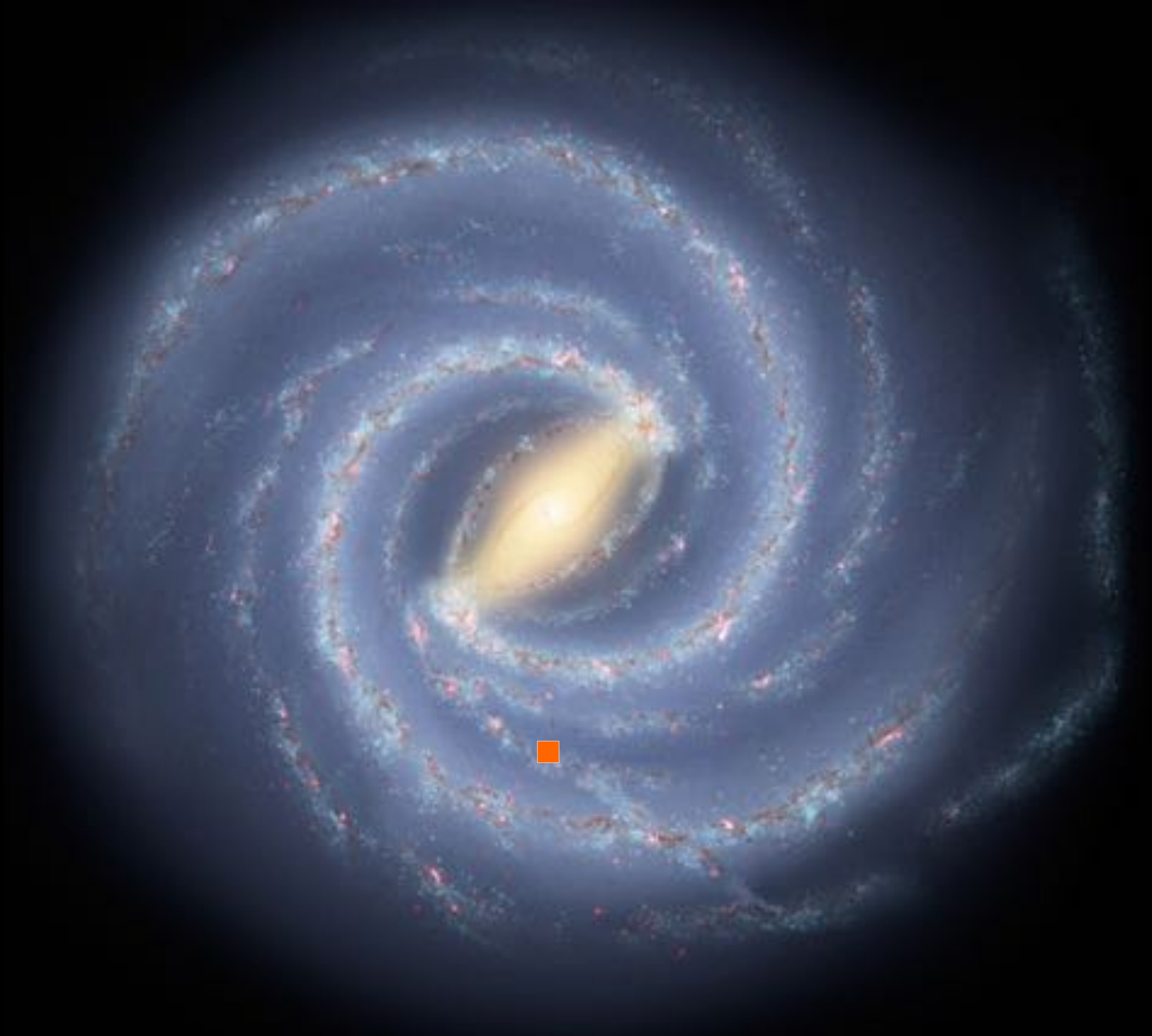
Numerical Information Field Theory
a NIFTy tutorial

Philipp Arras, Torsten Enßlin, Philipp Frank,
Sebastian Hutschenreuter, Reimar Leike
MPI for Astrophysics

IFT Team: Philipp Arras, Michael Bell, Vanessa Böhm, Sebastian Dorn, Martin Dupont, Mona Frommert, Philipp Frank, Mahsa Ghaempanah, Maksim Greiner, Philipp Haim, Sebastian Hutschenreuter, Henrik Junklewitz, Francisco-Shu Kitaura, Jakob Knollmüller, Christoph Lienhard, Reimar Leike, Ancla Müller, Johannes Oberpriller, Niels Oppermann, Natalia Porquerese, Daniel Pumpe, Tiago Ramalho, Martin Reinecke, Julia Stadler, Marco Selig, Theo Steininger, Valentina Vacca, Cornelius Weig, Margret Westerkamp, & many more



Galactic Tomography

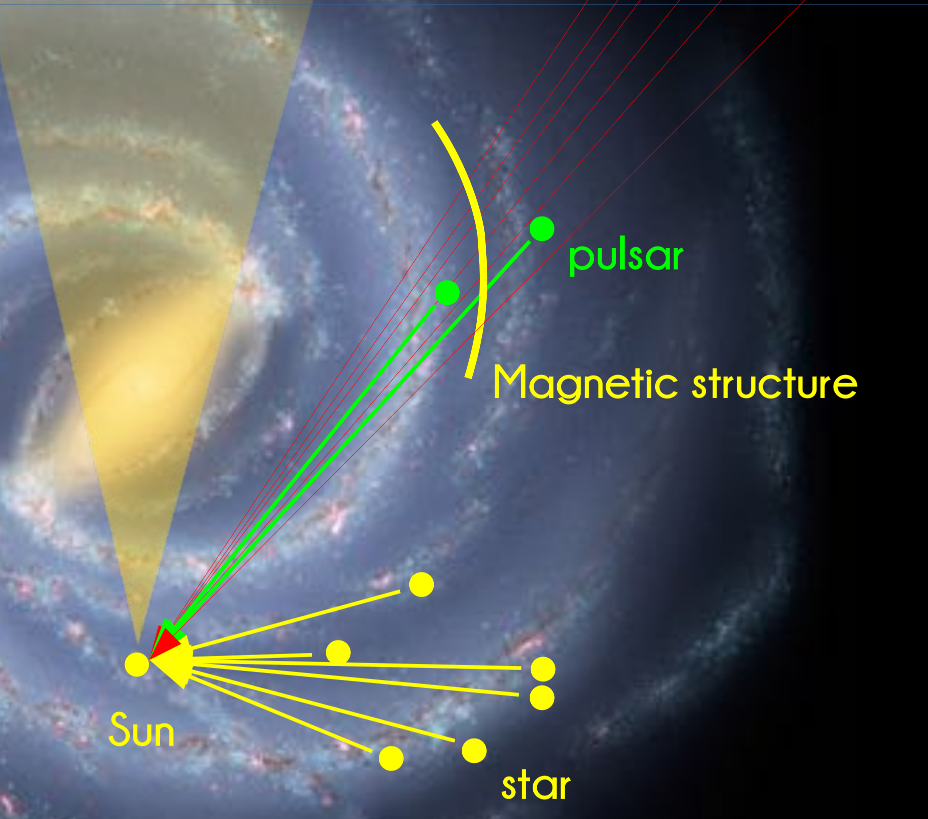
Pulsars:
Dispersion Measure → electron density
Rotation Measure → magnetic field x el. dens.
Scintillation Measure → el. dens. x turbulence

Extragalactic sources:
Rotation Measure → magnetic field x el. dens.
Ultra High Energy Cosmic Rays → mag. fields

Stars:
Dust reddening → dust density & properties
Positions → stellar density & radiation field
Kinematics → gravitational potential

Emission Processes:
Dust emission → dust density & radiation field
Synchrotron → relativistic el. x mag. Fields
Bremsstrahlung → thermal, rel. el. x gas density
Inverse Compton → rel. el. x radiation field
Hadronic interactions → rel. nuclei x gas density
Lines (21 cm, CO, ...) → gas density & kinematics

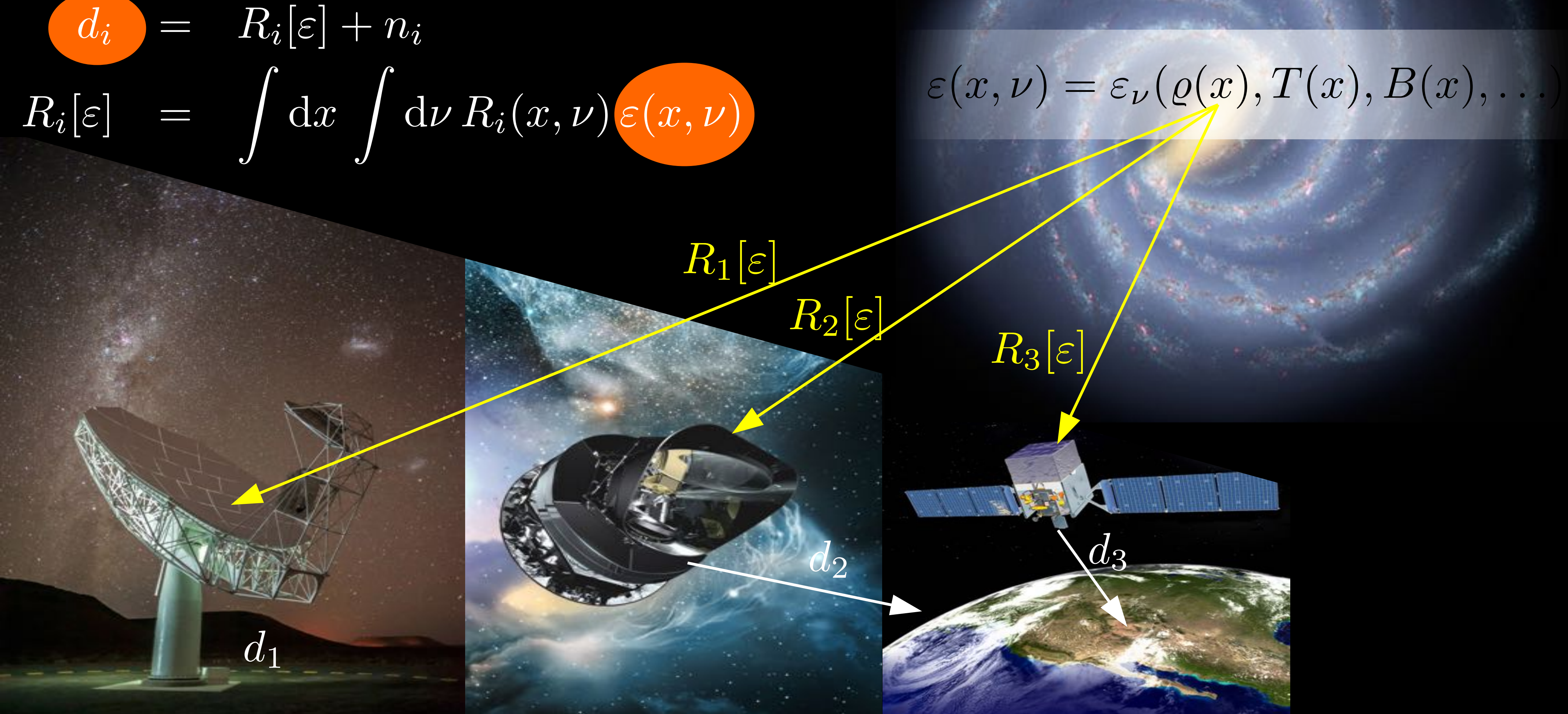
Other information sources:
Correlation structures (auto- & cross-correlations)
Approximate symmetries
Physical laws
Empirical laws, ...



Data Fusion

$$d_i = R_i[\varepsilon] + n_i$$

$$R_i[\varepsilon] = \int \mathrm{d}x \int \mathrm{d}\nu \, R_i(x,\nu)\, \varepsilon(x,\nu)$$



Information theory

$$
\begin{aligned}
\mathcal{P}(s|d) &= \frac{\mathcal{P}(d,s)}{\mathcal{P}(d)} = \frac{e^{-\mathcal{H}(d,s)}}{\mathcal{Z}(d)} \\
\mathcal{H}(d,s) &= -\log \mathcal{P}(d,s) \\
\mathcal{Z}(d) &= \mathcal{P}(d) \\
&= \int \mathcal{D}s\, \mathcal{P}(d,s) \\
\mathcal{P}(d,s) &= \mathcal{P}(d|s)\, \mathcal{P}(s) \\
\mathcal{H}(d_1, d_2, s) &= \mathcal{H}(d_1|s) + \mathcal{H}(d_2|s) + \mathcal{H}(s)
\end{aligned}
$$

Information

is additive

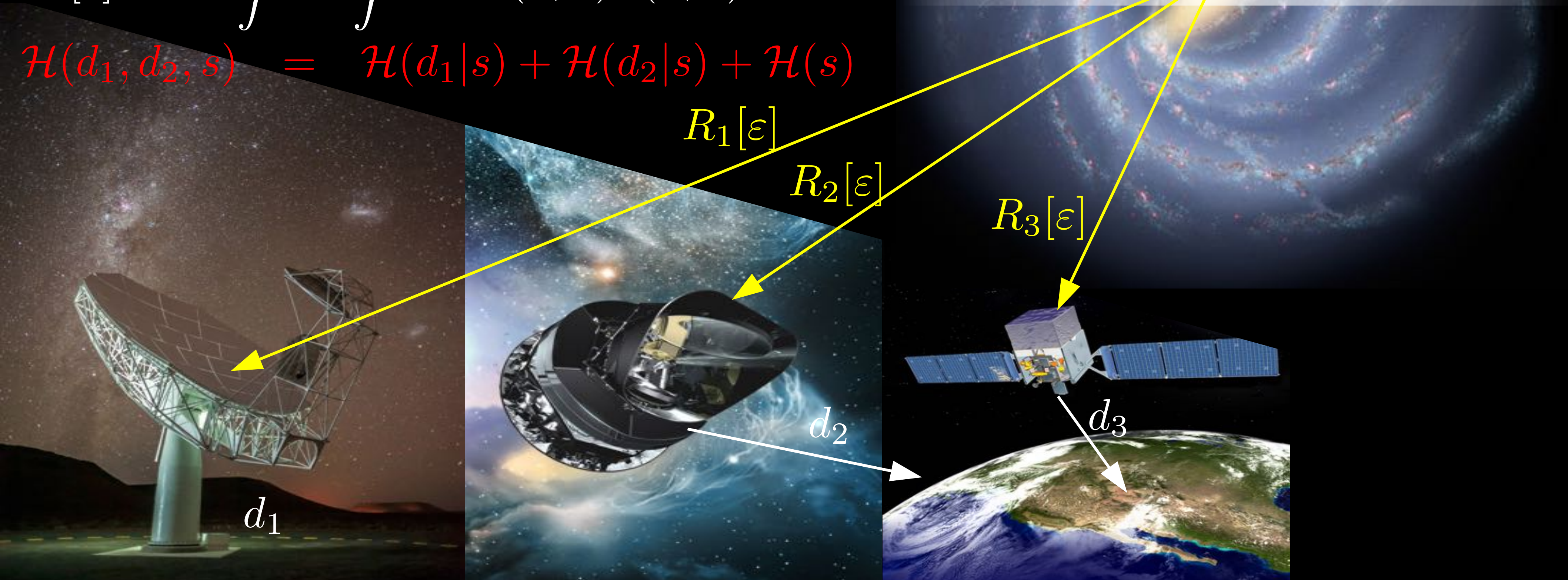
Data Fusion

$$d_i = R_i[\varepsilon] + n_i$$

$$R_i[\varepsilon] = \int \mathrm{d}x \int \mathrm{d}\nu \, R_i(x,\nu)\,\varepsilon(x,\nu)$$

$$\mathcal{H}(d_1, d_2, s) = \mathcal{H}(d_1|s) + \mathcal{H}(d_2|s) + \mathcal{H}(s)$$

$$\varepsilon(x,\nu) = \varepsilon_\nu(\varrho(x), T(x), B(x), \ldots)$$



Probability & Information

$$\mathcal{P}(s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{s^2}{2\sigma^2}}$$

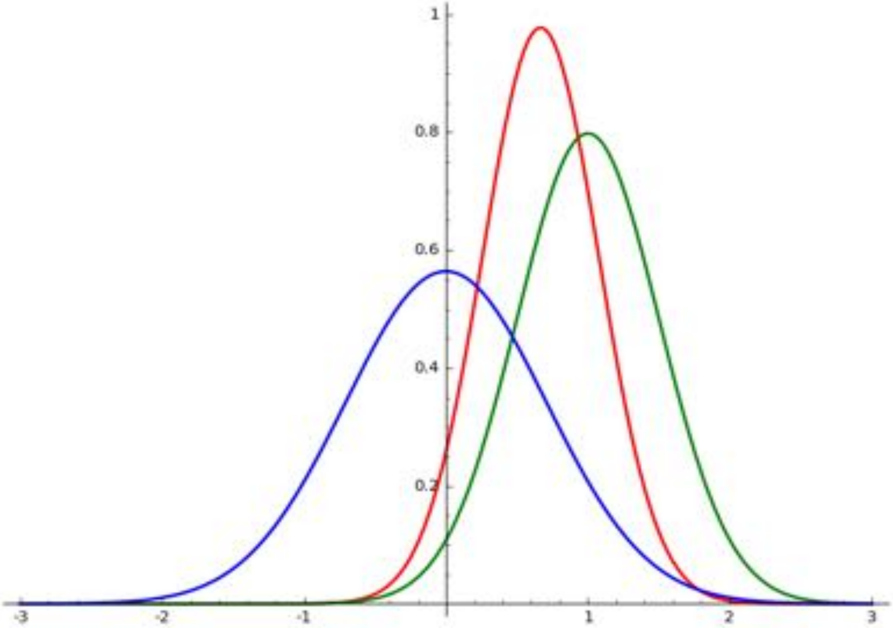
$$\mathcal{P}(d|s) \propto e^{-\frac{(s-d)^2}{2\sigma'^2}}$$

