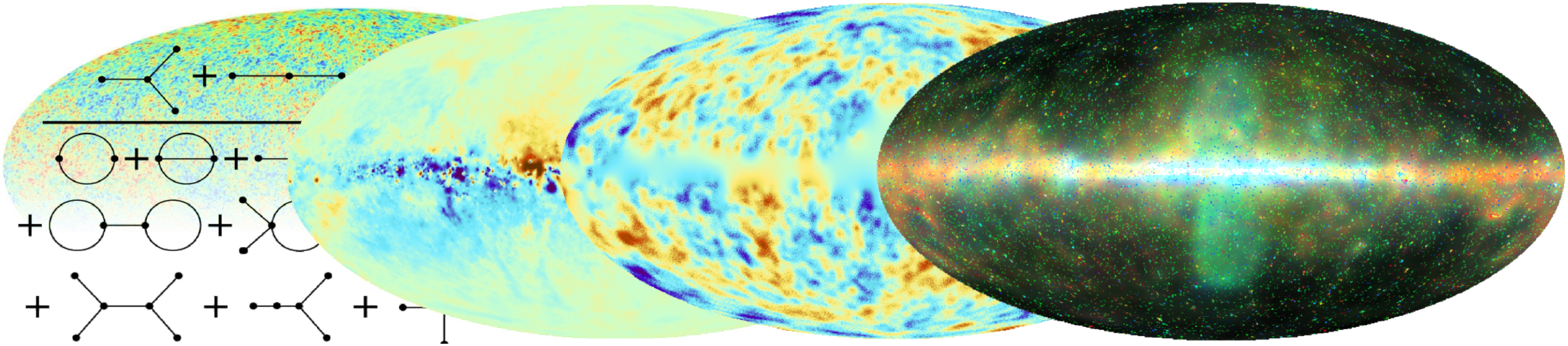
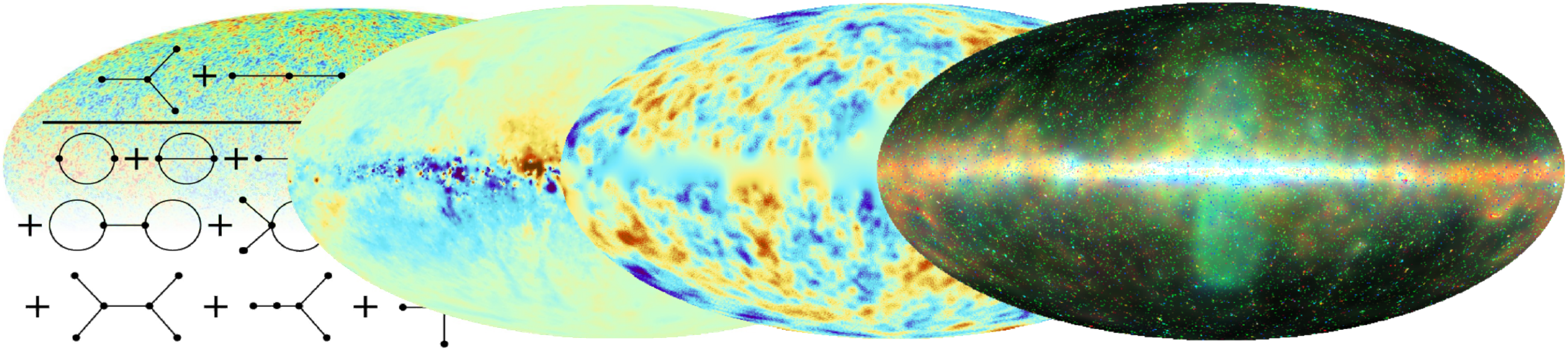


Information field theory

Torsten Enßlin
MPI for Astrophysics
Ludwig Maximilian University Munich



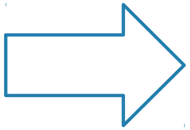
Information theory for fields

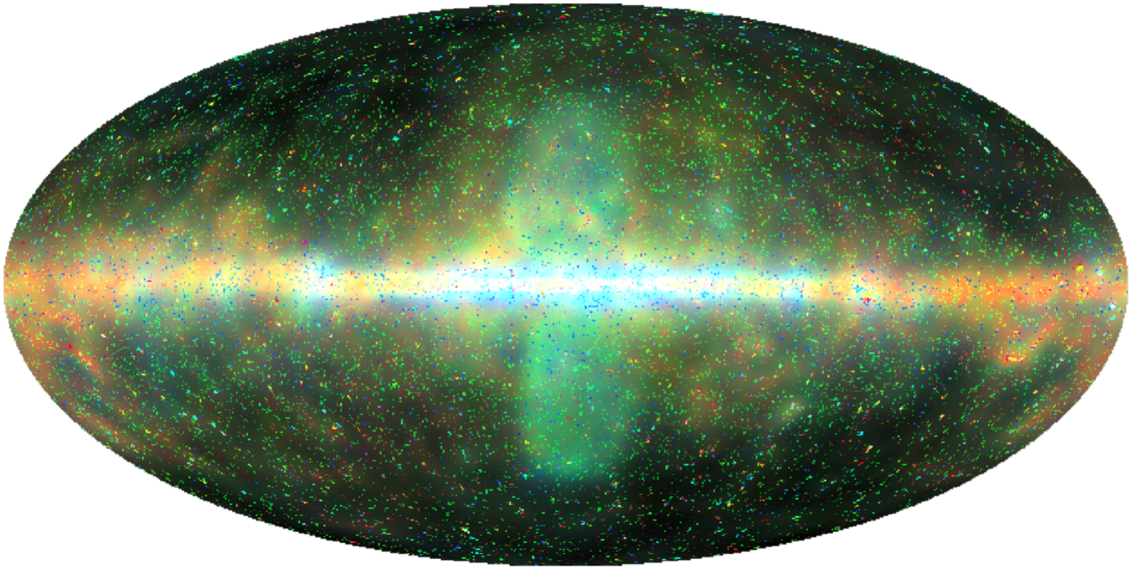
Torsten Enßlin
MPI for Astrophysics
Ludwig Maximilian University Munich

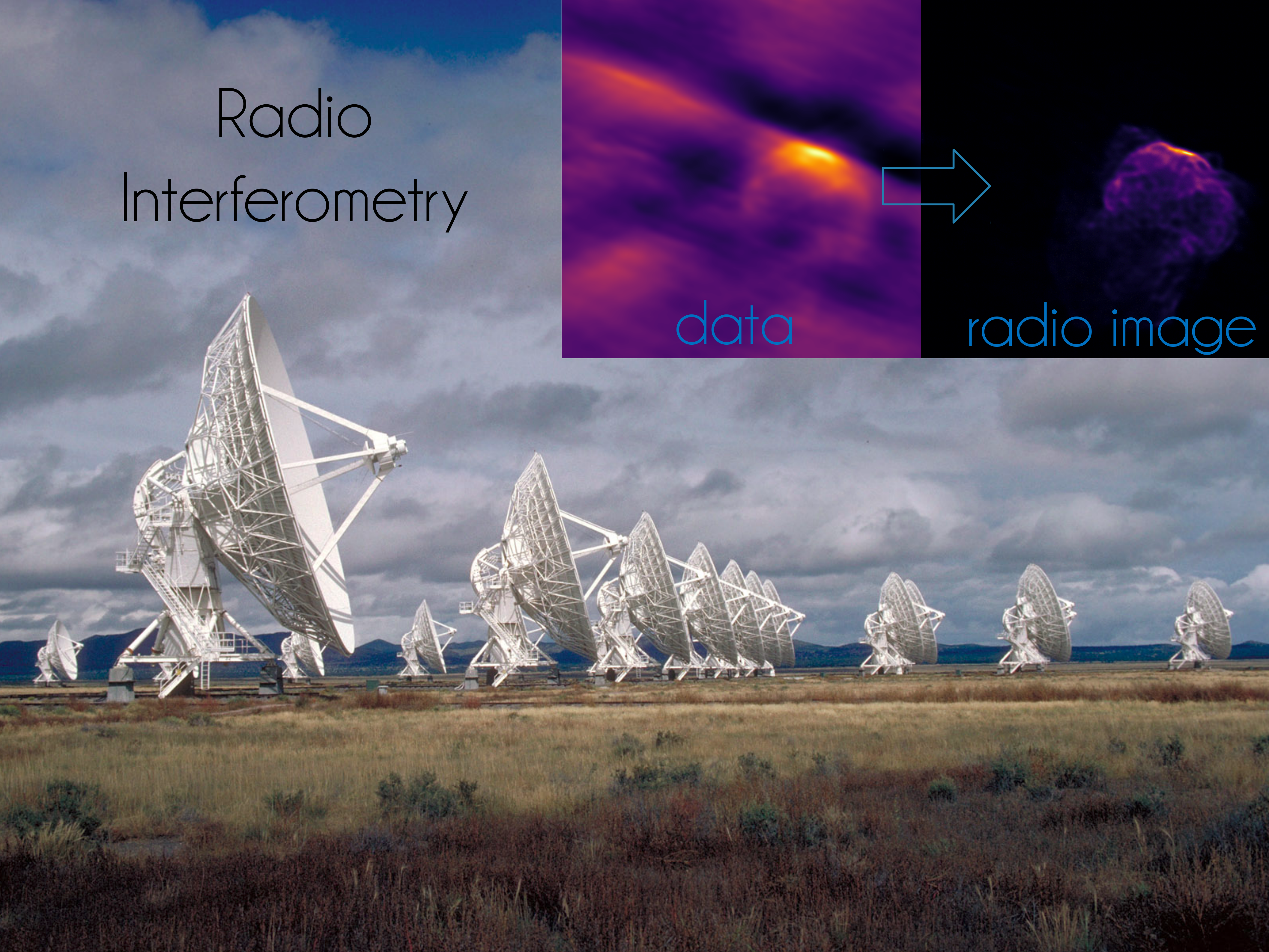
Photon count imaging

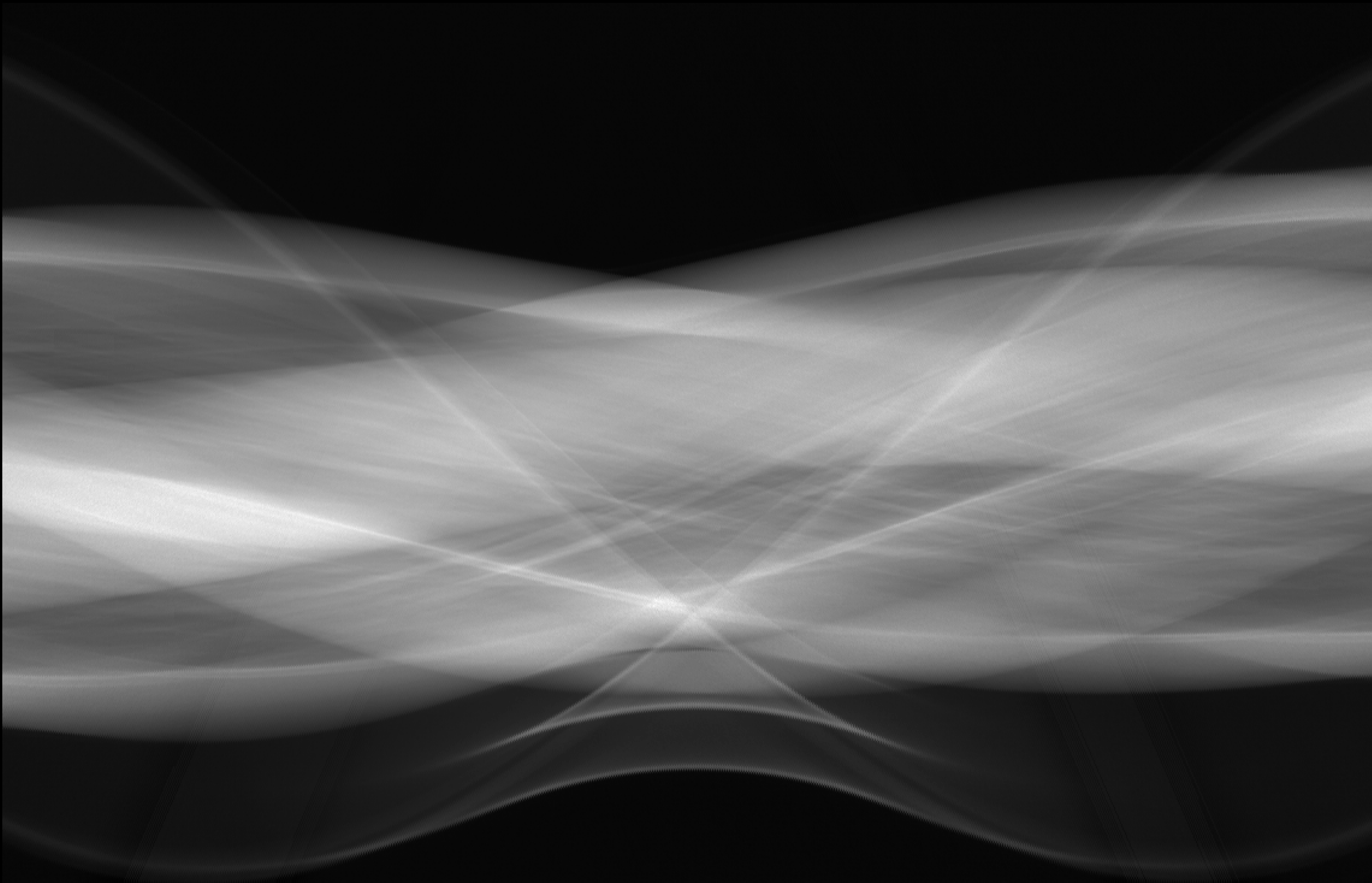
D³PO (Selig et al.)

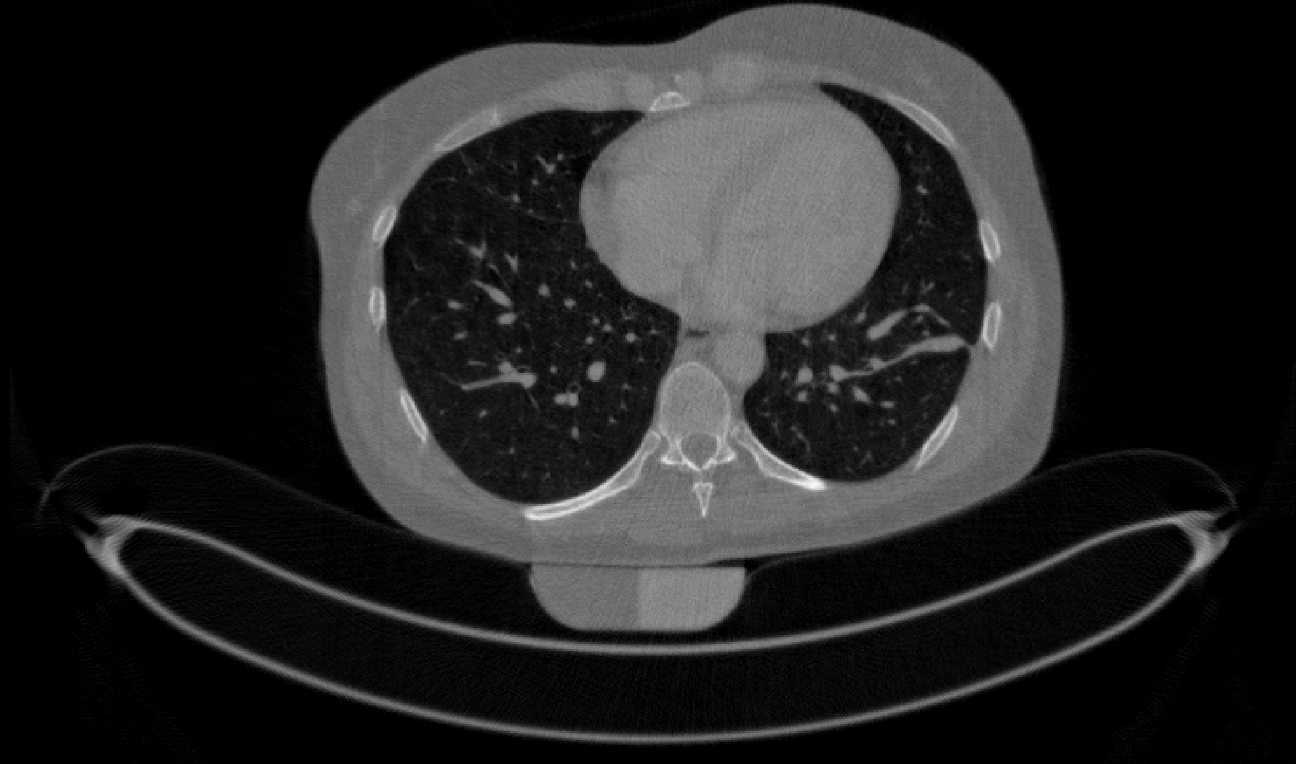
[Fermi - NASA/Kim Shiflett]



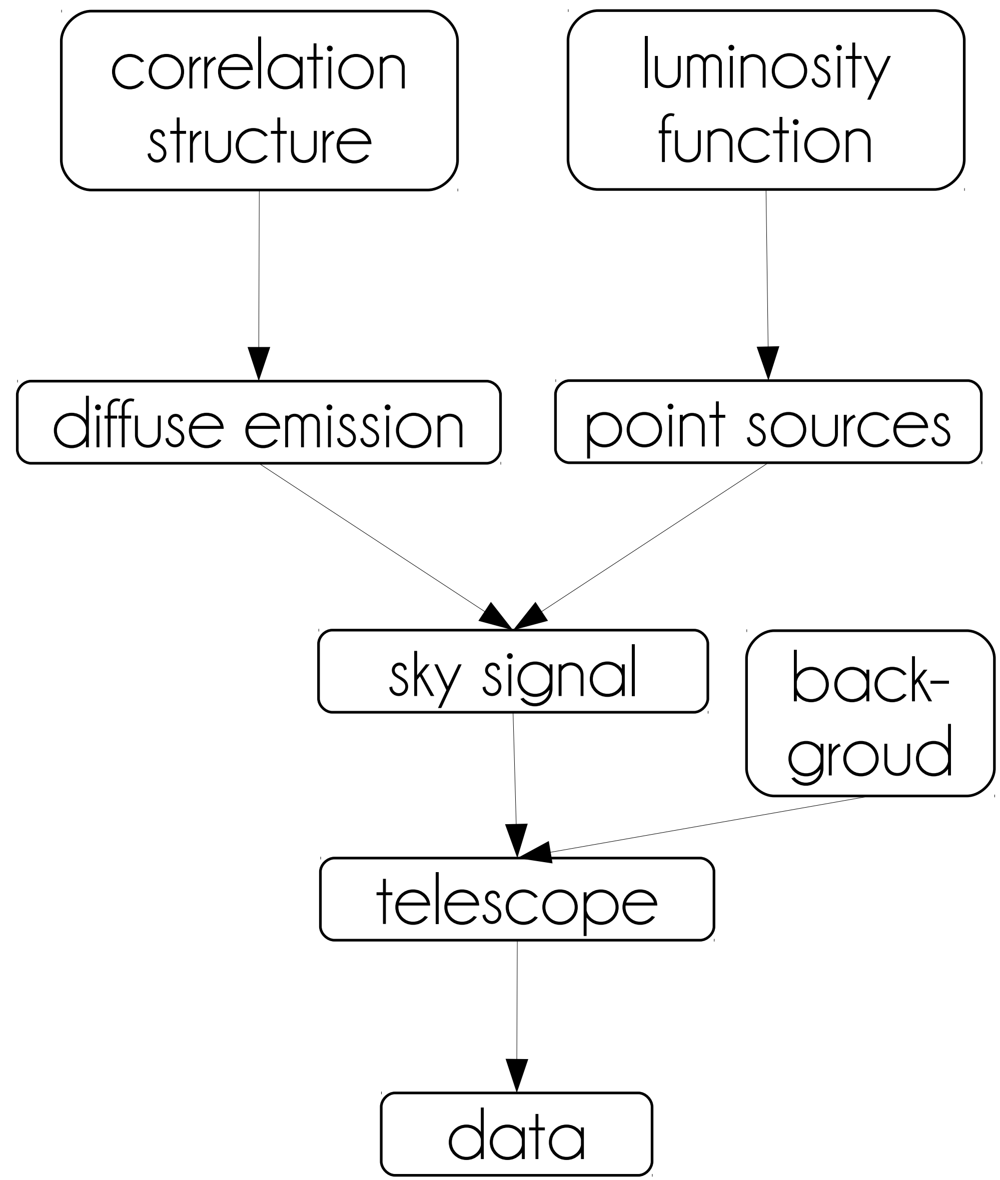



Radio
Interferometry
data
radio image



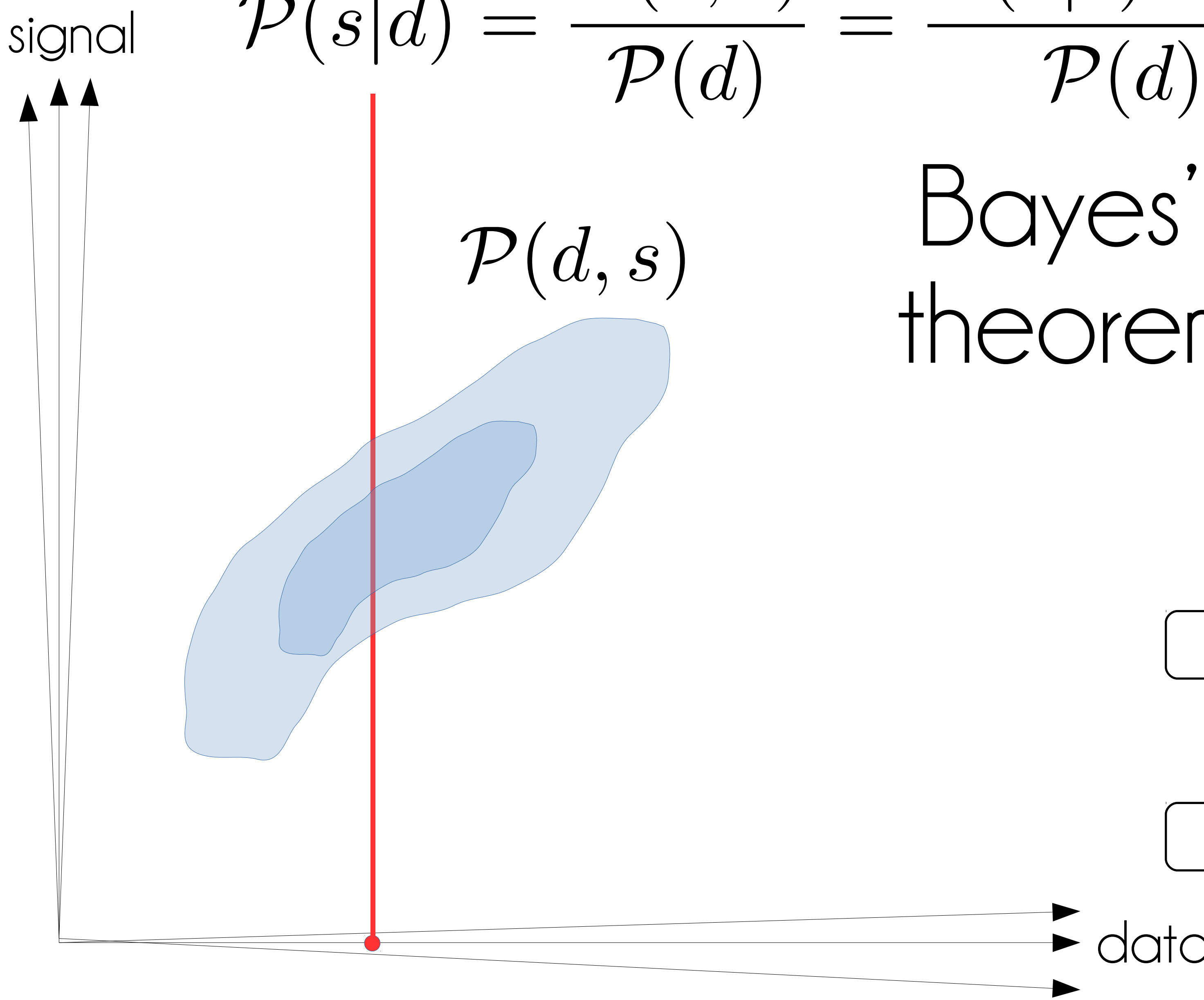


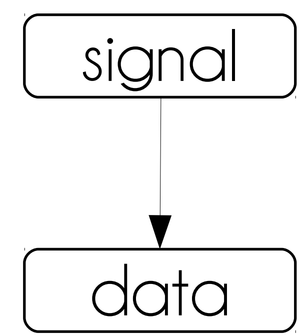
Measurement model



$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d,s)}{\mathcal{P}(d)} = \frac{\mathcal{P}(d|s)\,\mathcal{P}(s)}{\mathcal{P}(d)}$$

Bayes' theorem





Information theory

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d,s)}{\mathcal{P}(d)}$$

$$\mathcal{H}(d,s) = -\log \mathcal{P}(d,s)$$ Information

$$\mathcal{Z}(d) = \mathcal{P}(d)$$

$$= \int \mathcal{D}s \, \mathcal{P}(d,s)$$

$$\mathcal{P}(d,s) = \mathcal{P}(d|s)\,\mathcal{P}(s)$$

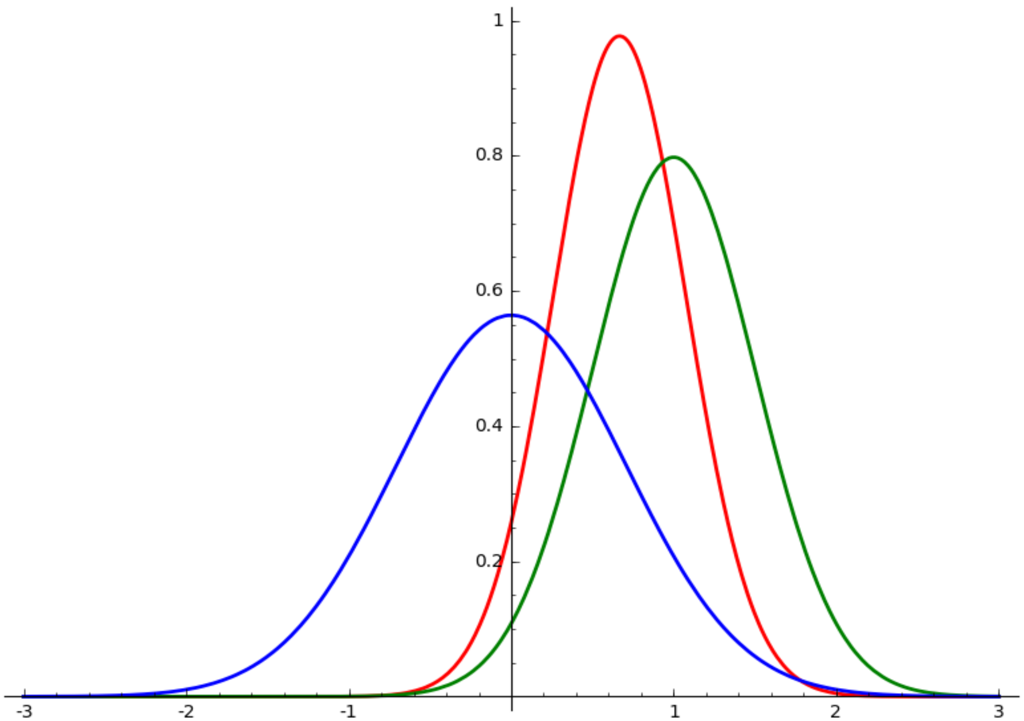
$$\mathcal{H}(d,s) = \mathcal{H}(d|s) + \mathcal{H}(s)$$ is additive

