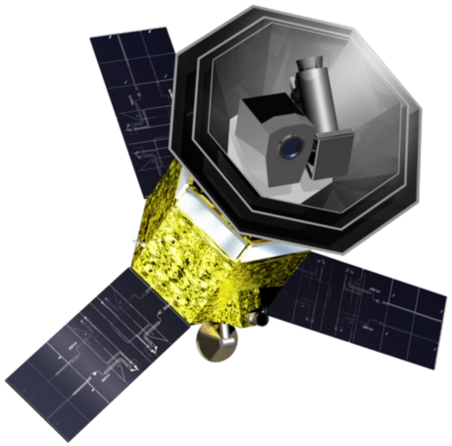
Analysis techniques for CMB B-mode experiments

Benjamin D. Wandelt

with Marius Millea, Ethan Anderes, Doogesh Kodi Ramanah, Justin Alsing, Tom Charnock, Suvodip Mukherjee, Guilhem Lavaux, Stephen Feeney…

The first challenge:

design and
construct CMB
experiment!



(with data analysis
in mind)

Challenges in B-mode experiment data analysis

1. Systematics

2. Systematics

3. Systematics

Challenges in B-mode experiment data analysis

1. Instrumental and observational systematics

2. Secondary anisotropies

3. Foregrounds

Challenges in B-mode experiment data analysis

1. E and B mode mixing due to observational and instrumental effects

2. Lensing

3. Non-Gaussian, anisotropic data model, e.g. due to foregrounds.

+ a bonus at the end if there is time…

Optimal B-mode filtering and purification with anisotropic and correlated noise

with Doogesh Kodi Ramanah, Guilhem Lavaux

How to purify a B-mode map

- E and B only separate easily on the full sky.
- It was thought that optimally separating E and B modes on an incomplete and noisy sky requires solving for a special basis of pure E, pure B and ambiguous modes (e.g. Bunn, Zaldarriaga, Tegmark de Oliveira-Costa 2011)
- In Bunn & Wandelt 2017 (arXiv:1610.03345) we showed that instead it can be cast as the solution to a Wiener filtering problem

Key ingredient: optimal, Wiener filtering

$$d = s + n$$

Not sparse in any easily accessible basis

$$\boxed{\left(\mathbf{S}^{-1} + \mathbf{N}^{-1}\right)} s_{\mathrm{WF}} = \mathbf{N}^{-1} d$$

Sparse in Fourier space

Sparse in pixel space

Usual solution strategies

- Iterative conjugate gradients (optimal for SPD matrices)
- Preconditioner (diagonal in Fourier space, multi-grid,…) (Smith et al.; Eriksen, Wehus, et al., Seljebotn)
- Problems:
  - Even for Planck: ***extremely ill-conditioned***, condition number $>10^9$
  - Preconditioner not universal
  - Stability issues (Jacobi smoother in multi-grid)

The Messenger method: Wiener Filtering *without* preconditioner

- Introduce auxiliary field (*messenger* field) t with covariance $\mathbf{T}$. Then

$$\left(\bar{\mathbf{N}}^{-1} + (\ \mathbf{T})^{-1}\right) t = \bar{\mathbf{N}}^{-1} d + (\ \mathbf{T})^{-1} s$$
$$\left(\mathbf{S}^{-1} + (\ \mathbf{T})^{-1}\right) s = (\ \mathbf{T})^{-1} t\,,$$

$$\bar{\mathbf{N}} \equiv \mathbf{N} - \mathbf{T}$$

is solved by the Wiener filter.

Can solve each of these equations *exactly* algebraically.

Iterate. Easy to show that this converges and is unconditionally stable.

Elsner & BDW, arXiv:1210.4931

Kodi-Ramanah, Lavaux & BDW, arXiv:1702.08852

The Messenger method: Wiener Filtering *without* preconditioner

- Introduce auxiliary field (*messenger* field) t with covariance $\mathbf{T}$ and a parameter $\lambda \geq 1$. Then

$$\left(\bar{\mathbf{N}}^{-1} + (\boxed{\lambda}\mathbf{T})^{-1}\right) t = \bar{\mathbf{N}}^{-1} d + (\boxed{\lambda}\mathbf{T})^{-1} s$$
$$\left(\mathbf{S}^{-1} + (\boxed{\lambda}\mathbf{T})^{-1}\right) s = (\boxed{\lambda}\mathbf{T})^{-1} t\,,$$

$$\bar{\mathbf{N}} \equiv \mathbf{N} - \mathbf{T}$$

is solved by the Wiener filter **for λ=1**.