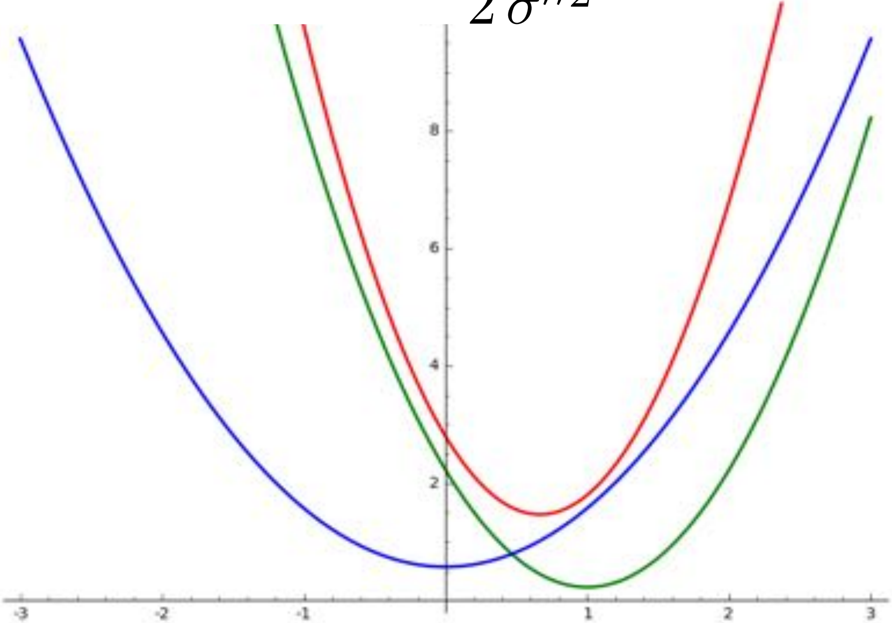
$$\mathcal{P}(s|d) \propto e^{-\frac{(s-m)^2}{2\sigma''^2}}$$

$$\mathcal{H}(s) \widehat{=} \frac{s^2}{2\,\sigma^2}$$

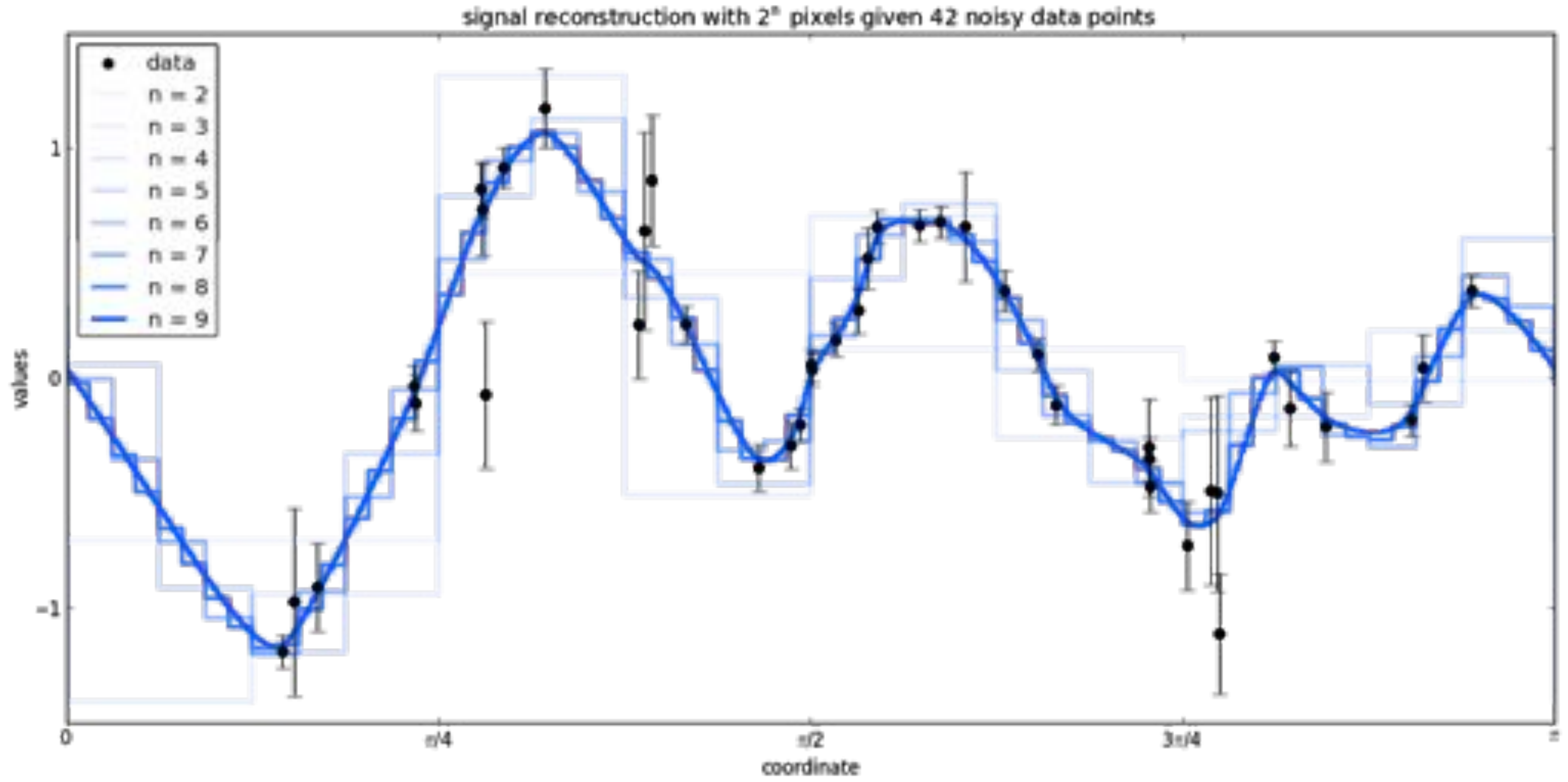
$$\mathcal{H}(d|s) \widehat{=} \frac{(s-d)^2}{2\,\sigma'^2}$$

$$\mathcal{H}(d,s) \widehat{=} \frac{(s-m)^2}{2\,\sigma''^2}$$






Correlations

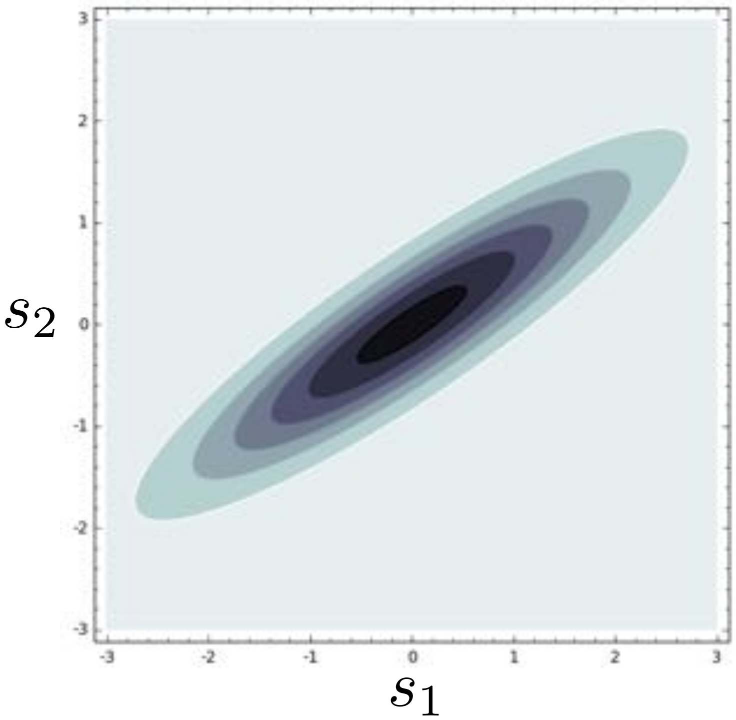


signal reconstruction with 2ⁿ pixels given 42 noisy data points
data
n = 2
n = 3
n = 4
n = 5
n = 6
n = 7
n = 8
n = 9
values
1
0
−1
0
π/4
π/2
3π/4
π
coordinate

Correlations

$\mathcal{P}(s)$

$$s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$



Correlations

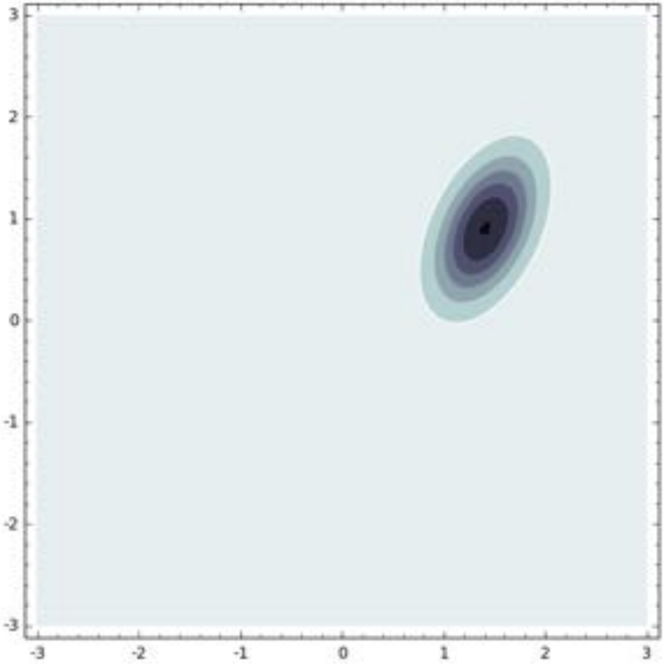
$\mathcal{P}(s|d)$

$$s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

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

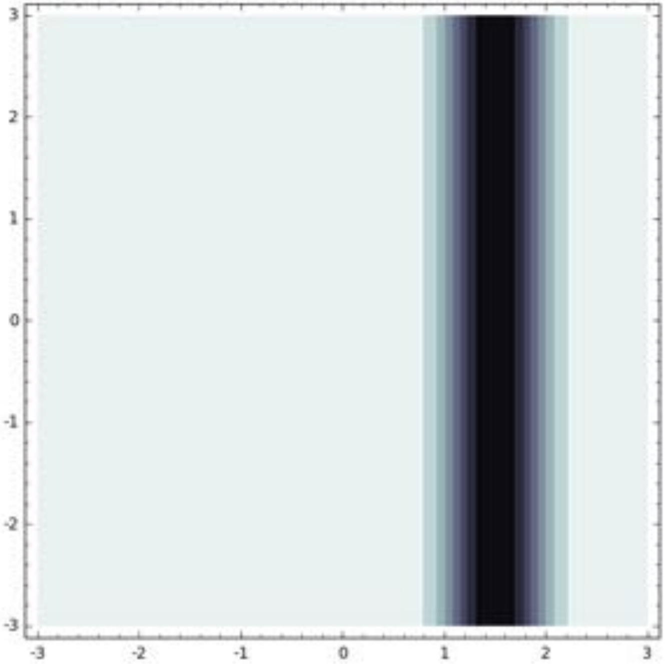
$$d = s_1 + n$$

s_2



s_1

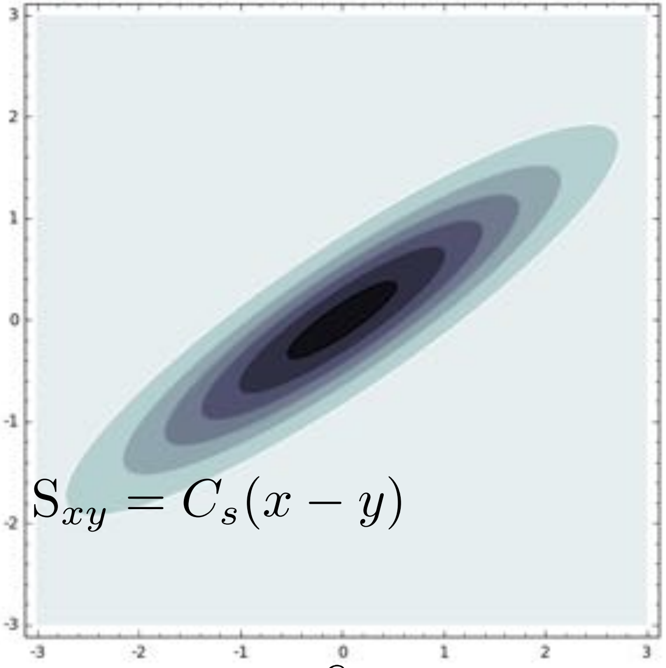
s_2



s_1

Correlations

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

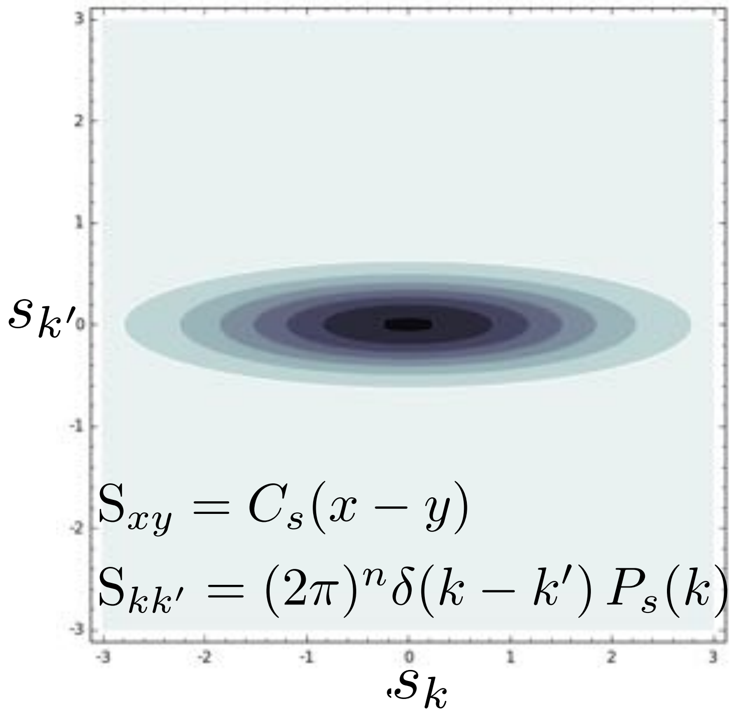


$$\begin{aligned}
\mathcal{P}(s) &= \mathcal{G}(s, S) \\
&= \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right) \\
S &= \begin{pmatrix} \langle s_1 s_1 \rangle & \langle s_1 s_2 \rangle \\ \langle s_2 s_1 \rangle & \langle s_2 s_2 \rangle \end{pmatrix} \quad \text{2-dim.} \\
S_{ij}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

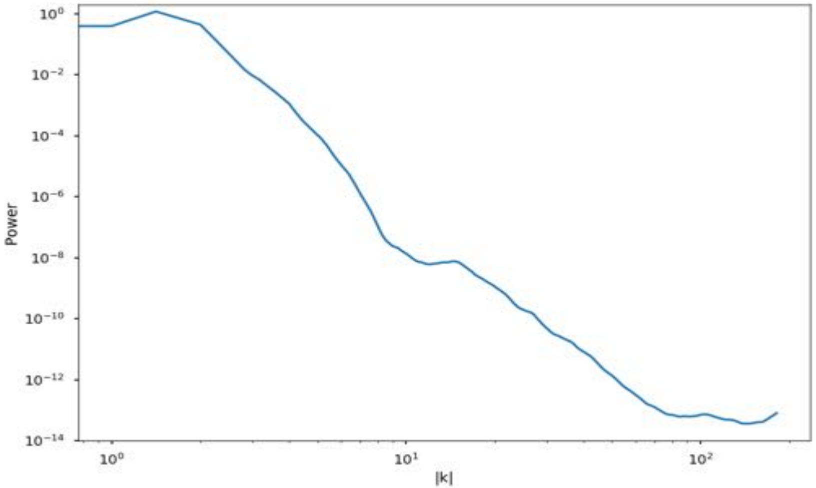
Correlations

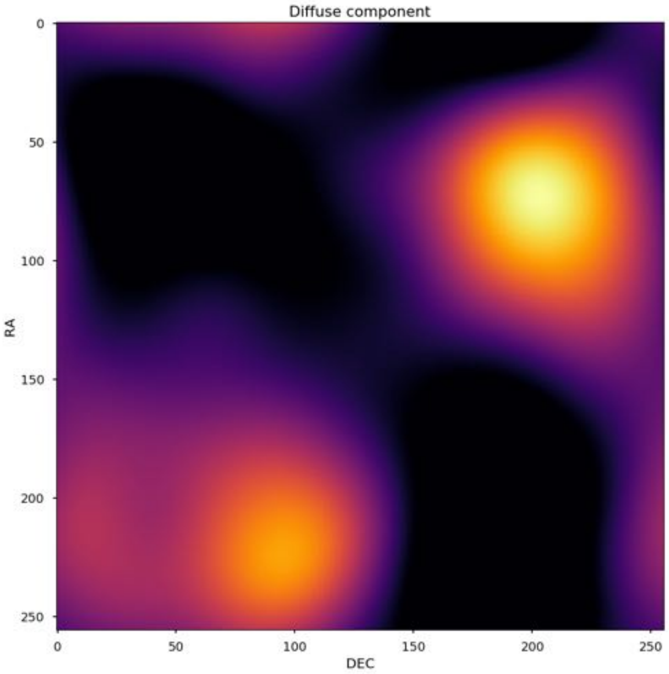
$$\mathcal{P}(s)$$

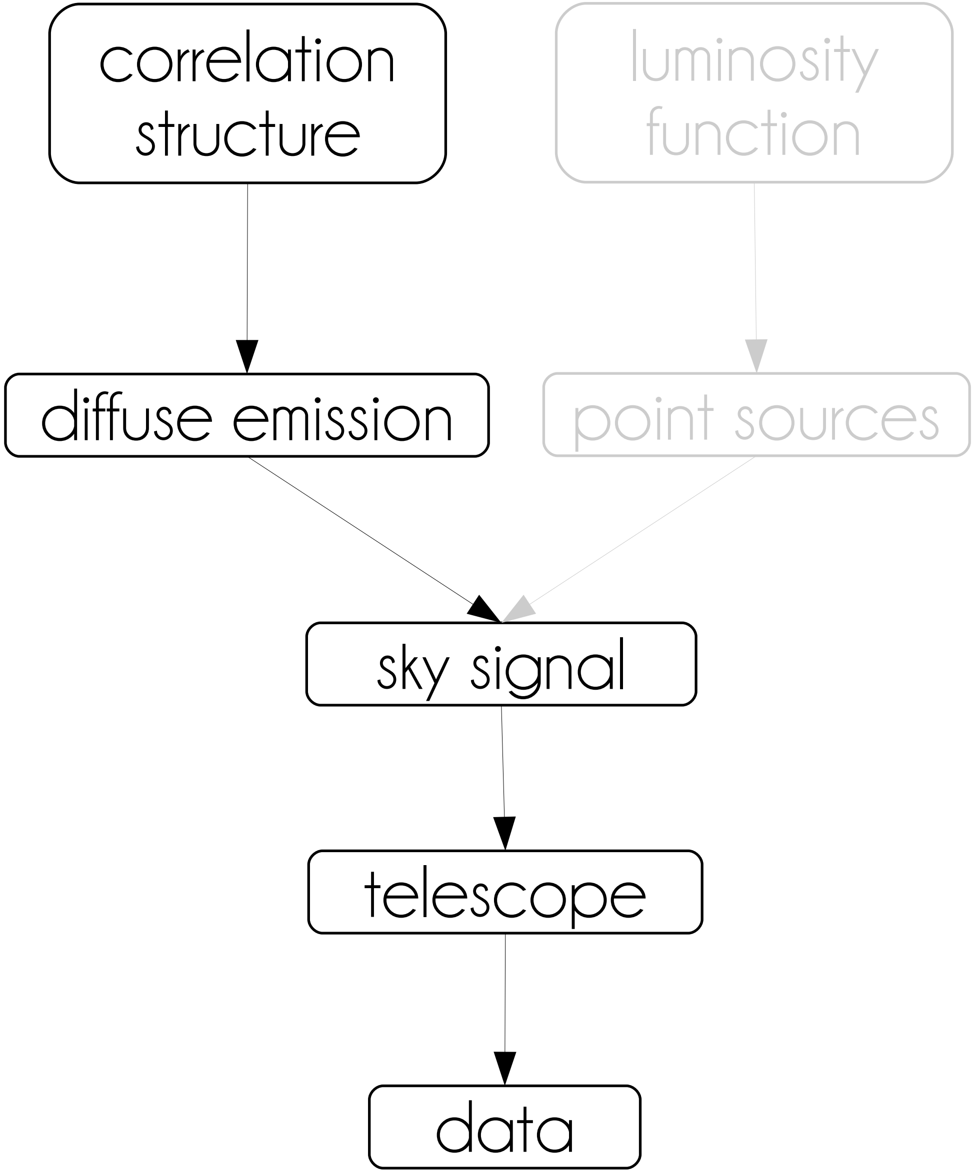
$$s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$