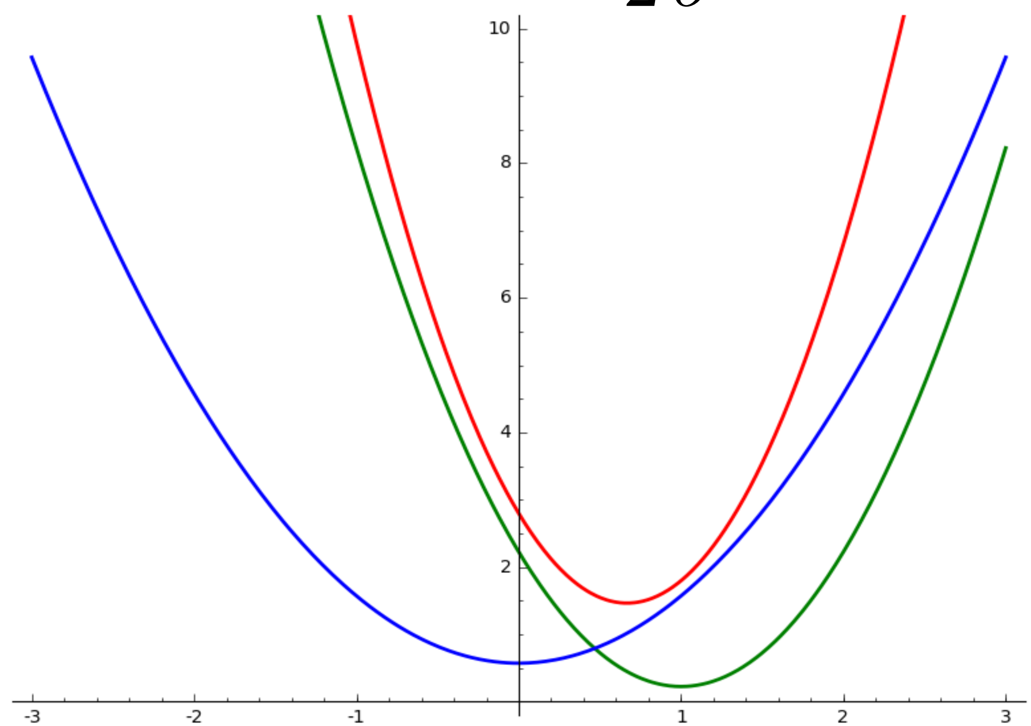
Probability & Information

$$\mathcal{P}(s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{s^2}{2\sigma^2}} \qquad \mathcal{H}(s) \;\widehat{=}\; \frac{s^2}{2\,\sigma^2}$$

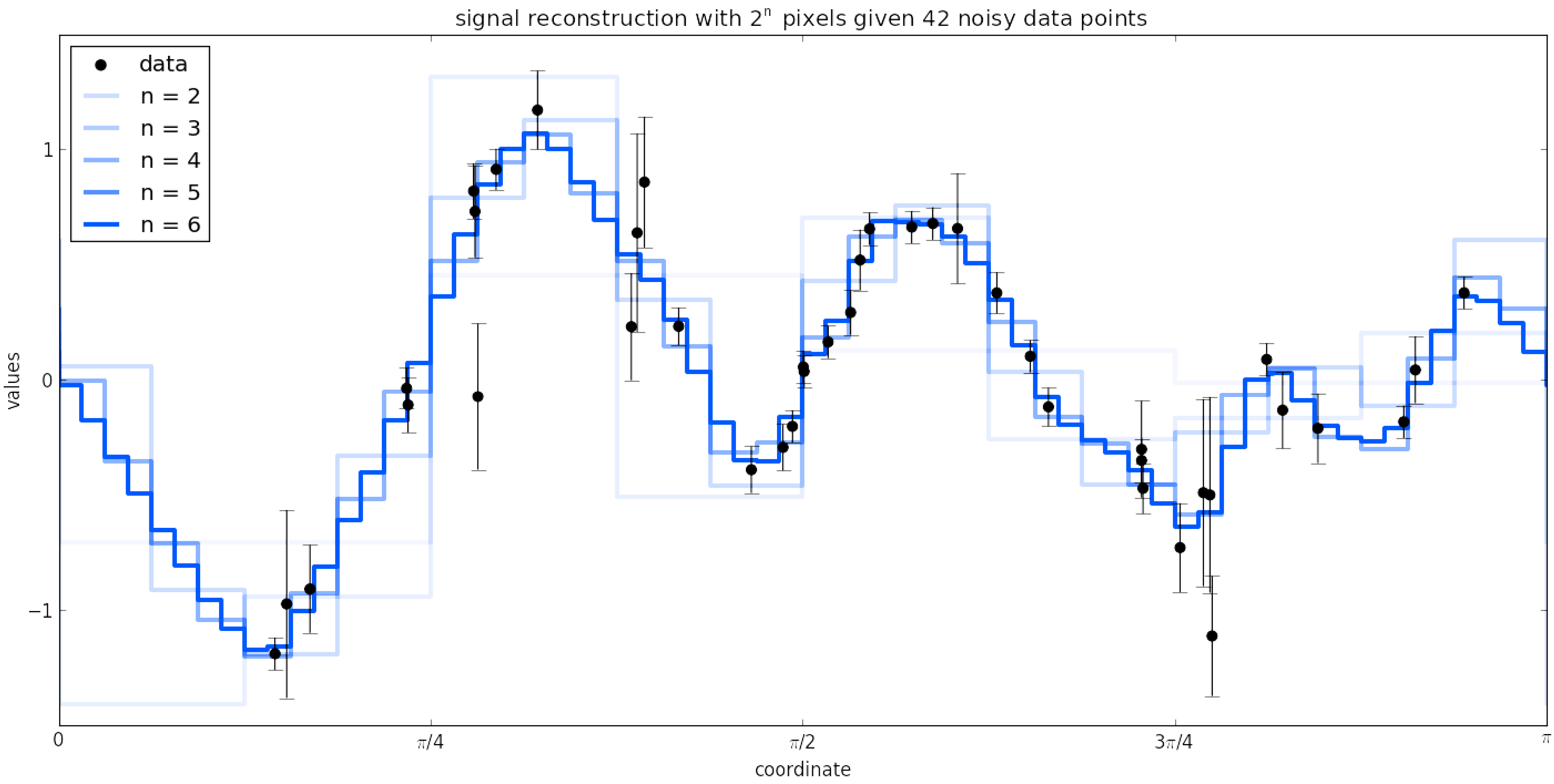
$$\mathcal{P}(d|s) \propto e^{-\frac{(s-d)^2}{2\sigma'^2}} \qquad \mathcal{H}(d|s) \;\widehat{=}\; \frac{(s-d)^2}{2\,\sigma'^2}\sigma^2)$$

$$\mathcal{P}(s|d) \propto e^{-\frac{(s-m)^2}{2\sigma''^2}} \qquad \mathcal{H}(d,s) \;\widehat{=}\; \frac{(s-m)^2}{2\,\sigma''^2}$$



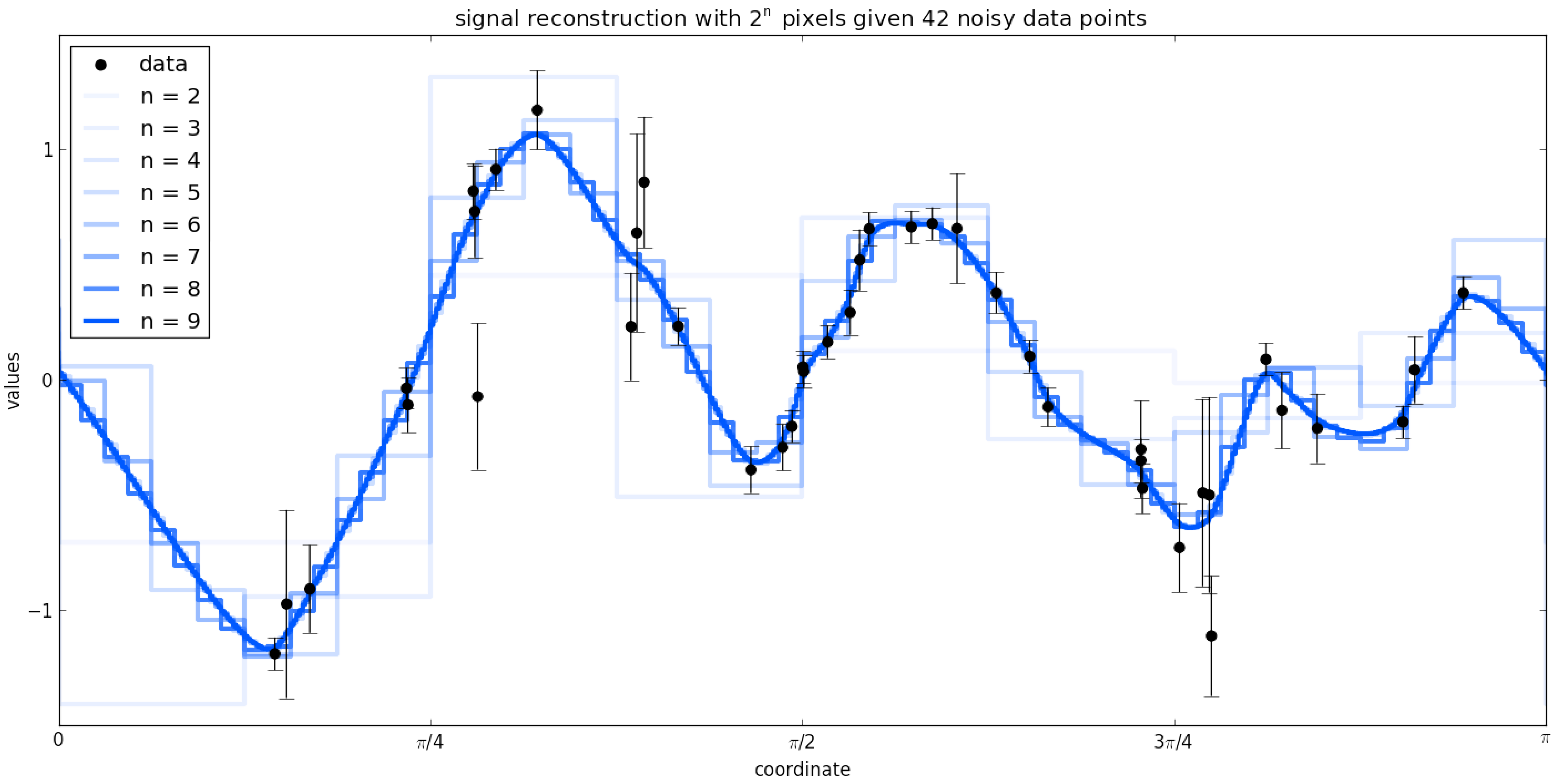



Correlations



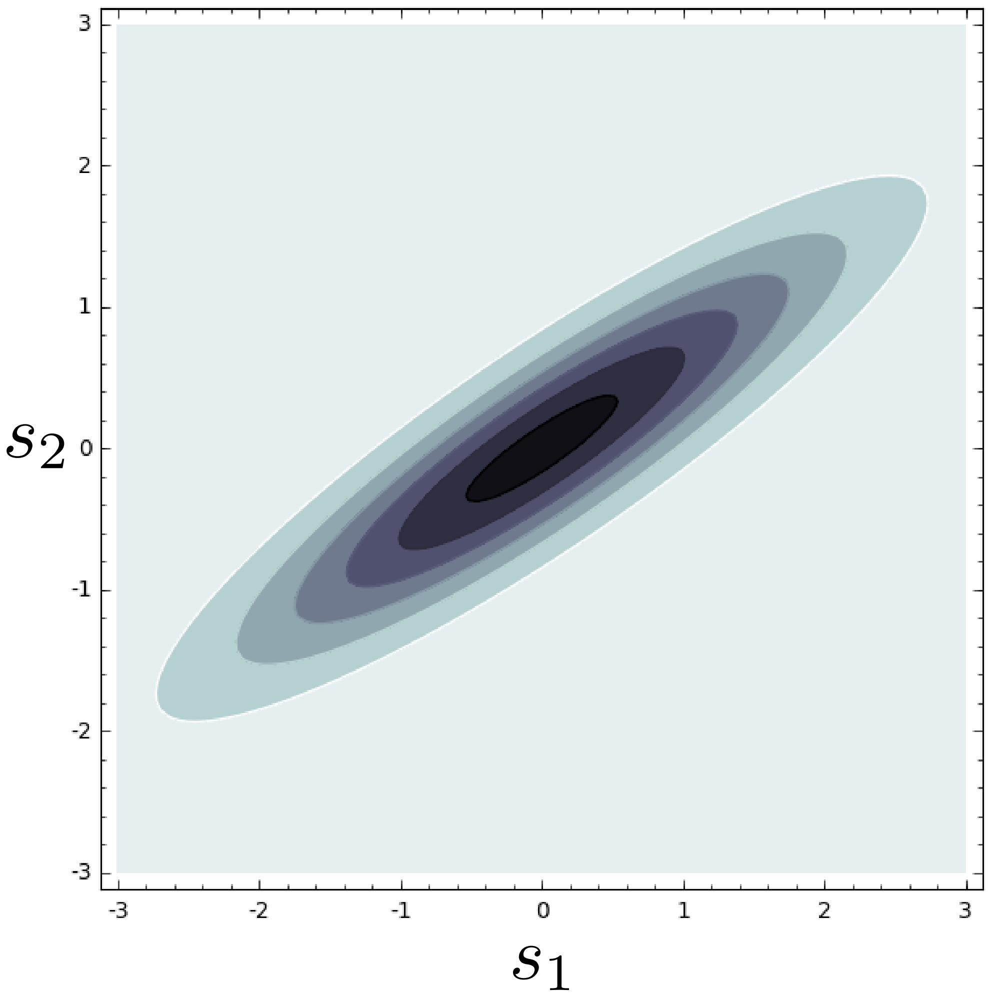
signal reconstruction with 2^n pixels given 42 noisy data points
data
n = 2
n = 3
n = 4
n = 5
n = 6
1
0
−1
values
0
π/4
π/2
3π/4
π
coordinate

Correlations



Correlations

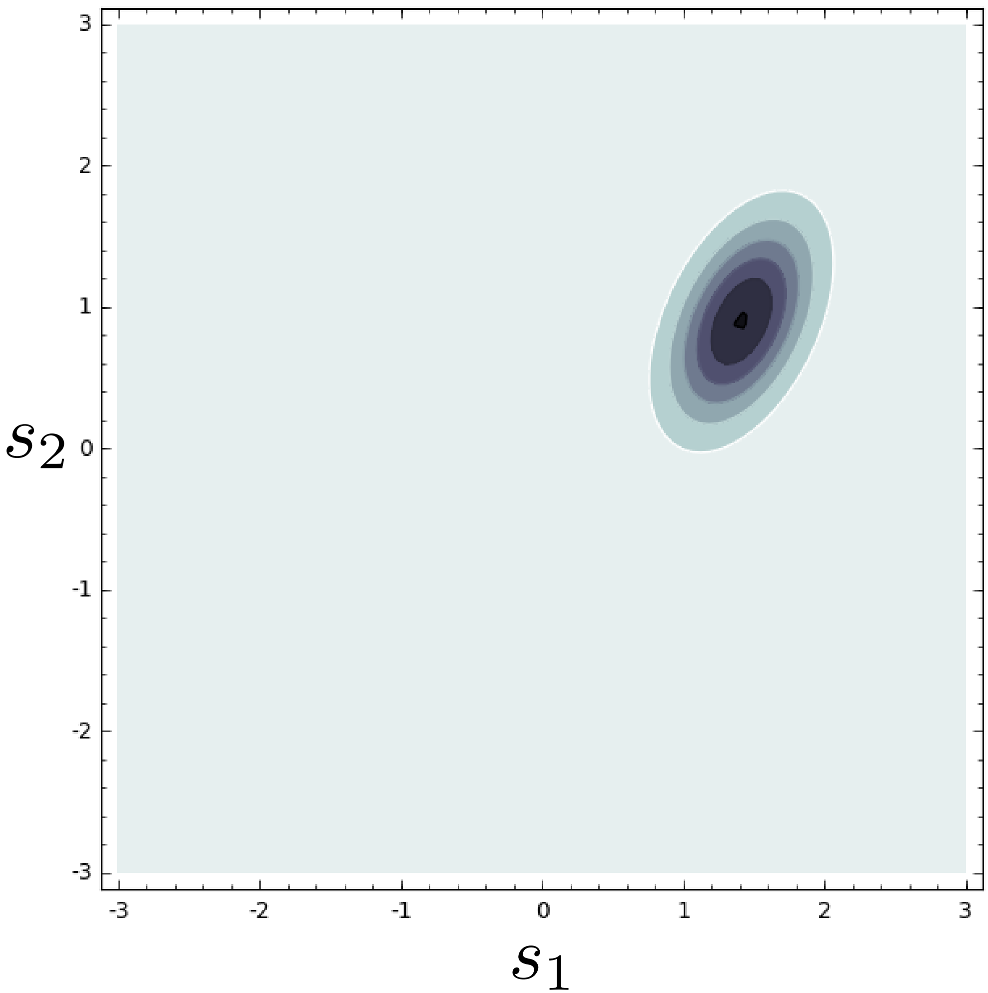
$$\mathcal{P}(s) \qquad\qquad s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

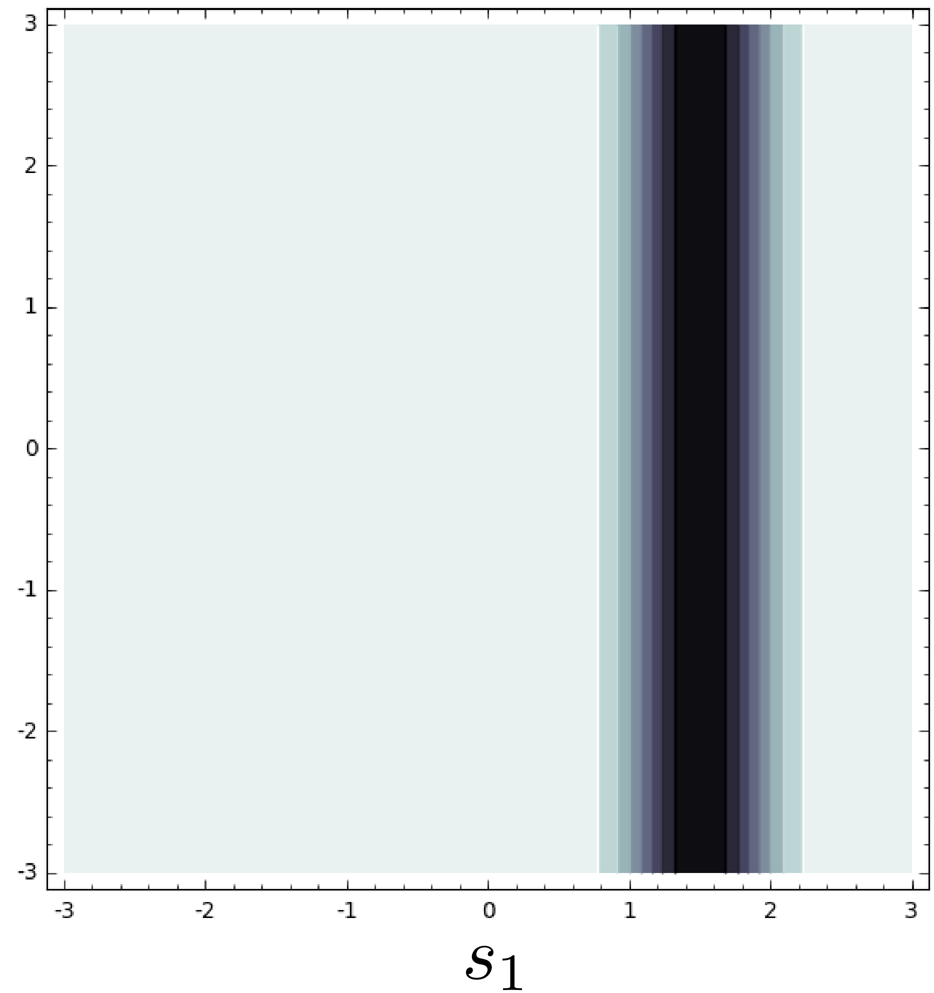


Correlations

$$\mathcal{P}(s|d) \qquad s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \qquad \mathcal{P}(d|s)$$

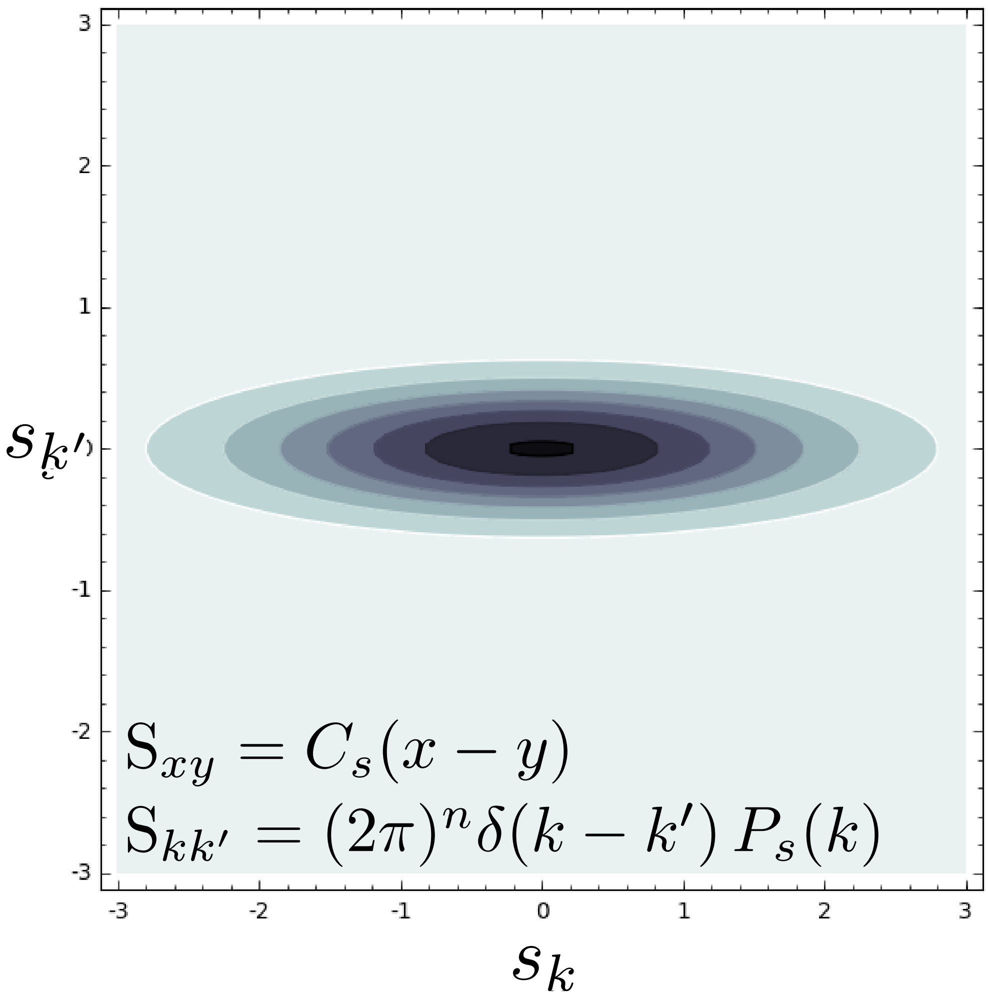
$$d = s_1 + n$$





Correlations

$$\mathcal{P}(s) \qquad s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$



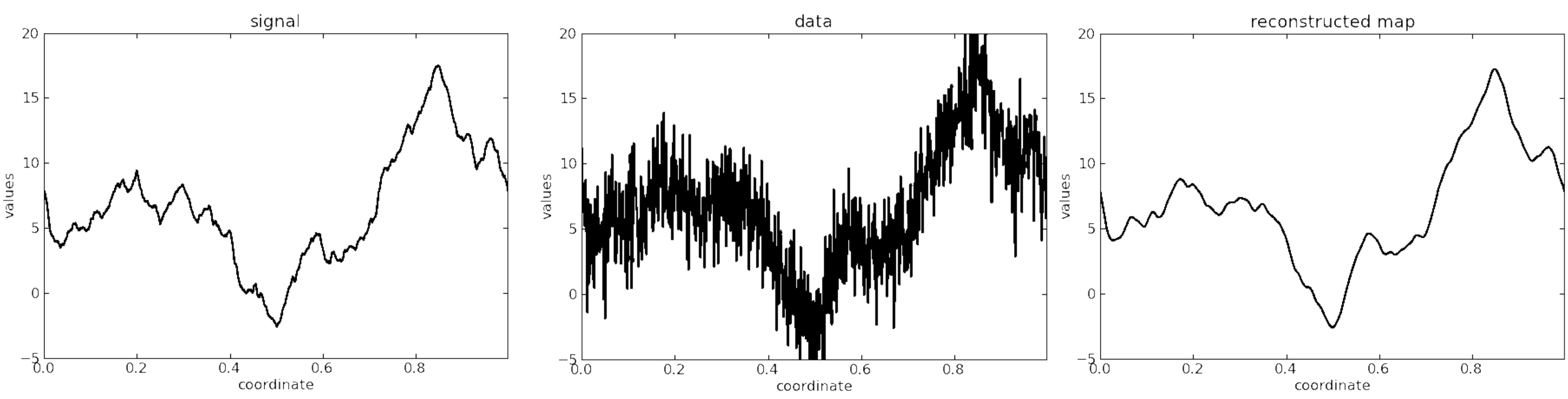
$$\begin{aligned}
\mathcal{P}(s) &= \mathcal{G}(s, S) \\
&= \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right) \\
S &= \begin{pmatrix} \langle s_1 s_1 \rangle & \langle s_1 s_2 \rangle \\ \langle s_2 s_1 \rangle & \langle s_2 s_2 \rangle \end{pmatrix} \quad \text{2-dim.} \\
S_{ij} &= \langle s_i s_j \rangle \quad n\text{-dim.} \\
S_{xy} &= \langle s_x s_y \rangle,\ x \in \mathbb{R}^n \quad \infty\text{-dim.} \\
s^\dagger S^{-1} s &= \int dx \int dy\, s_x\, (S^{-1})_{xy}\, s_y \\
s^\dagger j &= \int dx\, s_x\, j_x
\end{aligned}$$

Wiener filter

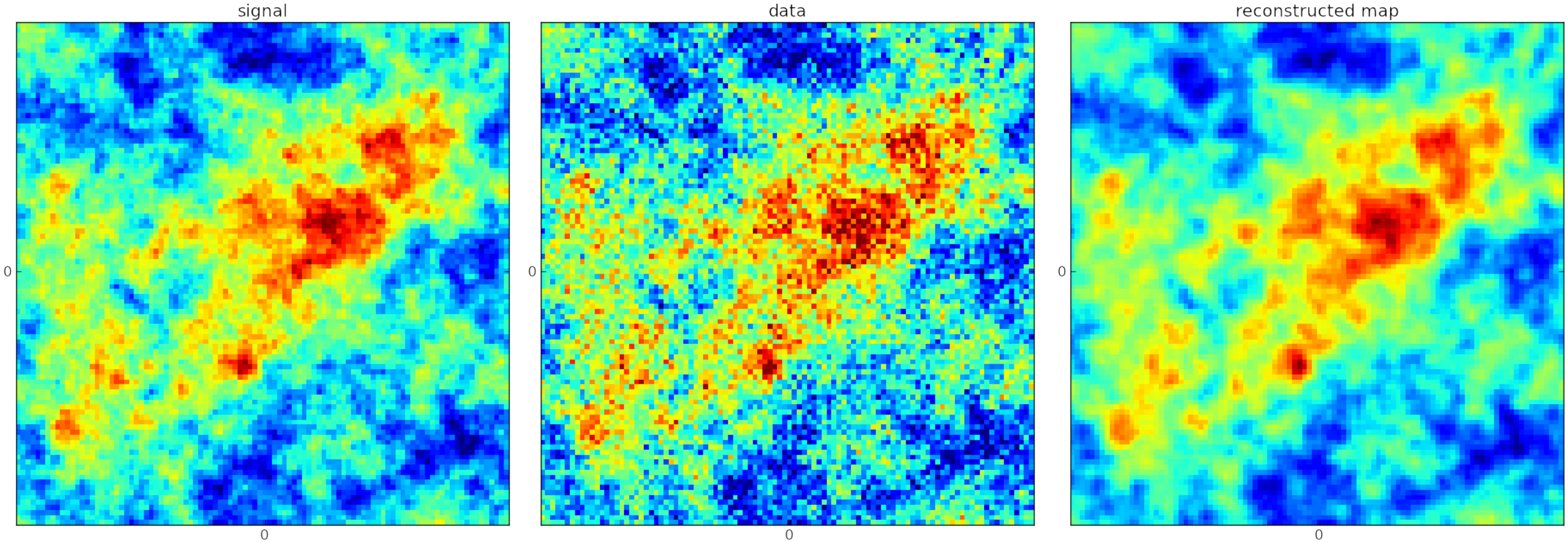
$$
\begin{aligned}
d &= R\,s + n && \text{data}\\
\mathcal{P}(d,s|R,S,N) &= \mathcal{G}(s,S)\,\mathcal{G}(d-R\,s,N) && \text{prior \& likelihood}\\
\mathcal{P}(s|d,R,S,N) &= \mathcal{G}(s-m,D) && \text{posterior}\\
\mathcal{H}(d,s|R,S,N) &\widehat{=} \frac{1}{2}\,s^\dagger S^{-1} s + \frac{1}{2}\,(d-R\,s)^\dagger N^{-1}(d-R\,s)\\
&\widehat{=} \frac{1}{2}\,[s^\dagger\,\underbrace{(S^{-1}+R^\dagger N^{-1}R)}_{=D^{-1}}\,s + s\,\underbrace{R^\dagger N^{-1} d}_{=j} + \underbrace{d^\dagger N^{-1} R}_{=j^\dagger}\,s]\\
&= \frac{1}{2}\,[s^\dagger D^{-1} s + s^\dagger j + j^\dagger s]\\
&= \frac{1}{2}\,[s^\dagger D^{-1} s + s^\dagger D^{-1}\,\underbrace{D\,j}_{=m} + j^\dagger D\,D^{-1} s]\\
&\widehat{=} \frac{1}{2}\,[(s-m)^\dagger D^{-1}(s-m)]
\end{aligned}
$$

