are also called interferometers as they each comprise a large number of individual elements working together to form a single large telescope.
Location: Australia
Frequency range:
50 MHz to 350 MHz
131,072
antennas spread between 512 stations
Total collecting area:
0.4km²
Maximum distance between stations:
>65km
SKA-low
Data transfer rate:
7.2 Terabits
per second
Image quality of SKA-low (left) versus the best current facility operating in the same frequency range, the LOw Frequency ARray (LOFAR), in the Netherlands (right). SKA-low's resolution will be similar to LOFAR.
Compared to LOFAR Netherlands, the current best similar instrument in the world
25% better resolution
8x more sensitive
135x the survey speed
www.skatelescope.org
@SKAO
SKA Observatory
SKA Observatory
SKA Observatory
@skaobservatory

SKA data rates



8.5 Exabytes over the 15-year lifetime of initial high-priority science programmes (Scaife 2020).

All 3 computational challenges (high-dimensional, big-data, expensive operator).

$\Rightarrow$ **MCMC sampling infeasible.**

Recover point estimator by optimisation

Consider **MAP point estimator** by solving variation regularisation problem:

$$\hat{x}_{\text{MAP}} = \arg\max_x \Big[\log p(x \mid y)\Big] = \arg\min_x \Big[\underbrace{\log p(y \mid x)}_{\text{data fidelity}} + \underbrace{\lambda R(x)}_{\text{regulariser}}\Big]$$

Recover point estimator by optimisation

Consider **MAP point estimator** by solving variation regularisation problem:

$$\hat{x}_{\text{MAP}} = \arg\max_{x} \Big[\log p(x \,|\, y) \Big] = \arg\min_{x} \Big[\underbrace{\log p(y \,|\, x)}_{\text{data fidelity}} + \underbrace{\lambda\, R(x)}_{\text{regulariser}} \Big]$$

- **Log-likelihood (data fidelity) encodes** physics through measurement operator $\boldsymbol{\Phi}$ and statistical acquisition model $\mathbb{P}$.
- **Regulariser encodes** prior.

Recover point estimator by optimisation

Consider **MAP point estimator** by solving variation regularisation problem:

$$\hat{x}_{\text{MAP}} = \arg\max_x \left[\log p(x \mid y) \right] = \arg\min_x \left[\underbrace{\log p(y \mid x)}_{\text{data fidelity}} + \underbrace{\lambda R(x)}_{\text{regulariser}} \right]$$

- **Log-likelihood (data fidelity) encodes physics** through measurement operator $\mathbf{\Phi}$ and statistical acquisition model $\mathbb{P}$.
- **Regulariser encodes prior.**

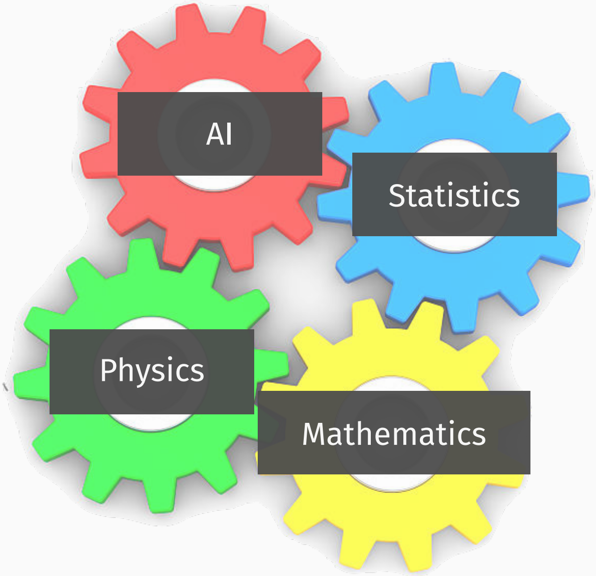
✖ But fails to capture uncertainty.

✖ Hand-crafted priors (not expressive) considered traditionally.

Goals

- Computationally efficient (optimisation).
- Physics-informed (robust and interpretable).
- Expressive data-driven AI priors (enhance reconstruction fidelity).
- Quantify uncertainties (for scientific inference).

Interdisciplinary solution

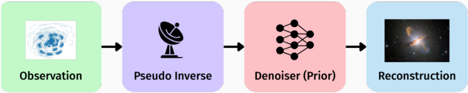


Outline

1. Physics + AI
2. Physics + AI + UQ
3. Physics + AI + UQ + Calibration

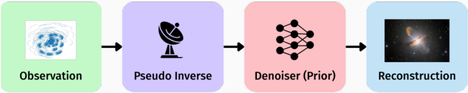
Physics + AI

Learned inverse imaging

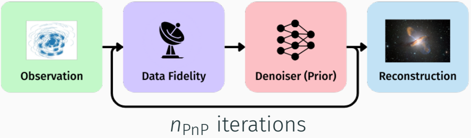


Learned post-processing

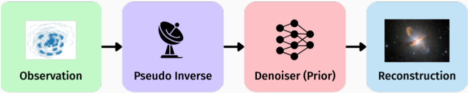
Learned inverse imaging



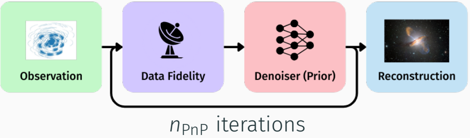
Learned post-processing



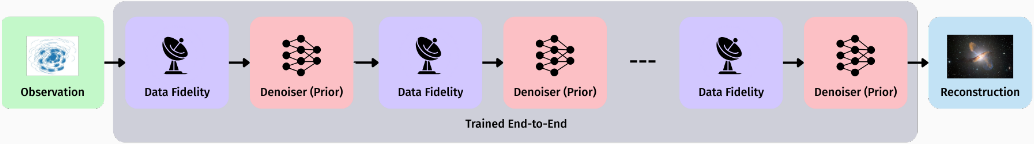
Plug-and-Play (PnP)