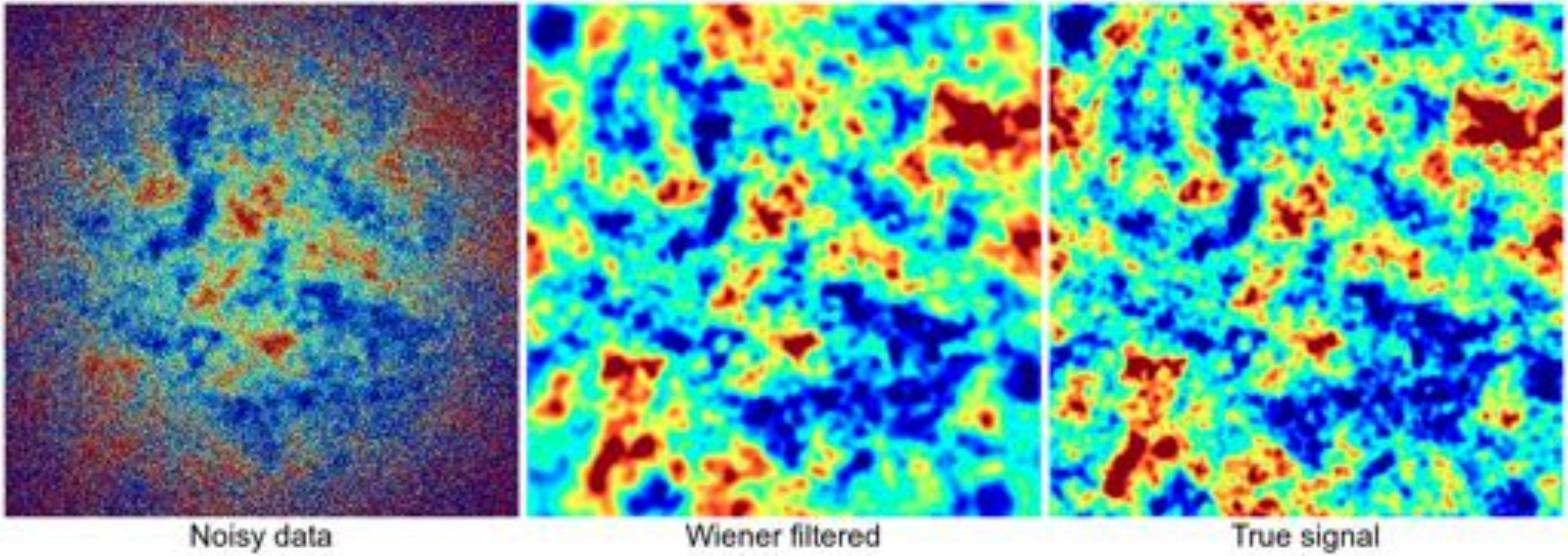
$$\begin{aligned}
\mathcal{P}(s) &= \mathcal{G}(s, S) \\
&= \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right) \\
S &= \begin{pmatrix} \langle s_1 s_1 \rangle & \langle s_1 s_2 \rangle \\ \langle s_2 s_1 \rangle & \langle s_2 s_2 \rangle \end{pmatrix} \quad \text{2-dim.} \\
S_{ij}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


Power
|k|


Diffuse component
RA
DEC


correlation structure
luminosity function
diffuse emission
point sources
sky signal
telescope
data

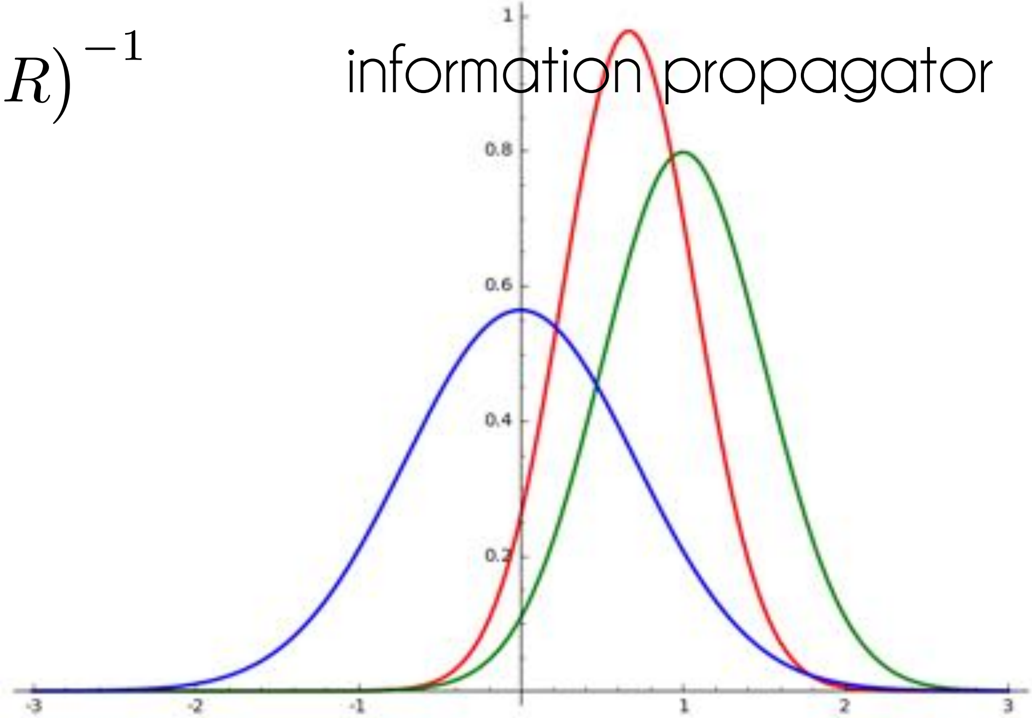
Wiener Filter



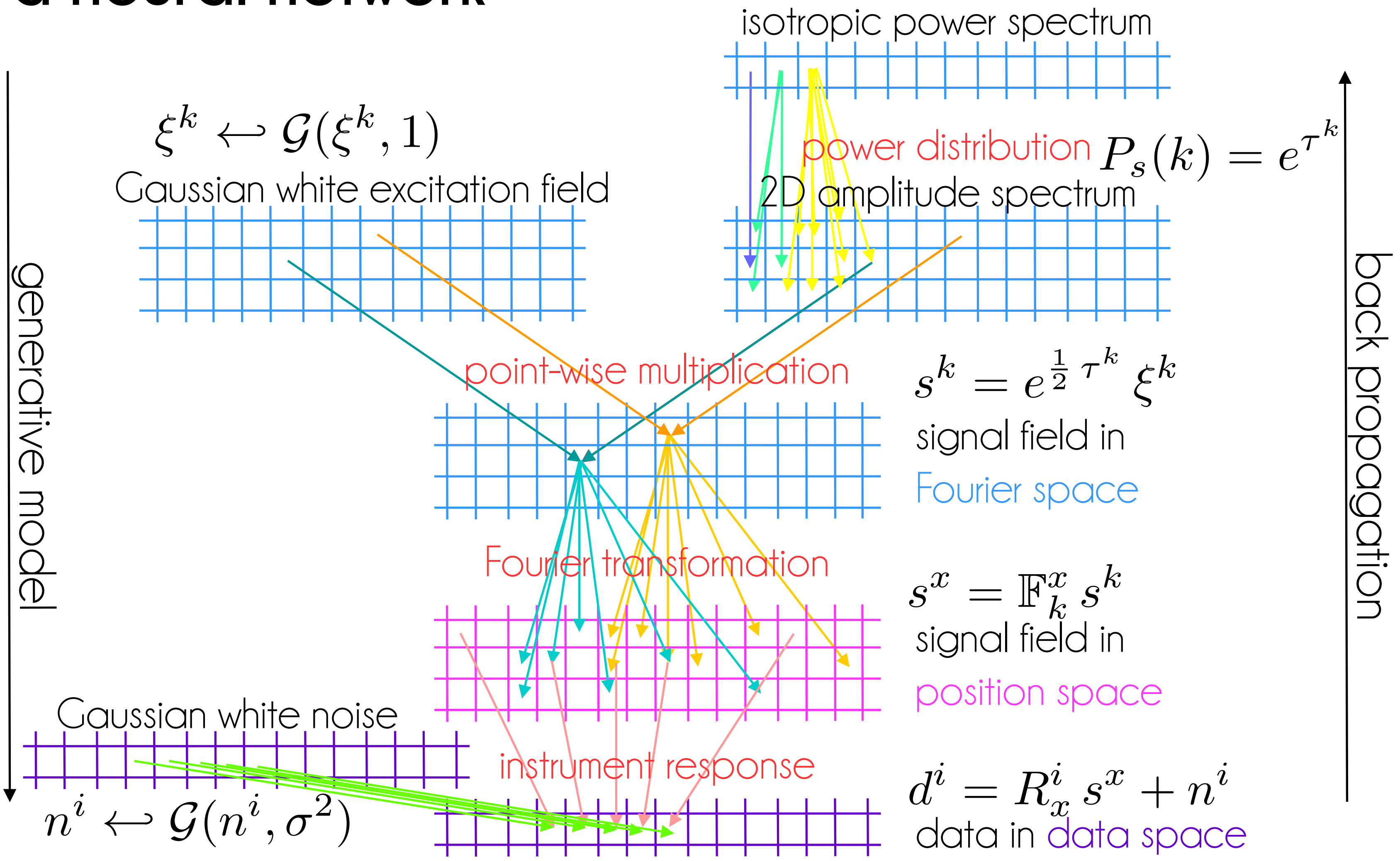
https://en.wikipedia.org/wiki/Generalized_Wiener_filter

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

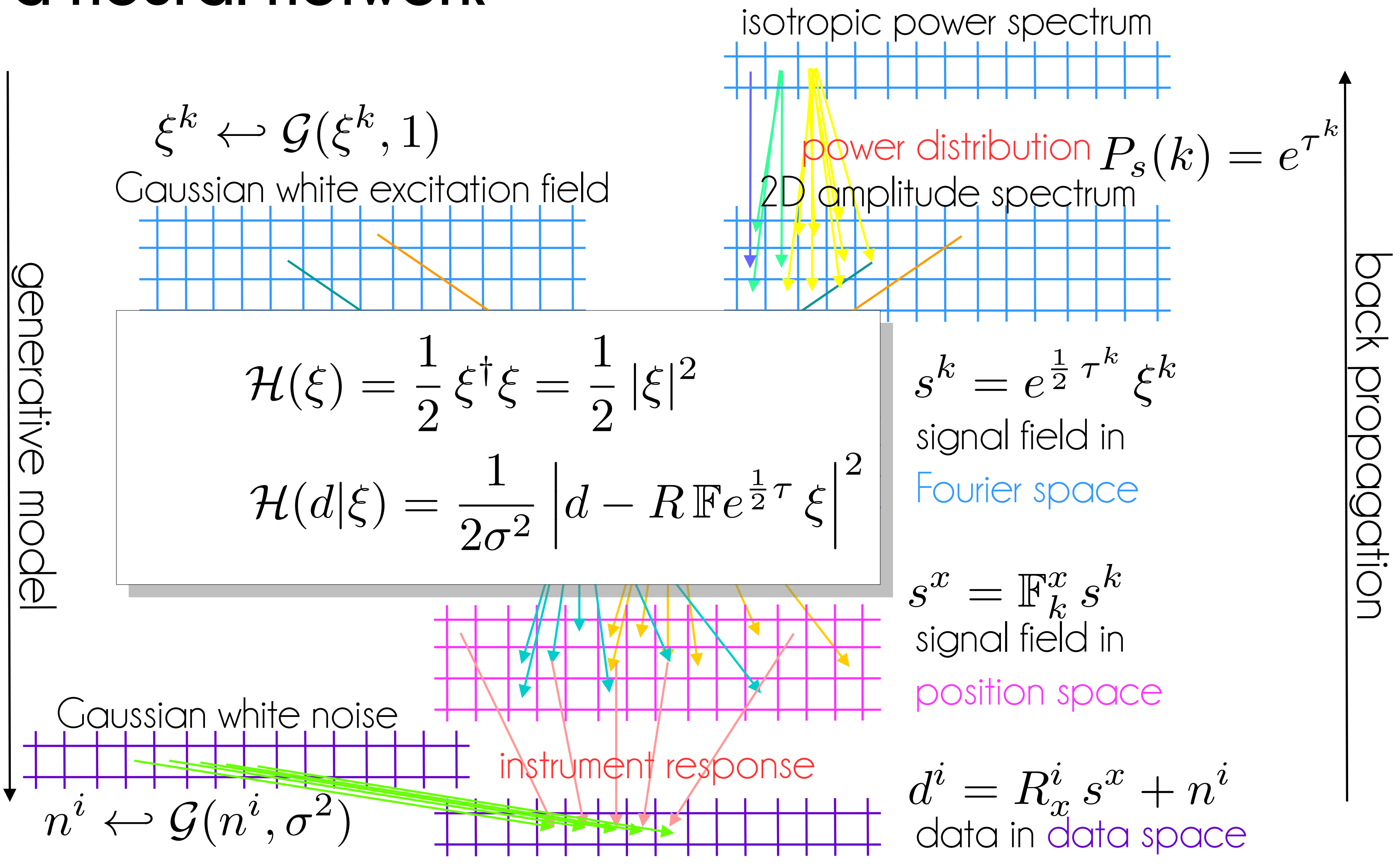
$$
\begin{aligned}
d &= R\,s + n && \text{data} \\
\mathcal{P}(d, s|R, S, N) &= \mathcal{G}(s, S)\,\mathcal{G}(d - R\,s, N) && \text{prior \& likelihood} \\
\mathcal{P}(s|d, R, S, N) &= \mathcal{G}(s - m, D) && \text{posterior} \\
m &= D\,j && \text{posterior mean} \\
j &= R^{\dagger} N^{-1} d && \text{information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