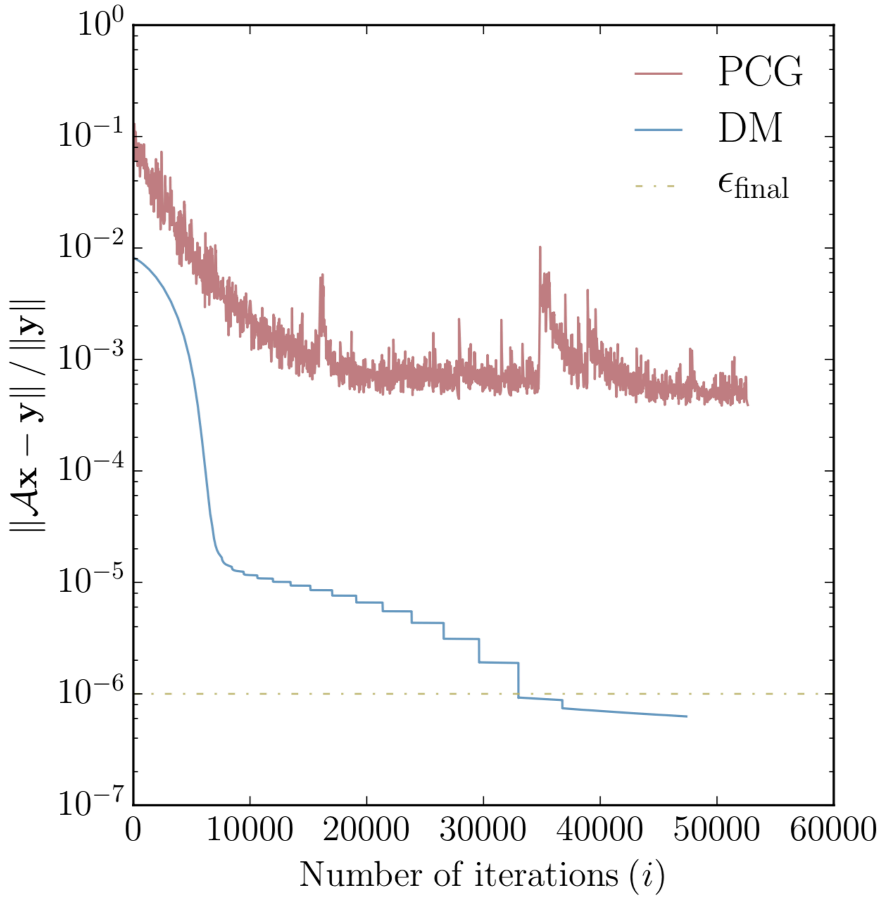
Can solve each of these equations *exactly* algebraically.

Iterate. Easy to show that this converges and is unconditionally stable.

Elsner & BDW, arXiv:1210.4931
Kodi-Ramanah, Lavaux & BDW, arXiv:1702.08852

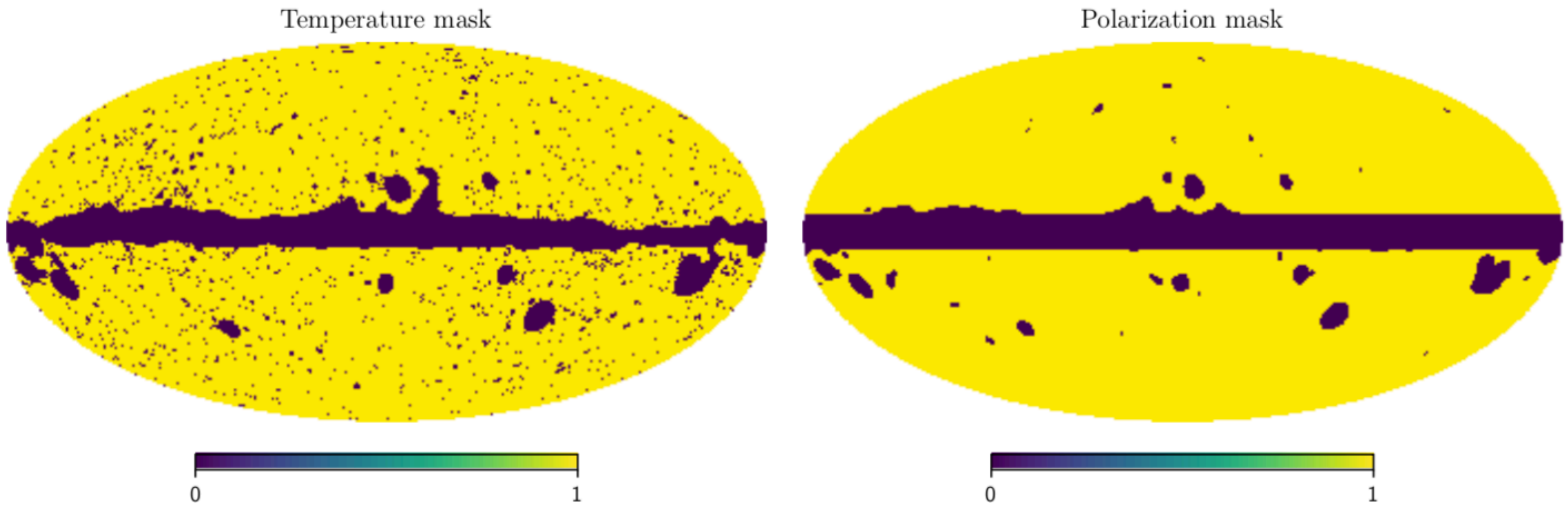
Messenger is fast and unconditionally convergent