Learned post-processing



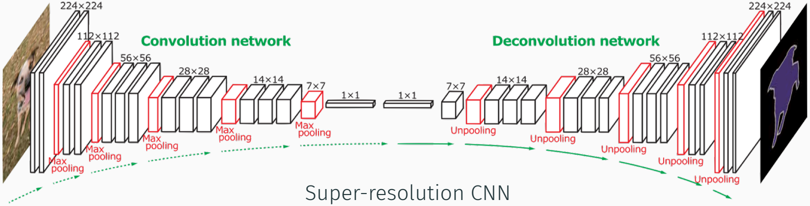
Plug-and-Play (PnP)



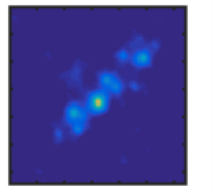
Unrolled ($n_{\text{unrolled}} \ll n_{\text{PnP}}$)

Learned post-processing: pre-UNet

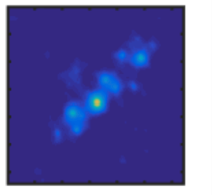
- Allam Jn & McEwen (2016): RI imaging using super-resolution CNN with fixed measurement operator (uv coverage)



Super-resolution CNN



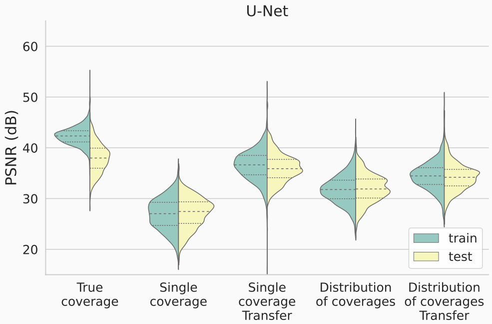
Backprojected dirty image (SNR=7.8dB)



Reconstructed image (SNR=12.3dB)

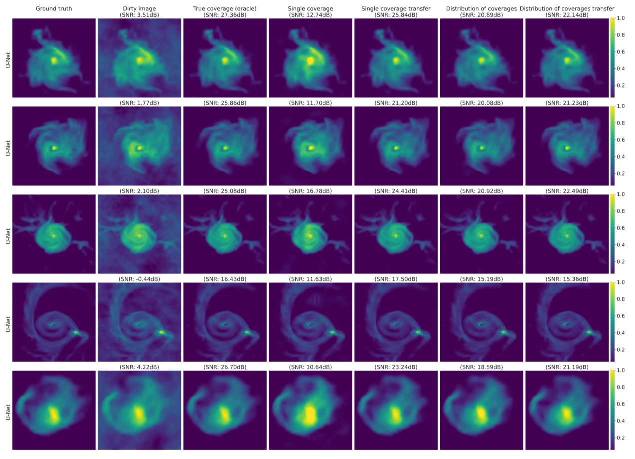
Learned post-processing: post-UNet

- Terris *et al.* (2019): RI imaging using UNet
- Mars, Betcke & McEwen (2024): RI imaging using UNet with varying measurement operator (varying coverage)



PSNR for different strategies to adapt to varying operator (uv coverage).

Learned post-processing: post-UNet



Gallery of UNet reconstructions for different strategies to adapt to varying operator (uv coverage).

- Venkatakrishnan *et al.* (2013), Ryu *et al.* (2019)
- Terris *et al.* (2022, 2024): introduced AIRI
- Aghabiglou *et al.* (2022, 2024): R2D2 series of networks trained sequentially
- McEwen *et al.* papers in prep.: Optimus Primal, QuantifAI (Python), PURIFY (distributed, C++)

PURIFY

CI passing | codecov 86% | DOI 10.5281/zenodo.2555252

Description

PURIFY is an open-source collection of routines written in `C++` available under the license below. It implements different tools and high-level to perform radio interferometric imaging, *i.e.* to recover images from the Fourier measurements taken by radio interferometric telescopes.

GitHub: `https://github.com/astro-informatics/purify`

Sparse OPTimisation Library

CMake passing | codecov 96% | DOI 10.5281/zenodo.2584256

Description

SOPT is an open-source `C++` package available under the license below. It performs Sparse OPTimisation using state-of-the-art convex optimisation algorithms. It solves a variety of sparse regularisation problems, including the Sparsity Averaging Reweighted Analysis (SARA) algorithm.

GitHub: `https://github.com/astro-informatics/sopt`

Unrolled

Unrolled approaches (Gregor & LeCun 2010):

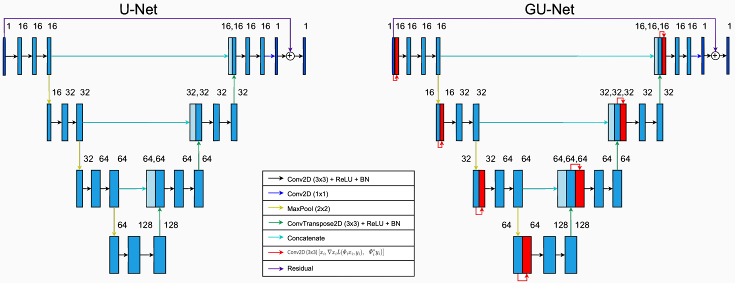
- ✔ **Trained end-to-end** $\rightarrow$ excellent performance.
- ✖ Typically **many expensive measurement operator applications** during training.
- ✖ Require **differentiable measurement operator**.