Wiener filter

$$
\begin{aligned}
d &= R\,s + n && \text{data} \\
\mathcal{P}(d,s|R,S,N) &= \mathcal{G}(s,S)\,\mathcal{G}(d - R\,s, N) && \text{prior \& likelihood} \\
\mathcal{P}(s|d,R,S,N) &= \mathcal{G}(s-m, D) && \text{posterior} \\
m &= D\,j && \text{posterior mean} \\
j &= R^{\dagger}N^{-1}d && \text{information source} \\
D &= \left(S^{-1} + R^{\dagger}N^{-1}R\right)^{-1} && \text{information propagator}
\end{aligned}
$$

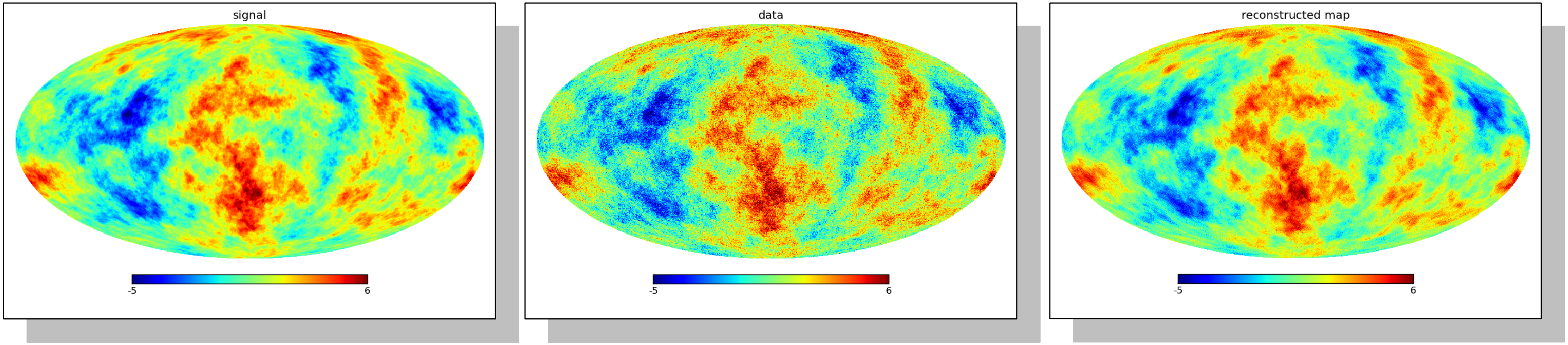


NIFTy – Numerical Information Field Theory



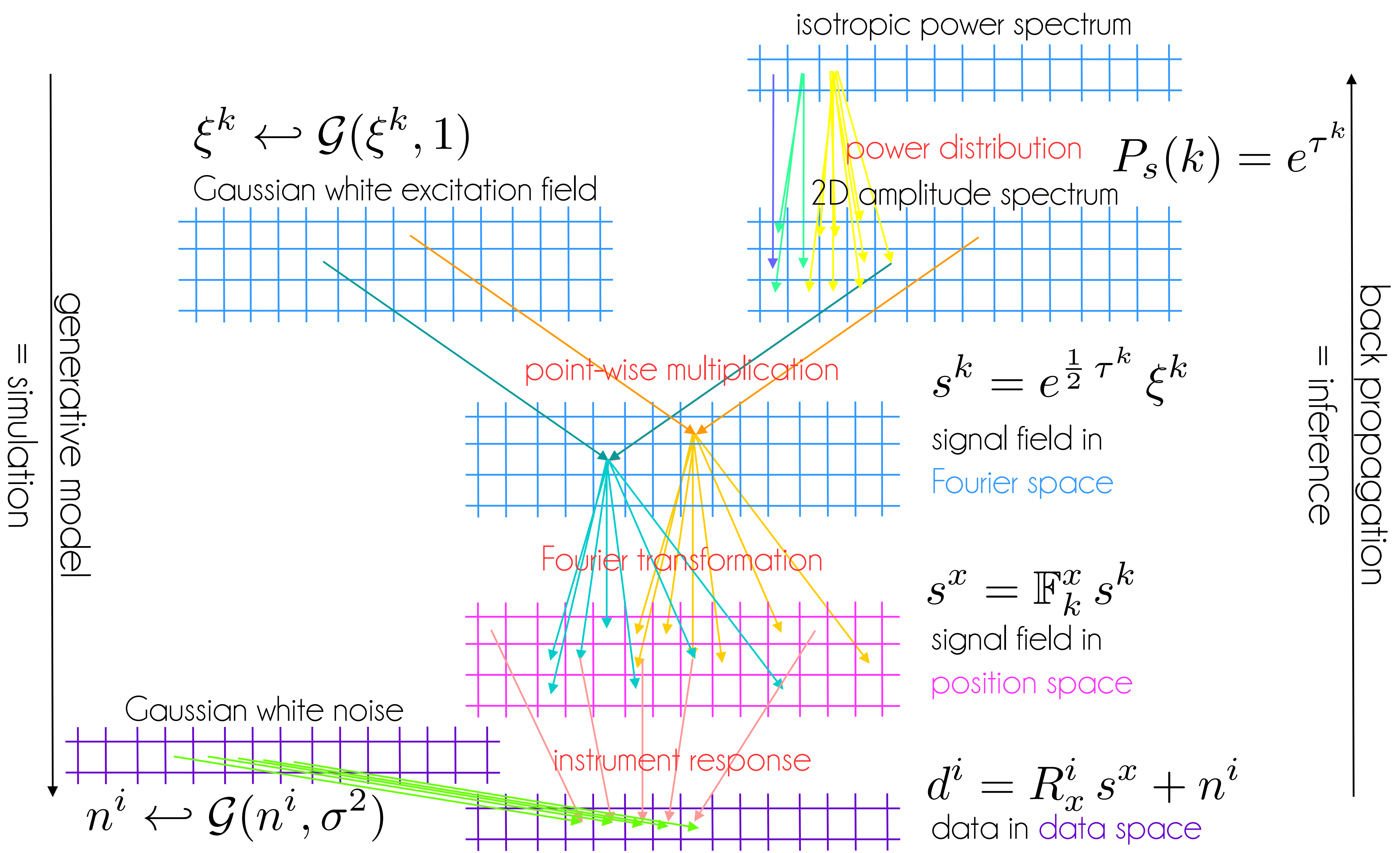

```
import nifty5 as ift
...
s_space = ift.RGSpace([N,N])
...
m = D(j)
```

NIFTy – Numerical Information Field Theory

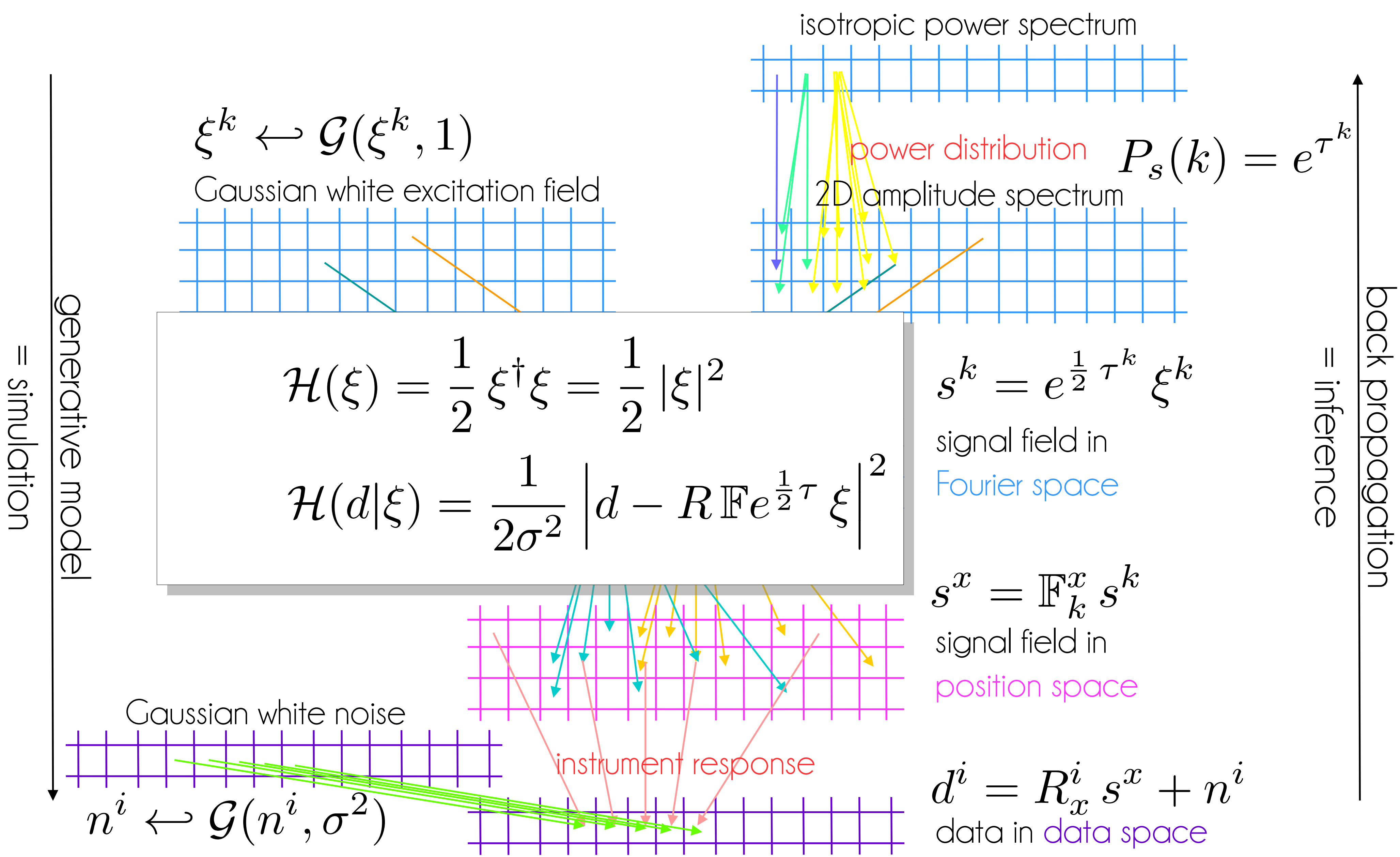



```
import nifty5 as ift
...
s_space = ift.HPSpace(NSide)
...
m = D(j)
```

Wiener filter as a neural network



Wiener filter as a neural network

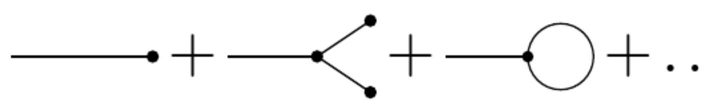


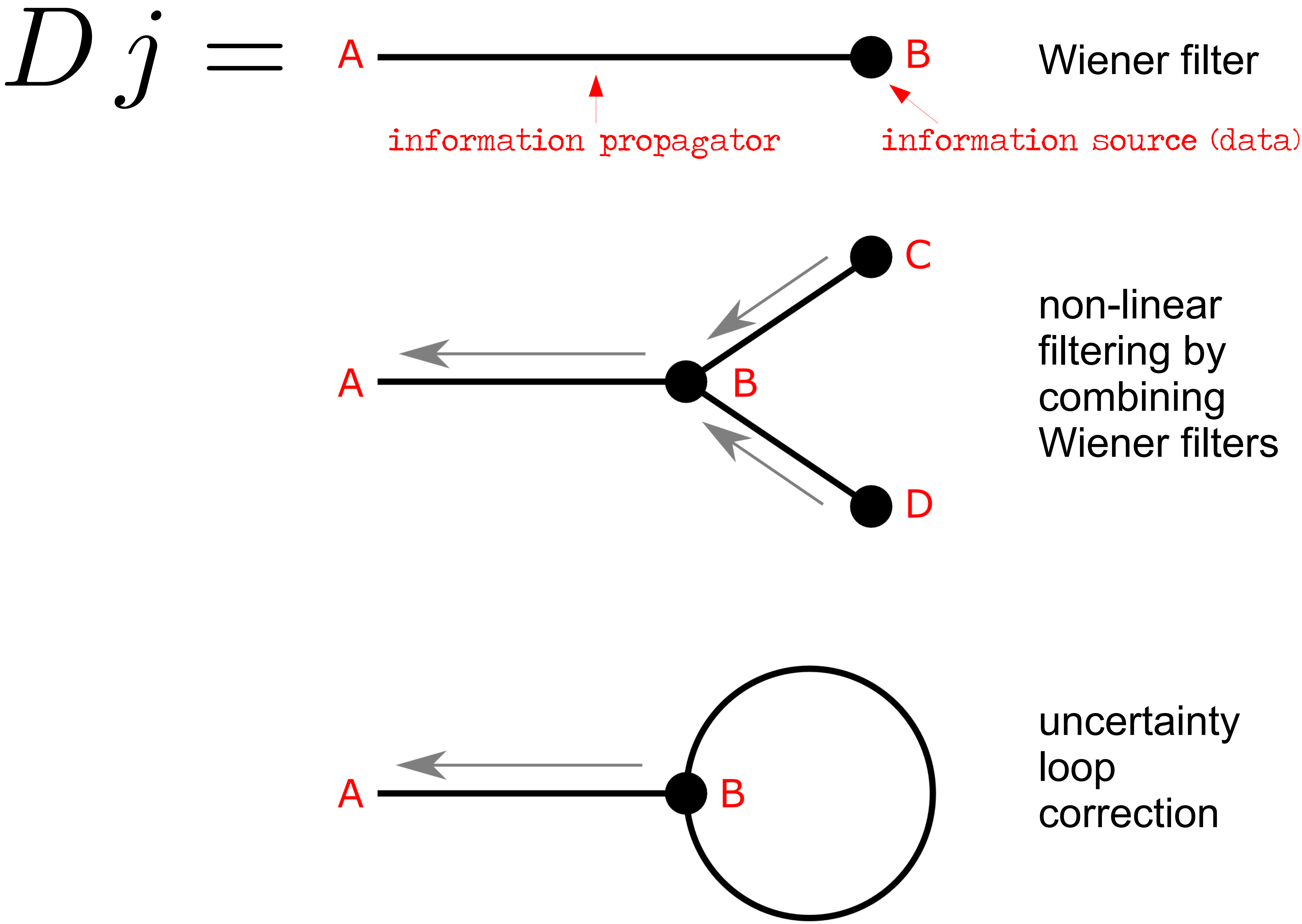
Interacting Theory

non-Gaussian signal, noise, or non-linear response

$$
\begin{aligned}
H(d,s) &= -\ln \mathcal{P}(d,s) \\
&= \underbrace{H_0 - j^\dagger s + \frac{1}{2} s^\dagger D^{-1} s}_{H_{\text{free}}} + \underbrace{\sum_{i=3}^{\infty} \int dx_1 \cdots \int dx_i \; \Lambda^{(i)}_{x_1 \ldots x_i} \; s_{x_1} \cdots s_{x_i}}_{H_{\text{int}}}
\end{aligned}
$$

$$
m = \langle s \rangle_{(s|d)} =
$$

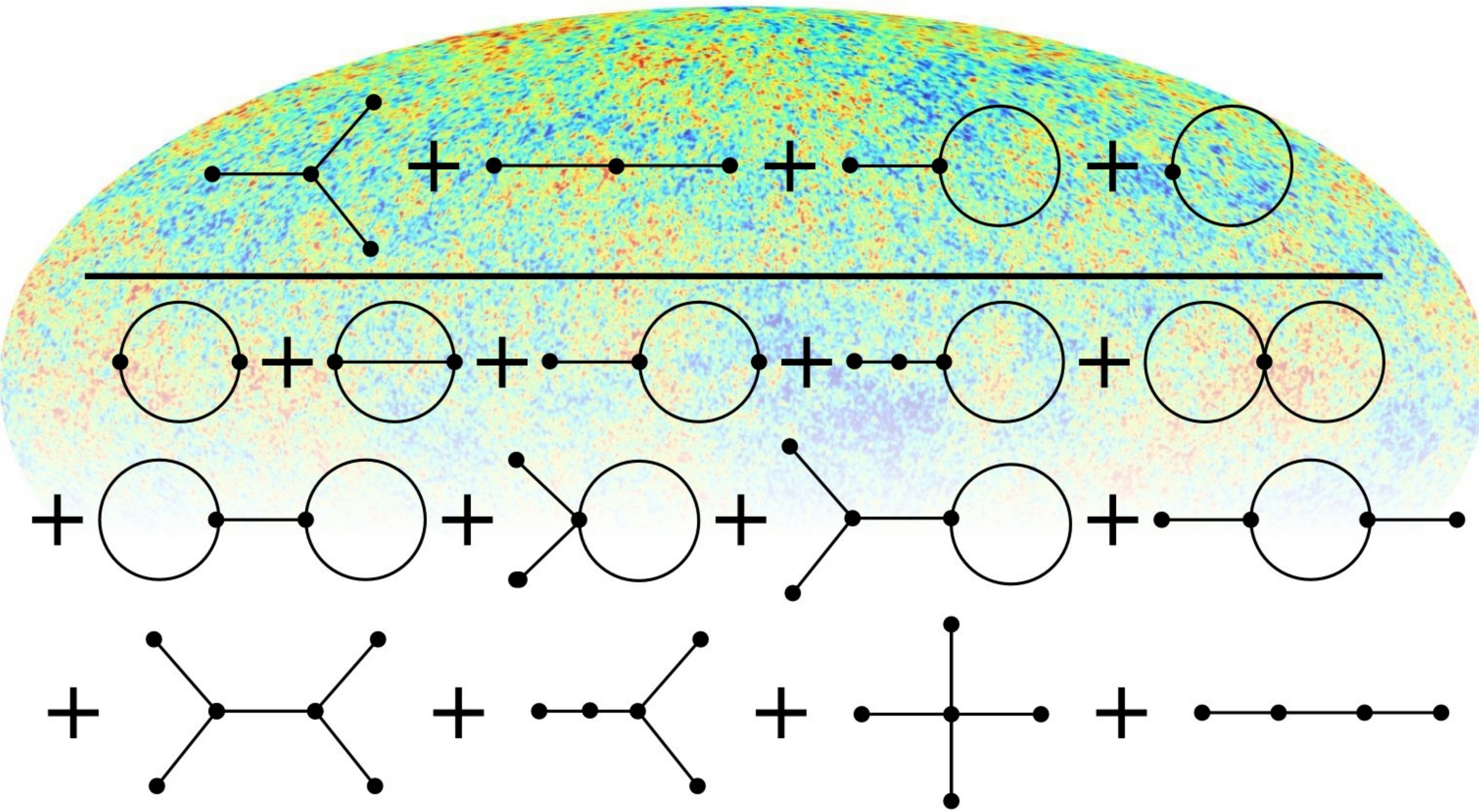




$D\,j =$
A
B
Wiener filter
information propagator
information source (data)
A
B
C
D
non-linear filtering by combining Wiener filters
A
B
uncertainty loop correction

Primordial non-Gaussianity measurement

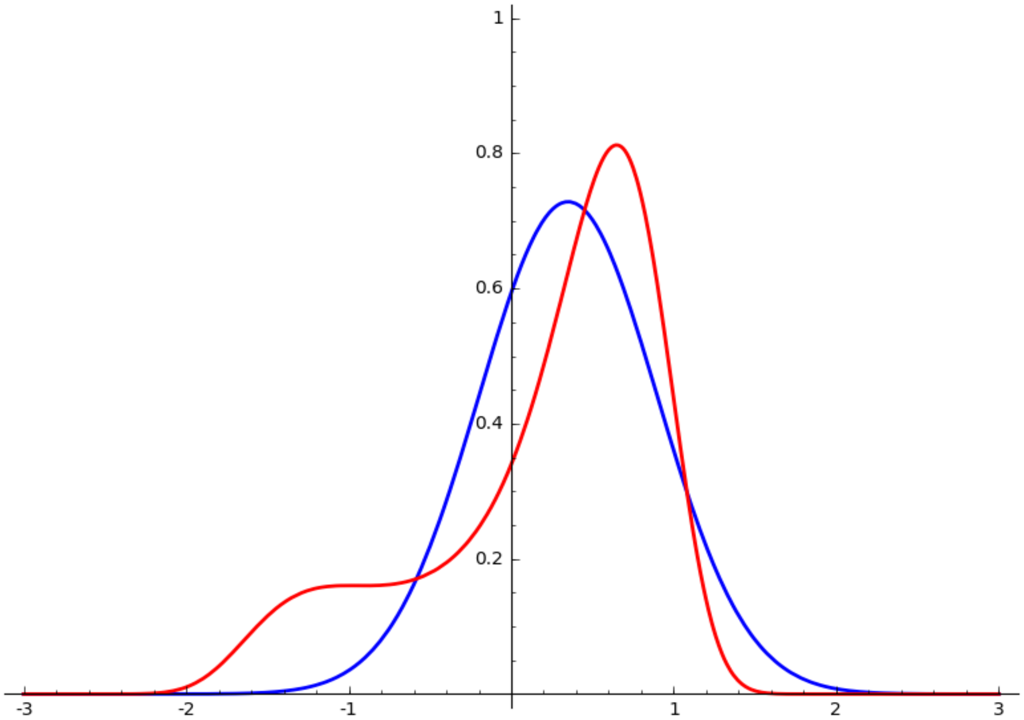
$$\phi = \varphi + f\left(\varphi^2 - \langle \varphi^2 \rangle\right)$$



Variational Bayes

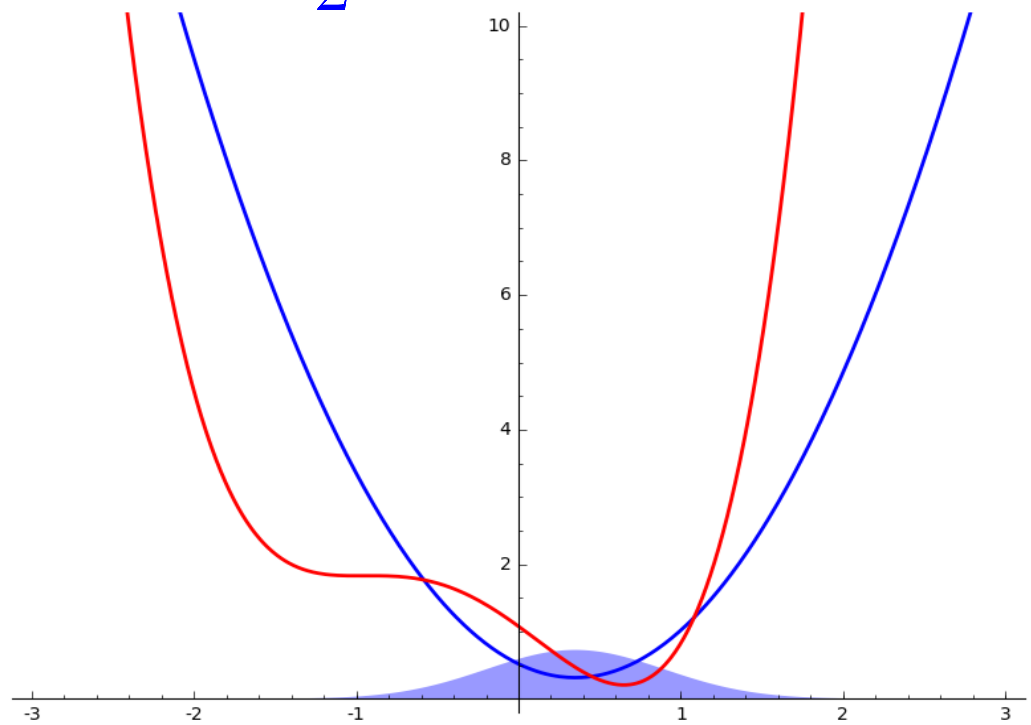
$\mathcal{P}(s|d)$

$\widetilde{\mathcal{P}}(s|d) = \mathcal{G}(s-m, D)$



$\mathcal{H}(s|d)$

$\widetilde{\mathcal{H}}(s|d) \widehat{=} \frac{1}{2}(s-m)^\dagger D^{-1}(s-m)$



$$\mathrm{KL}(\widetilde{\mathcal{P}}, \mathcal{P}) = \int \mathcal{D}s\, \widetilde{\mathcal{P}}(s|d) \left[\mathcal{H}(s|d) - \widetilde{\mathcal{H}}(s|d)\right]$$

Metric Gaussian Variational Bayes

$\mathcal{P}(s|d)$

$\widetilde{\mathcal{P}}(s|d) = \mathcal{G}(s-m, D)$

$\mathcal{H}(s|d)$

$\widetilde{\mathcal{H}}(s|d) \widehat{=} \frac{1}{2}(s-m)^\dagger D^{-1}(s-m)$

$$D \approx \left\langle \frac{\partial \mathcal{H}(d,s)}{\partial s} \, \frac{\partial \mathcal{H}(d,s)}{\partial s}^\dagger \right\rangle^{-1}_{(d|s=m)}$$

$$\mathrm{KL}(\widetilde{\mathcal{P}}, \mathcal{P}) = \frac{1}{n} \sum_{\substack{i \\ s_i \hookleftarrow \mathcal{G}(s_i - m, D)}} \left[\mathcal{H}(s_i|d) - \widetilde{\mathcal{H}}(s_i|d) \right] (s|d) \Big]$$

Denoising, Deconvolving, and Decomposing Photon Observations

Selig et al. (2014)
www.mpa-garching.mpg.de/ift/d3po

D^3PO

```
      /-\
     (oo )
      \=/
    /     \
   //|/o\|\\
  //  \ /  ||
 ||  |\ /|  ||
 ('  \_ _/  ')
     | | |
     | | |
     []|[]
     | | |
    /_] [_\
```