Residual



Kodi-Ramanah, Lavaux & Wandelt 2017, arXiv:1702.08852

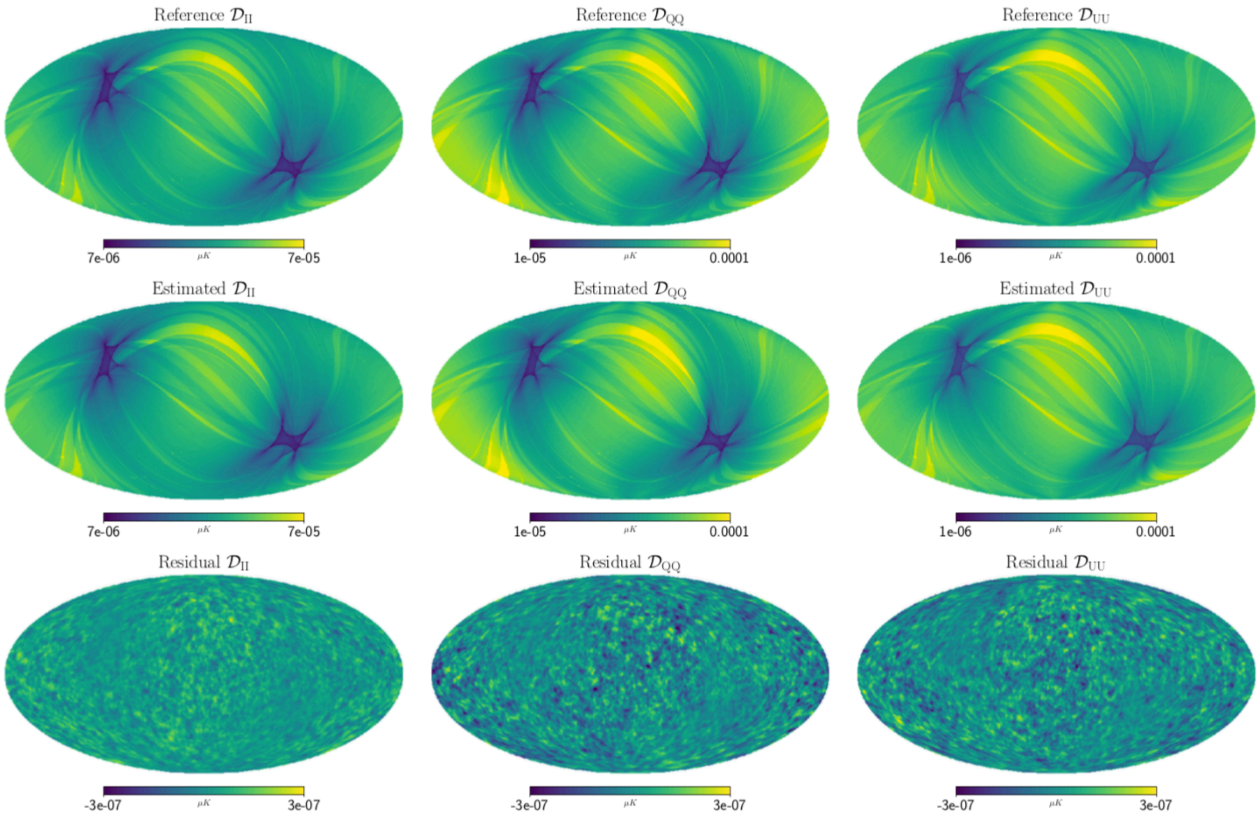
Curved sky case, with IQU, masks, correlated and anisotropic noise