Unrolled

Unrolled approaches (Gregor & LeCun 2010):

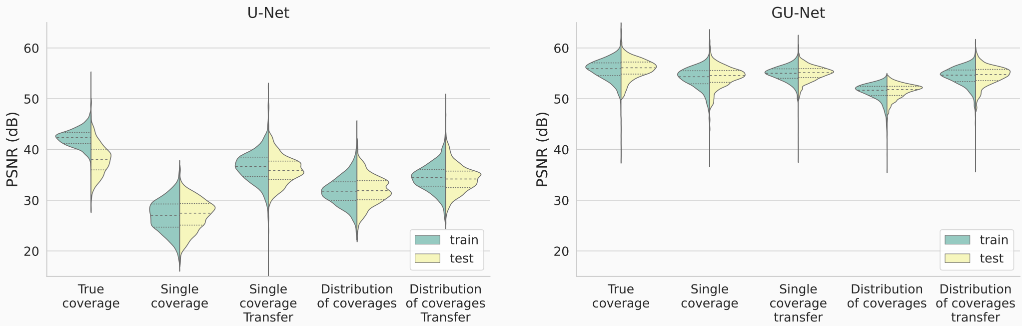
- ✔ **Trained end-to-end** → excellent performance.
- ✖ Typically **many expensive measurement operator applications** during training.
- ✖ Require **differentiable measurement operator**.

Introduce **Gradient UNet (GUNet)** to **solve scalability of unrolled approaches**, with a multi-resolution measurement operator (Mars *et al.* 2024, 2025).



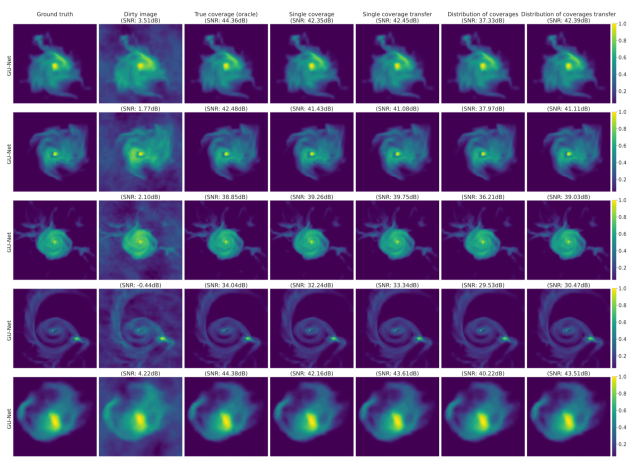
Unrolled

Post-processing (UNet) → Unrolled (GUNet): significantly improves reconstruction fidelity and robustness to varying measurement operator (visibility coverage).



PSNR for different strategies to adapt to varying operator (uv coverage).

Unrolled



Gallery of GUNet reconstructions for different strategies to adapt to varying operator (uv coverage).

Physics $+$ AI $+$ UQ

UQ outline

1. Direct UQ estimation
2. PnP UQ estimation
3. Unrolled generative UQ estimation

Direct UQ estimation

Estimating UQ summary statistics

Train a network to estimate a summary statistic:

- ▷ Magnitude of residual: train a network to estimate residuals.
- ▷ Gaussian per pixel: train a network to estimate the standard deviation.
- ▷ Classification for regression ranges: train a classifier with softmax output to estimate distribution of pixel values.
- ▷ Pixelwise quantile regression: train network to estimate lower/upper quantiles for $1-\alpha$ uncertainty level, using quantile (pinball) loss.

Heuristic $\rightarrow$ no statistical guarantees.

PnP UQ estimation

Convex probability concentration for uncertainty quantification

Posterior credible region:

$$p(\boldsymbol{x} \in C_\alpha | \boldsymbol{y}) = \int_{\boldsymbol{x} \in \mathbb{R}^N} p(\boldsymbol{x}|\boldsymbol{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\boldsymbol{x} = 1 - \alpha.$$

Consider the **highest posterior density (HPD) region**

$$C^*_\alpha = \left\{\boldsymbol{x} : -\log p(\boldsymbol{x}) \leq \gamma_\alpha\right\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\boldsymbol{x} \in C^*_\alpha | \boldsymbol{y}) = 1 - \alpha \text{ holds.}$$

Convex probability concentration for uncertainty quantification

Posterior credible region:

$$p(x \in C_\alpha | y) = \int_{x \in \mathbb{R}^N} p(x|y) \mathbb{1}_{C_\alpha} \mathrm{d}x = 1 - \alpha.$$

Consider the **highest posterior density (HPD) region**

$$C^*_\alpha = \{x : -\log p(x) \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(x \in C^*_\alpha | y) = 1 - \alpha \text{ holds.}$$

Bound of HPD region for log-concave distributions (Pereyra 2017)

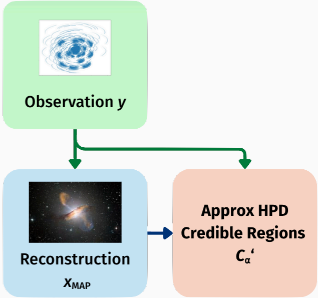
Suppose the posterior $\log p(x|y) \propto \log \mathcal{L}(x) + \log \pi(x)$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\mathrm{e}^[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{x : \log \mathcal{L}(x) + \log \pi(x) \leq \hat{\gamma}_\alpha = \log \mathcal{L}(\hat{x}_{\text{MAP}}) + \log \pi(\hat{x}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

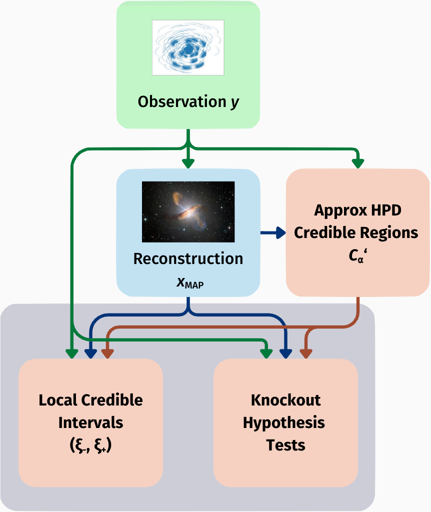
with a positive constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$ independent of $p(x|y)$.

Need **only evaluate** $\log \mathcal{L} + \log \pi$ **for the MAP estimate** $\hat{x}_{\text{MAP}}$!