Data model

$$d = R\,e^{s} + n$$

known

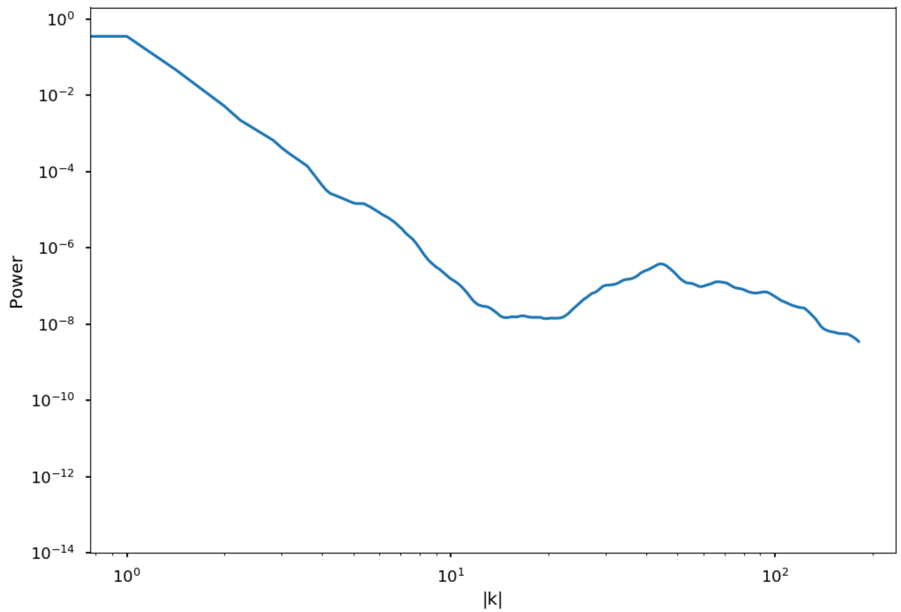
known response

unknown

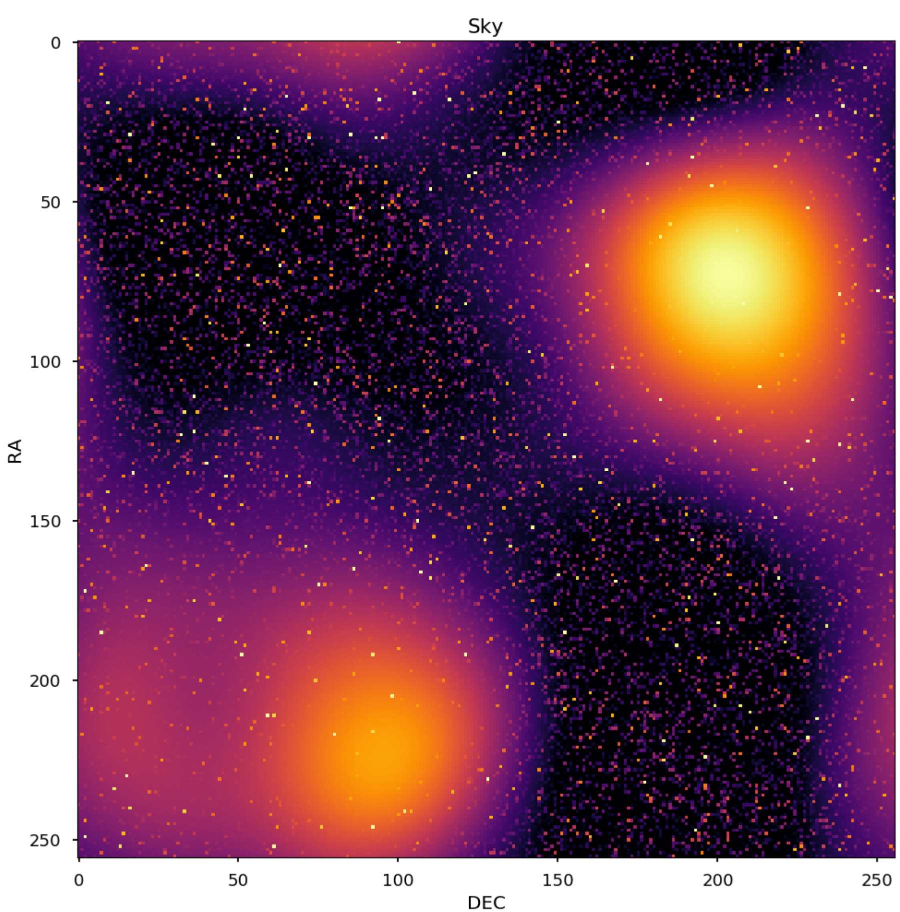
$$\lambda = R\,e^{s}$$

$$\mathcal{P}(s) = \mathcal{G}(s,\ S)$$ unknown

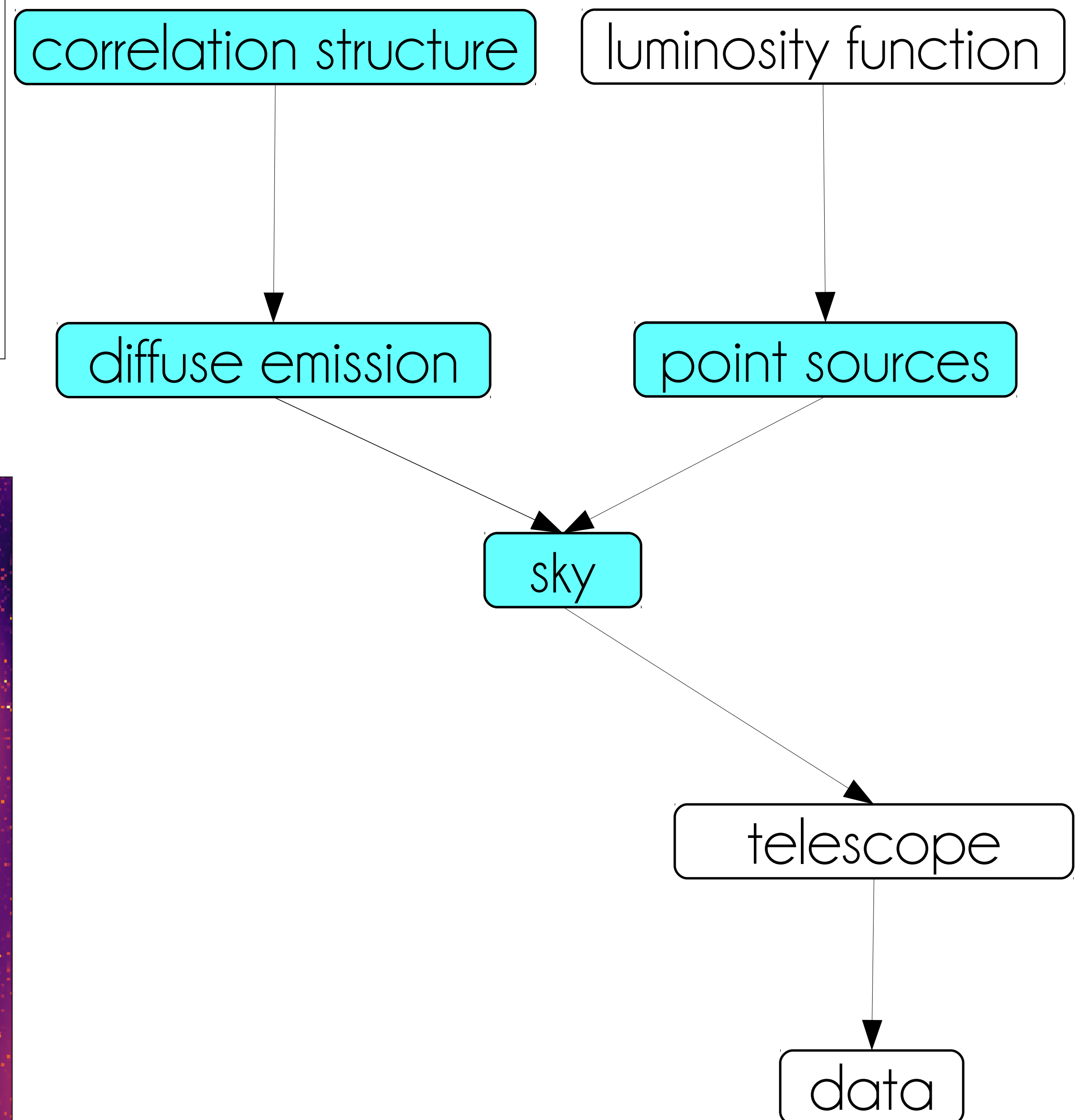
$$\mathcal{P}(d|\lambda) = \prod_i \frac{\lambda_i^{d_i}}{d_i!}\, e^{-\lambda_i}$$



Power
|k|
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{0}
10^{1}
10^{2}



Sky
RA
DEC
0
50
100
150
200
250



correlation structure
luminosity function
diffuse emission
point sources
sky
telescope
data

Information

$$\mathcal{H}(\boldsymbol{d}, \boldsymbol{s}, \boldsymbol{\tau}) = -\log \mathcal{P}(\boldsymbol{d}, \boldsymbol{s}, \boldsymbol{\tau})$$

likelihood

$$= 1^\dagger \left[\log(d!) + \boldsymbol{R}\,(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}})\right] - \boldsymbol{d}^\dagger \log\left[\boldsymbol{R}\,(\mathrm{e}^{\boldsymbol{s}} + \mathrm{e}^{\boldsymbol{u}})\right]$$

prior / regularization

$$+ \frac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s} + \frac{1}{2}\log\left(\det\left[\boldsymbol{S}\right]\right)$$

hyper-prior / intelligence

$$+ (\boldsymbol{\alpha} - \boldsymbol{1})^\dagger \boldsymbol{\tau} + \boldsymbol{q}^\dagger \mathrm{e}^{-\boldsymbol{\tau}} + \frac{1}{2}\boldsymbol{\tau}^\dagger \boldsymbol{T} \boldsymbol{\tau}$$

point prior

$$+ (\boldsymbol{\beta} - \boldsymbol{1})^\dagger \boldsymbol{u} + \boldsymbol{\eta}^\dagger \mathrm{e}^{-\boldsymbol{u}}$$