Kodi Ramanah, Lavaux & Wandelt, (arXiv:1906.10704)

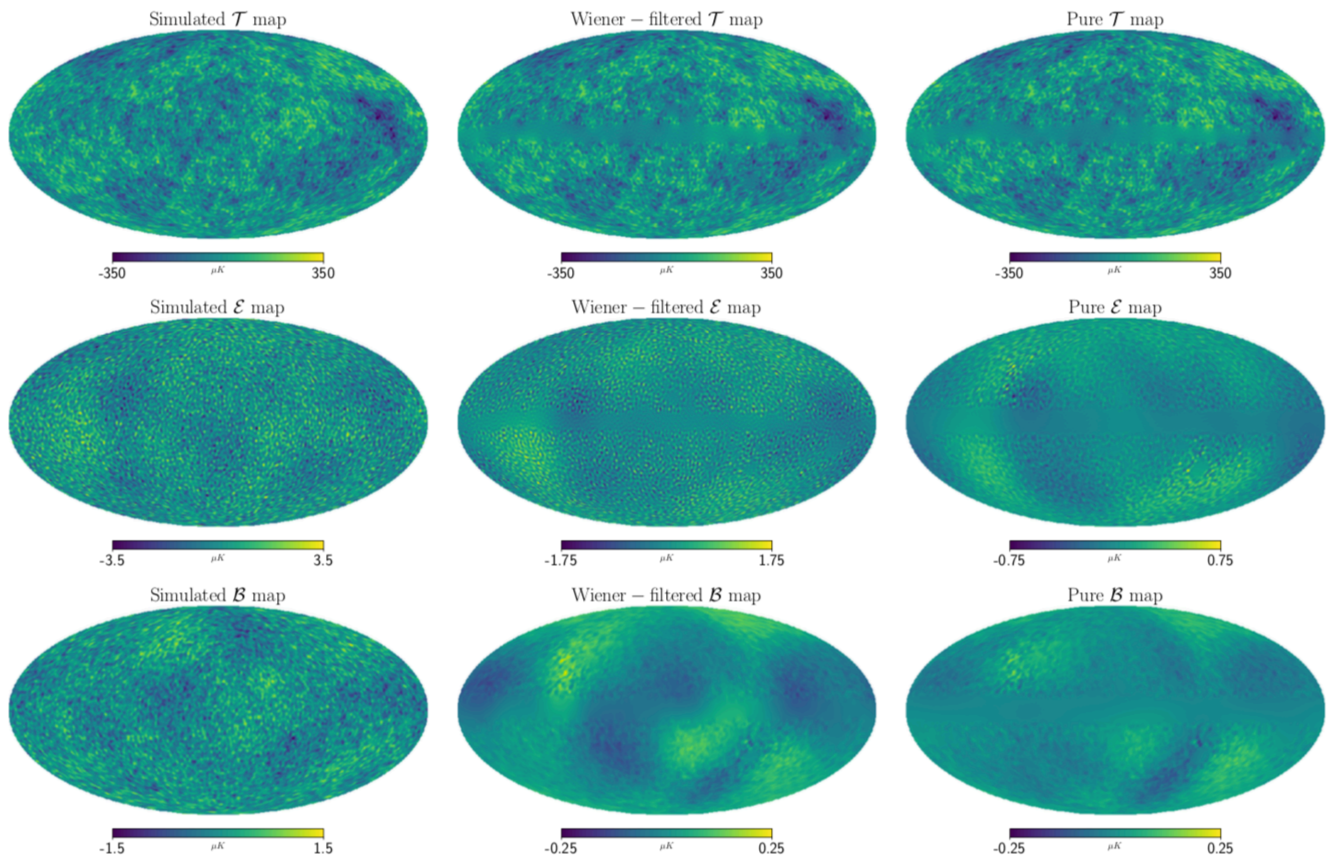
Fast inference of an anisotropic, correlated noise model from noise simulations

After 5 iterations!