Leverging the approximate HPD region for UQ



Leverging the approximate HPD region for UQ



Hypothesis testing

Hypothesis testing of physical structure

(Pereyra 2017; Cai, Pereyra & McEwen 2018a)

1. Remove structure of interest from recovered image $x^\star$.
2. Inpaint background (noise) into region, yielding surrogate image x'.
3. Test whether $x' \in C_\alpha$:
   - If $x' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, *i.e.* **structure most likely physical**.
   - If $x' \in C_\alpha$ uncertainly too high to draw strong conclusions about the physical nature of the structure.

Local Bayesian credible intervals for sparse reconstruction

(Cai, Pereyra & McEwen 2018b)

Let Ω define the area (or pixel) over which to compute the credible interval $(\tilde{\xi}_-, \tilde{\xi}_+)$ and ζ be an index vector describing Ω (*i.e.* $\zeta_i = 1$ if $i \in \Omega$ and 0 otherwise).

Consider the test image with the Ω region replaced by constant value ξ:

$$x' = x^\star(\mathcal{I} - \zeta) + \xi\zeta \, .$$

Given $\tilde{\gamma}_\alpha$ and $x^\star$, compute the credible interval by

$$\tilde{\xi}_- = \min_{\xi} \left\{ \xi \mid \log \mathcal{L}(x') + \log \pi(x') \leq \tilde{\gamma}_\alpha, \ \forall \xi \in [-\infty, +\infty) \right\} ,$$

$$\tilde{\xi}_+ = \max_{\xi} \left\{ \xi \mid \log \mathcal{L}(x') + \log \pi(x') \leq \tilde{\gamma}_\alpha, \ \forall \xi \in [-\infty, +\infty) \right\} .$$

Convex data-driven AI prior

Adopt **neural-network-based convex regulariser** R (Goujon *et al.* 2022; Liaudat *et al.* McEwen 2024):

$$R(x) = \sum_{n=1}^{N_C} \sum_{k} \psi_n \left((\boldsymbol{h}_n * x) [k] \right) ,$$

- ▷ ψ_n are learned convex profile functions with Lipschitz continuous derivative;
- ▷ N_C learned convolutional filters $\boldsymbol{h}_n$.

Convex data-driven AI prior

Adopt **neural-network-based convex regulariser** R
(Goujon *et al.* 2022; Liaudat *et al.* McEwen 2024):

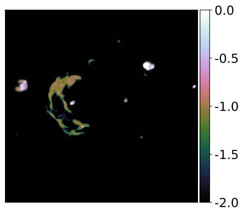
$$R(\boldsymbol{x}) = \sum_{n=1}^{N_C} \sum_{k} \psi_n \left((\boldsymbol{h}_n * \boldsymbol{x}) [k] \right),$$

- ψ_n are learned convex profile functions with Lipschitz continuous derivative;
- N_C learned convolutional filters $\boldsymbol{h}_n$.

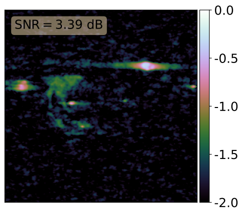
Properties:

1. **Convex** + **explicit potential** $\Rightarrow$ leverage convex UQ theory.
2. **Smooth regulariser with known Lipschitz constant** $\Rightarrow$ theoretical convergence guarantees.

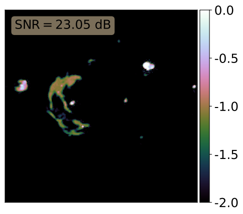
Reconstructed images



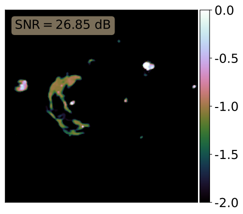
Ground truth



Dirty image
SNR=3.39 dB

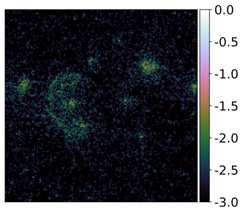


Reconstruction (classical)
SNR=23.05 dB

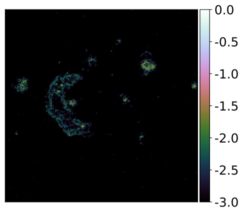


Reconstruction (learned)
SNR= 26.85 dB

(Liaudat *et al.* McEwen 2024)

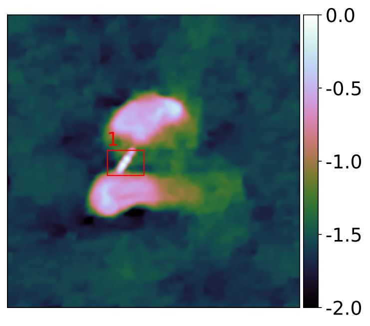


Error (classical)



Error (learned)

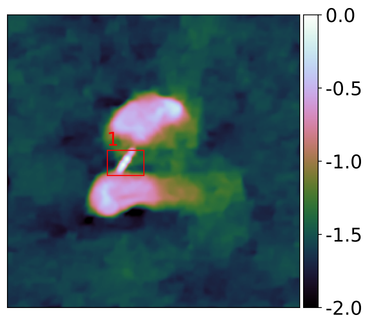
Hypothesis testing of structure



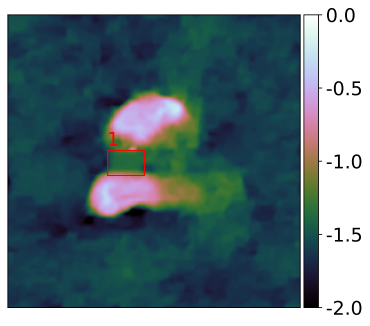
Reconstructed image

(Liaudat *et al.* McEwen 2024)