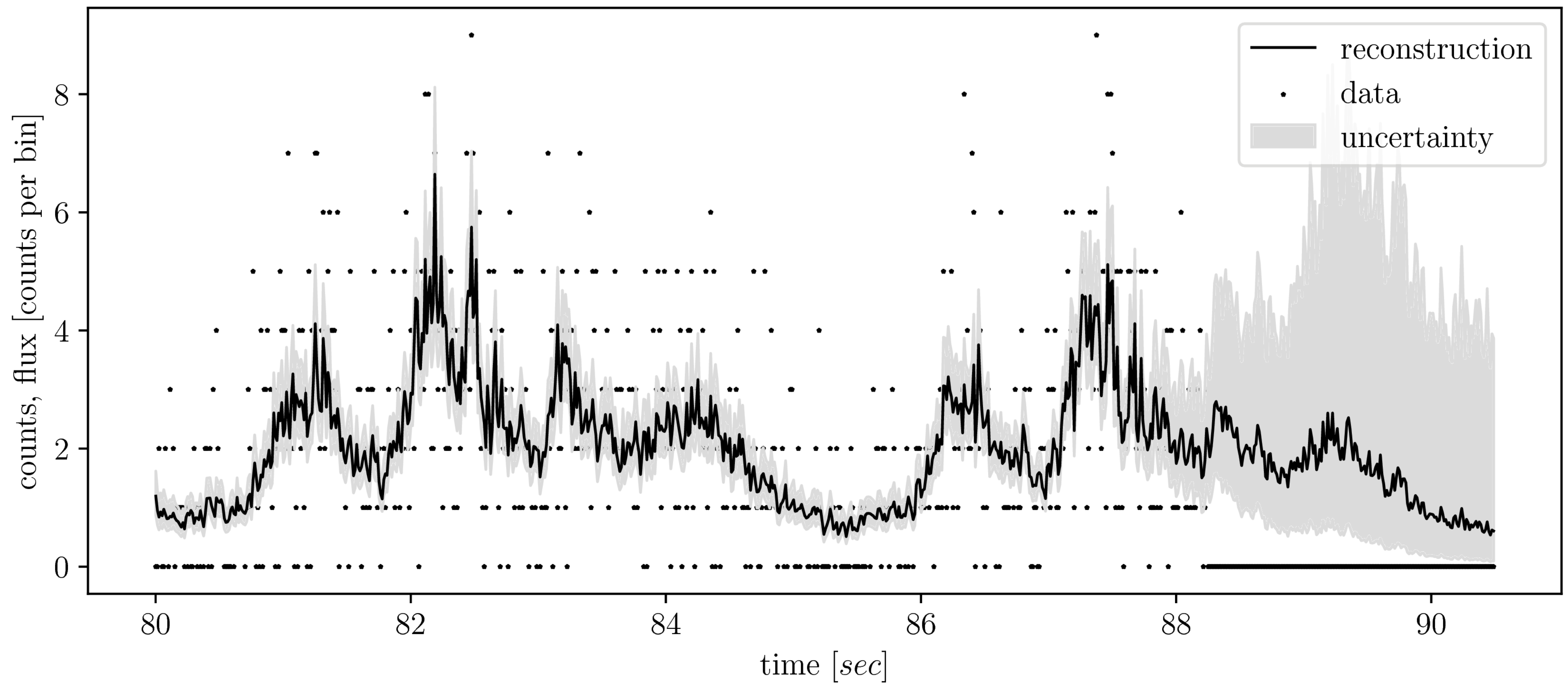
correlation structure

$$\boldsymbol{S} = \sum_k \mathrm{e}^{\tau_k} \boldsymbol{S}_k$$

D³PO in 1D & QPOs

Magnetar flare SGR 1900+14

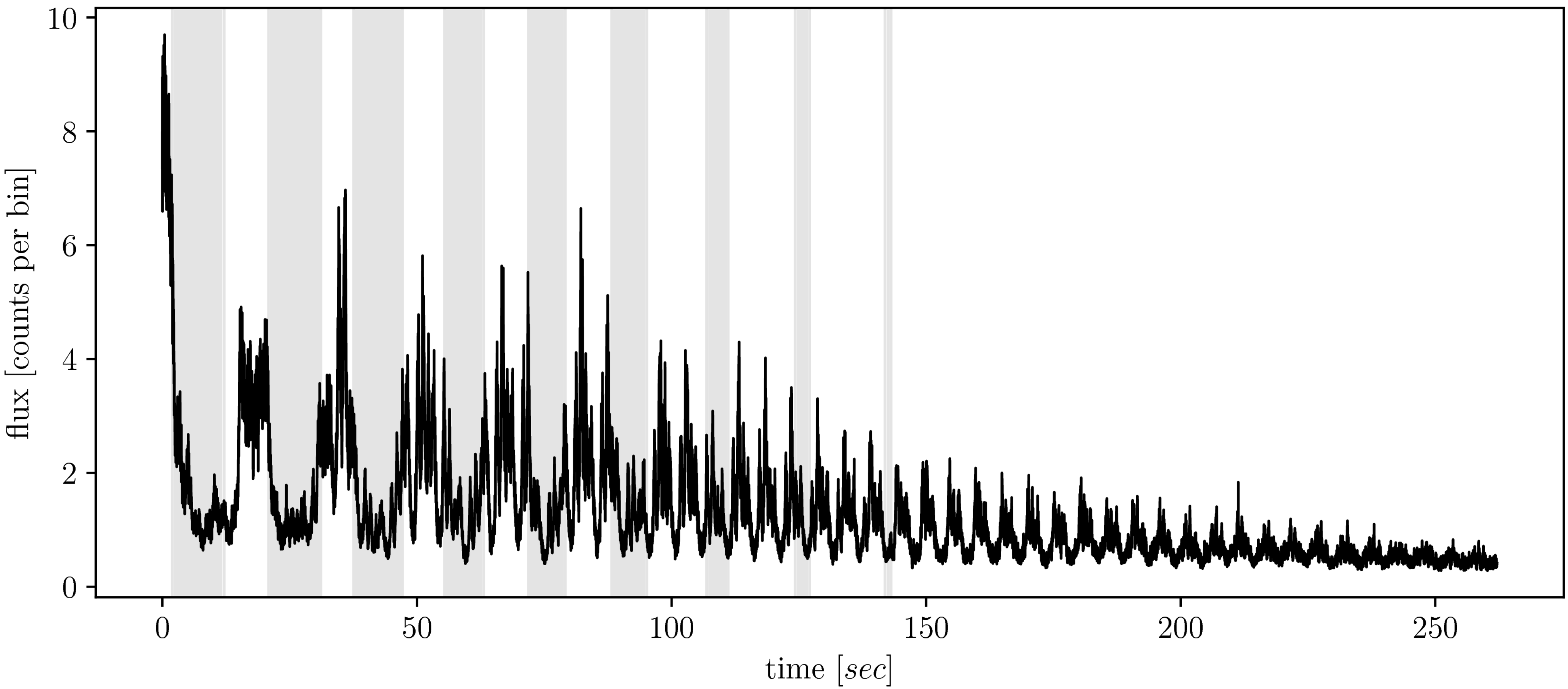
Pumpe et al. arXiv:1708.05702



D³PO in 1D & QPOs

Magnetar flare SGR 1900+14

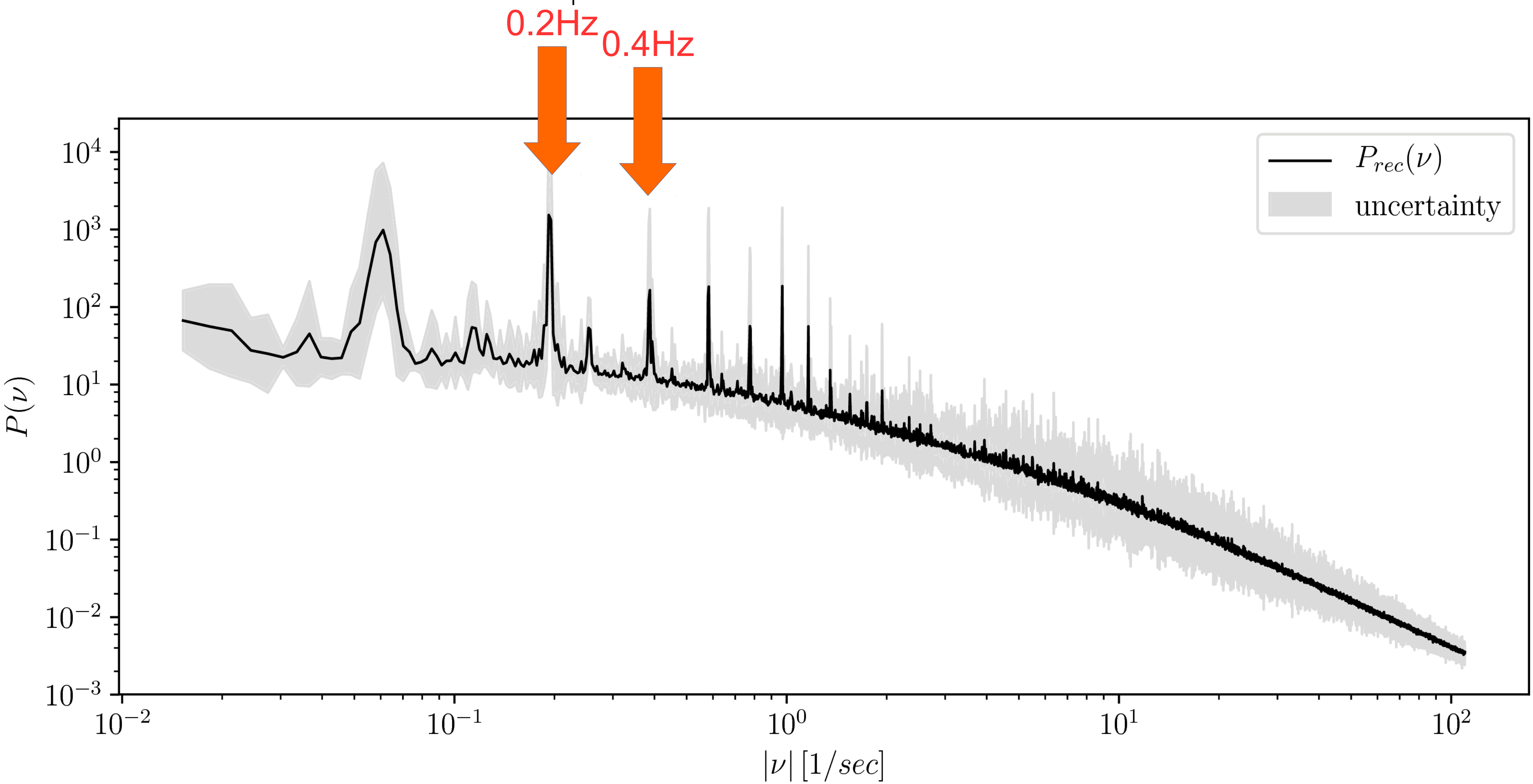
Pumpe et al. arXiv:1708.05702

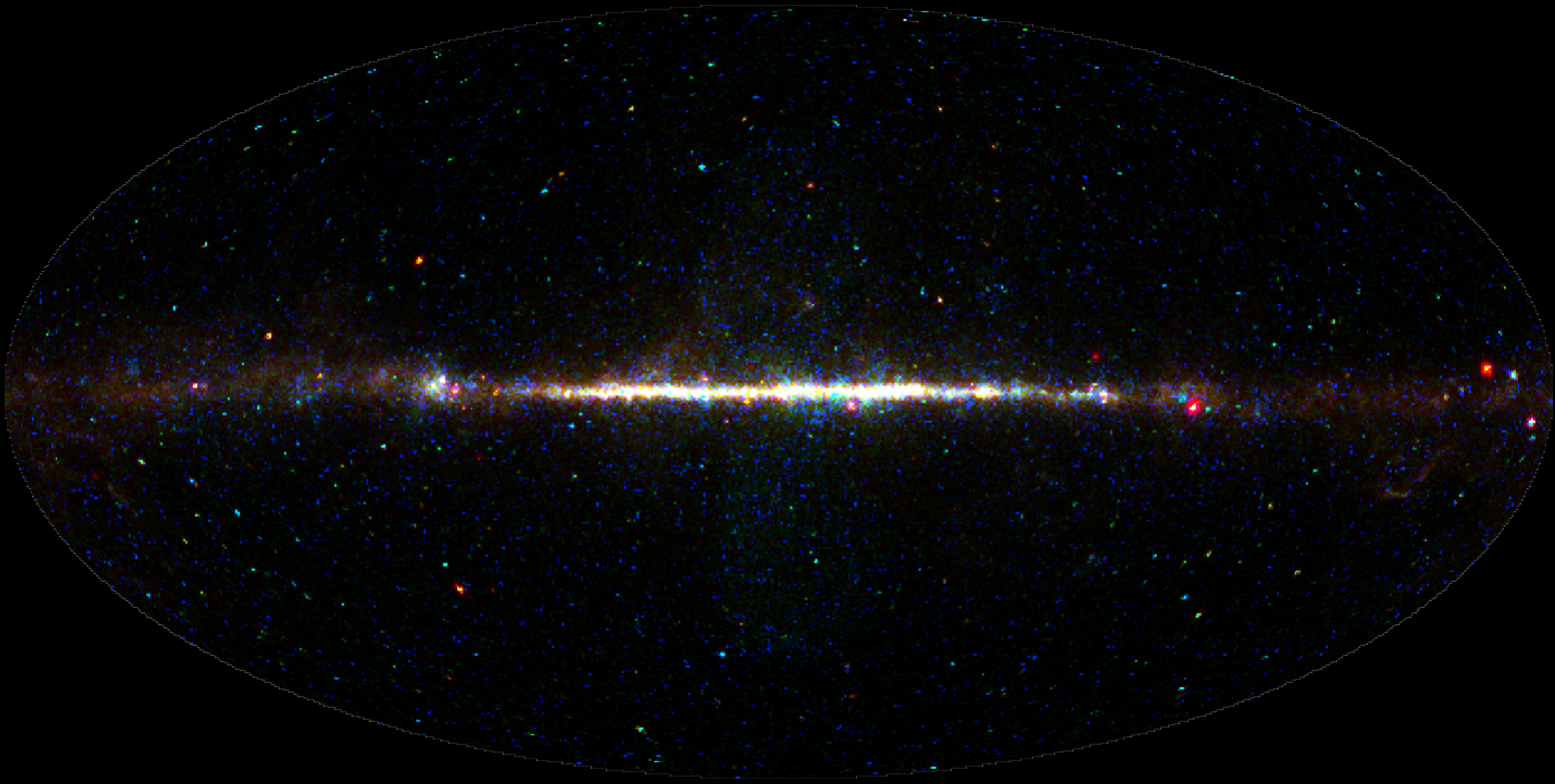


D³PO in 1D & QPOs

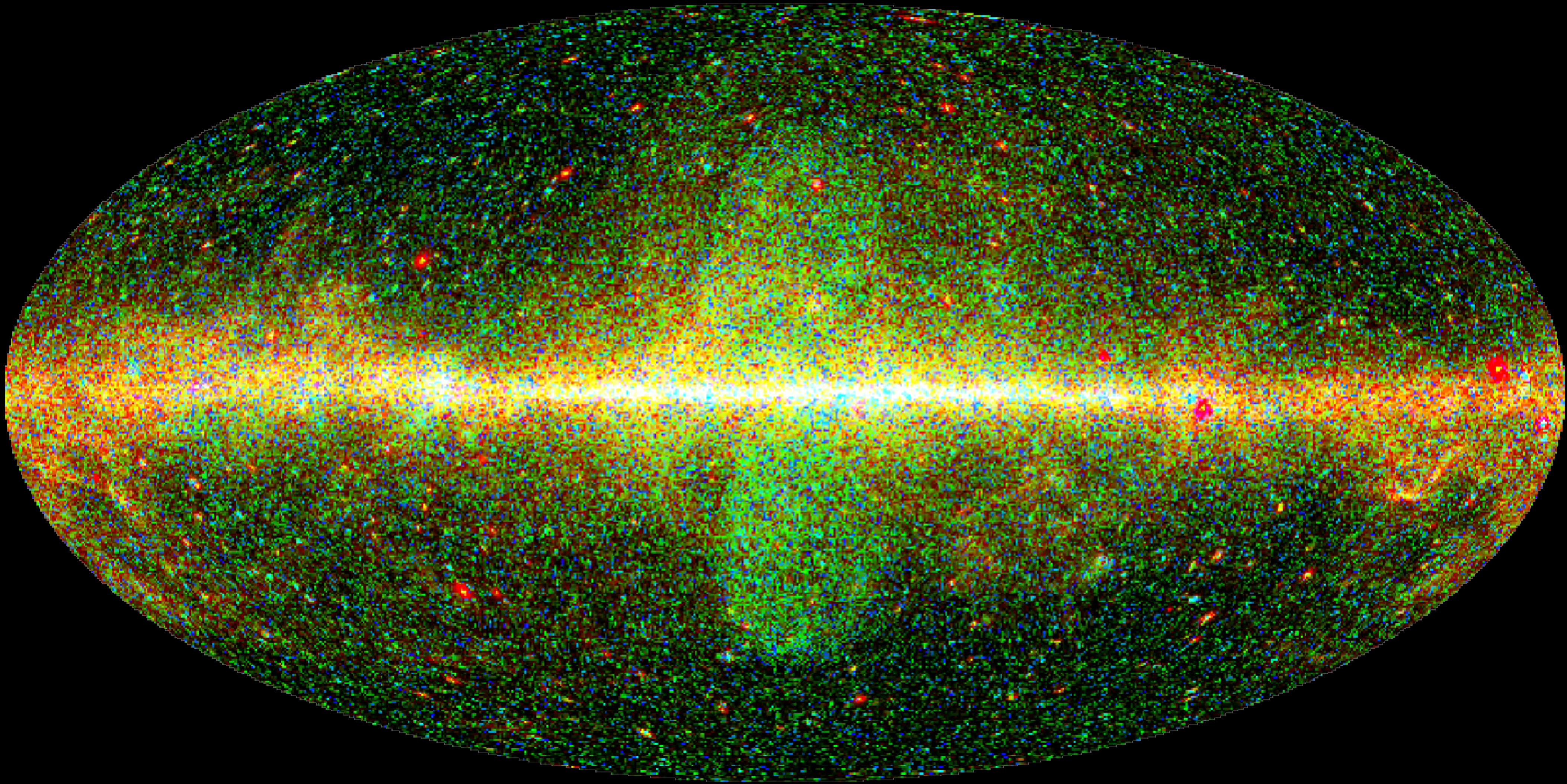
Magnetar flare SGR 1900+14

Pumpe et al. arXiv:1708.05702

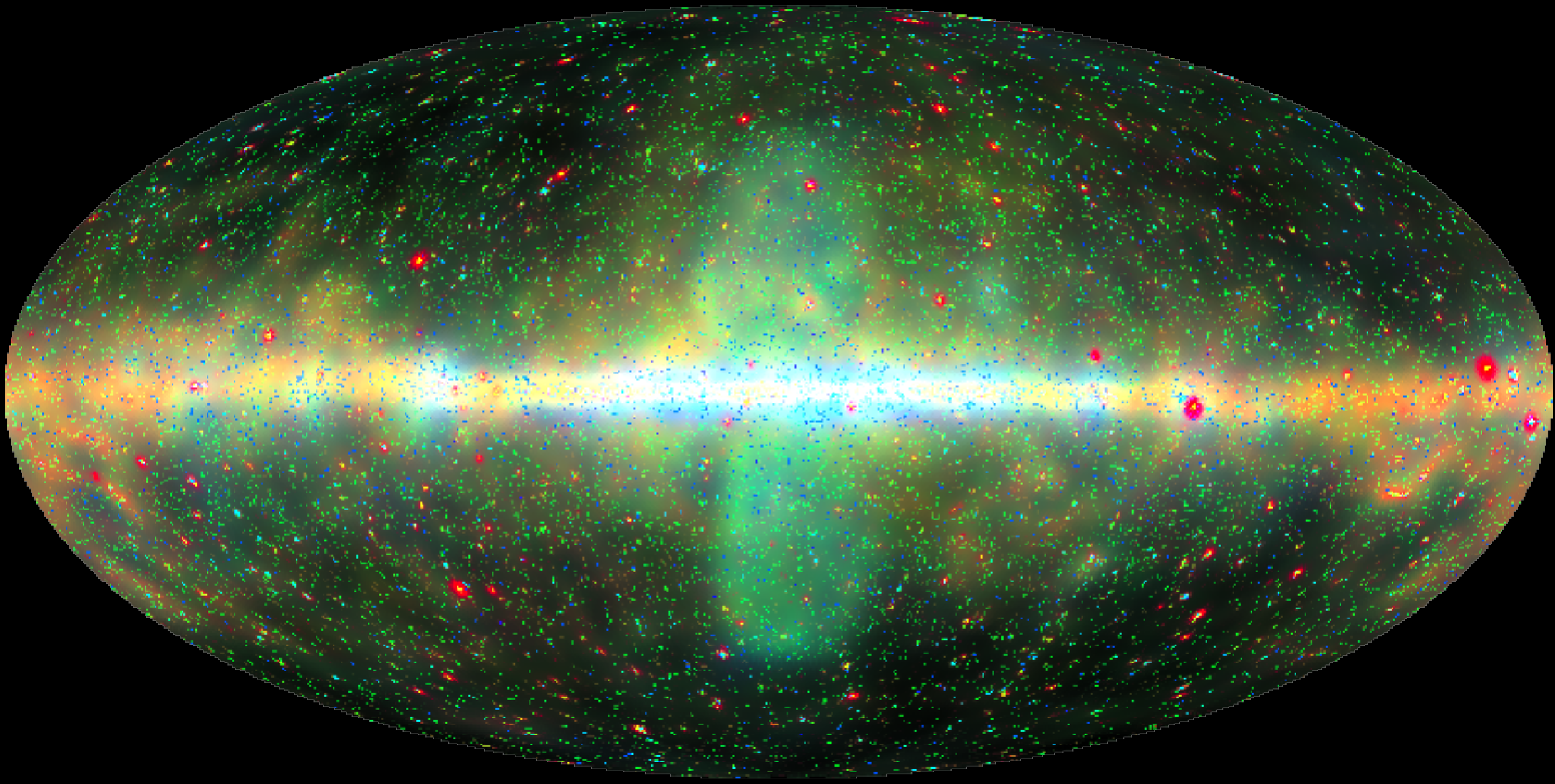




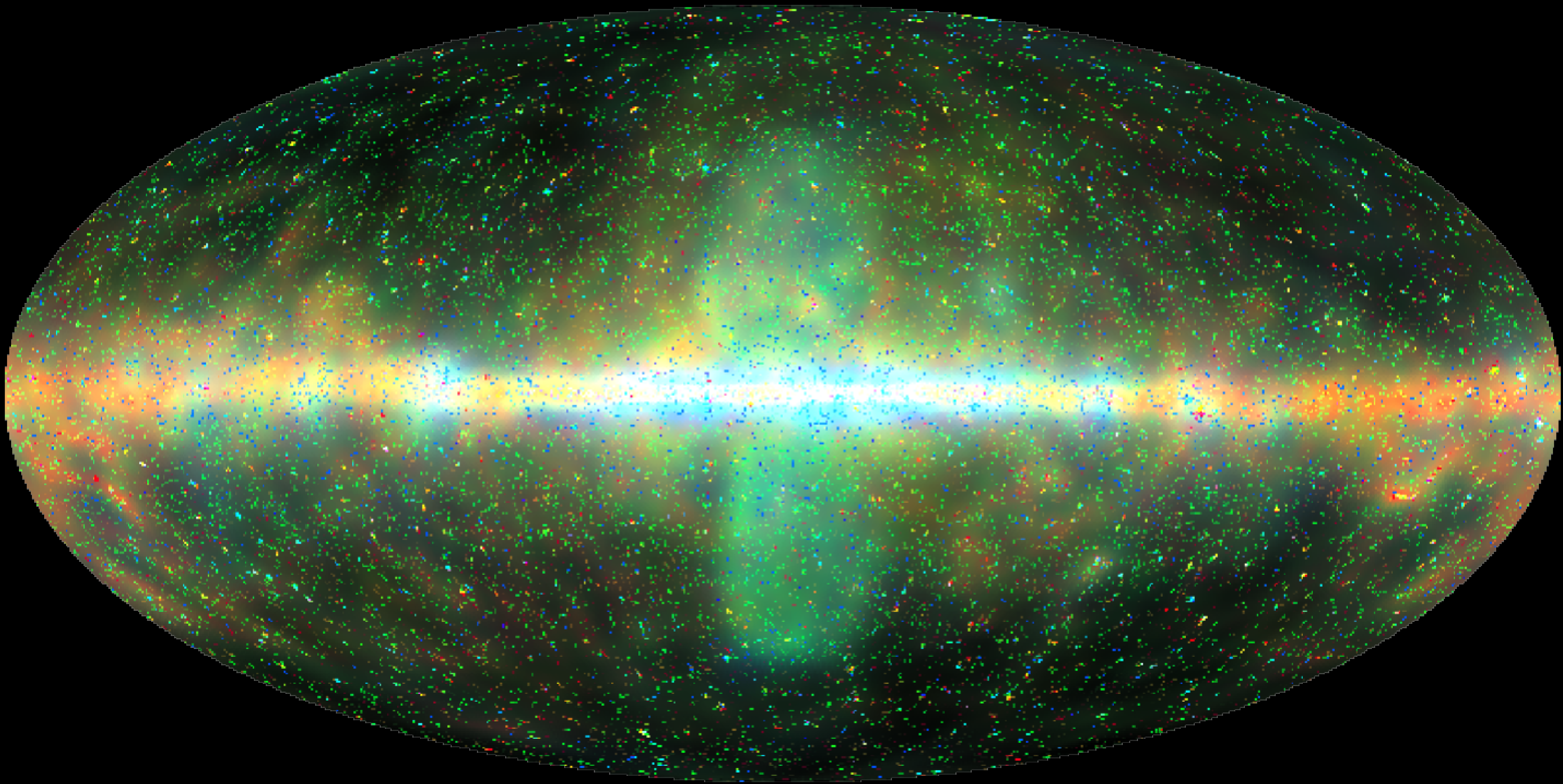
data



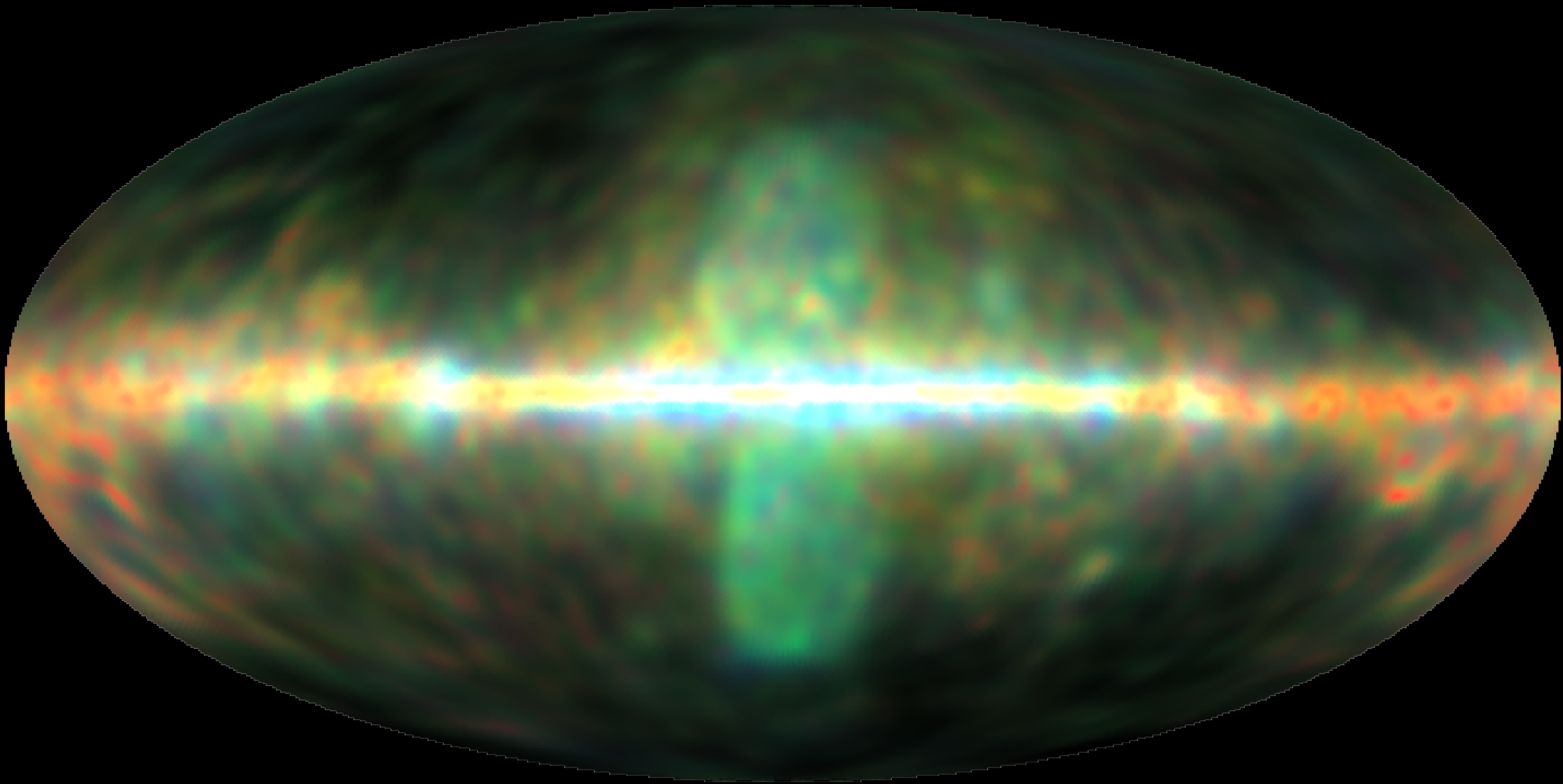
log-data



log-data … denoised

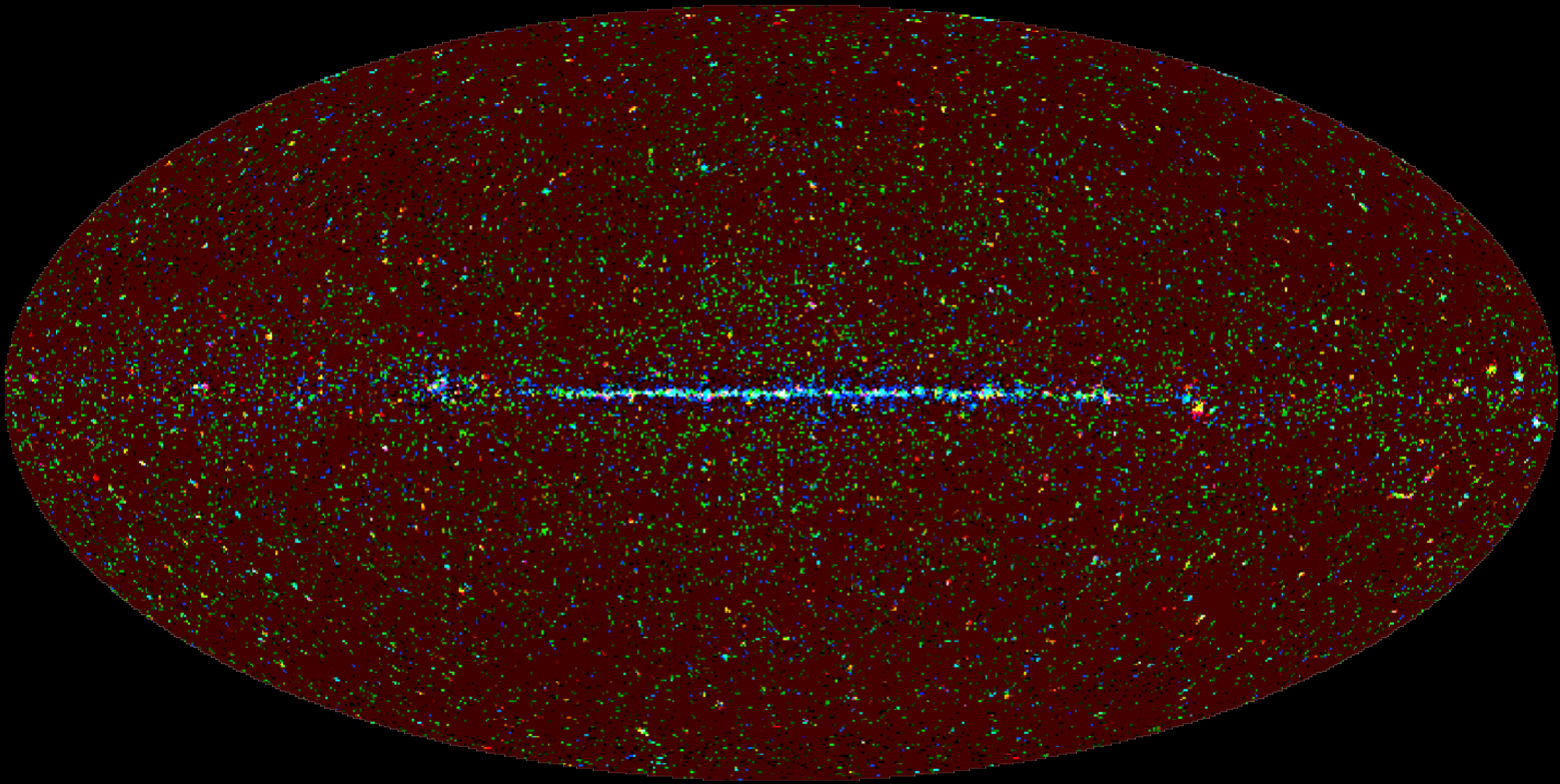


log-data … denoised … deconvolved



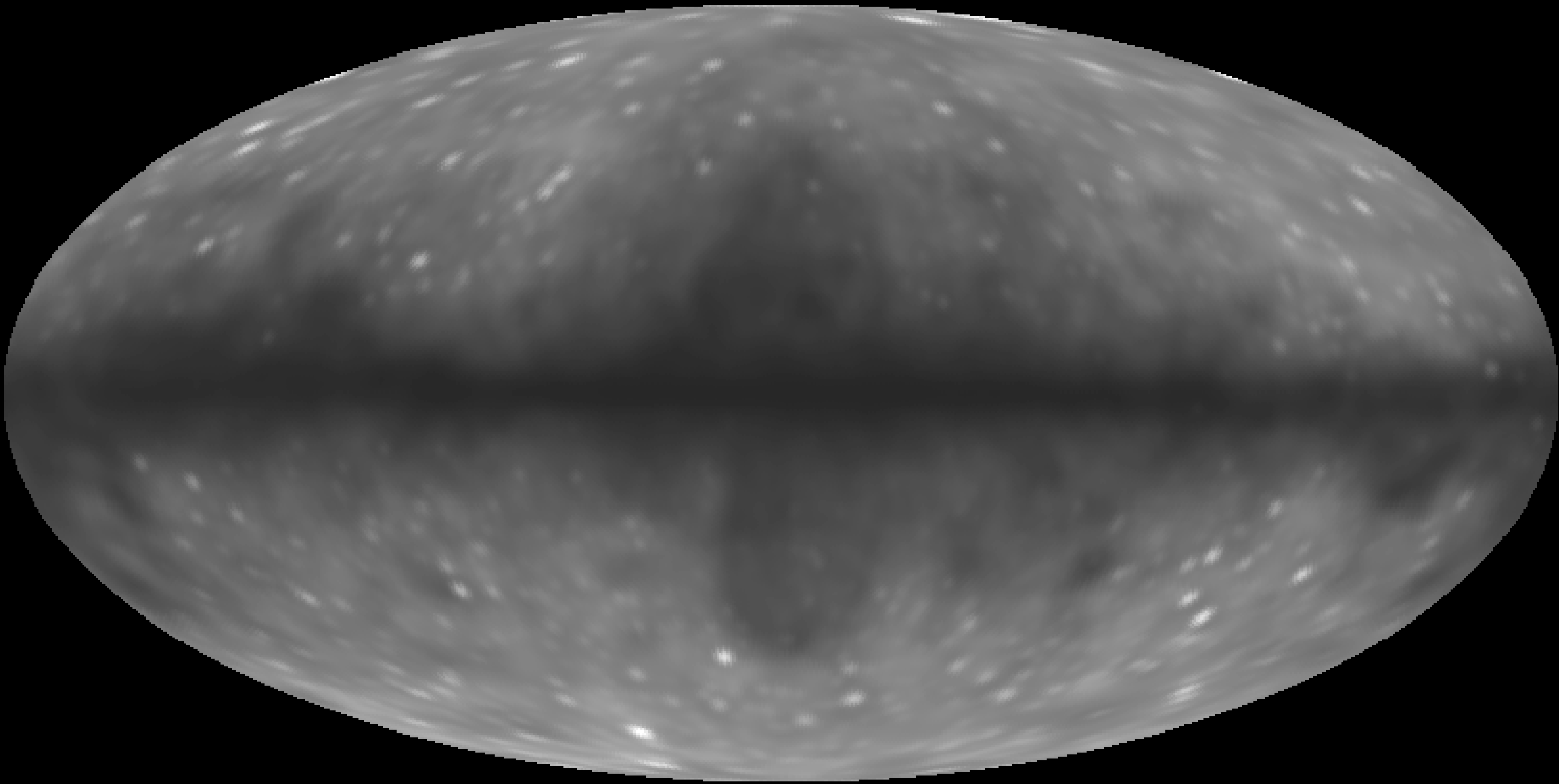
log-data … denoised … deconvolved … decomposed

Selig, Vacca, Oppermann, Enßlin (2015)



log-data … denoised … deconvolved … decomposed

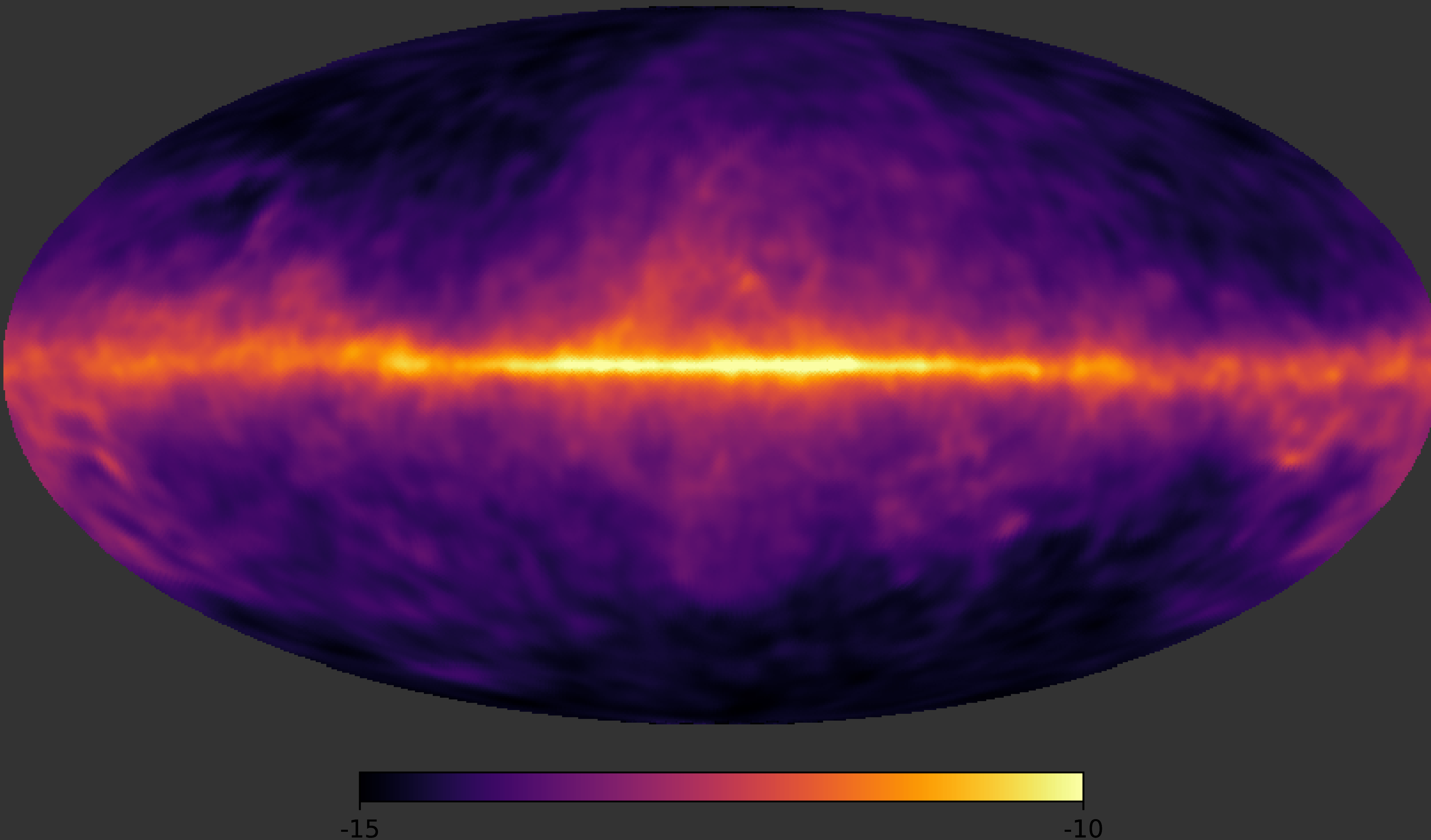
Selig, Vacca, Oppermann, Enßlin (2015)



relative uncertainty of diffuse emission

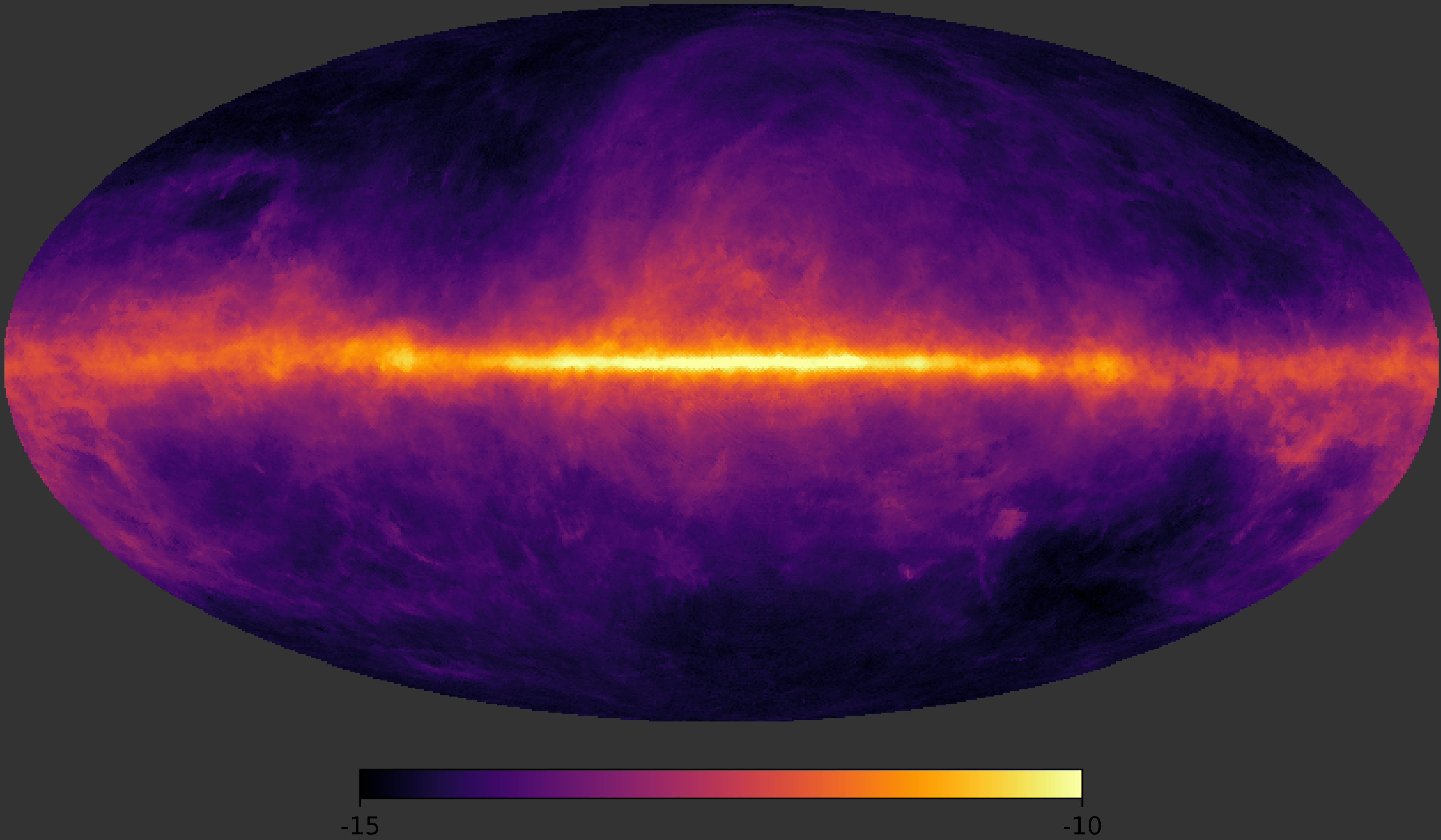
Sharpening up Galactic all-sky maps with complementary data

Ancla Müller et al (2018)



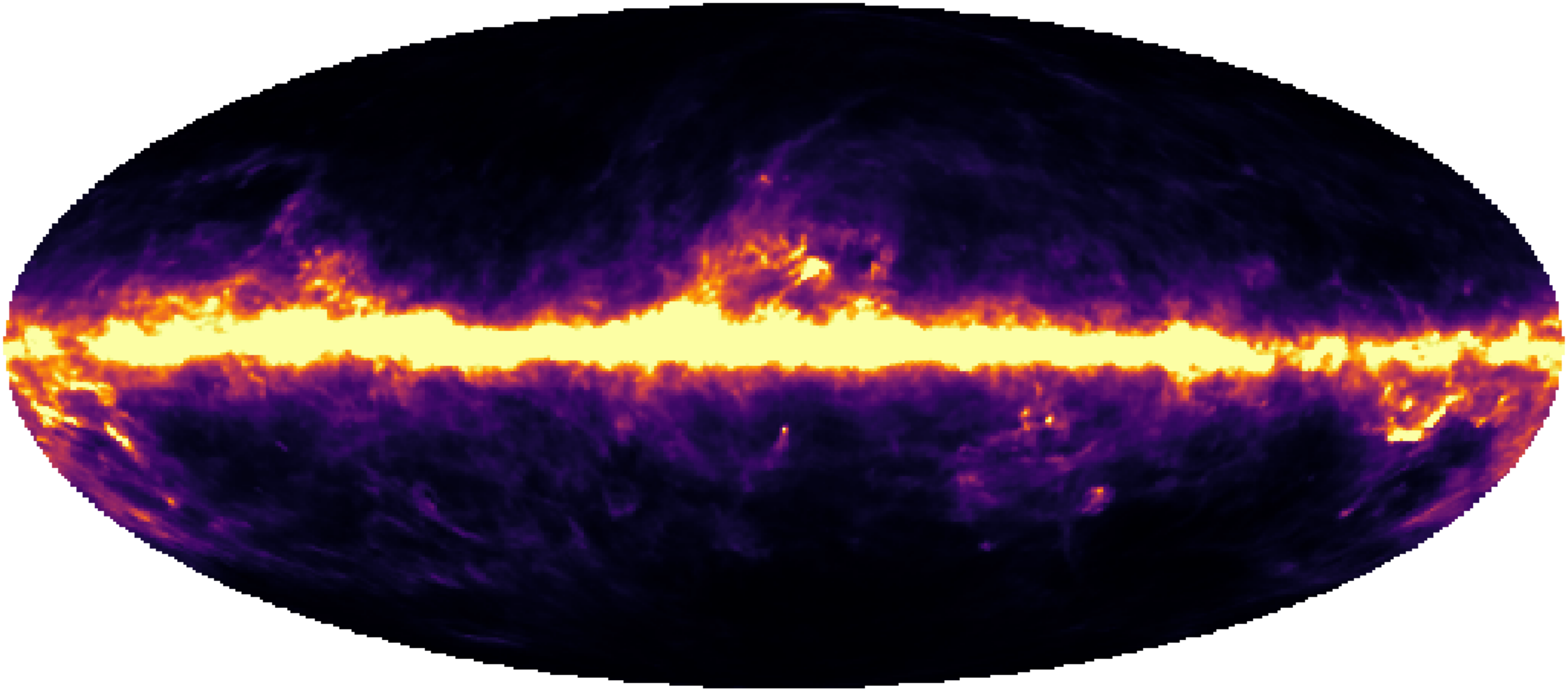
Sharpening up Galactic all-sky maps with complementary data