IFT as a neural network

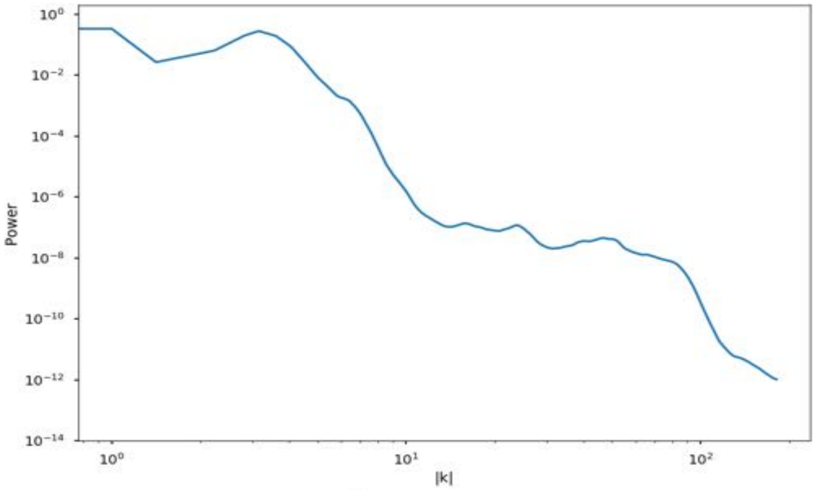


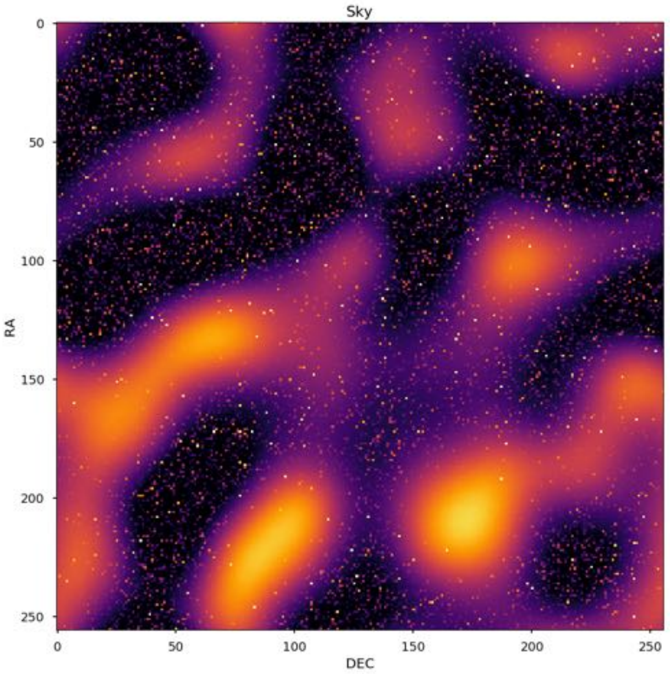
IFT as a neural network

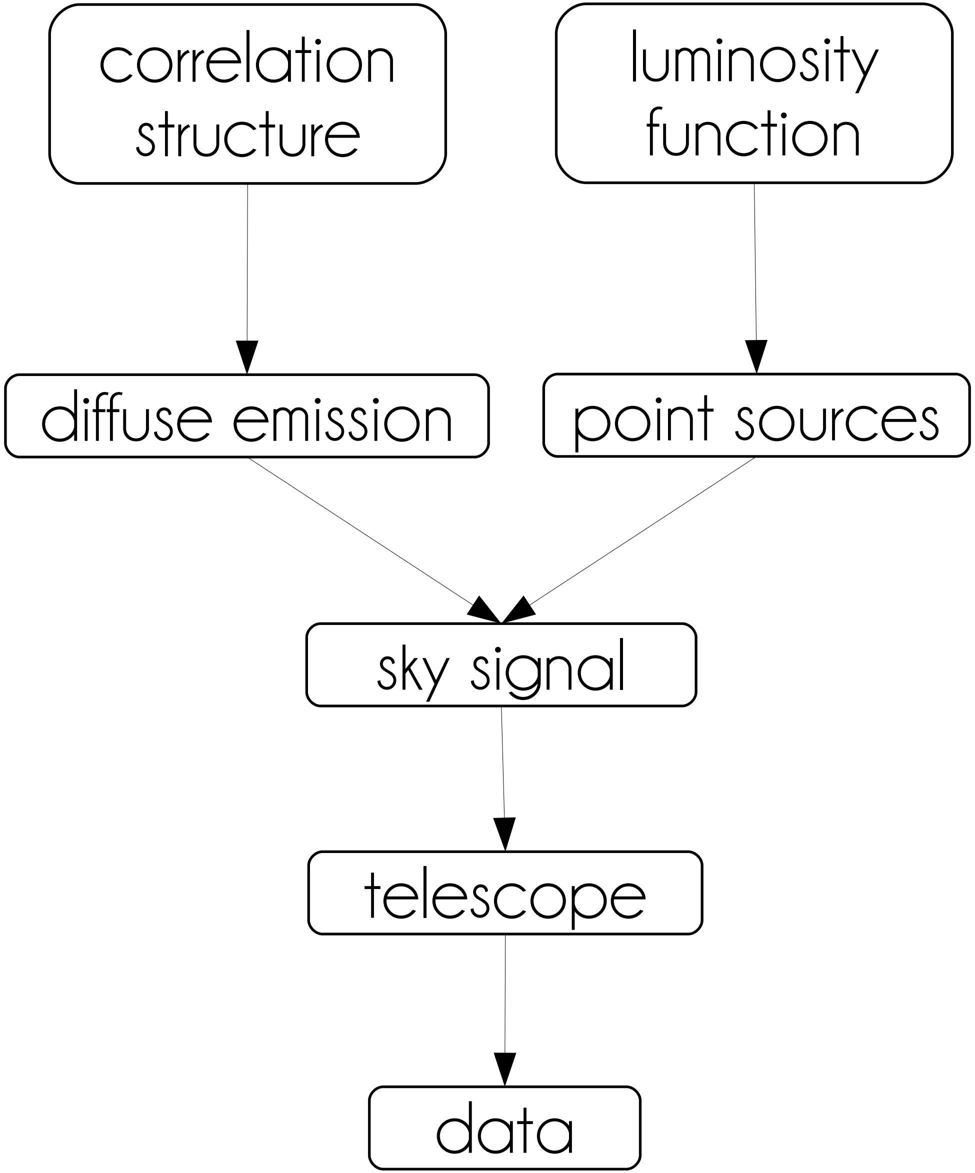


NIFTy tutorial part 1

linear reconstructions


Power
|k|


Sky
RA
DEC


correlation structure
luminosity function
diffuse emission
point sources
sky signal
telescope
data

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

known → $d = R\, e^{s} + n$

unknown → $\lambda = R\, e^{s}$

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

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

Information

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

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

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

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

$$+ (\boldsymbol{\beta} - \boldsymbol{1})^\dagger \ldots$$

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

- Convert into **generative model**
- Compress information into Gaussian via **Metric Gaussian Variational Inference**

Variational Bayes

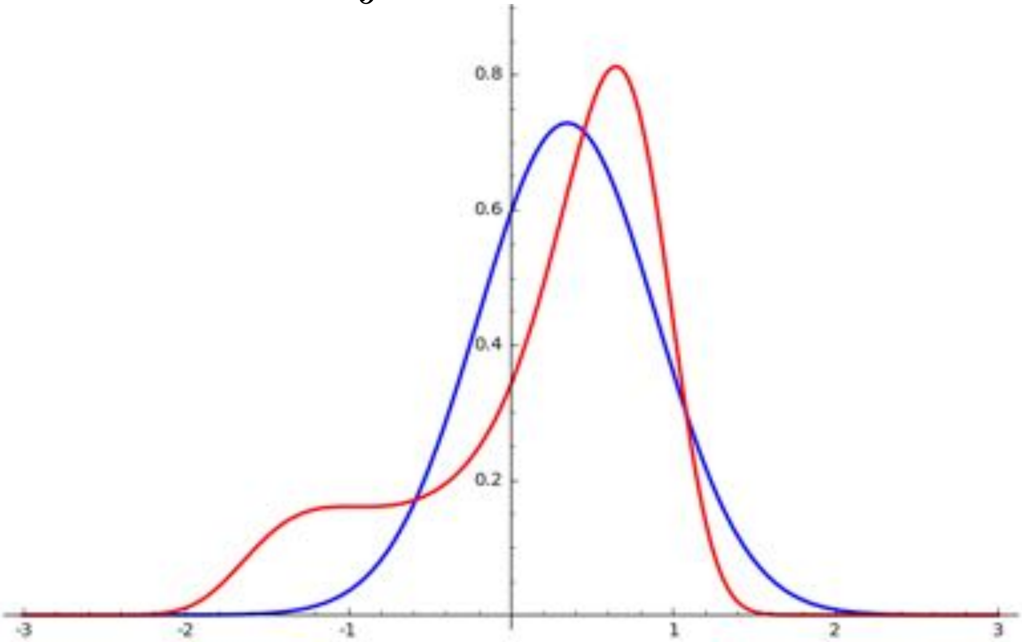
$\mathcal{P}(s|d)$

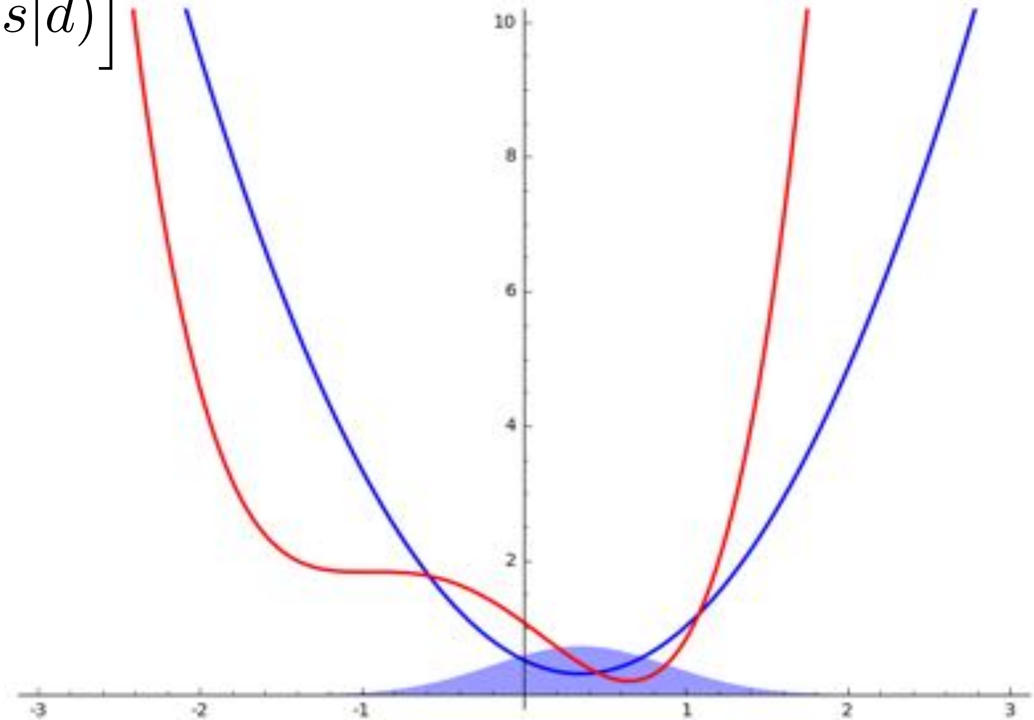
$\widetilde{\mathcal{P}}(s|d) = \mathcal{G}(s-m, D)$

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

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

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





Metric Gaussian Variational Bayes

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

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

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

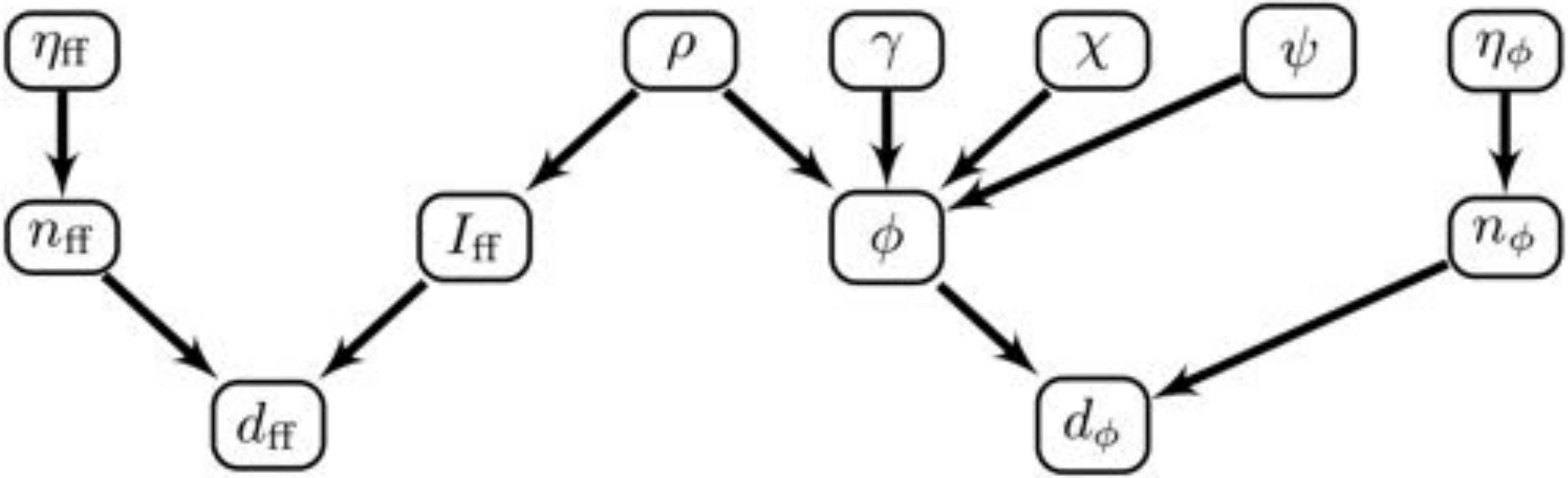
$\widetilde{\mathcal{H}}(s|d) \widehat{=} \frac{1}{2}(s-m)^\dagger D^{-1}(s-m)$

Knollmüller & Enßlin (2019)

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

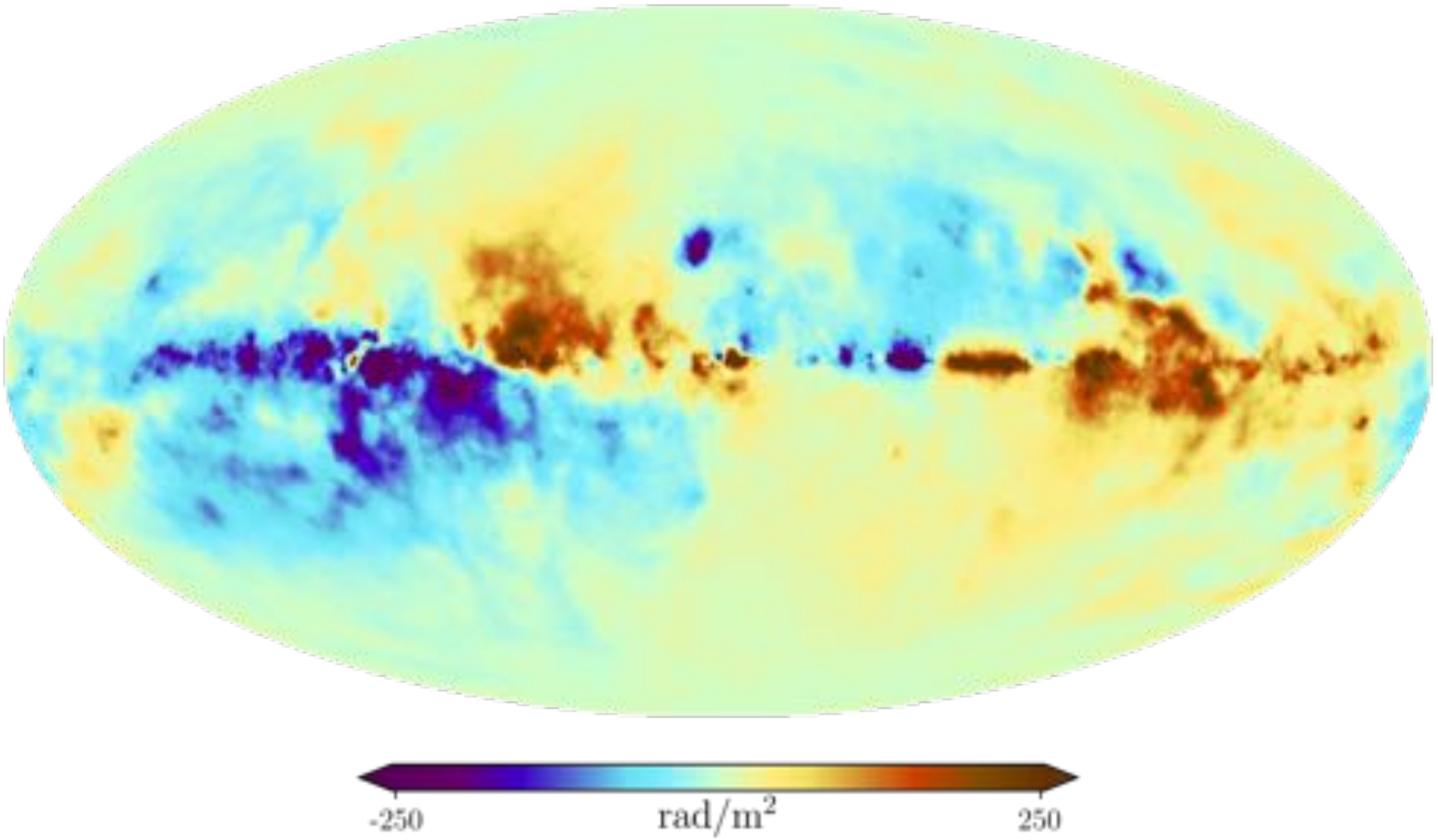
$$D \approx \left\langle \frac{\partial \mathcal{H}(d,s)}{\partial s} \frac{\partial \mathcal{H}(d,s)}{\partial s}^\dagger \right\rangle_{(d|s=m)}^{-1}$$

Hierarchical Bayesian Model

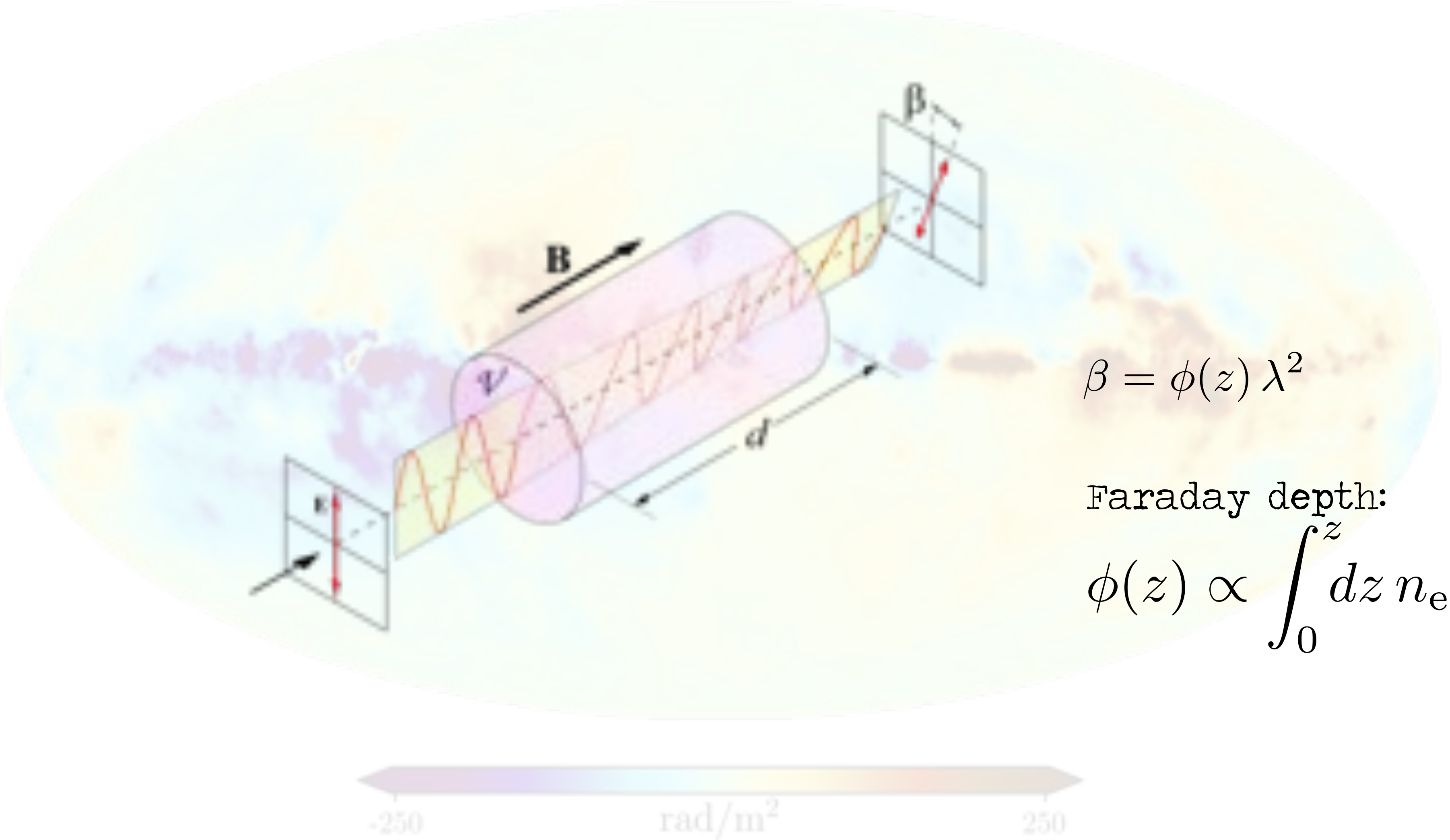


Galactic Faraday Sky

Hutschenreuter & Enßlin (2019)