Hypothesis testing of structure

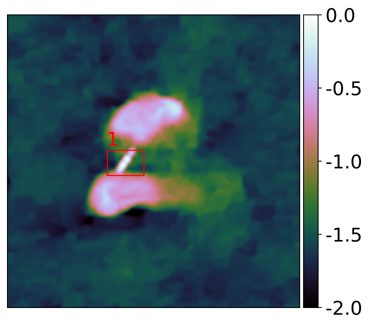


Reconstructed image

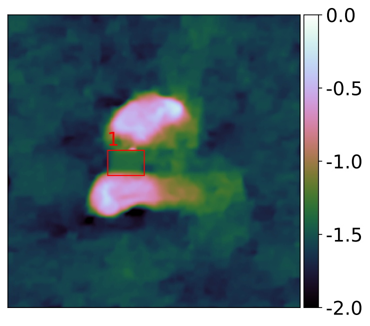


Surrogate test image (region removed)

(Liaudat *et al.* McEwen 2024)



Reconstructed image



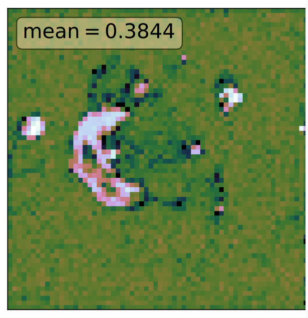
Surrogate test image (region removed)

Reject null hypothesis

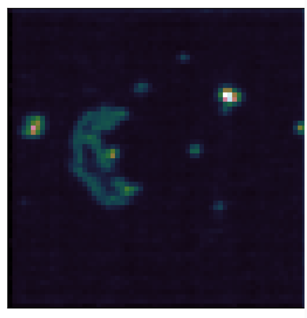
$\Rightarrow$ **structure physical**

(Liaudat *et al.* McEwen 2024)

Approximate local Bayesian credible intervals



LCI
(super-pixel size 4 × 4)



MCMC standard deviation
(super-pixel size 4 × 4)

$10^3 \times$ faster than MCMC sampling

(Liaudat *et al.* McEwen 2024)

QuantifAI code

GitHub quantifai | License GPL | arXiv 2312.00125

QuantifAI

`quantifai` is a PyTorch-based open-source radio interferometric imaging reconstruction package with scalable Bayesian uncertainty quantification relying on data-driven (learned) priors. This package was used to produce the results of Liaudat et al. 2023. The `quantifai` model relies on the data-driven convex regulariser from Goujon et al. 2022.

Github: https://github.com/astro-informatics/QuantifAI

PyTorch: Automatic differentiation (including instrument model) + GPU acceleration

Exascale imaging codes

PURIFY

CI passing | codecov 86% | DOI 10.5281/zenodo.2555252

Description

PURIFY is an open-source collection of routines written in `C++` available under the license below. It implements different tools and high-level to perform radio interferometric imaging, *i.e.* to recover images from the Fourier measurements taken by radio interferometric telescopes.

GitHub:
https://github.com/astro-informatics/purify

Sparse OPTimisation Library

CMake passing | codecov 96% | DOI 10.5281/zenodo.2584256

Description

SOPT is an open-source `C++` package available under the license below. It performs Sparse OPTimisation using state-of-the-art convex optimisation algorithms. It solves a variety of sparse regularisation problems, including the Sparsity Averaging Reweighted Analysis (SARA) algorithm.

GitHub:
https://github.com/astro-informatics/sopt

TensorFlow PyTorch ONNX Spack

Unrolled generative UQ estimation

Leveraging generative AI

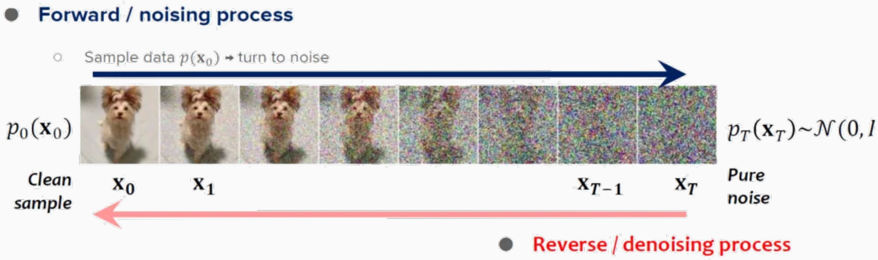
Bring generative AI to bear to **generate approximate posterior samples** but in a **physics-informed** manner.

Consider two approaches:

- Denoising diffusion models
- Generative adversarial networks (GANs)

Denoising diffusion models

Denoising diffusion models (Ho *et al.* 2020, Song & Ermon 2020).



Learn data distribution.

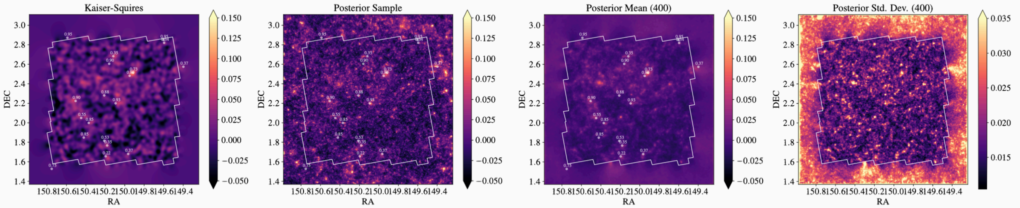
Consider as a **deep generative prior** for solving inverse problems.

Approximate posterior sampling with diffusion models / score matching

Combine generative prior with likelihood to solve inverse problems.

Probabilistic mass mapping with neural score estimation (Remy *et al.* 2023).

- Learn score $\nabla \log p_{\sigma_2}(x) = (D_{\sigma^2}(x) - x)/\sigma^2$.
- Combine with convolved likelihood $\log p_{\sigma_L^2}(y \,|\, x)$ and sample with annealed HMC approach.



Reconstructed mass maps of dark matter (Remy *et al.* 2023)

Diffusion posterior sampling

Diffusion posterior sampling is a highly active area of research (see Daras *et al.* 2024 for a recent survey).