Ancla Müller et al (2018)



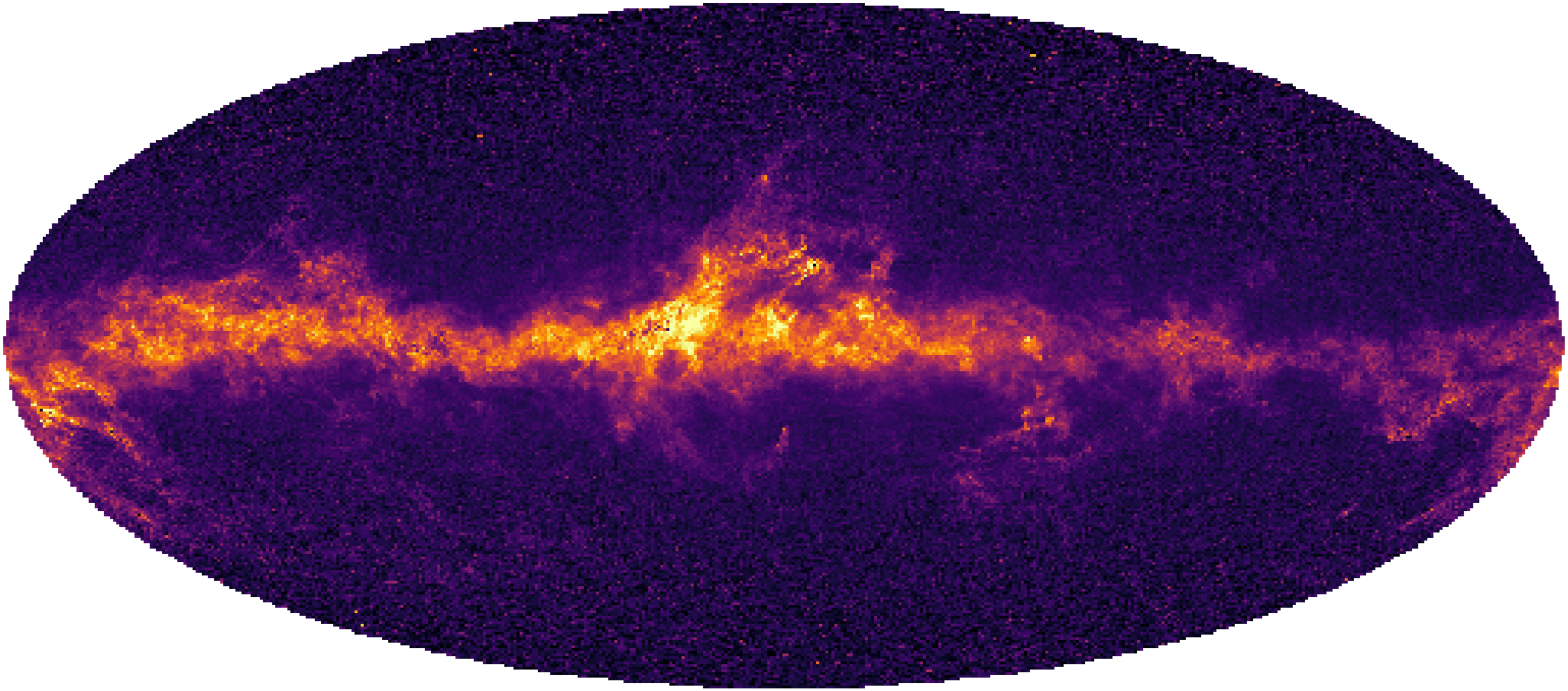
Planck Dust

Leike & Enßlin (2019)



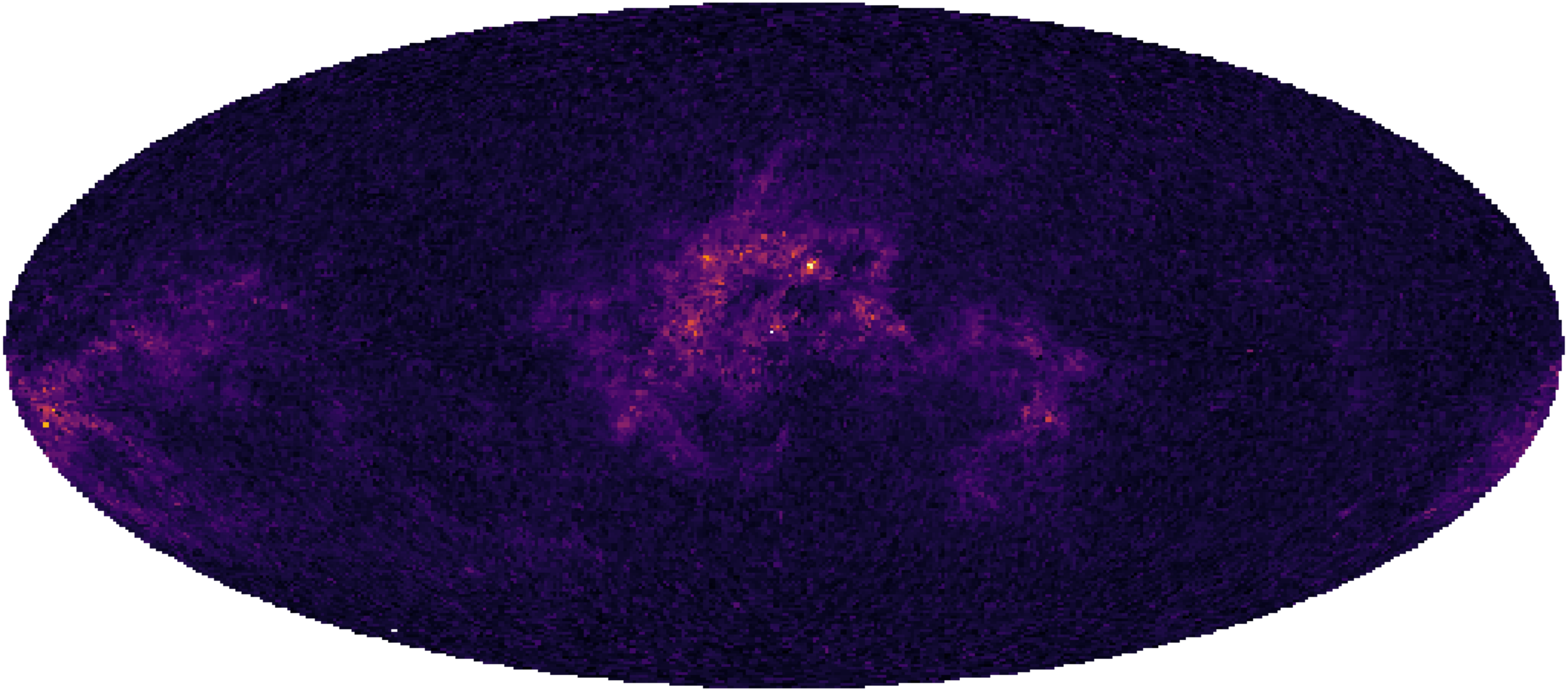
Gaia Dust

Leike & Enßlin (2019)



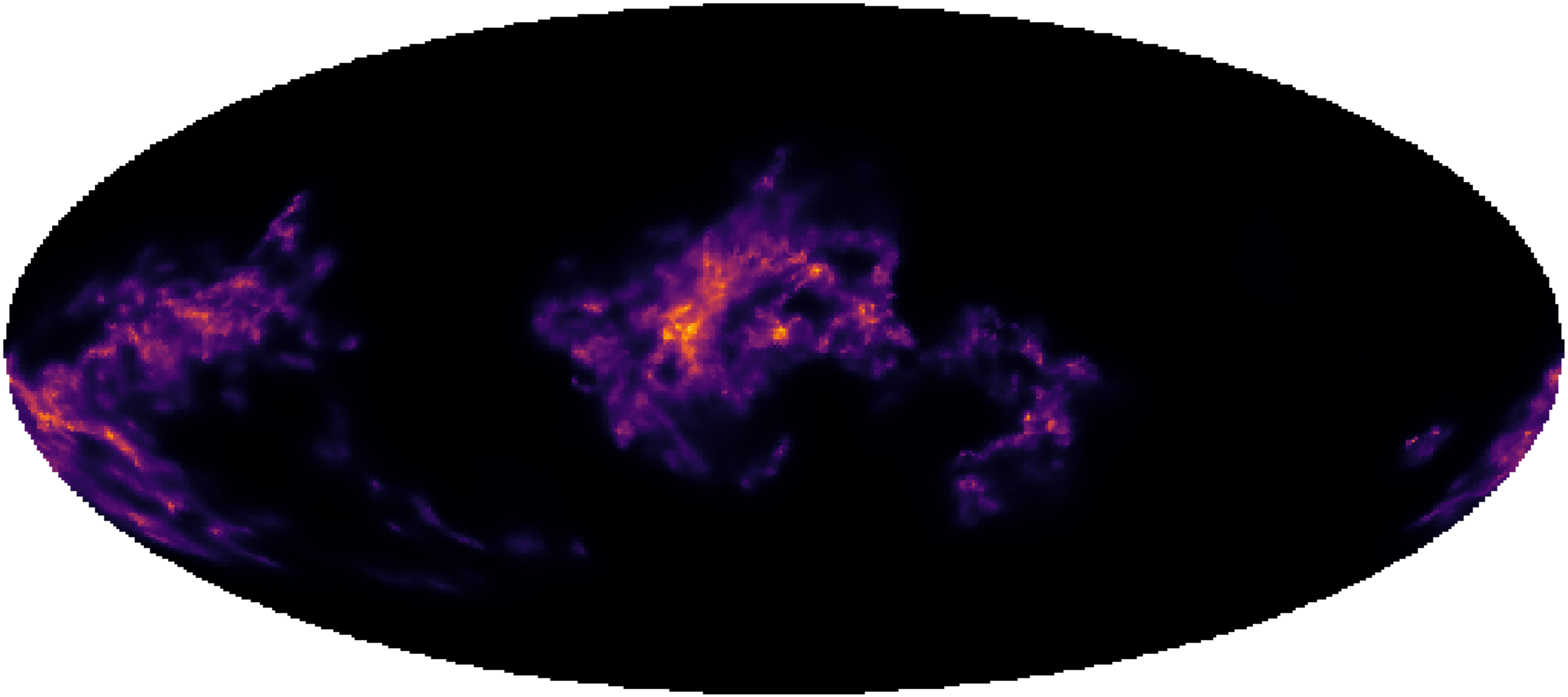
Gaia Dust

Leike & Enßlin (2019)



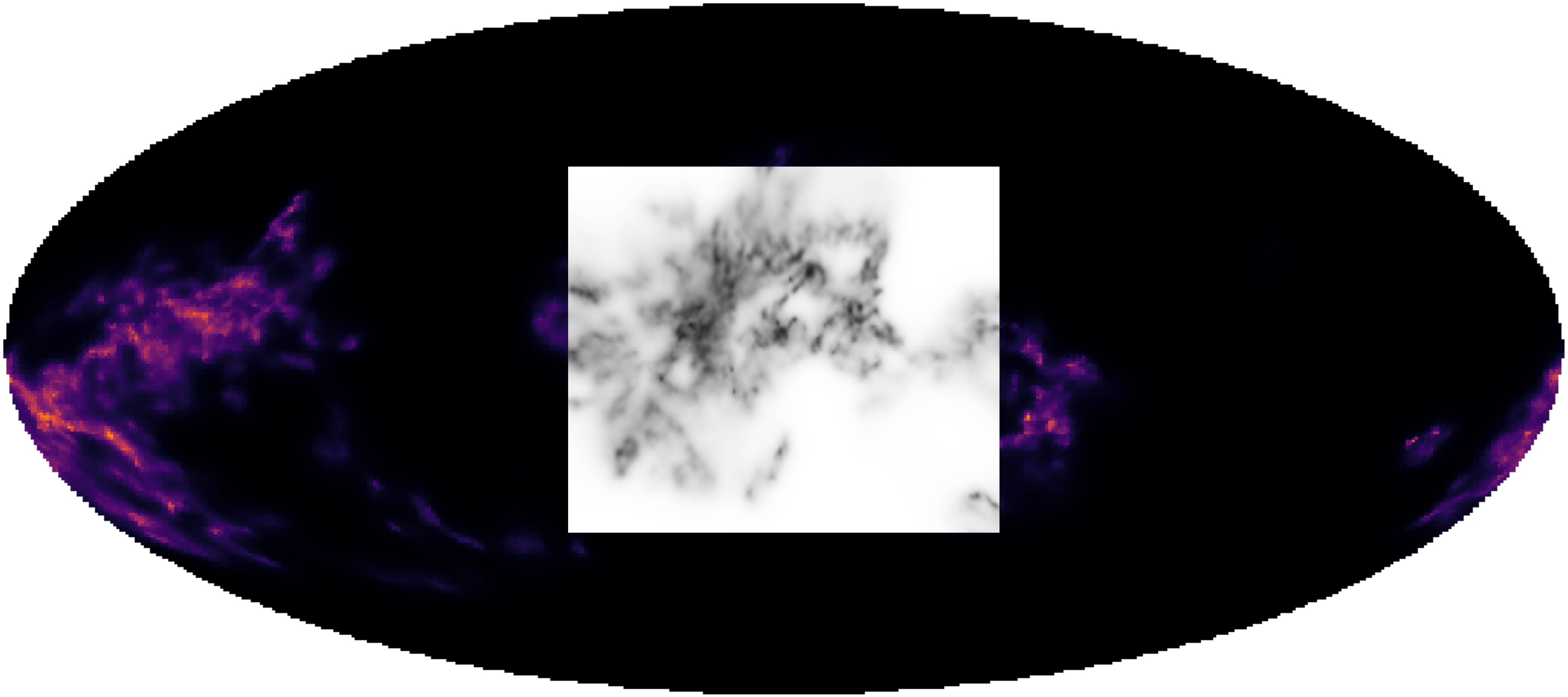
dust density

Reimar Leike et al. (in prep.)



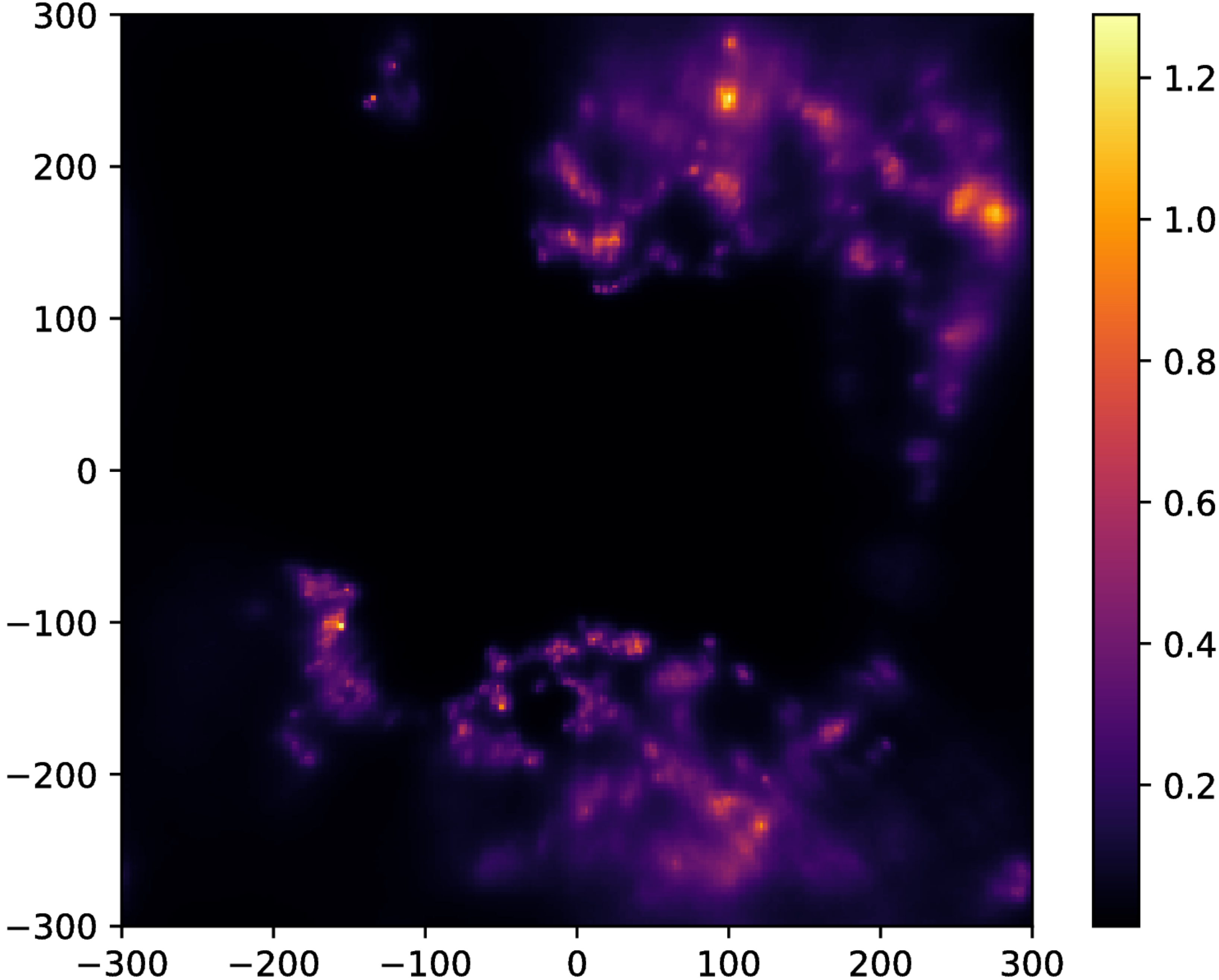
dust density

Reimar Leike et al. (in prep.)



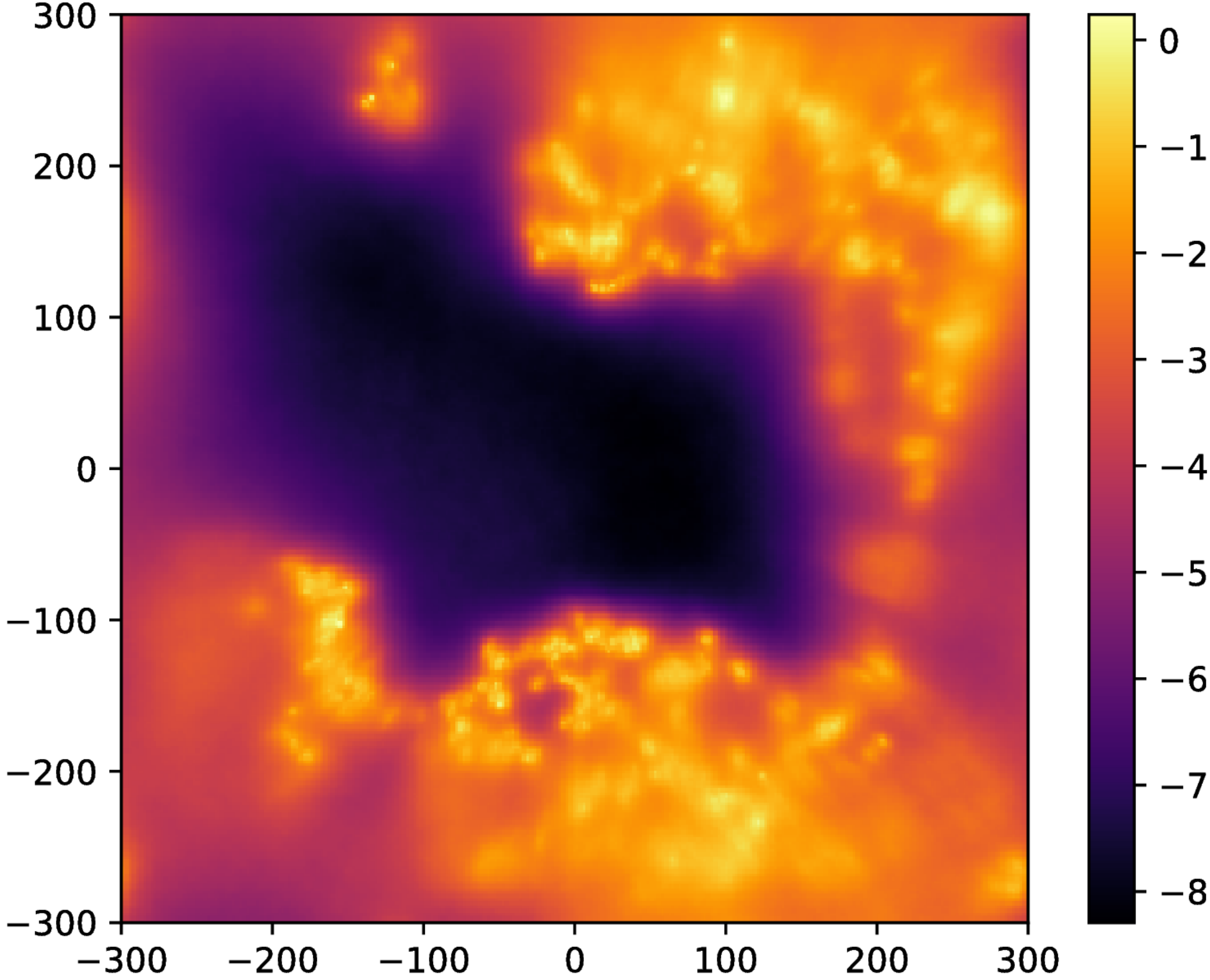
dust density

Leike & Enßlin (2019)



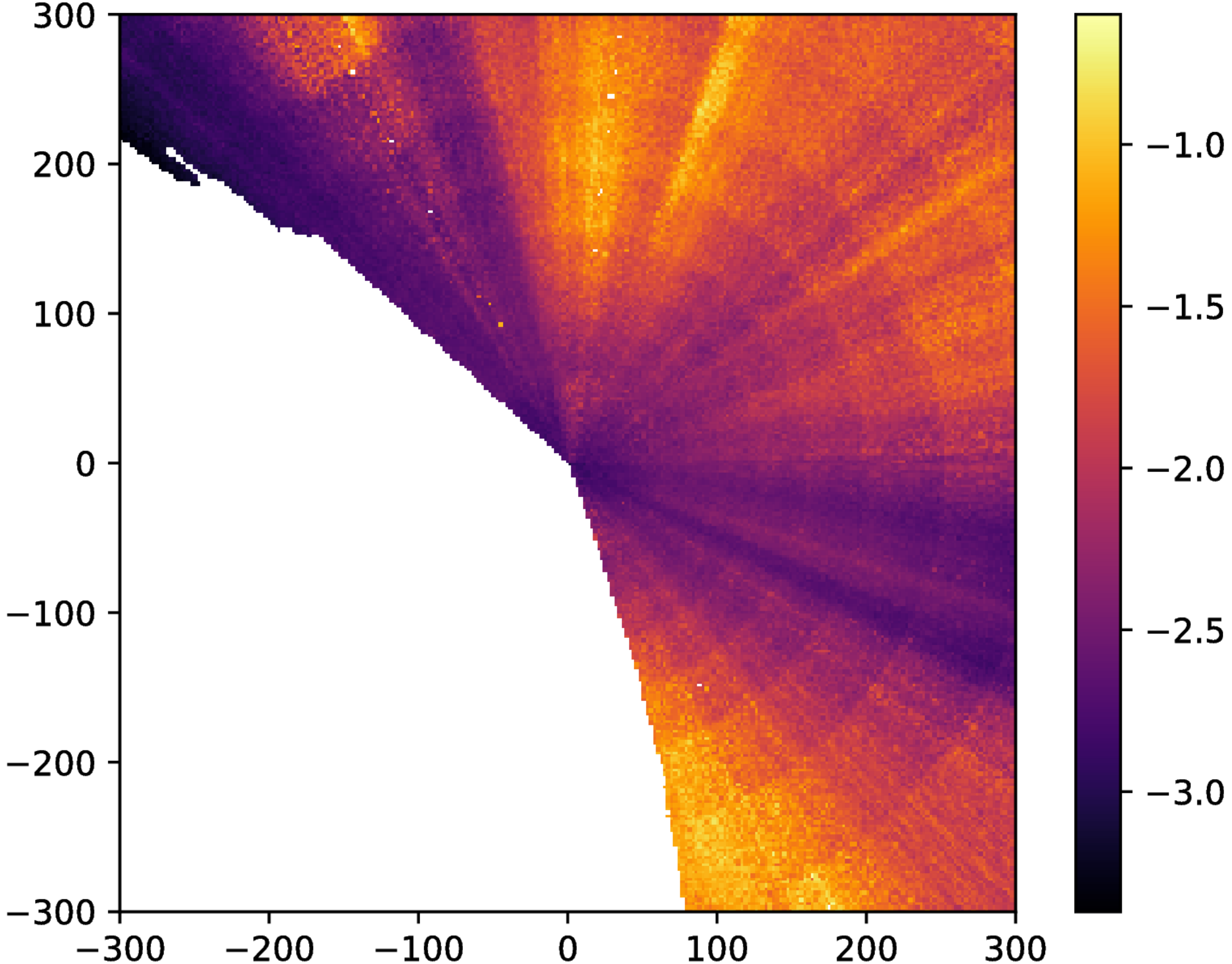
log dust density

Reimar Leike et al. (2019)



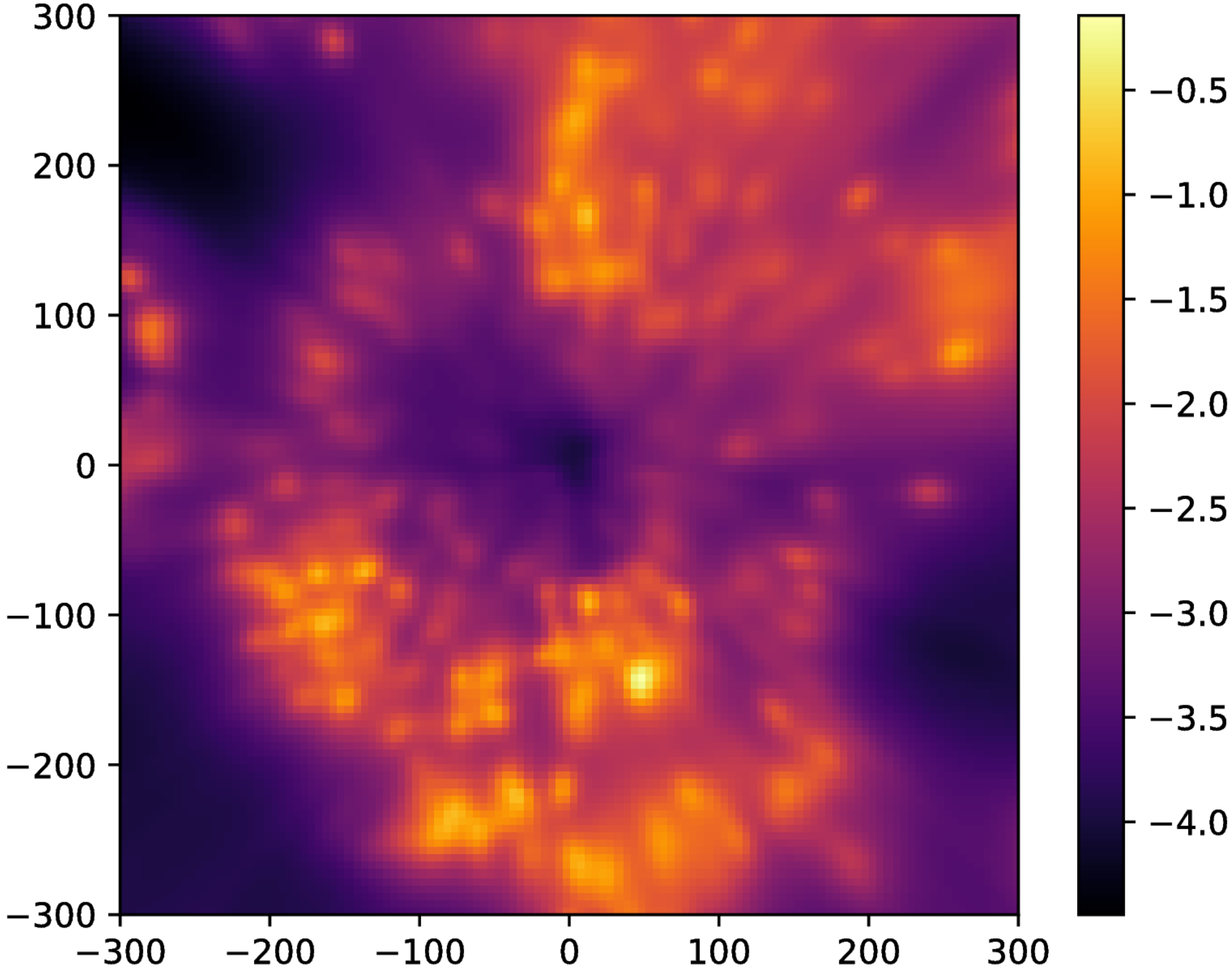
log dust density

Green et al. (2018.)