Faraday Effect



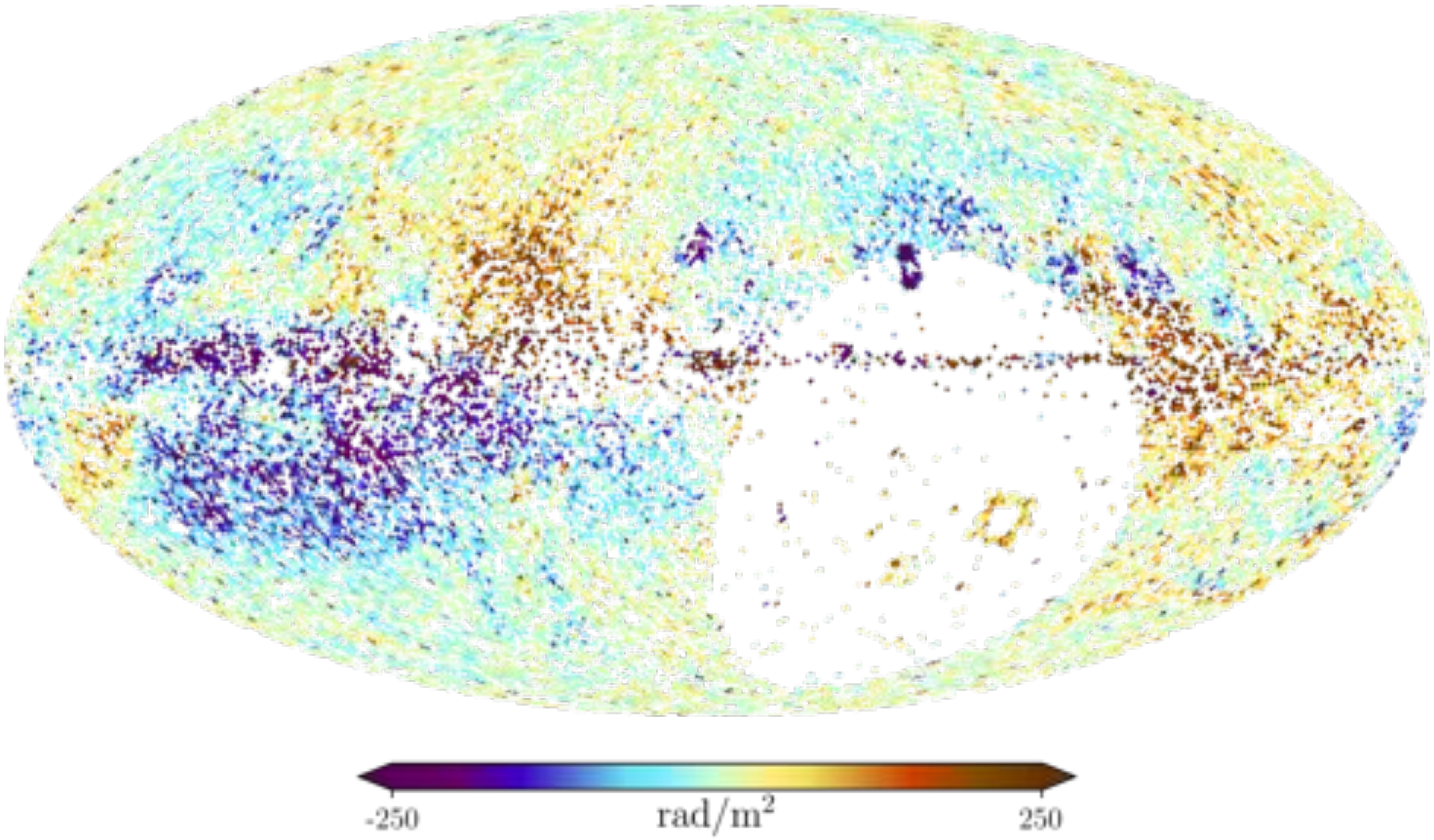
$$\beta = \phi(z)\,\lambda^2$$

Faraday depth:

$$\phi(z) \propto \int_0^z dz\, n_\mathrm{e}\, B_z$$

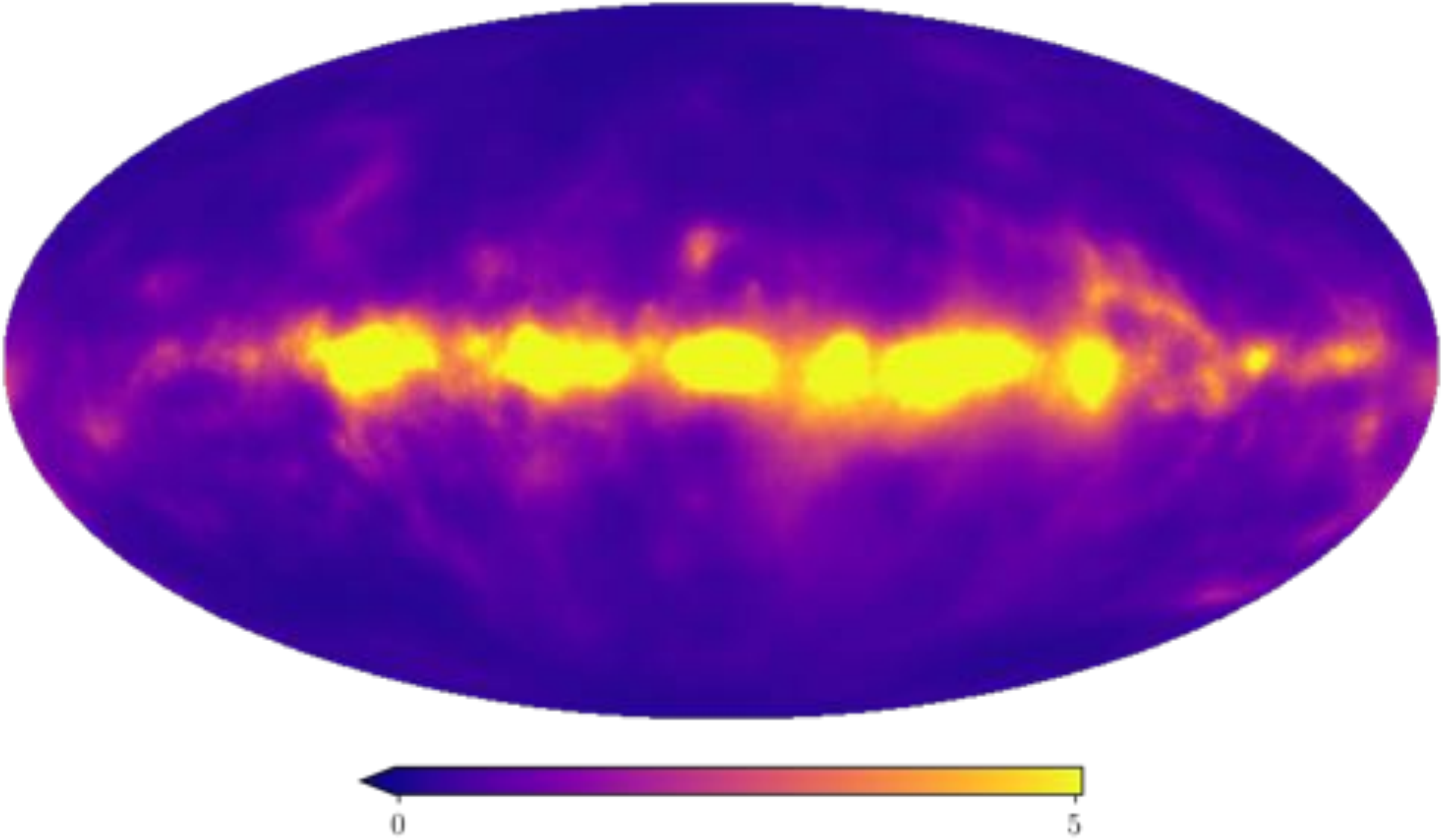
Faraday Data

Oppermann et al. (2012)



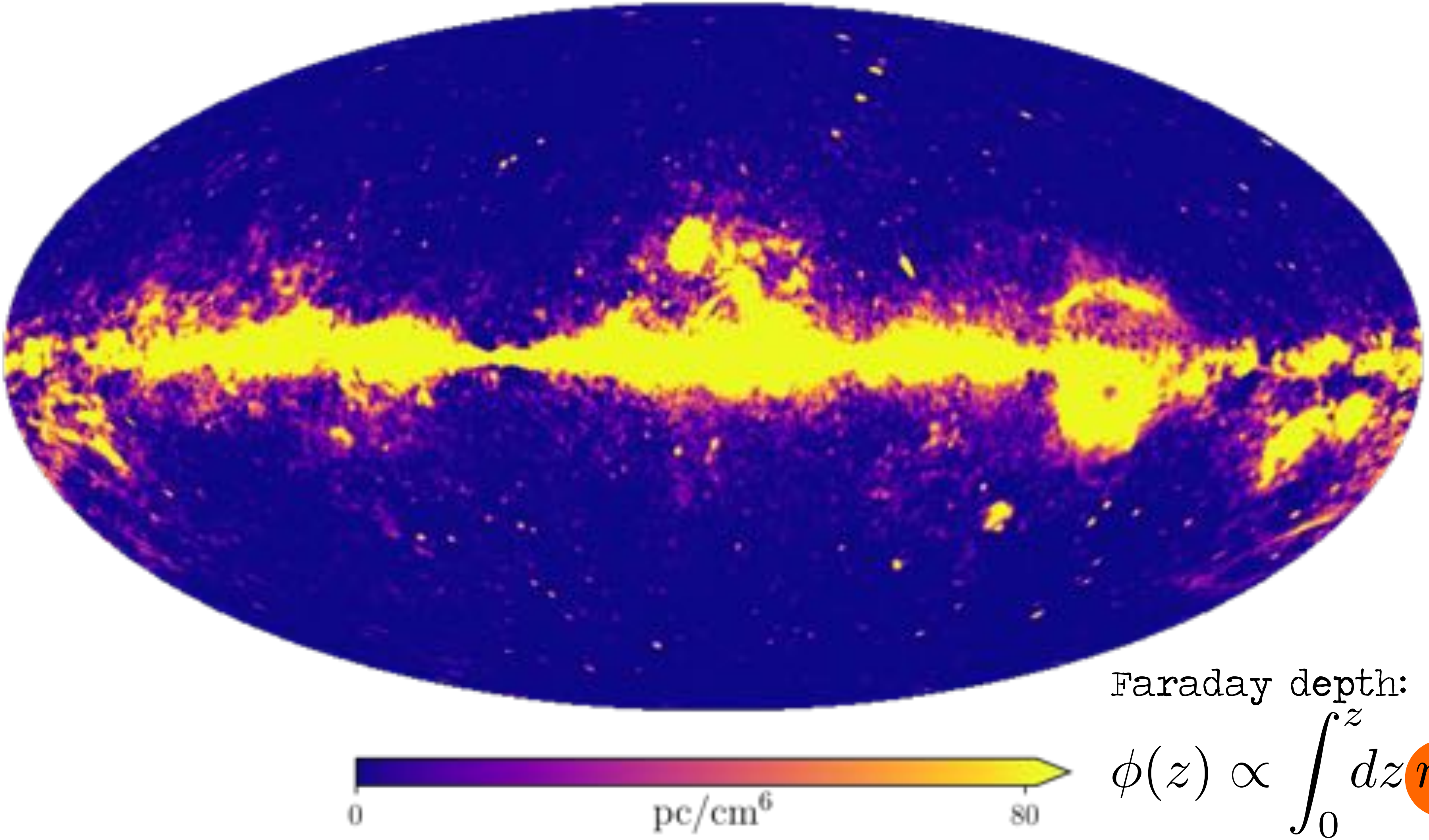
Faraday Amplitude Field

Hutschenreuter & Enßlin (2019)



Planck Free-Free Emission

Hutschenreuter & Enßlin (2019)

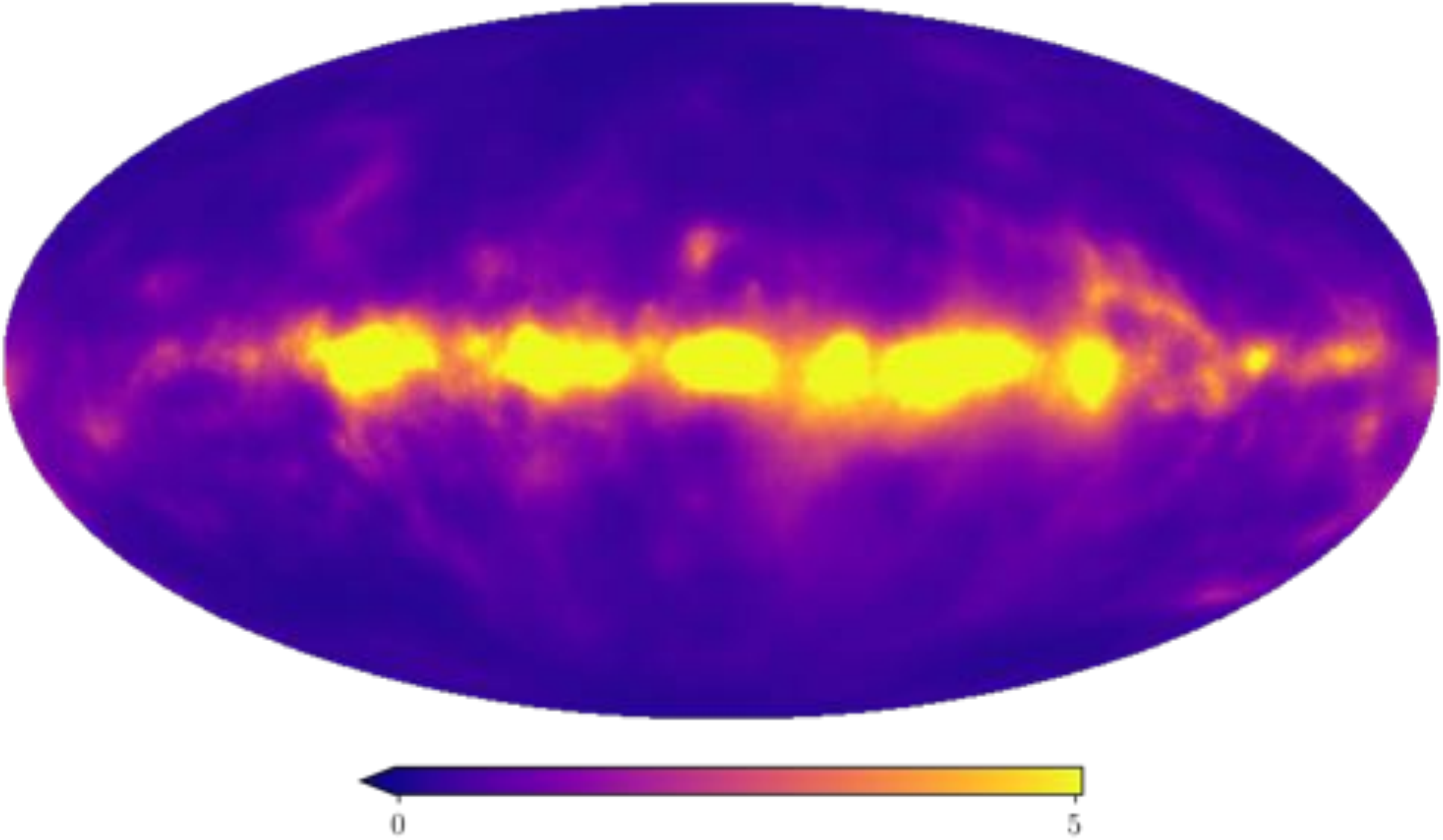


Faraday depth:

$$\phi(z) \propto \int_0^z dz\, n_\mathrm{e} B_z$$

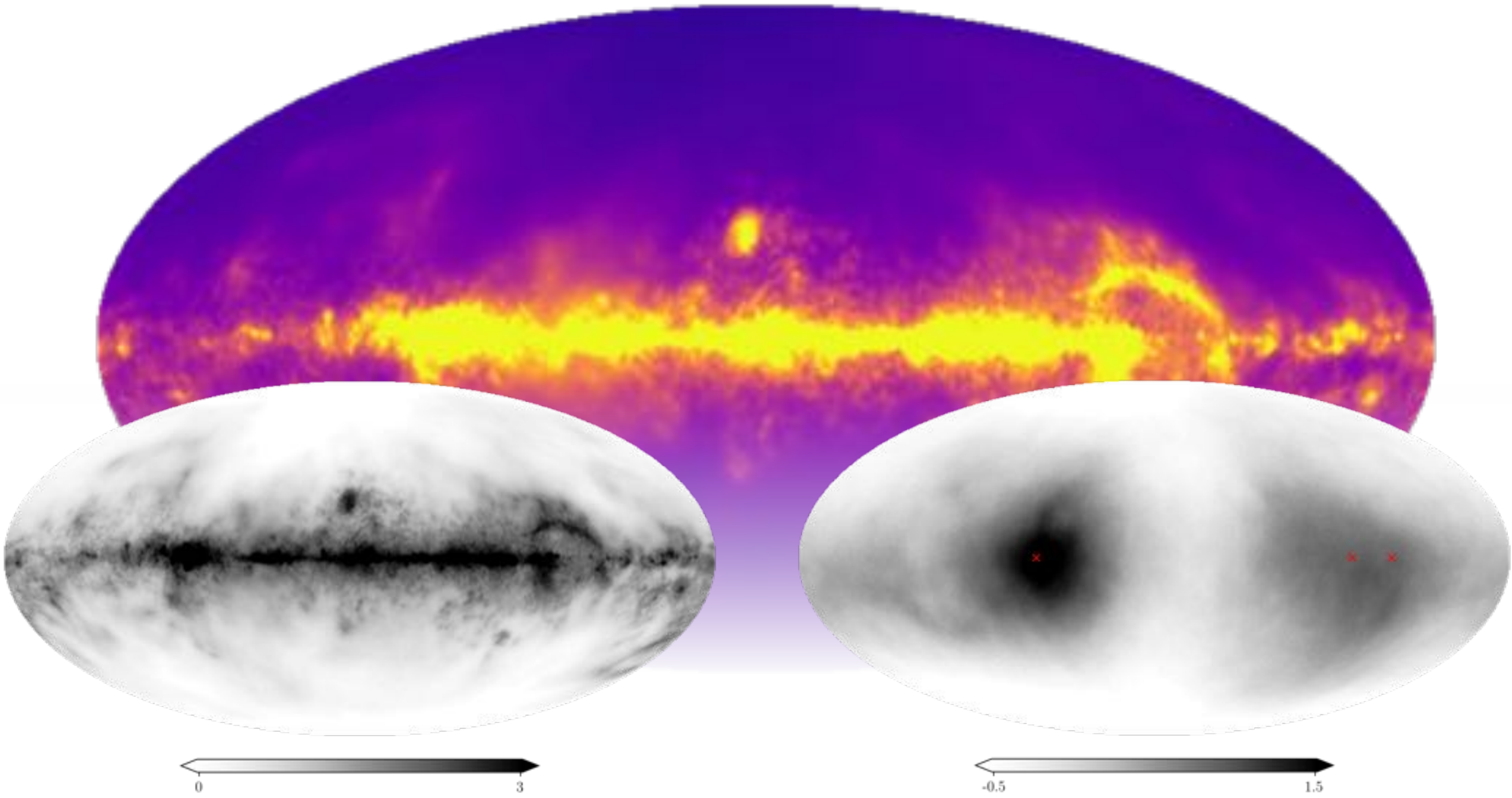
Faraday Amplitude Field

Hutschenreuter & Enßlin (2019)