Likelihood is analytically intractable due to dependence of diffusion process on time (Chung *et al.* 2022). Hence, various **approximations** considered.

- ✔ Diffusion models are highly expressive
- ✖ Slow
- ✖ Approximate posterior samples

GANs for approximate posterior sample generation

GANs very good for high-fidelity generation.

GANs for approximate posterior sample generation

GANs very good for high-fidelity generation.

Challenges:

- ✖ Difficult to train
- ✖ Suffer from mode collapse

GANs for approximate posterior sample generation

GANs very good for high-fidelity generation.

Challenges:

- ✖ Difficult to train
- ✖ Suffer from mode collapse

Solutions:

- ✔ Wasserstein loss (Arjovsky *et al.* 2017)
- ✔ Regularisation (Bendel *et al.* 2023)

Conditional regularised GANs

For inverse imaging problems, condition on observed data y.

Introduce regularisation to avoid mode collapse by **rewarding sampling diversity** (Bendel *et al.* 2023).

Conditional regularised GANs

For inverse imaging problems, condition on observed data $\boldsymbol{y}$.

Introduce regularisation to avoid mode collapse by **rewarding sampling diversity** (Bendel *et al.* 2023).

Add regularisation to loss:

$$\mathcal{L}_{\text{reg}}(\boldsymbol{\theta}) = \mathcal{L}_{1,P}(\boldsymbol{\theta}) - \beta \mathcal{L}_{\text{SD},P}(\boldsymbol{\theta}) \, ,$$

where

$$\mathcal{L}_{1,P}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{z}_1,\ldots,\boldsymbol{z}_P,\boldsymbol{y}} \| \boldsymbol{x} - \hat{\boldsymbol{x}}_{(P)} \|_1 \quad \text{and} \quad \mathcal{L}_{\text{SD},P}(\boldsymbol{x}) = \sqrt{\frac{\pi}{2P(P-1)}} \sum_{i=1}^{P} \mathbb{E}_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_P,\boldsymbol{y}} \| \hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_{(P)} \|_1 \, ,$$

and with $\hat{\boldsymbol{x}}_{(P)}$ denoting P-averaged samples.

Conditional regularised GANs

For inverse imaging problems, condition on observed data $\boldsymbol{y}$.

Introduce regularisation to avoid mode collapse by **rewarding sampling diversity** (Bendel *et al.* 2023).

Add regularisation to loss:

$$\mathcal{L}_{\text{reg}}(\boldsymbol{\theta}) = \mathcal{L}_{1,P}(\boldsymbol{\theta}) - \beta \mathcal{L}_{\text{SD},P}(\boldsymbol{\theta}) \, ,$$

where

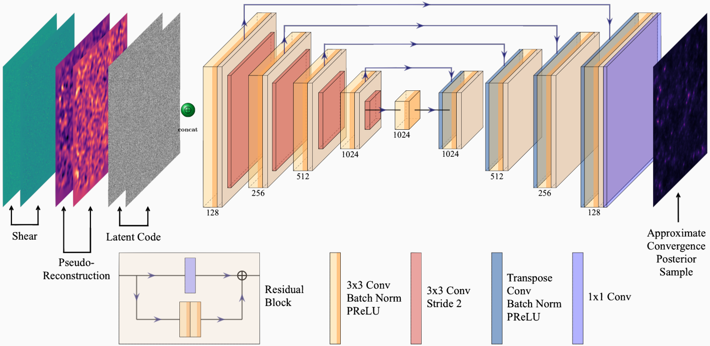
$$\mathcal{L}_{1,P}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{z}_1,\ldots,\boldsymbol{z}_P,\boldsymbol{y}} \| \boldsymbol{x} - \hat{\boldsymbol{x}}_{(P)} \|_1 \quad \text{and} \quad \mathcal{L}_{\text{SD},P}(\boldsymbol{x}) = \sqrt{\frac{\pi}{2P(P-1)}} \sum_{i=1}^{P} \mathbb{E}_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_P,\boldsymbol{y}} \| \hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_{(P)} \|_1 \, ,$$

and with $\hat{\boldsymbol{x}}_{(P)}$ denoting P-averaged samples.

Recover first two moments of true posterior (Bendel *et al.* 2023)

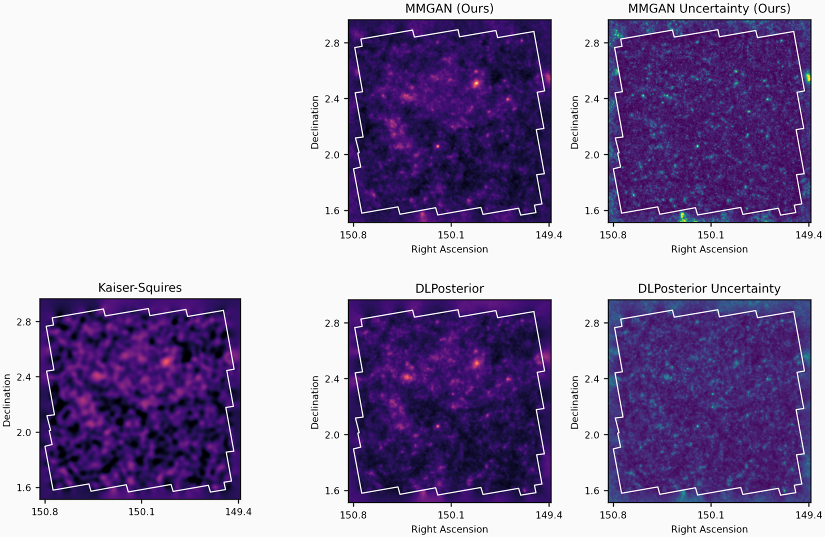
First two moments of the approximated posterior (mean and variance) match the true posterior (under Gaussian assumptions).

Adapted conditional regularised GANs to mass mapping dark matter (Whitney *et al.* McEwen 2025).