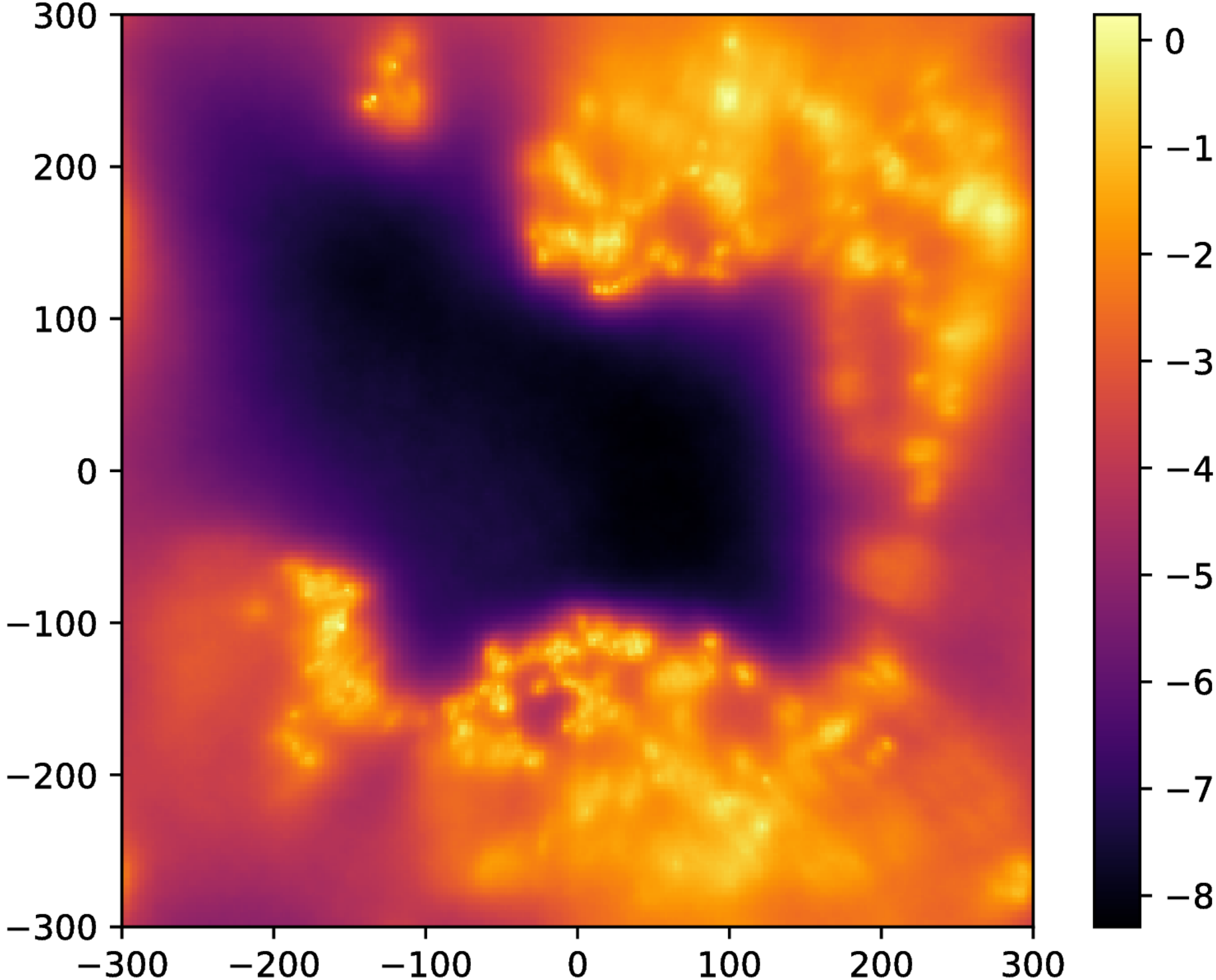
log dust density

Lallement et al. (2018)

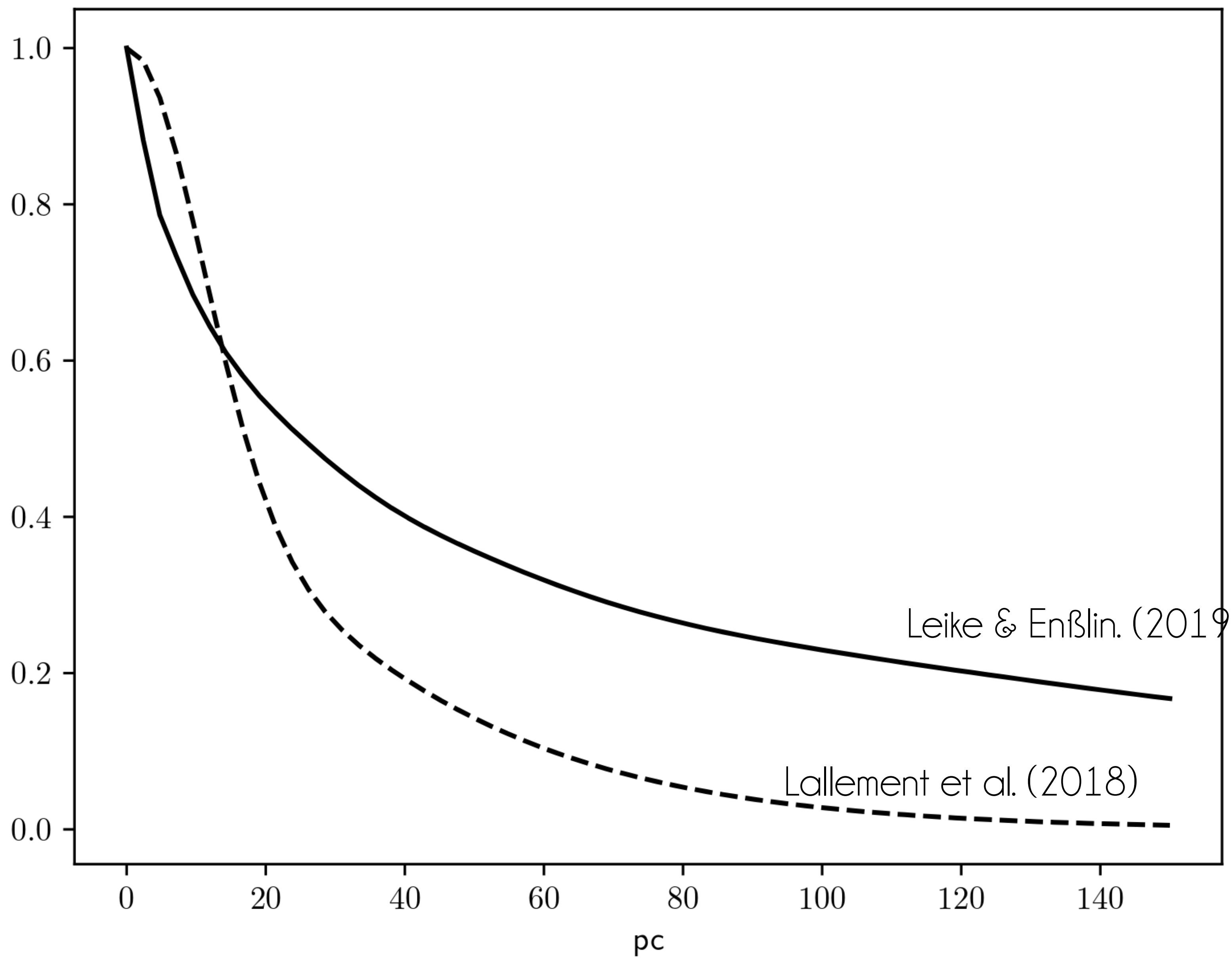


log dust density

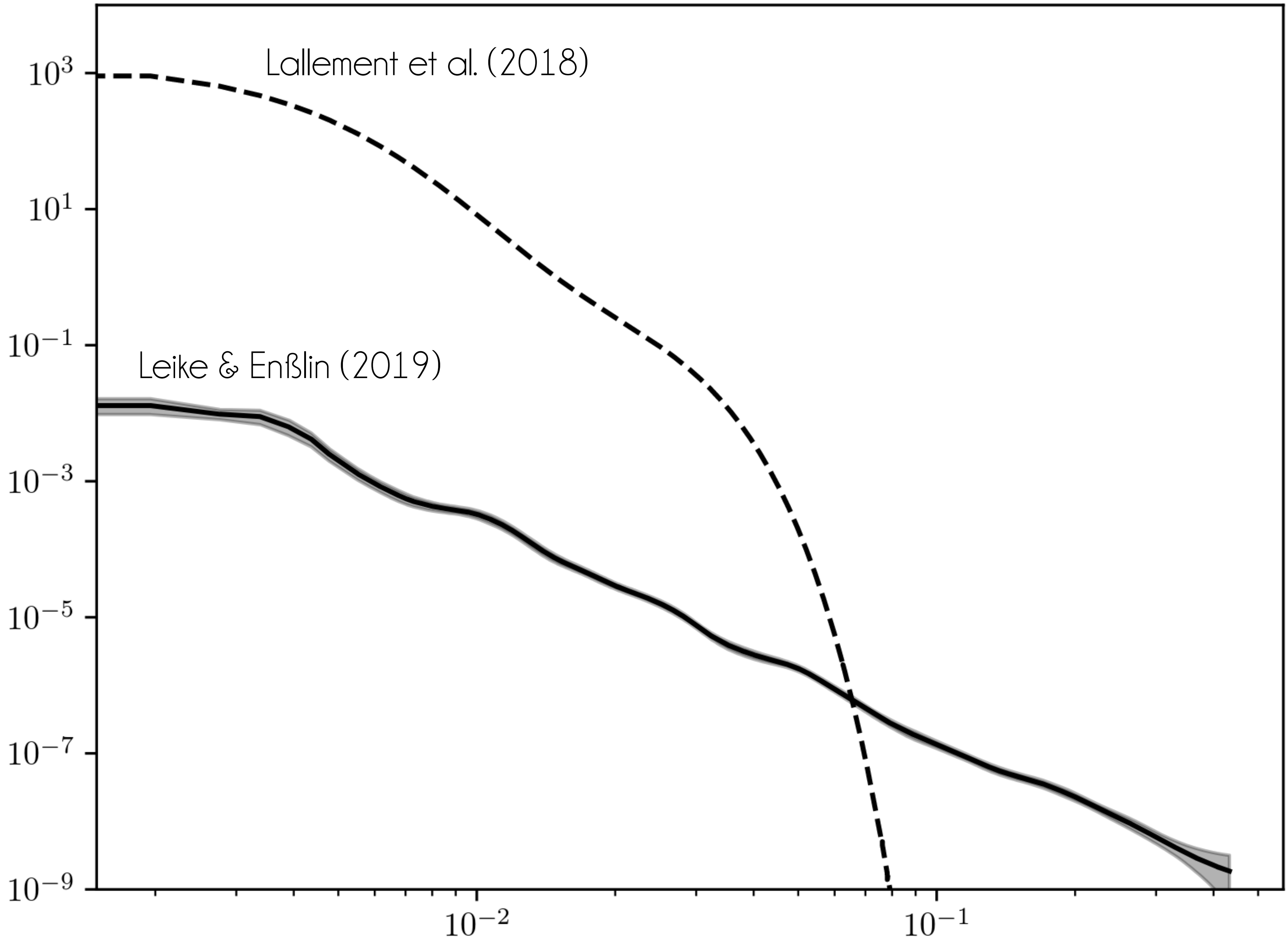
Reimar Leike et al. (2019)