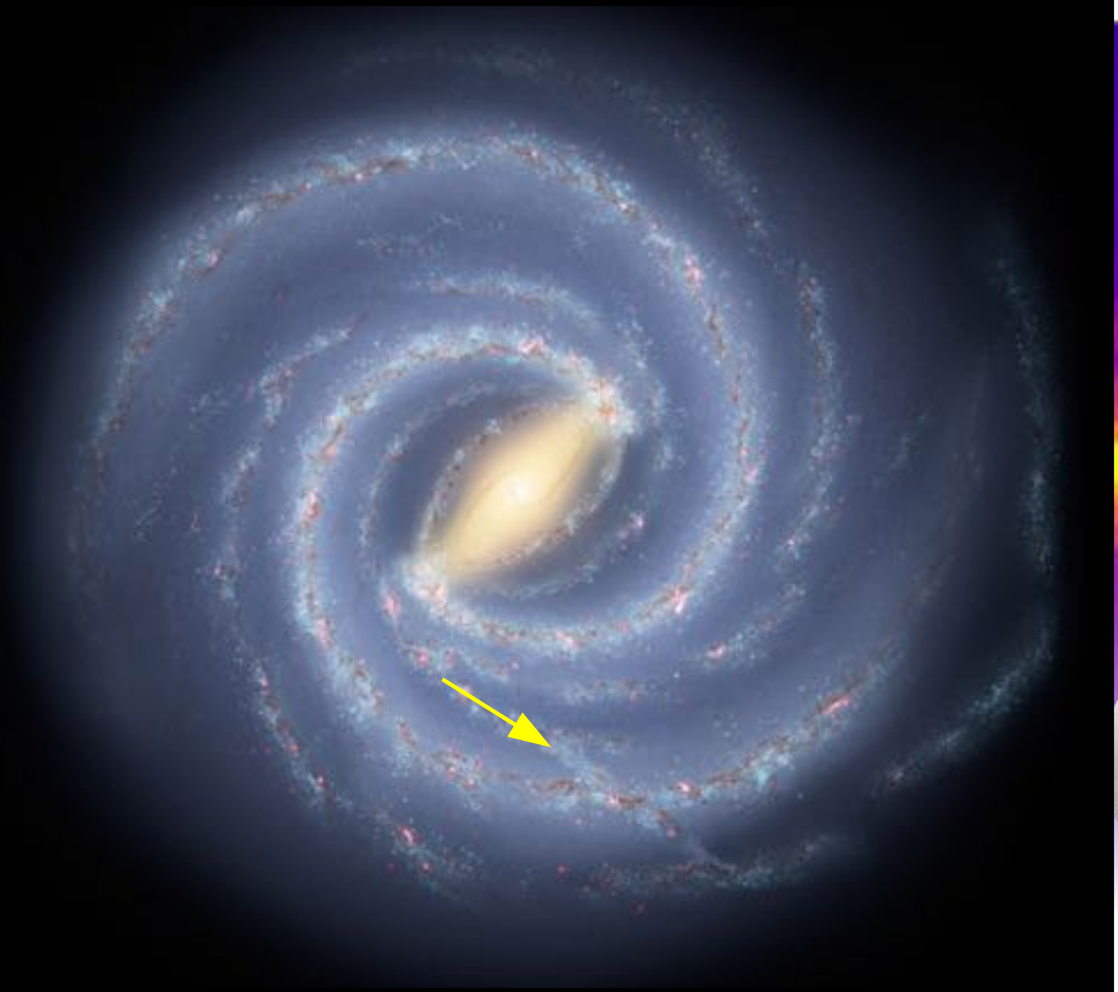
Faraday Amplitude Field

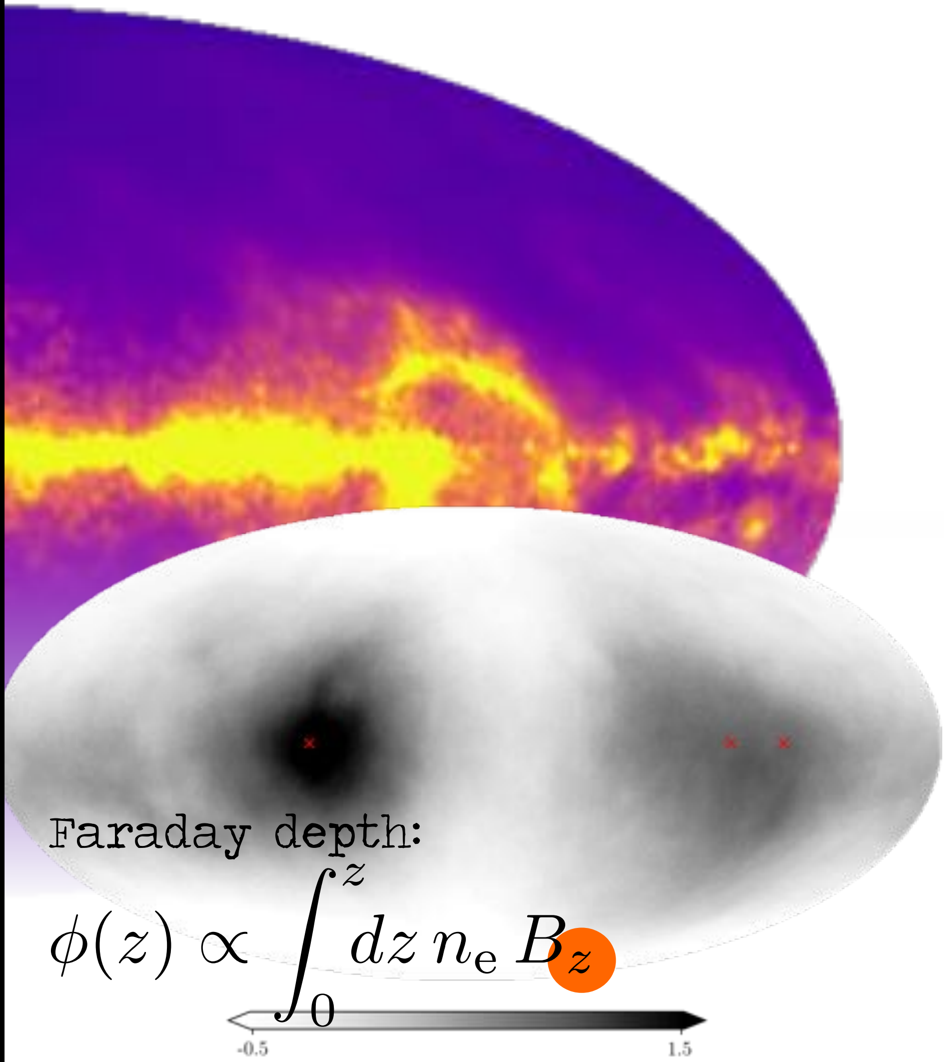
Hutschenreuter & Enßlin (2019)



…tude Field

Hutschenreuter & Enßlin (2019)



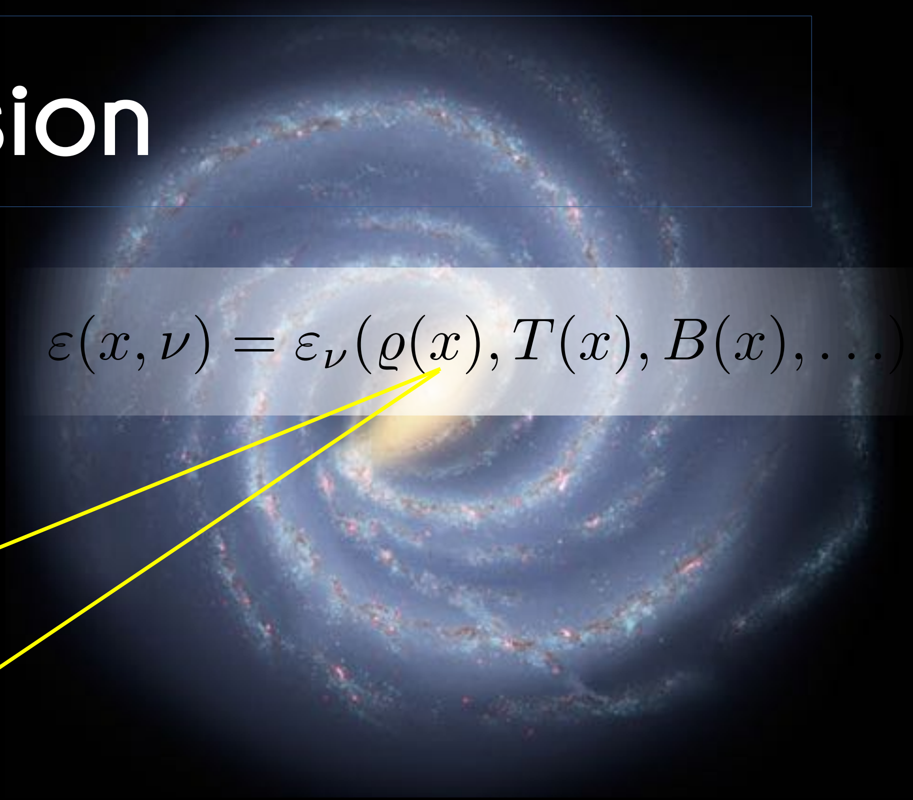


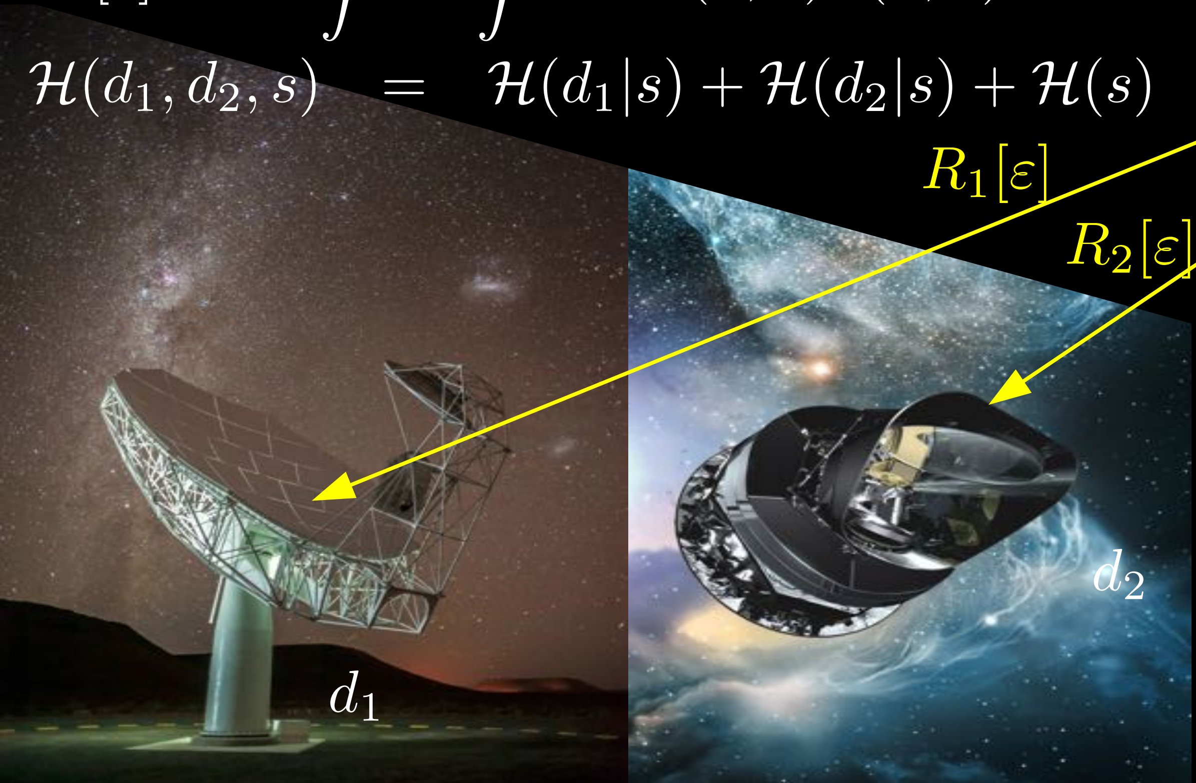
Faraday depth:

$$\phi(z) \propto \int_0^z dz\, n_e\, B_z$$

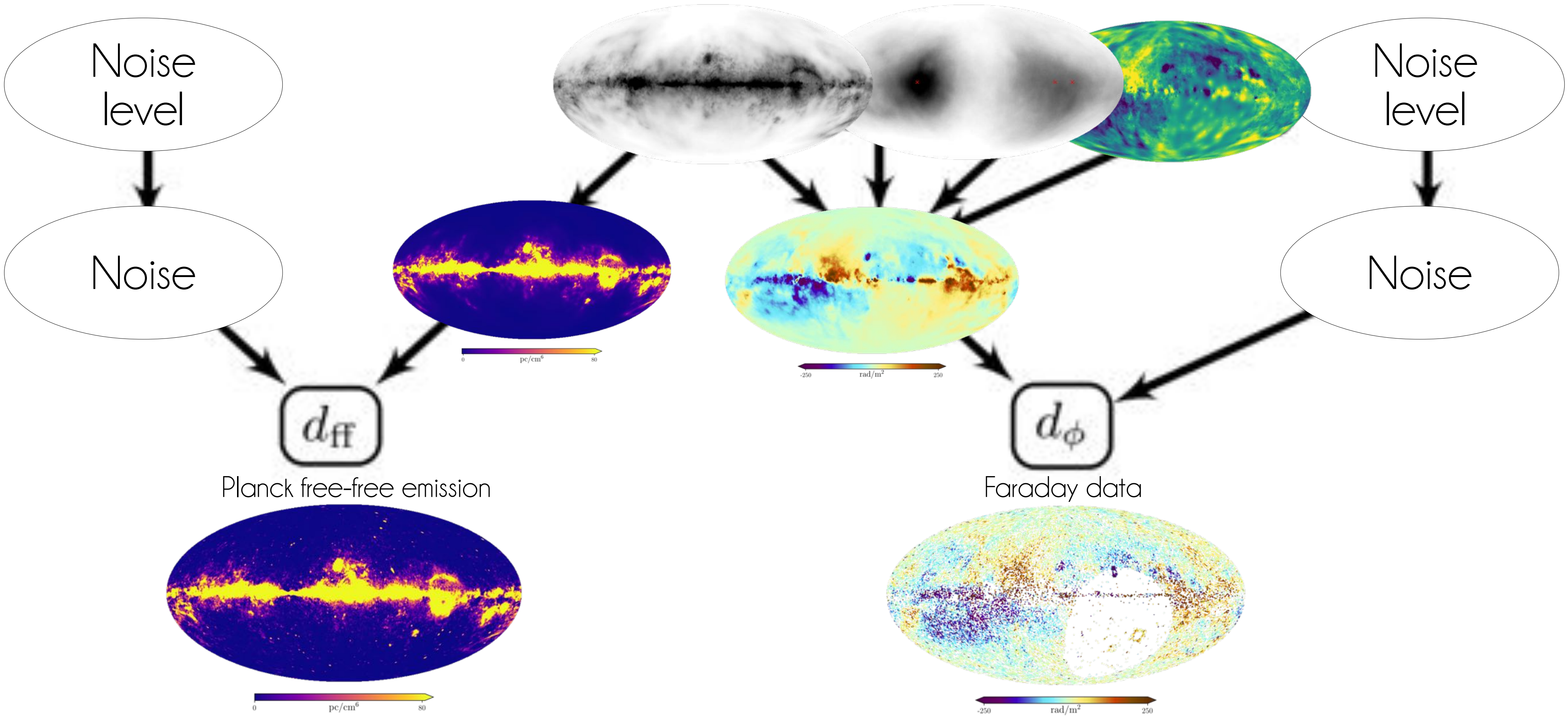
Data Fusion

$$
\begin{aligned}
d_i &= R_i[\varepsilon] + n_i \\
R_i[\varepsilon] &= \int \mathrm{d}x \int \mathrm{d}\nu \, R_i(x,\nu)\, \varepsilon(x,\nu) \\
\mathcal{H}(d_1, d_2, s) &= \mathcal{H}(d_1|s) + \mathcal{H}(d_2|s) + \mathcal{H}(s)
\end{aligned}
$$