MM-GAN for mass mapping dark matter

MM-GAN for mapping dark matter



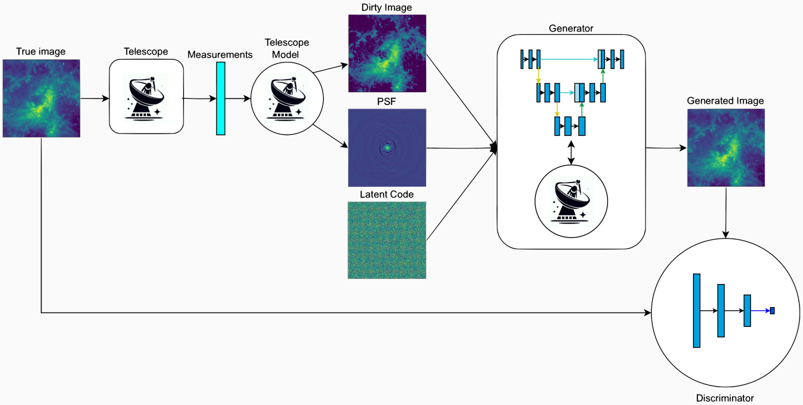
Classical case

Generative posterior samples

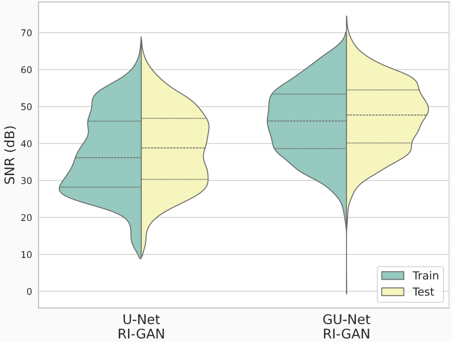
MM-GAN for mapping dark matter

	Pearson ↑	RMSE ↓	PSNR ↑
MMGAN (Ours)	**0.727**	**0.0197**	**34.106**
Kaiser-Squires	0.619	0.0229	32.803
Kaiser-Squires *	0.57	0.0240	-
Wiener filter *	0.61	0.0231	-
GLIMPSE *	0.42	0.0284	-
MCAlens *	0.67	0.0219	-
DeepMass *	0.68	0.0218	-
DLPosterior *	0.68	0.0216	-

Introduce **physical model of measurement operator** in architecture (Mars *et al.* McEwen 2025).



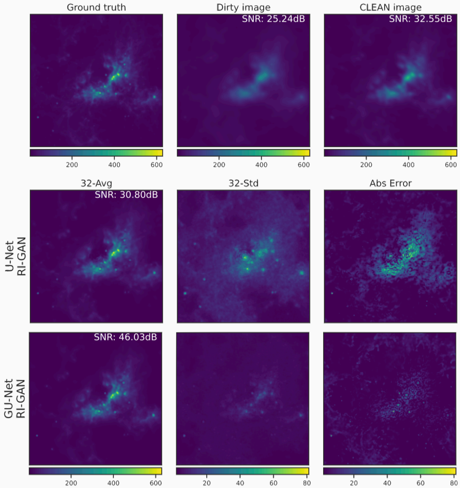
Physics-informed architecture improves reconstruction fidelity.



RI-GAN for radio interferometric imaging (left: UNet without physics; right: GUNet with physics)

RI-GAN for radio interferometric imaging

Physics-informed architecture improves reconstruction fidelity substantially for out-of-distribution settings.



Conditional regularised GANs for inverse imaging

- ✔ GANs are highly expressive
- ✔ Fast
- ✖ Guarantees for Gaussian case but otherwise approximate posterior samples

UQ overview

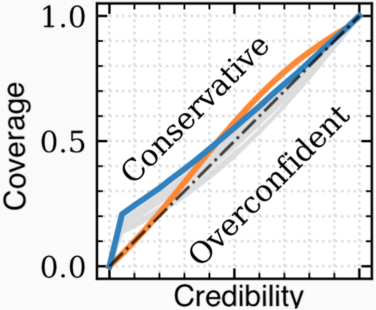
1. Direct UQ estimation
   - ✔ Fast
   - ✖ Heuristic with no statistical guarantees
2. PnP UQ estimation
   - ✔ Fast
   - ✔ Statistical guarantees by leveraging convexity
   - ✖ Restricted to HPD-related UQ
3. Unrolled generative UQ estimation
   - ✔ Fast (GANs); Slow (diffusion models)
   - ✖ Target posterior samples but no statistical guarantees (guarantees in Gaussian setting for GANs)

Physics + AI + UQ + Calibration

Coverage testing

Compute coverage plots to validate.

- Compute a credible interval.
- Check empirically the frequency that ground truth within interval.



Coverage analyses starting to be performed

Do Bayesian imaging methods report trustworthy probabilities? (Thong *et al.* 2024)

Coverage analyses starting to be performed

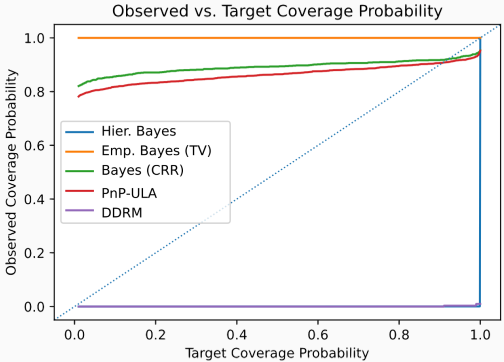
Do Bayesian imaging methods report trustworthy probabilities? (Thong *et al.* 2024)

No!

Coverage analyses starting to be performed

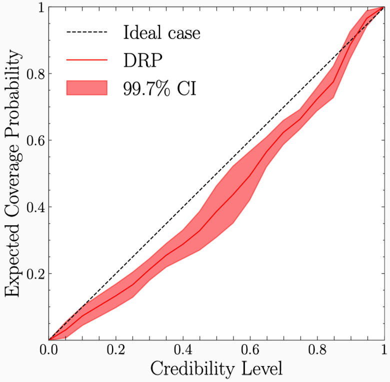
Do Bayesian imaging methods report trustworthy probabilities? (Thong *et al.* 2024)

No!