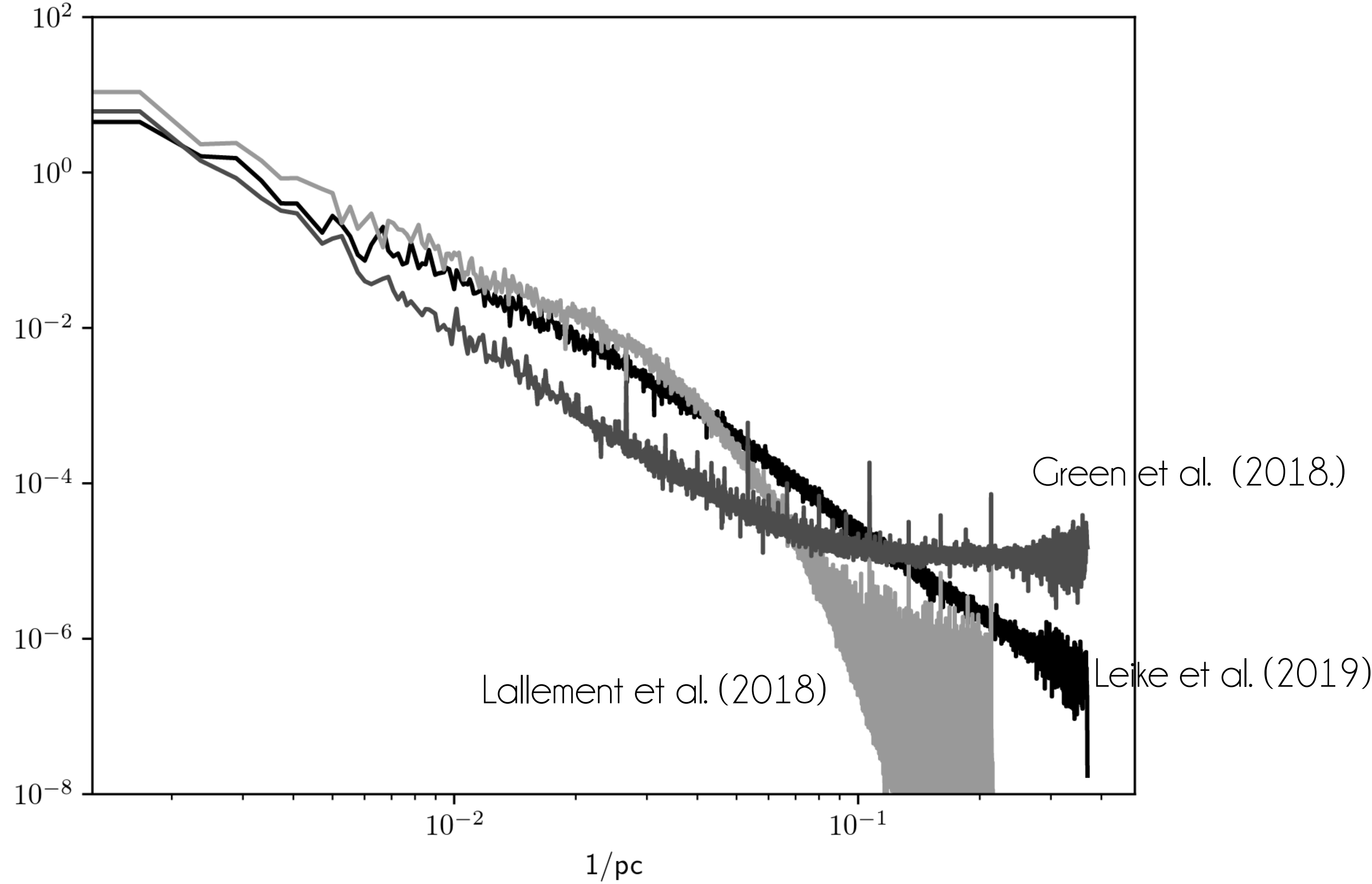
assumed 2-point correlation



assumed power spectra

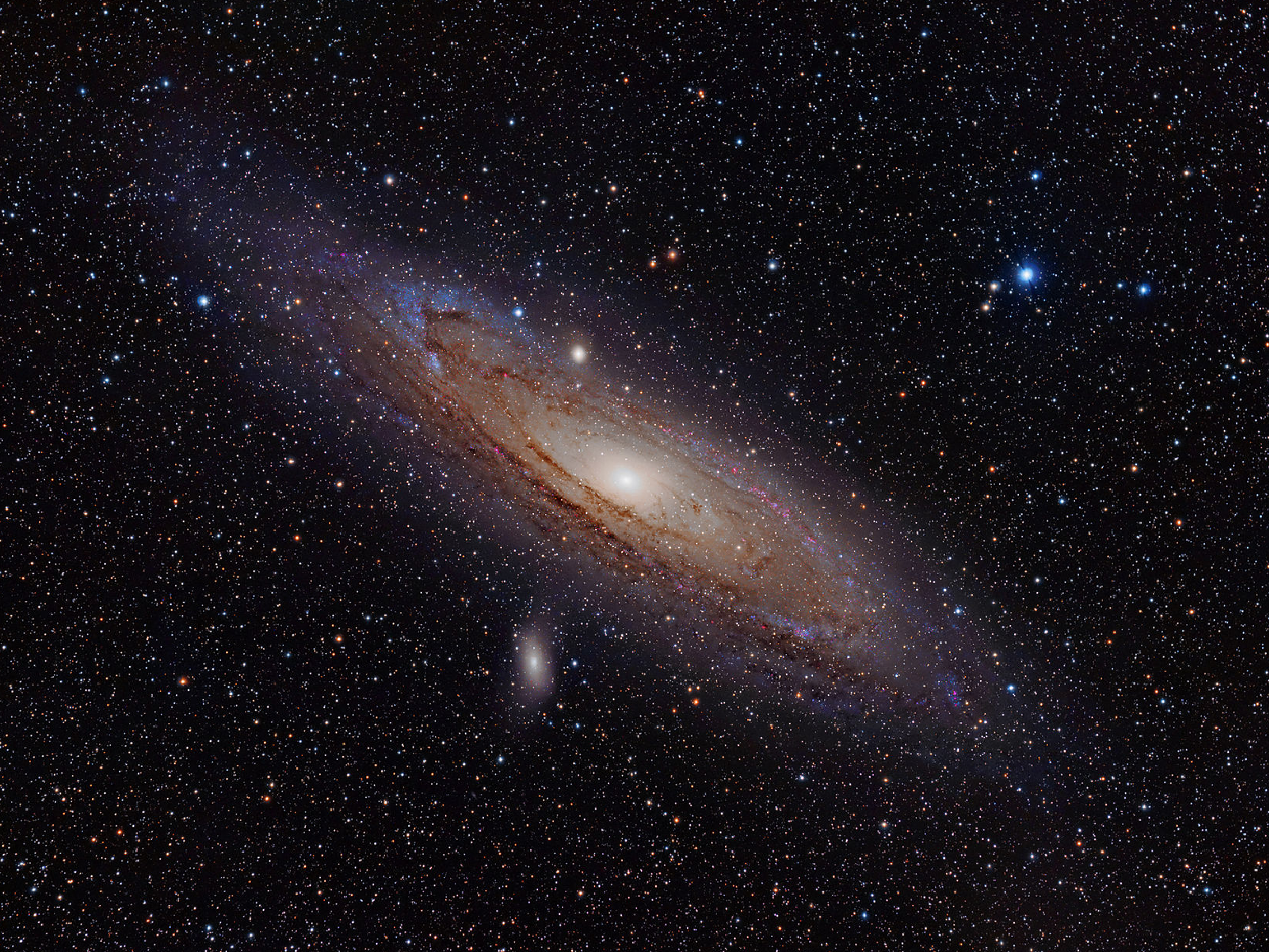


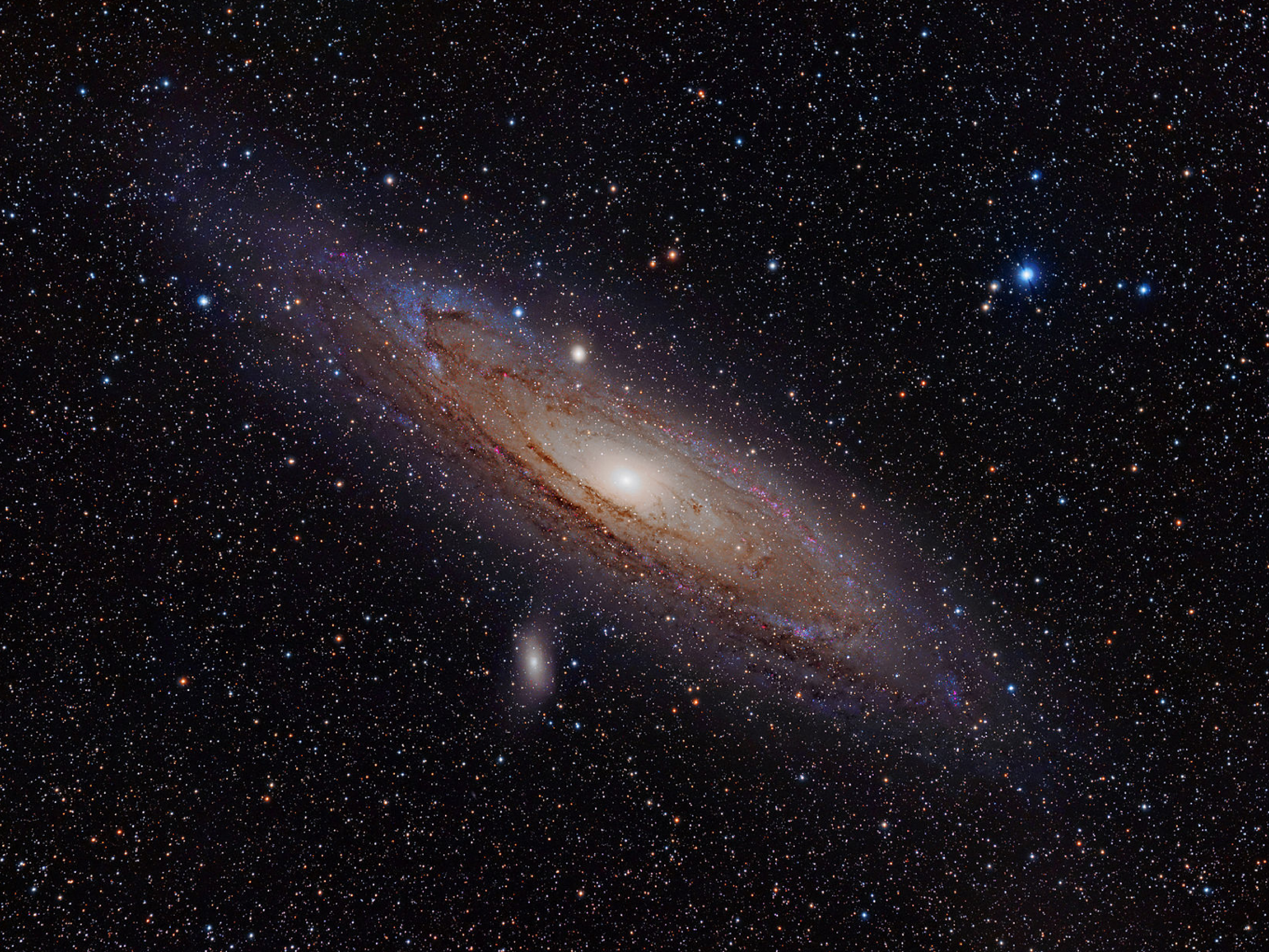
observed power spectra

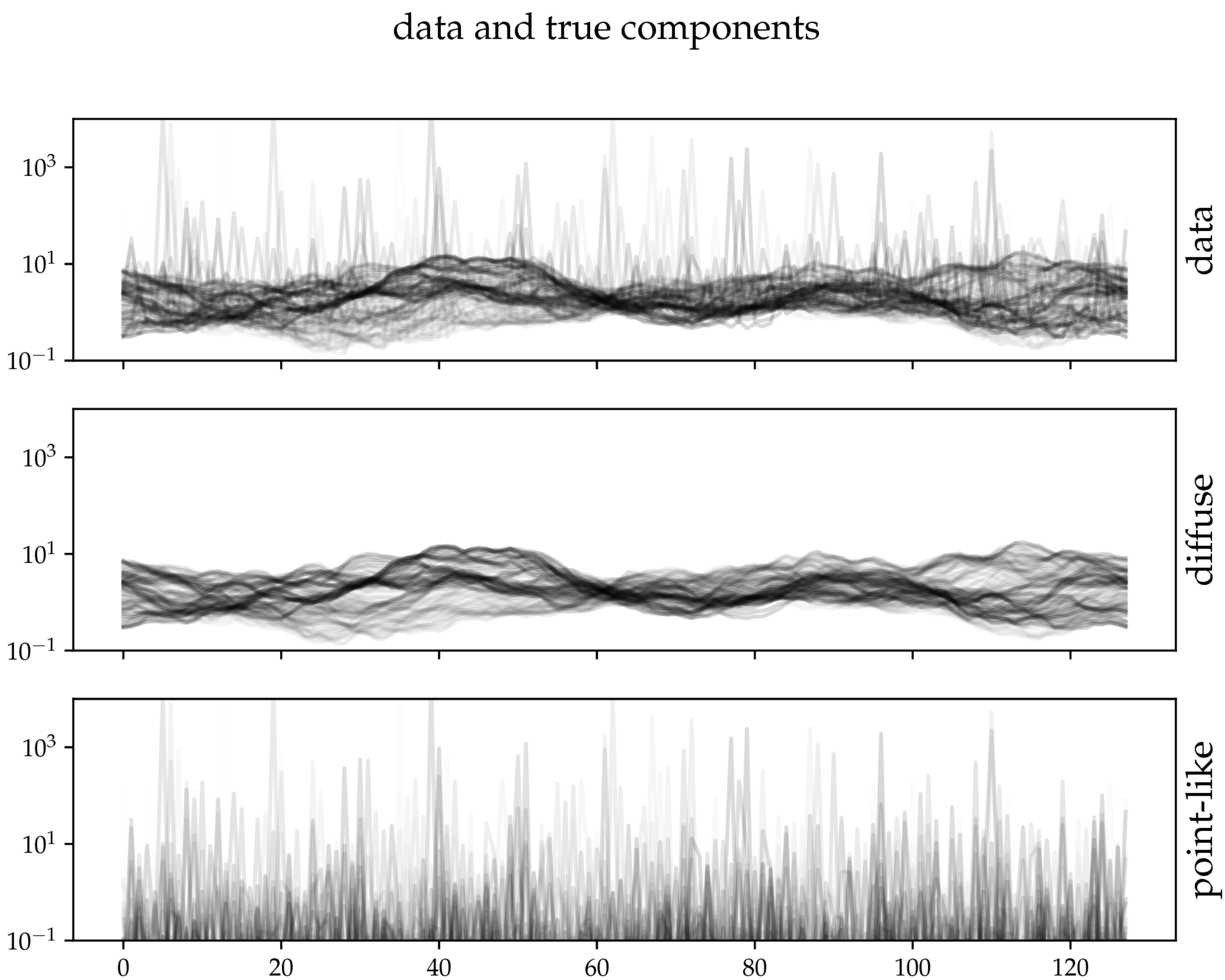


starblade

Jakob Knollmüller

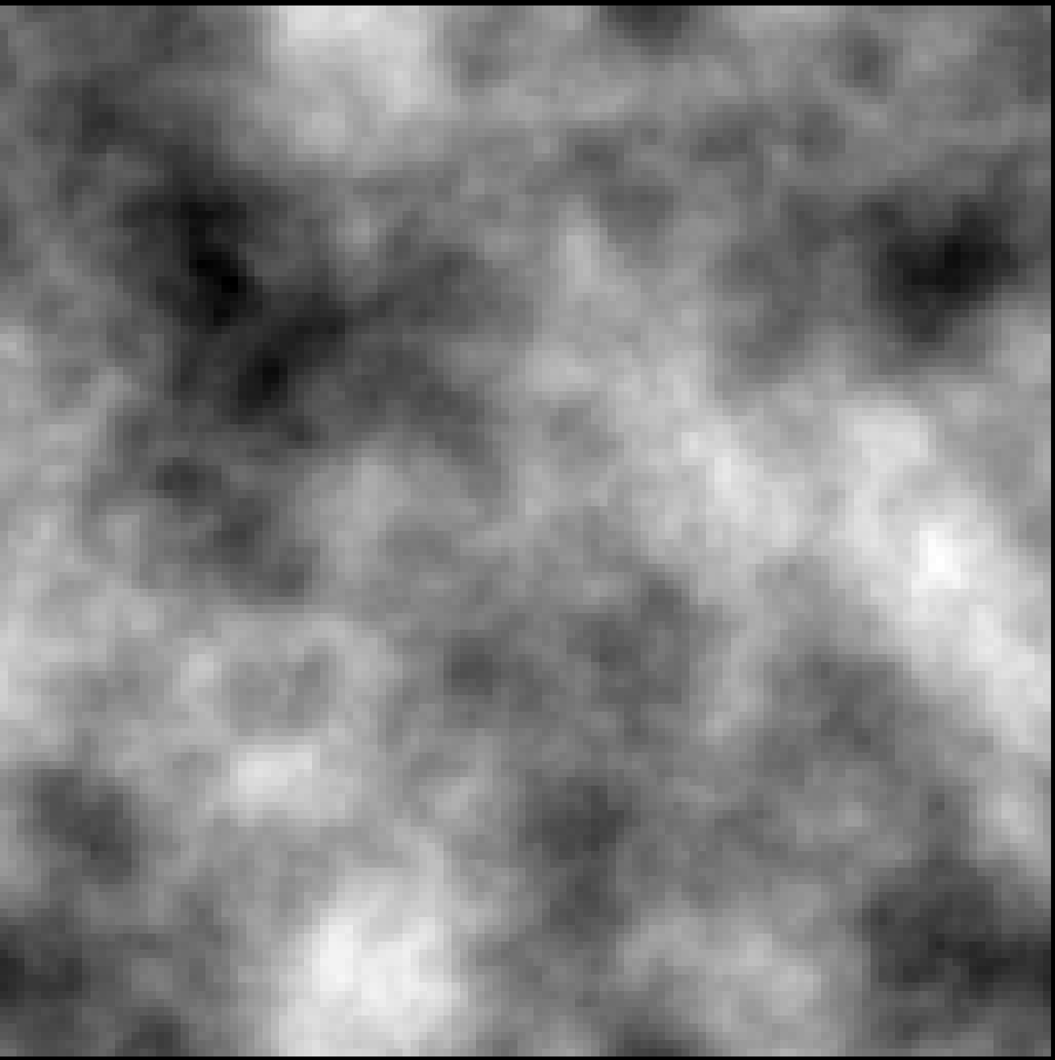




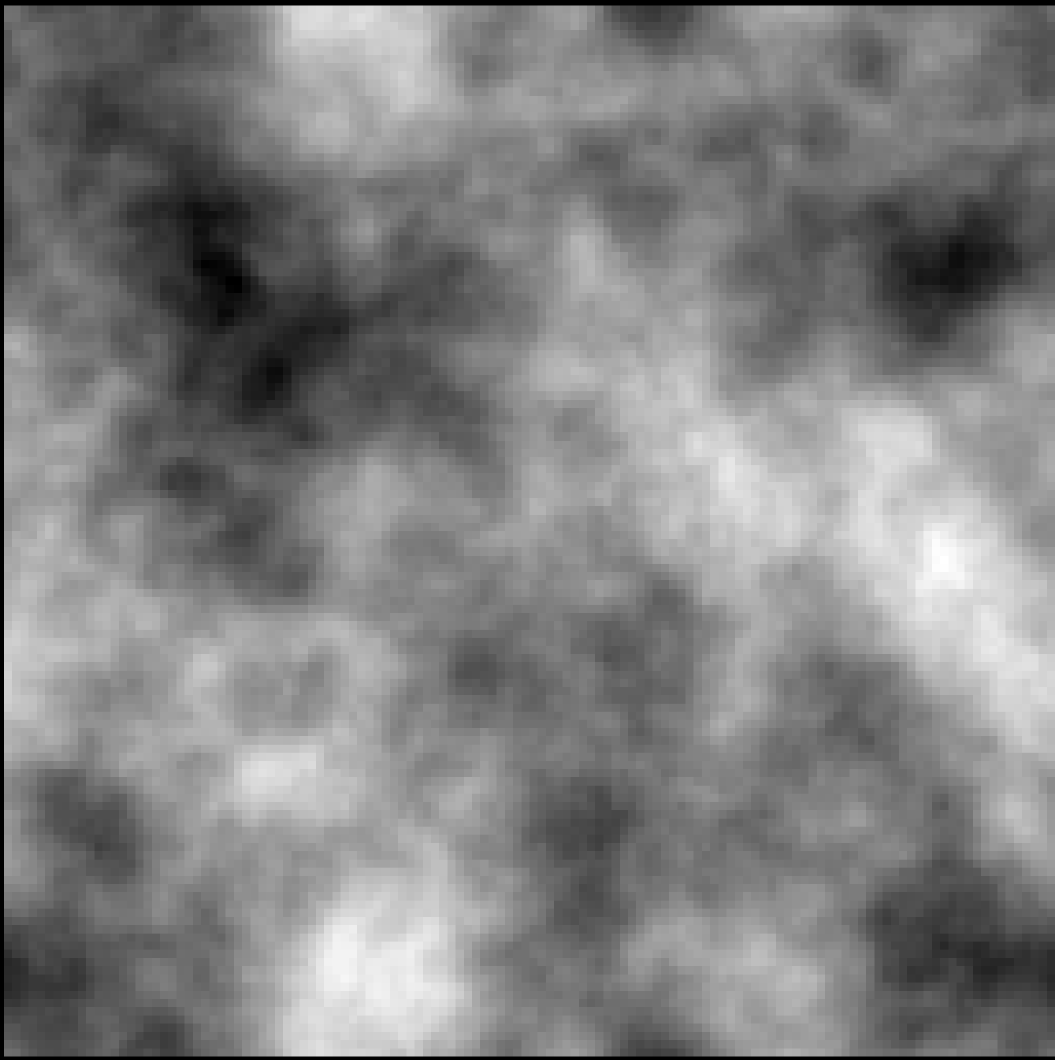


data and true components
data
diffuse
point-like
10^3
10^1
10^{-1}
0
20
40
60
80
100
120

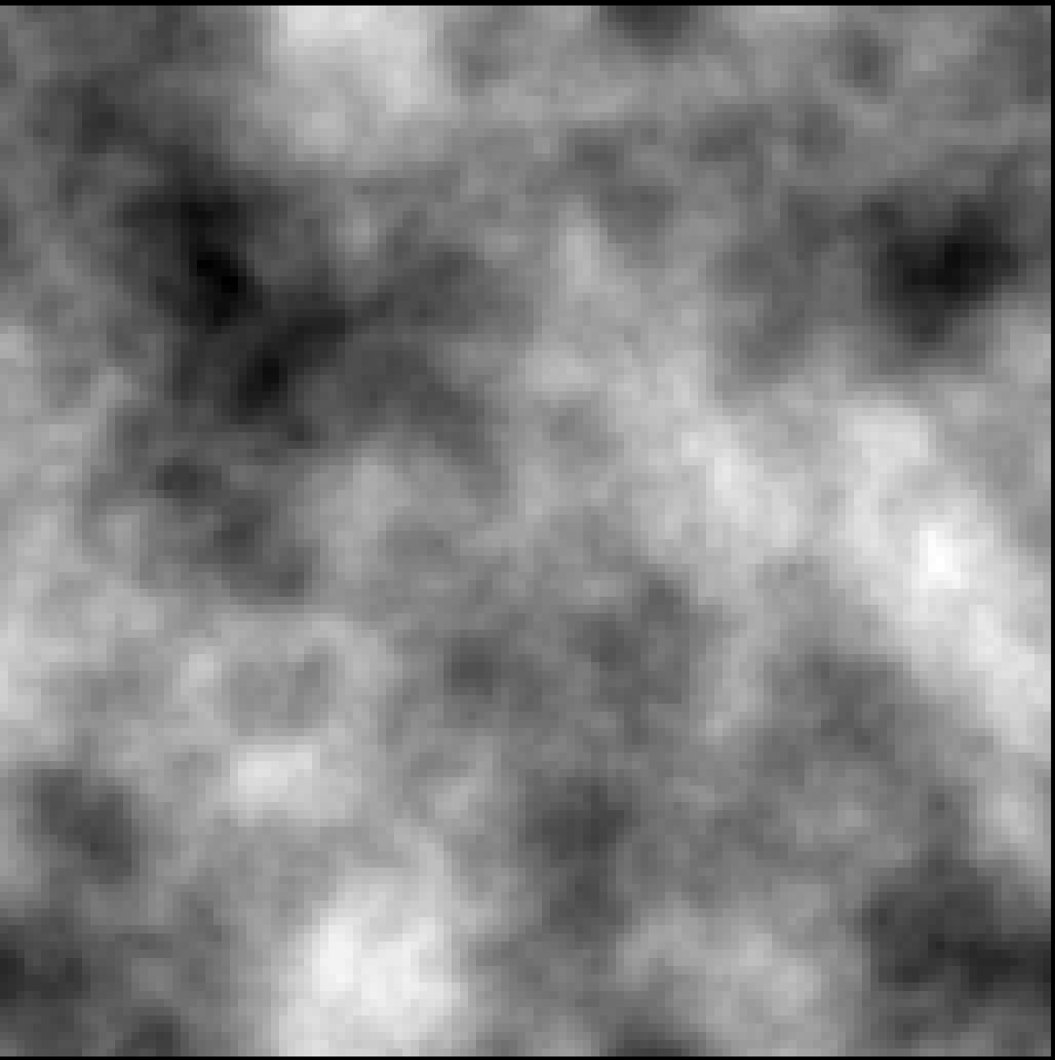
ground truth / starblade



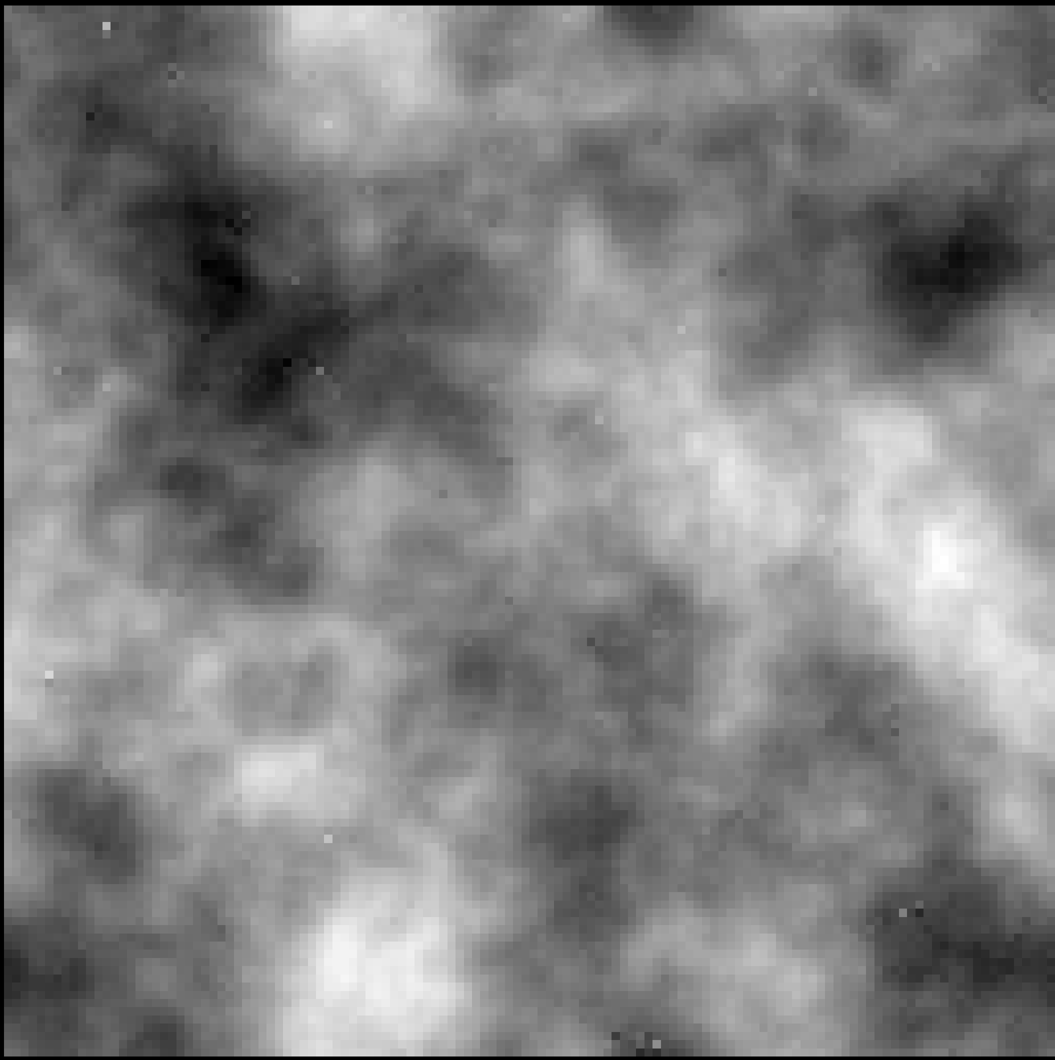
ground truth / autoencoder



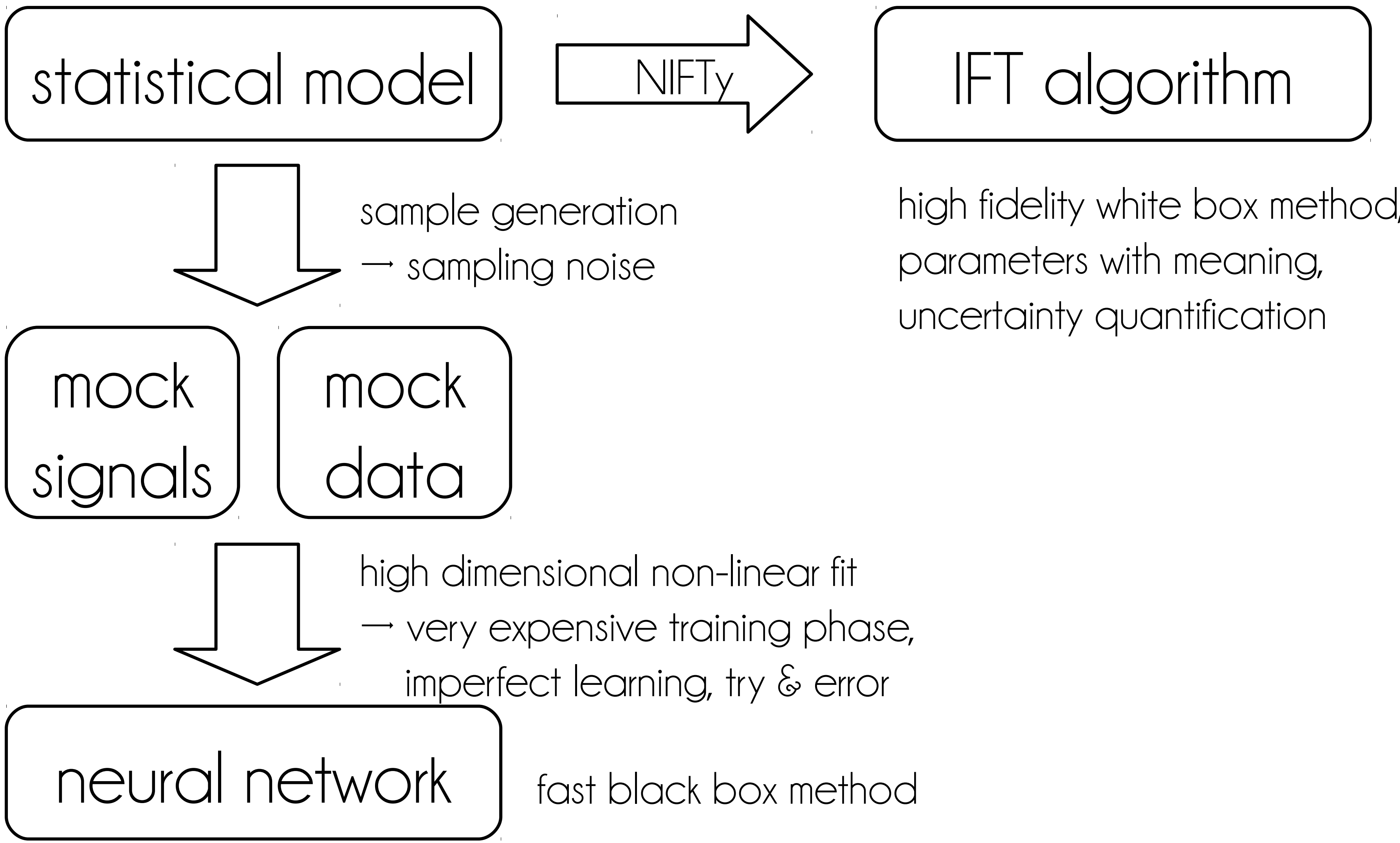
ground truth / starblade

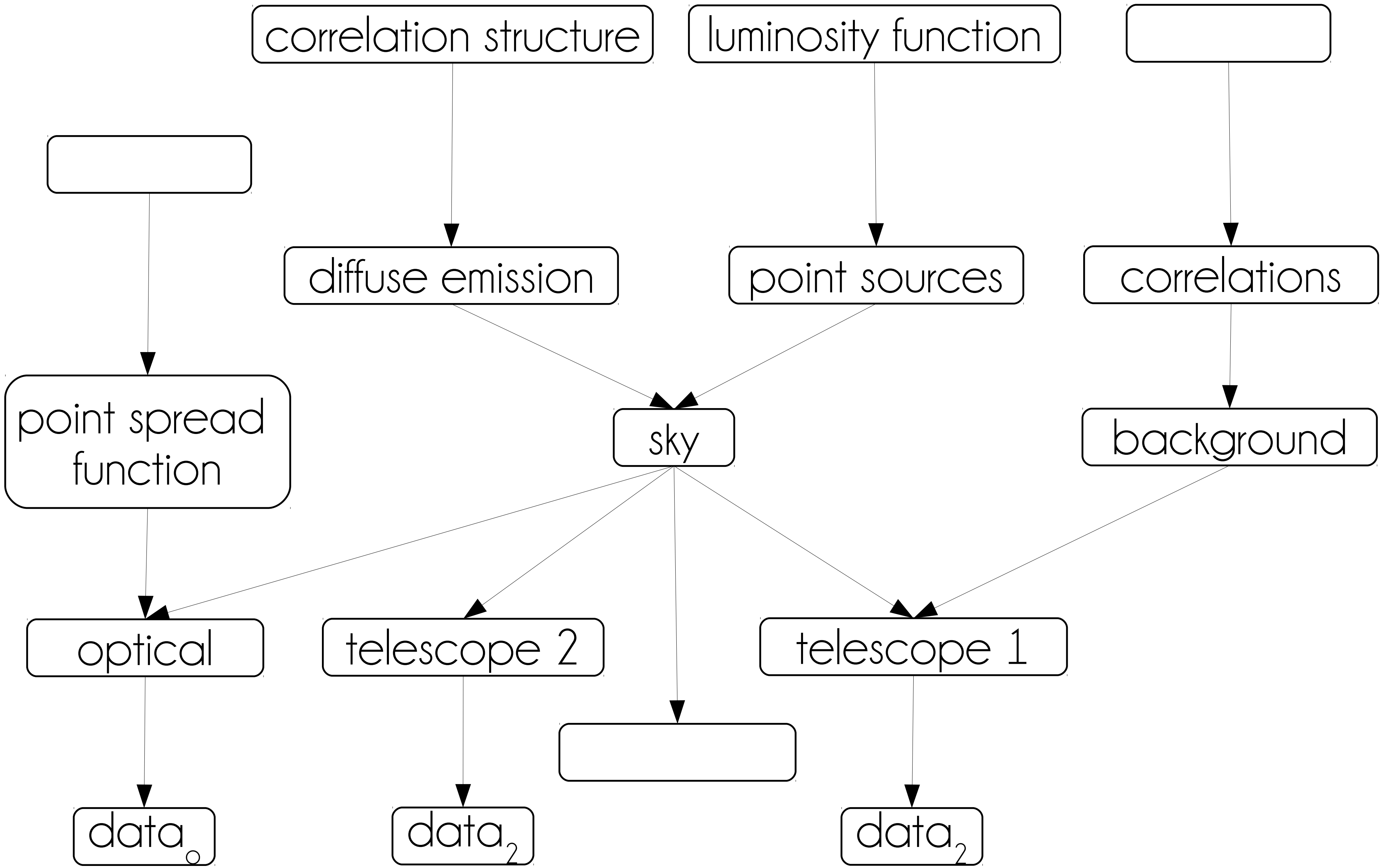


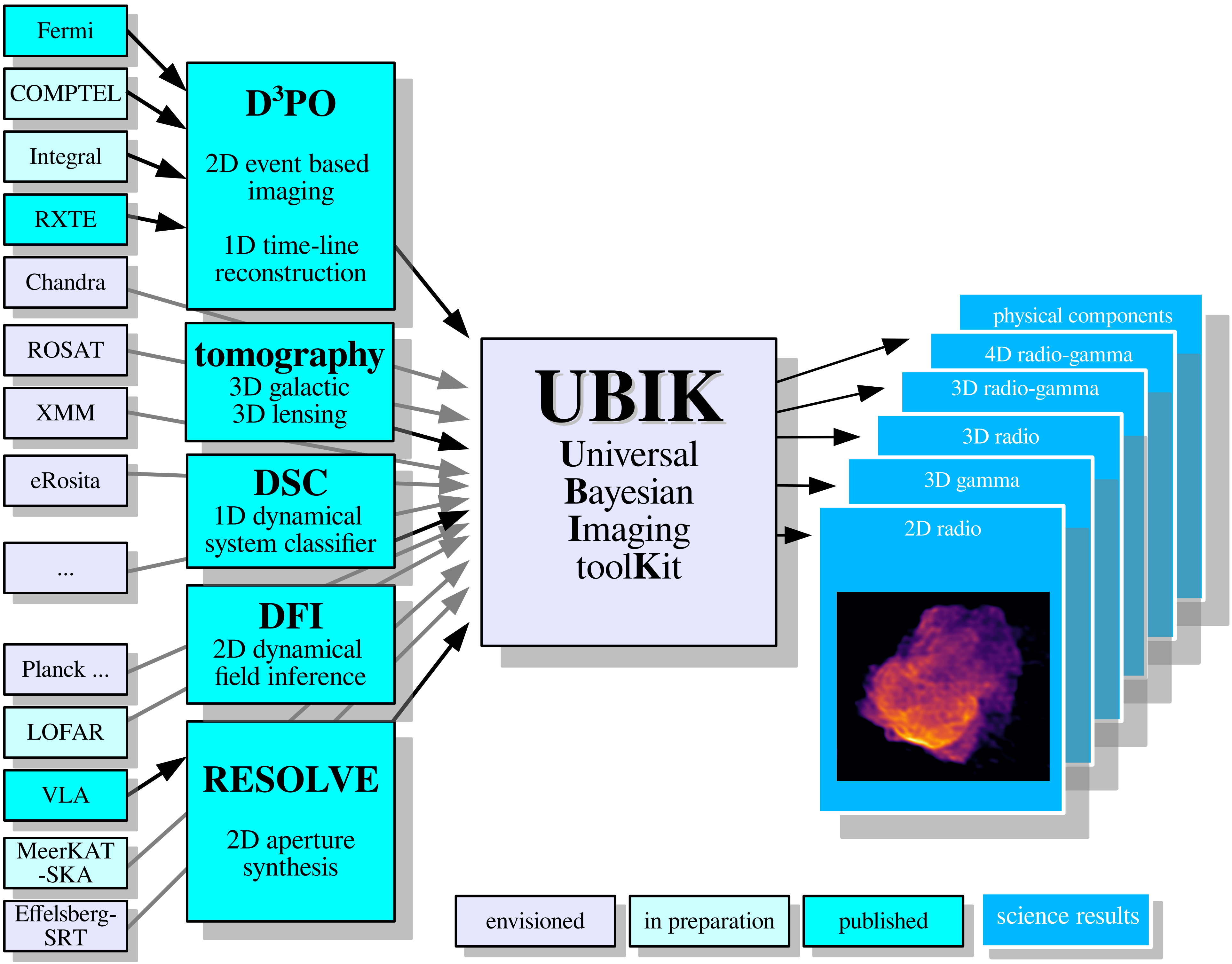
ground truth / autoencoder



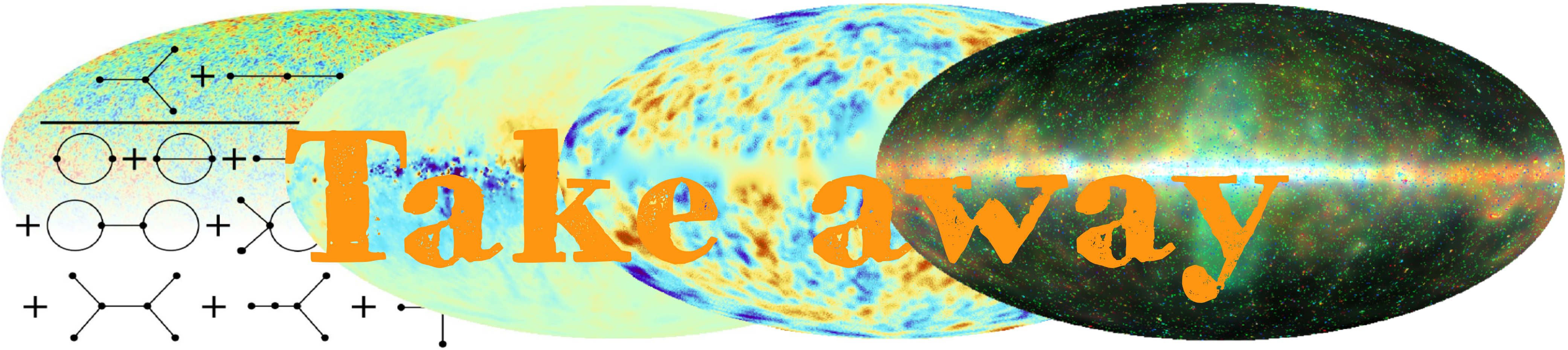
neural networks vs. information field theory




correlation structure
luminosity function
diffuse emission
point sources
correlations
point spread function
sky
background
optical
telescope 2
telescope 1
data₀
data₂
data₂


Fermi
COMPTEL
Integral
RXTE
Chandra
ROSAT
XMM
eRosita
...
Planck ...
LOFAR
VLA
MeerKAT -SKA
Effelsberg-SRT
D³PO
2D event based imaging
1D time-line reconstruction
tomography
3D galactic
3D lensing
DSC
1D dynamical system classifier
DFI
2D dynamical field inference
RESOLVE
2D aperture synthesis
UBIK
Universal Bayesian Imaging toolKit
physical components
4D radio-gamma
3D radio-gamma
3D radio
3D gamma
2D radio
envisioned
in preparation
published
science results

Take away



Imaging goes inference

IFT – **information field theory**

NIFTy – **numerical IFT in Python**
Feb 2019: **autodiff/variational inference**

UBIK – **Universal Bayesian Imaging toolKit** – under development