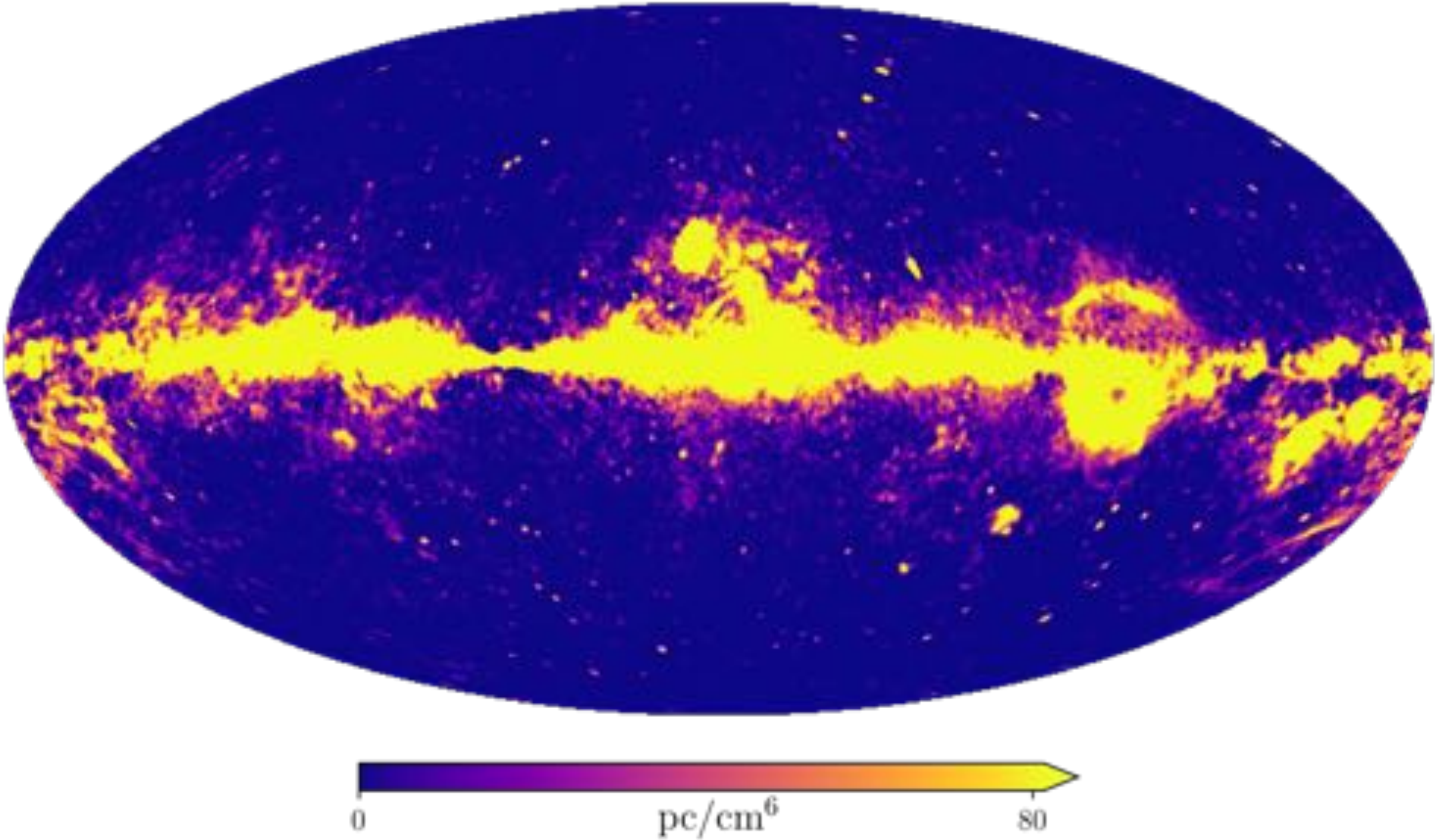


Hierarchical Bayesian Model

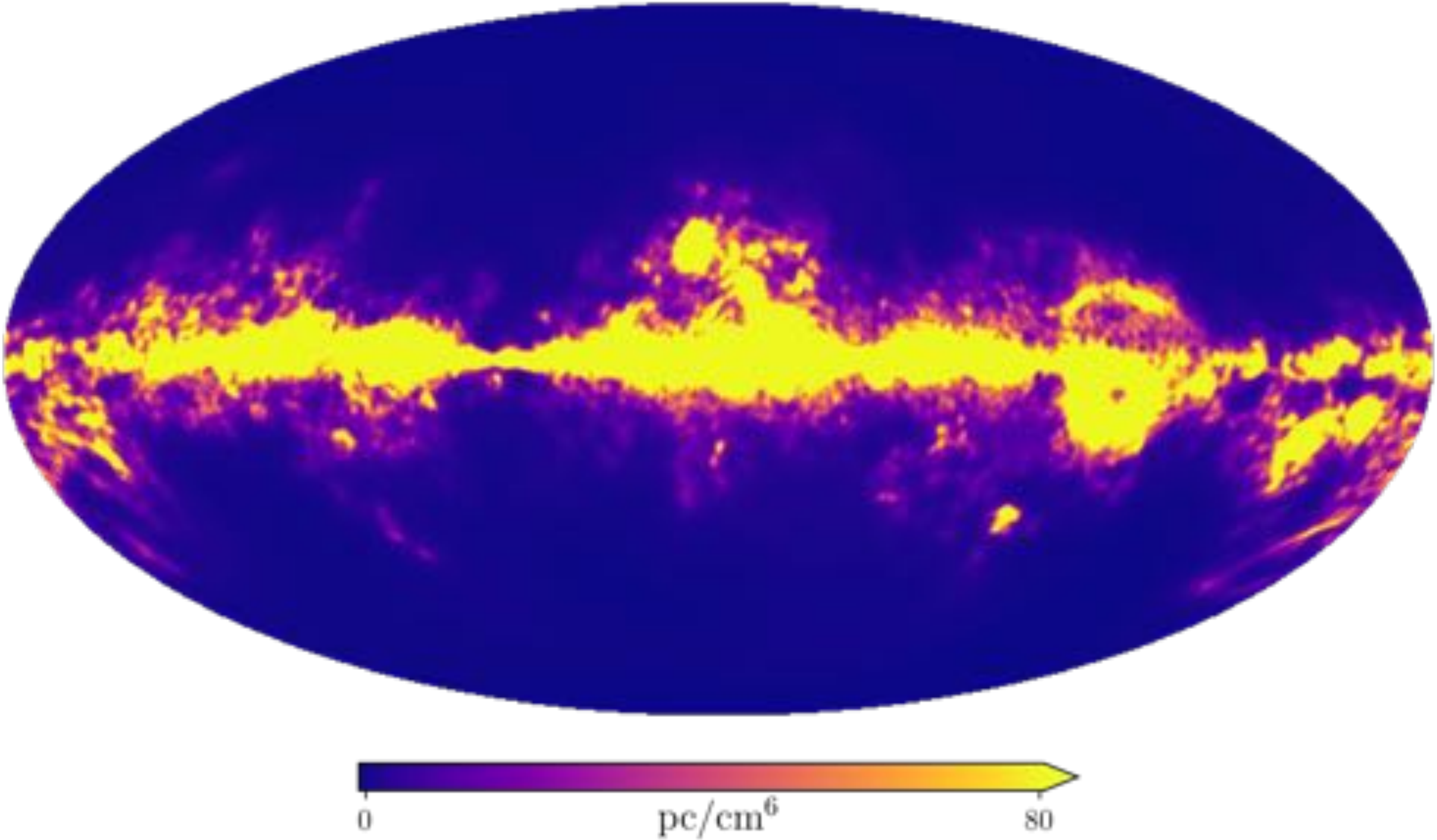


Planck free free map

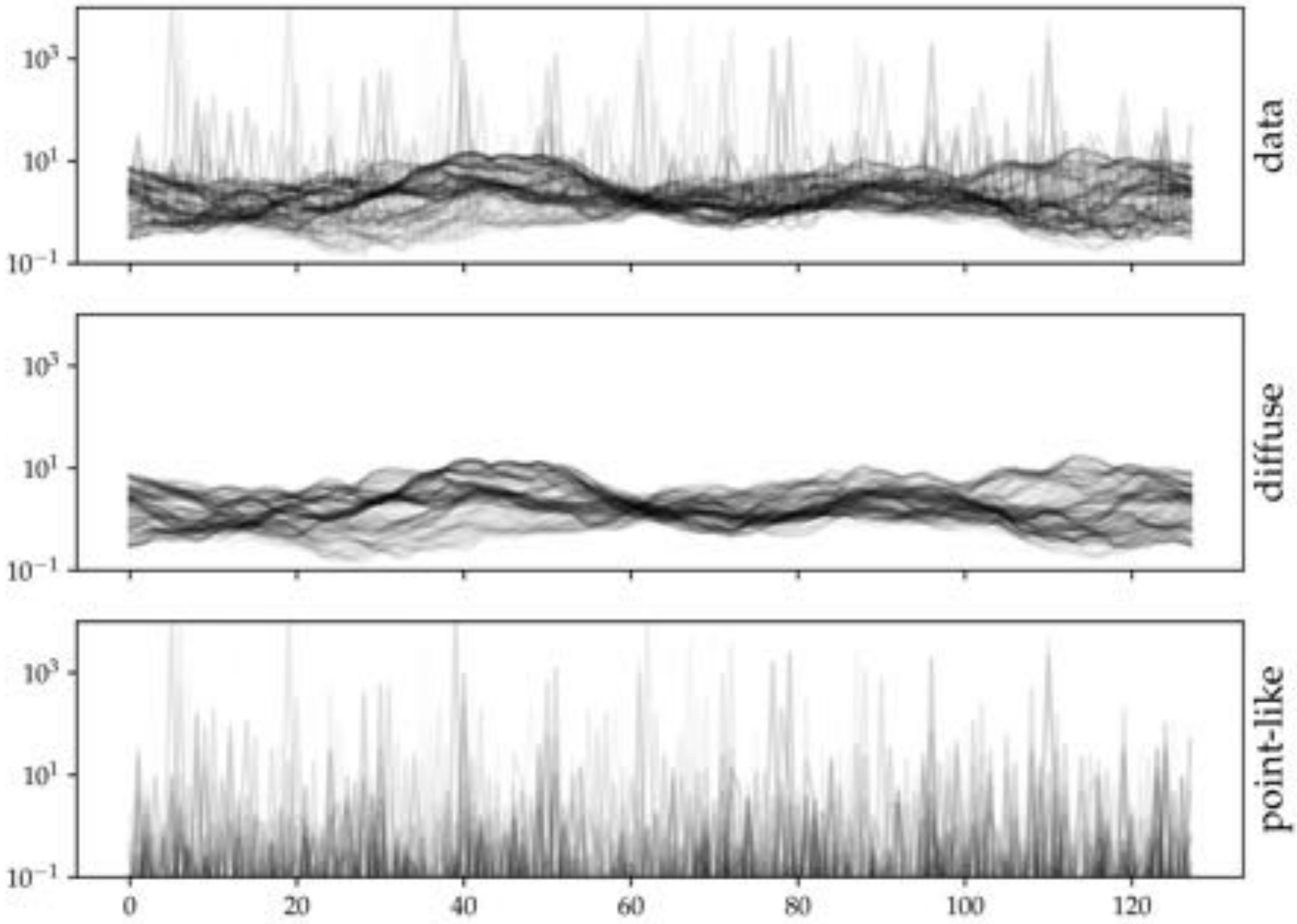
Hutschenreuter & Enßlin (2019)



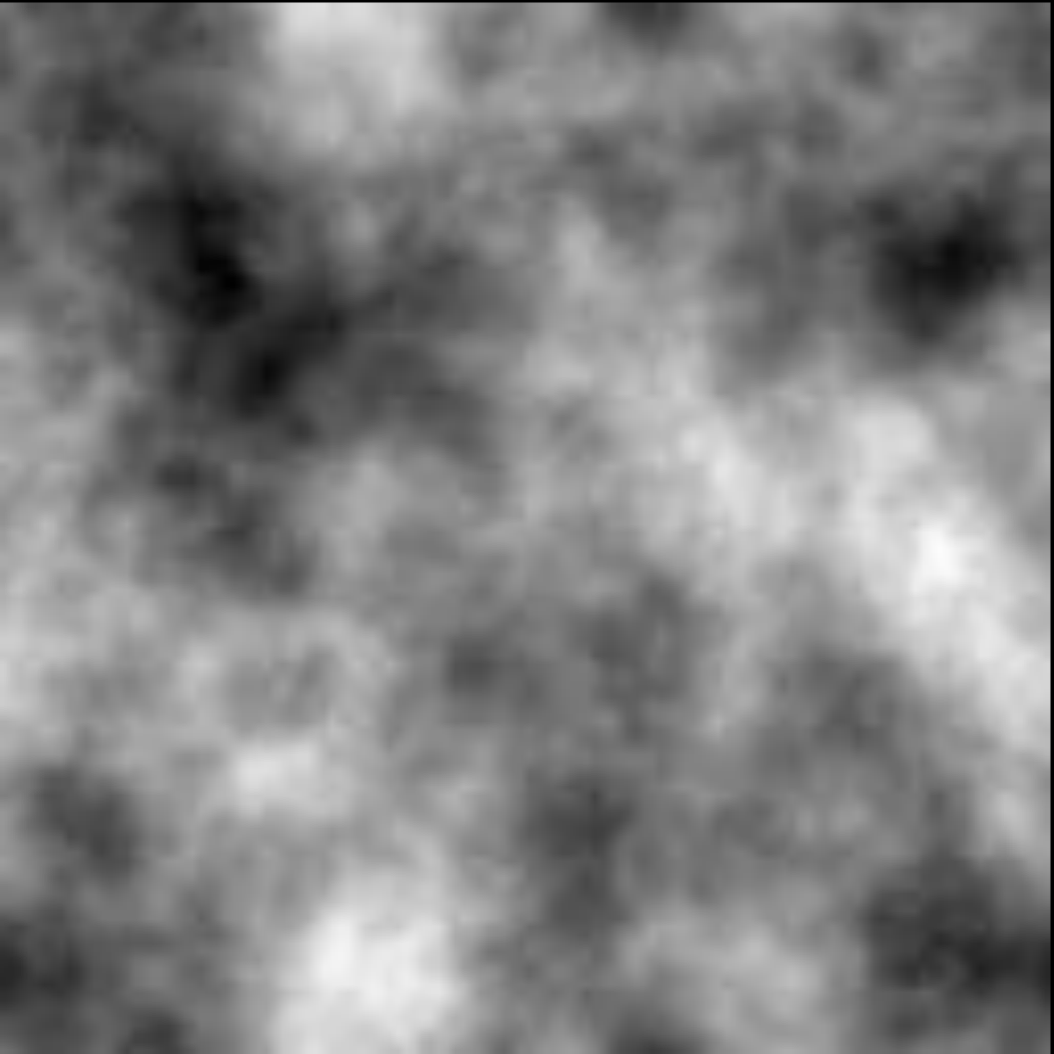
Inferred free free map Hutschenreuter & Enßlin (2019)