Coverage analysis for radio interferometry

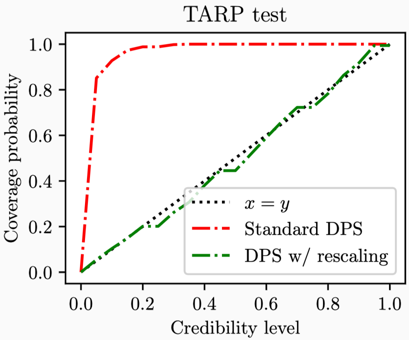
Bayesian imaging for **radio interferometry** with score-based priors (Dia *et al.* 2023).



Coverage analysis for mass mapping of dark matter

Mass mapping with diffusion posterior sampling (Anonymous submission to ML4PS, NeurIPs 2025).

- ▷ Introduce an ad hoc likelihood scaling approach to down weight the likelihood at early stages of diffusion.
- ▷ Works reasonably well but is ad hoc, with no statistical guarantees.



Calibrate uncertainties with conformal prediction

Conformal prediction with **Risk-Controlling Prediction Sets (RCPS)** (Bates *et al.* 2021, Angelopoulos *et al.* 2022).

Calibrate uncertainties with conformal prediction

Conformal prediction with **Risk-Controlling Prediction Sets (RCPS)** (Bates *et al.* 2021, Angelopoulos *et al.* 2022).

Given: estimator $\hat{f}(x)$; lower interval length $\hat{l}(x)$; upper interval length $\hat{u}(x)$.

Construct uncertainty intervals around each pixel (m, n):

$$\mathcal{T}_\lambda(x)_{(m,n)} = \left[\, \hat{f}(x)_{(m,n)} - \lambda \hat{l}(x)_{(m,n)}, \hat{f}(x)_{(m,n)} + \lambda \hat{u}(x)_{(m,n)} \,\right] .$$

Find λ to ensure interval contains the right number of pixels (exploiting Hoeffding's bound).

Calibrate uncertainties with conformal prediction

- ▷ **Distribution-free** uncertainty quantification with **statistical guarantees**.
- ▷ Guaranteed to be valid but not necessarily useful $\Rightarrow$ still **need good initial uncertainty estimates**.

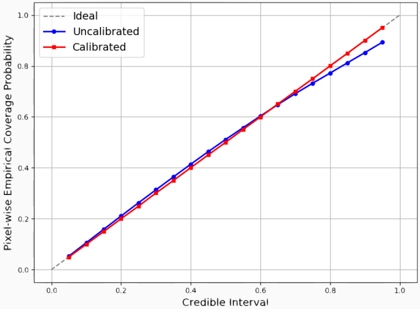
(Develop conformalised quantile regression for inverse problems and apply RCPS for mass-mapping in Leterme, Fadili & Starck 2025.)

Coverage tests with MM-GAN

Coverage testing and conformal prediction of MM-GAN for mass mapping of dark energy (Whitney, Liaudat & McEwen, in prep.).

Coverage tests with MM-GAN

Coverage testing and conformal prediction of MM-GAN for mass mapping of dark energy (Whitney, Liaudat & McEwen, in prep.).



- **Extremely good coverage (without RCPS)**
  $\rightarrow$ regularization and theoretical guarantee in idealised setting highly effective in practical setting.
- **Optimal coverage after calibration** with RCPS.

Summary

Summary

Inverse imaging problems typically **ill-conditioned** and **ill-posted**
$\Rightarrow$ **inject regularising prior, quantify uncertainty** $\Rightarrow$ **Bayesian inference**

Summary

Inverse imaging problems typically **ill-conditioned** and **ill-posted**

$\Rightarrow$ **inject regularising prior, quantify uncertainty** $\Rightarrow$ **Bayesian inference**

MCMC sampling computationally infeasible for many problems, motivating **goals**:

- ✔ Computationally efficient (optimisation).
- ✔ Physics-informed (robust and interpretable).
- ✔ Expressive data-driven AI priors (enhance reconstruction fidelity).
- ✔ Quantify uncertainties (for scientific inference).

Summary

Inverse imaging problems typically **ill-conditioned** and **ill-posted**
$\Rightarrow$ **inject regularising prior, quantify uncertainty** $\Rightarrow$ **Bayesian inference**

MCMC sampling computationally infeasible for many problems, motivating **goals**:

- ✔ Computationally efficient (optimisation).
- ✔ Physics-informed (robust and interpretable).
- ✔ Expressive data-driven AI priors (enhance reconstruction fidelity).
- ✔ Quantify uncertainties (for scientific inference).

PnP with convexity (Liaudat *et al.* McEwen 2024) goes some way towards these aims.

Summary

Inverse imaging problems typically **ill-conditioned** and **ill-posted**

$\Rightarrow$ **inject regularising prior, quantify uncertainty** $\Rightarrow$ **Bayesian inference**

MCMC sampling computationally infeasible for many problems, motivating **goals**:

- ✔ Computationally efficient (optimisation).
- ✔ Physics-informed (robust and interpretable).
- ✔ Expressive data-driven AI priors (enhance reconstruction fidelity).
- ✔ Quantify uncertainties (for scientific inference).

PnP with convexity (Liaudat *et al.* McEwen 2024) goes some way towards these aims.

Regularised conditional GAN with physics and UQ calibration (Whitney *et al.* McEwen 2025, Mars *et al.* McEwen 2025) achieves goals:

- ✔ Fast (many posterior samples in seconds).
- ✔ Physics can be integrated in generator architecture.
- ✔ High fidelity imaging since GANs are highly expressive.
- ✔ Excellent coverage (without calibration; RCPS for statistical guarantees).