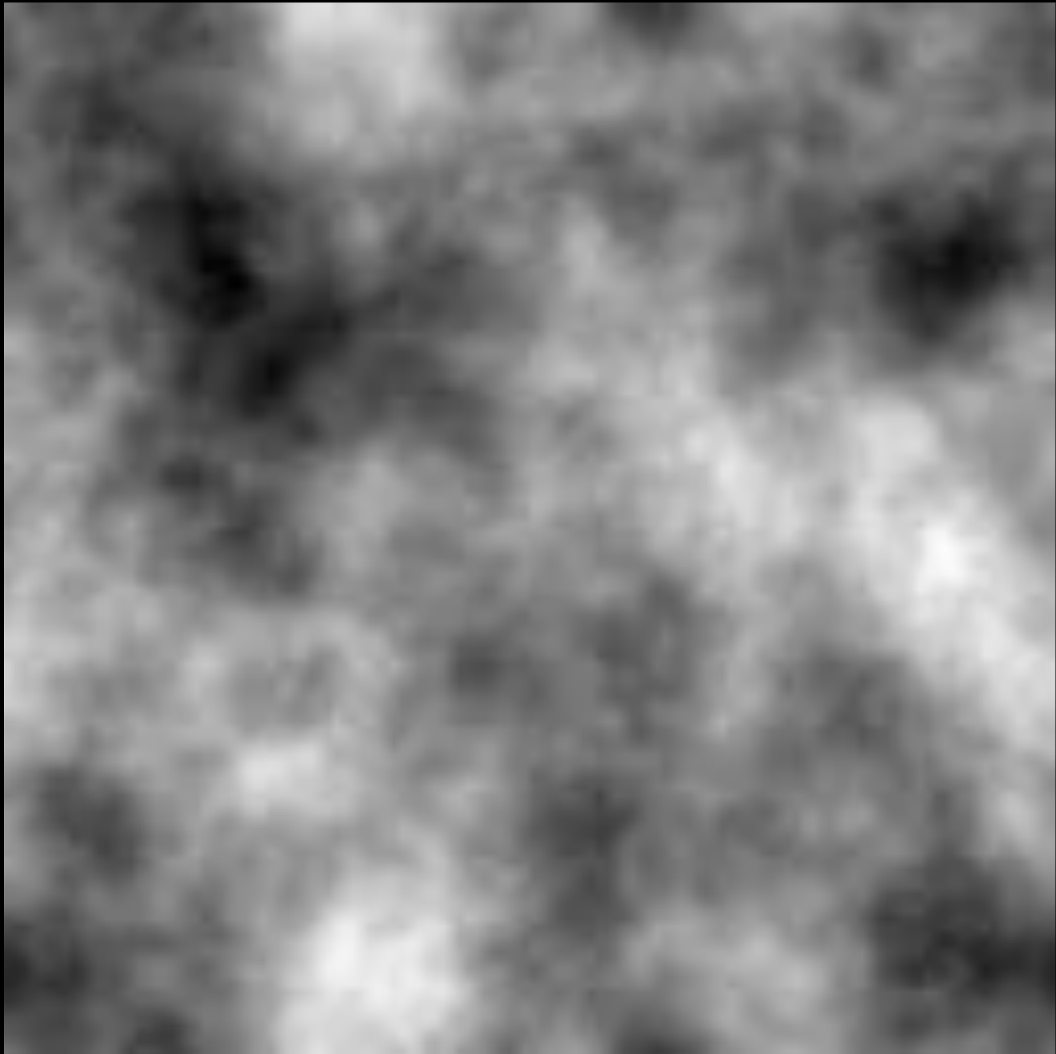
data and true components



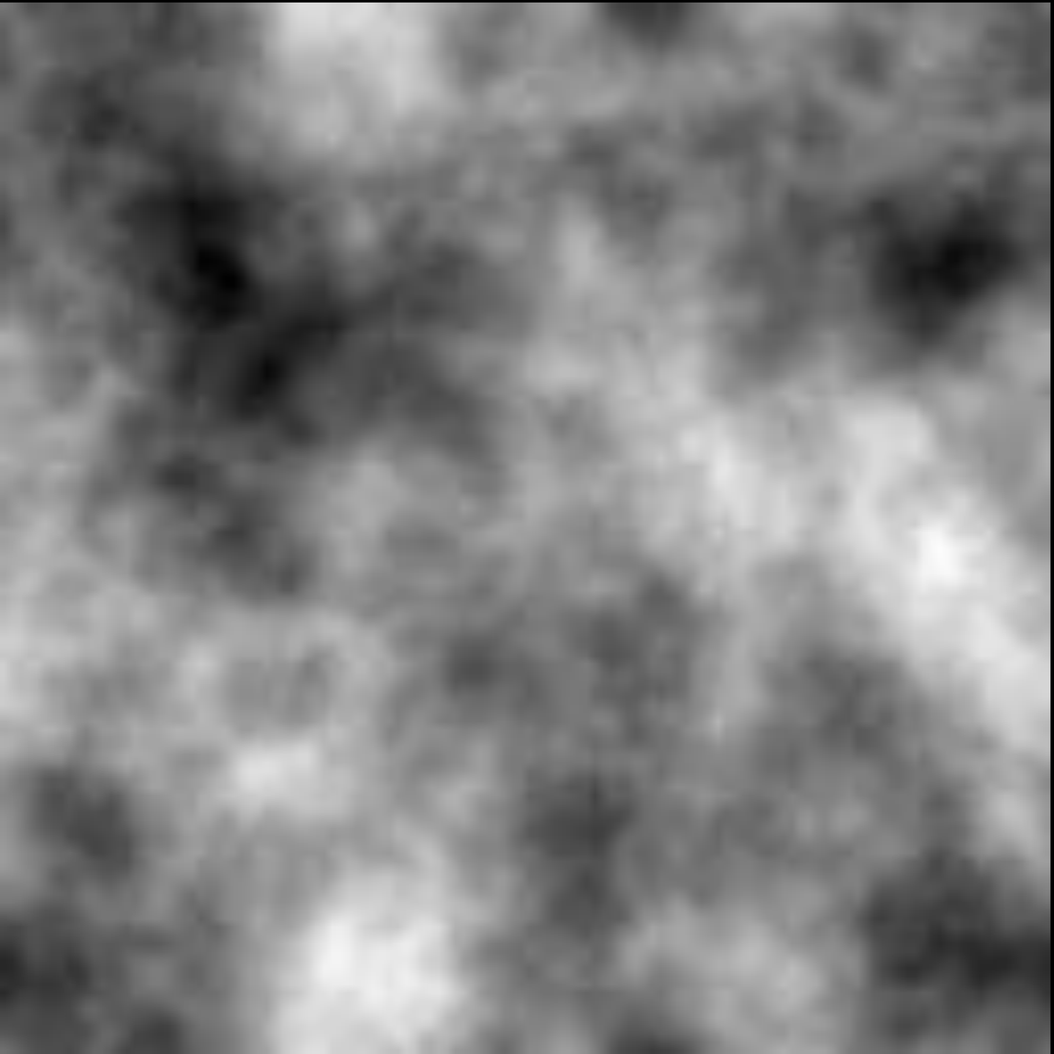
ground truth / starblade



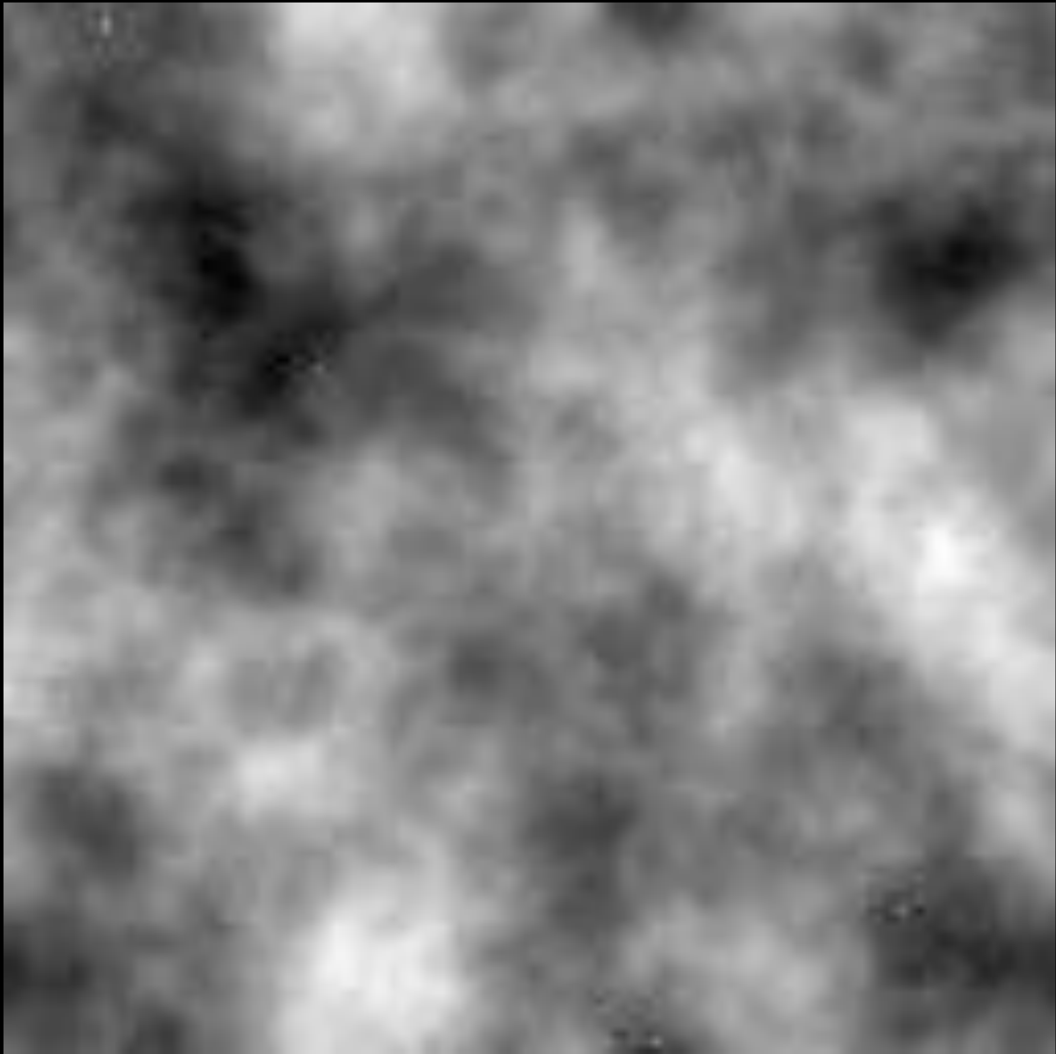
ground truth / autoencoder

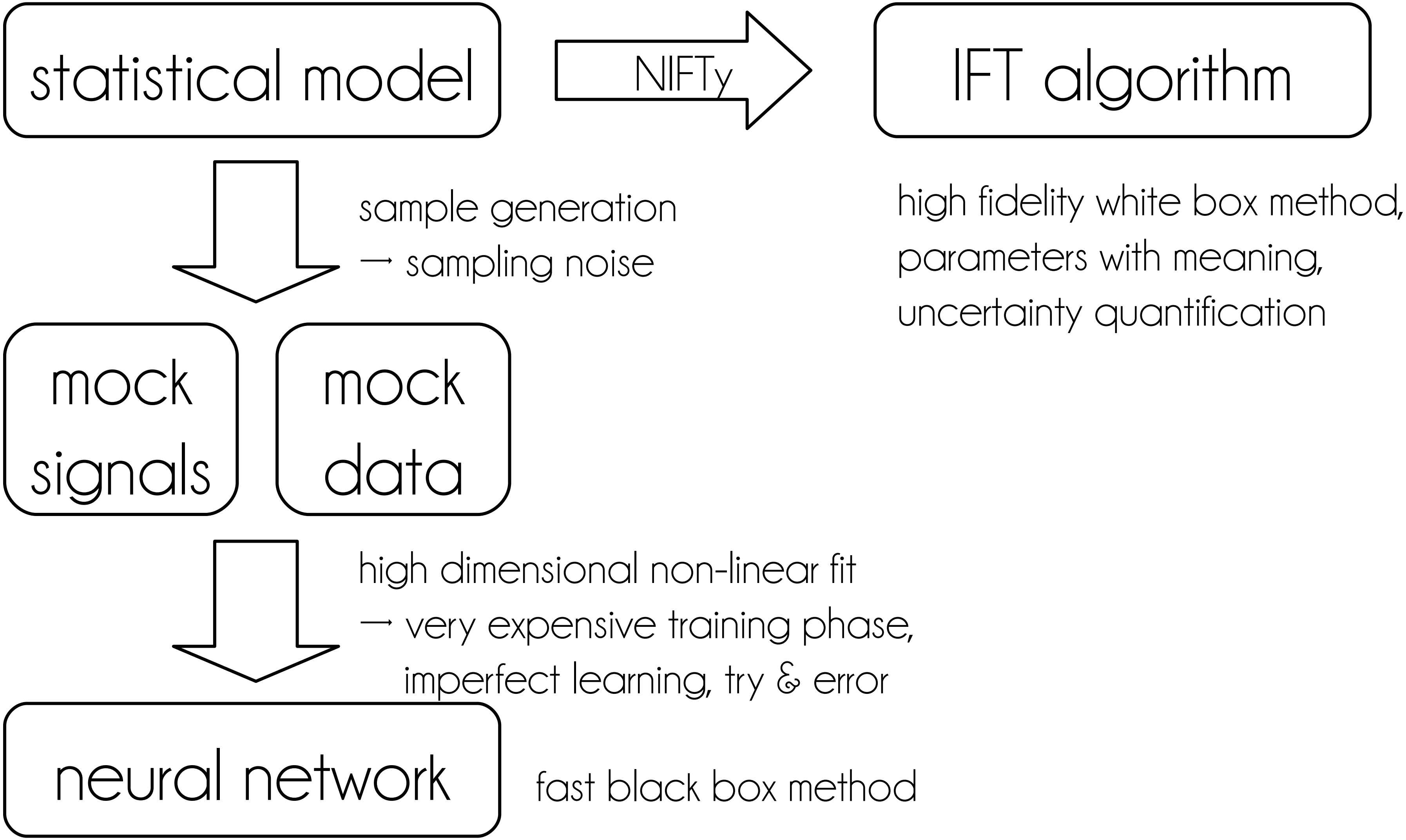


ground truth / starblade



ground truth / autoencoder

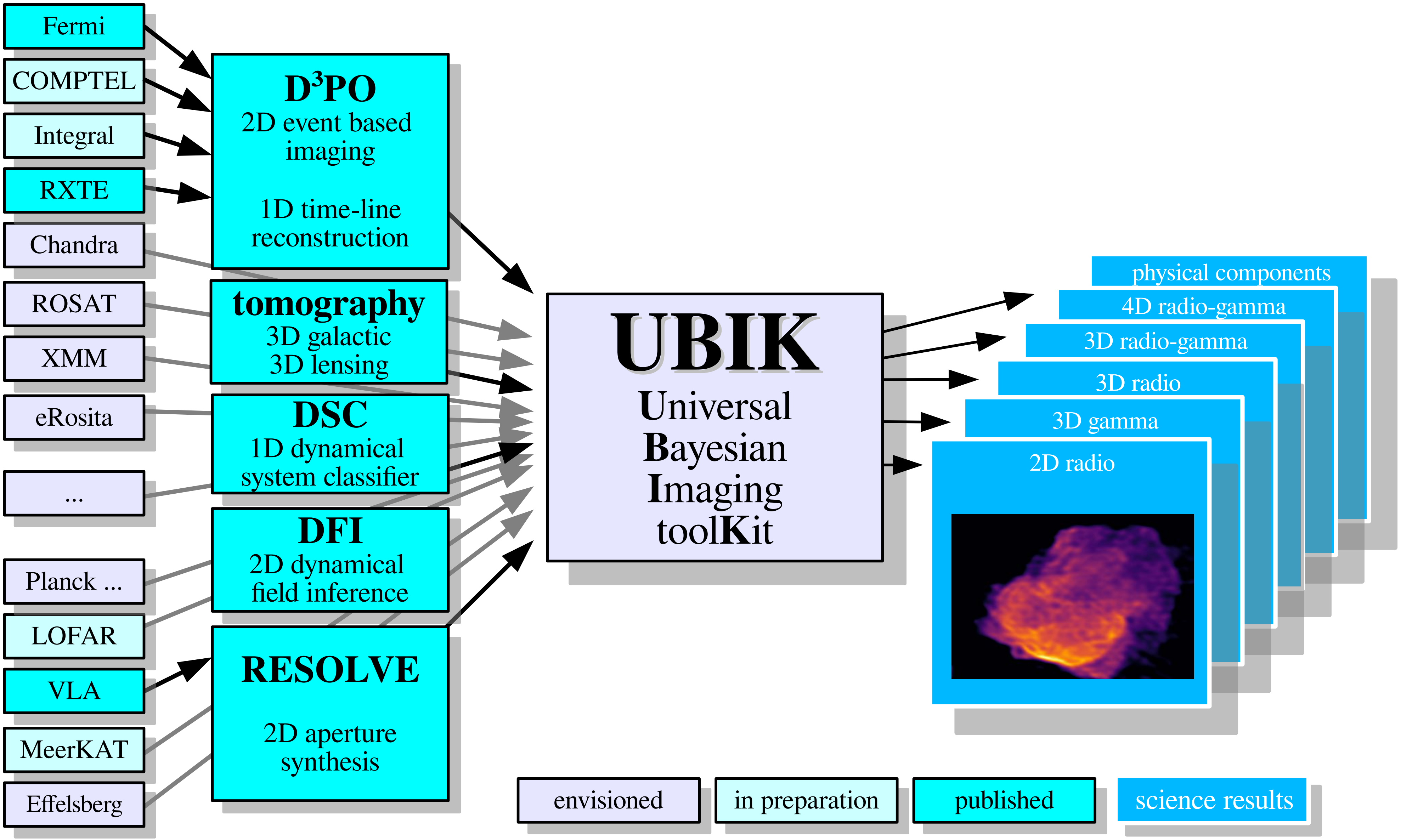




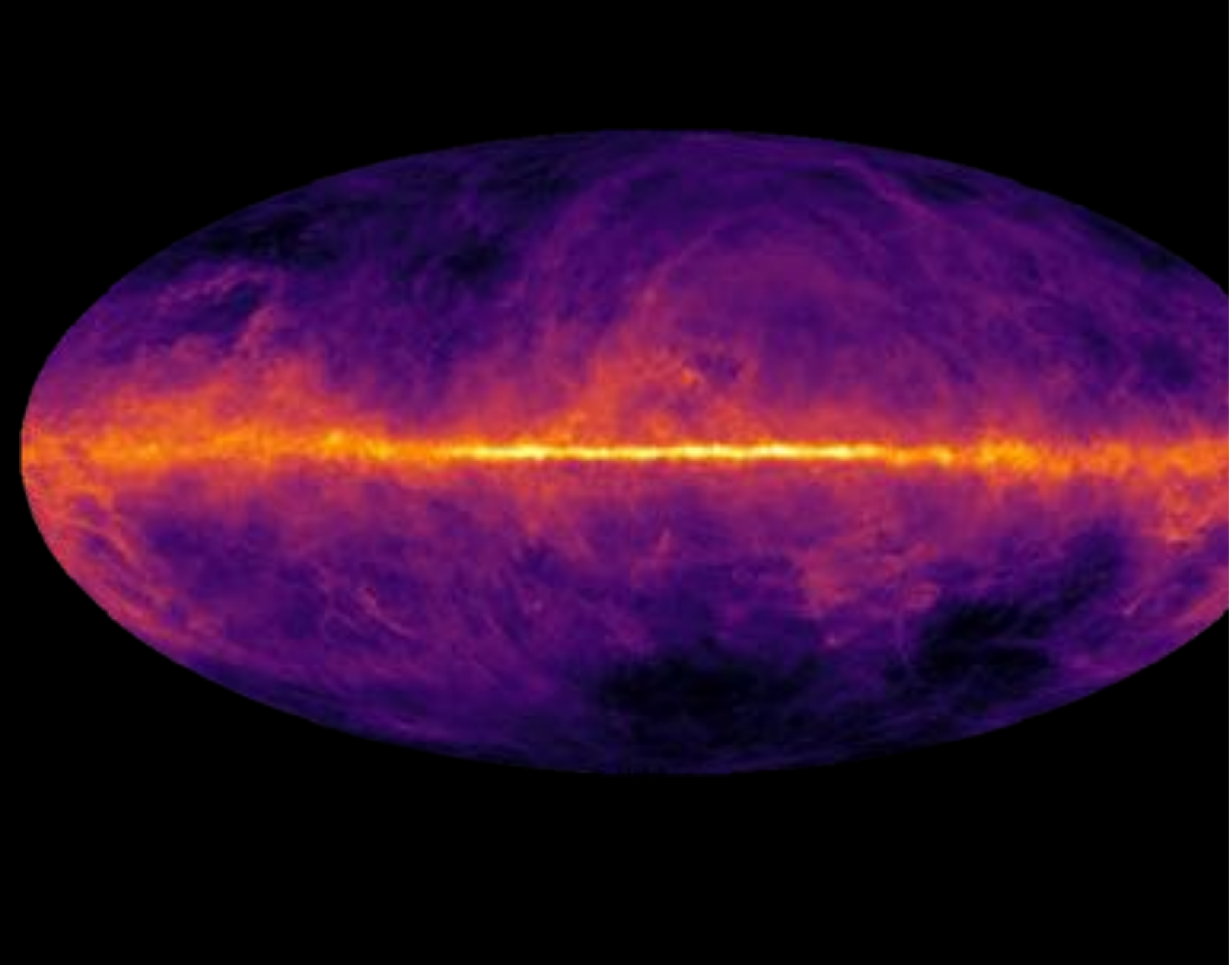
statistical model
NIFTy
IFT algorithm
high fidelity white box method,
parameters with meaning,
uncertainty quantification
sample generation
→ sampling noise
mock signals
mock data
high dimensional non-linear fit
→ very expensive training phase,
imperfect learning, try & error
neural network
fast black box method

NIFTy tutorial part 2

nonlinear reconstructions



Fermi
COMPTEL
Integral
RXTE
Chandra
ROSAT
XMM
eRosita
...
Planck ...
LOFAR
VLA
MeerKAT
Effelsberg
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



Thank you!

www.mpa-garching.mpg.de/ift