Kodi Ramanah, Lavaux, Wandelt, (arXiv:1906.10704)

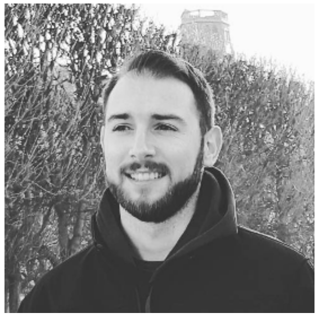
Optimal filtering and T, E, B purification

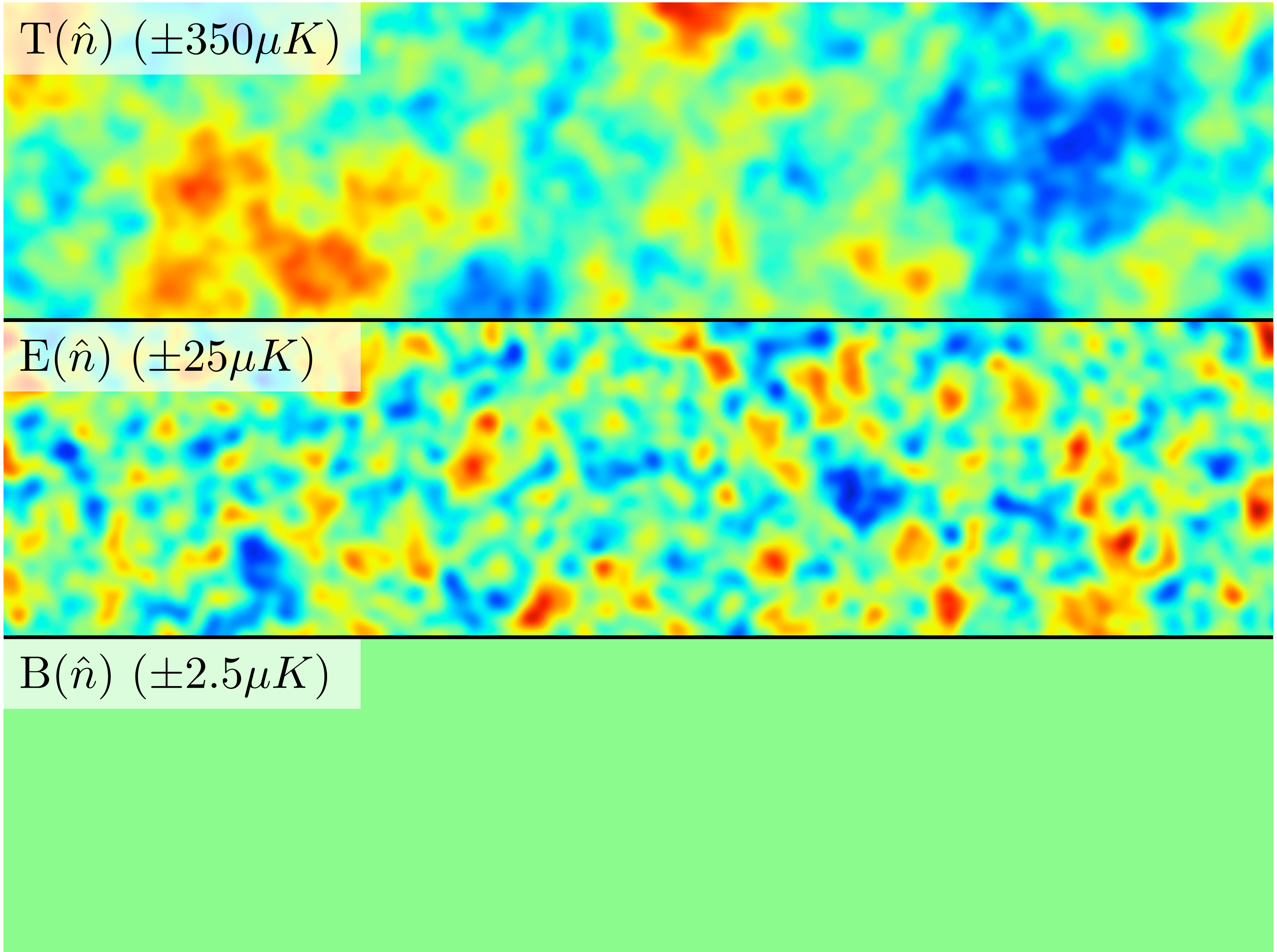


Kodi Ramanah, Lavaux, Wandelt, (arXiv:1906.10704)

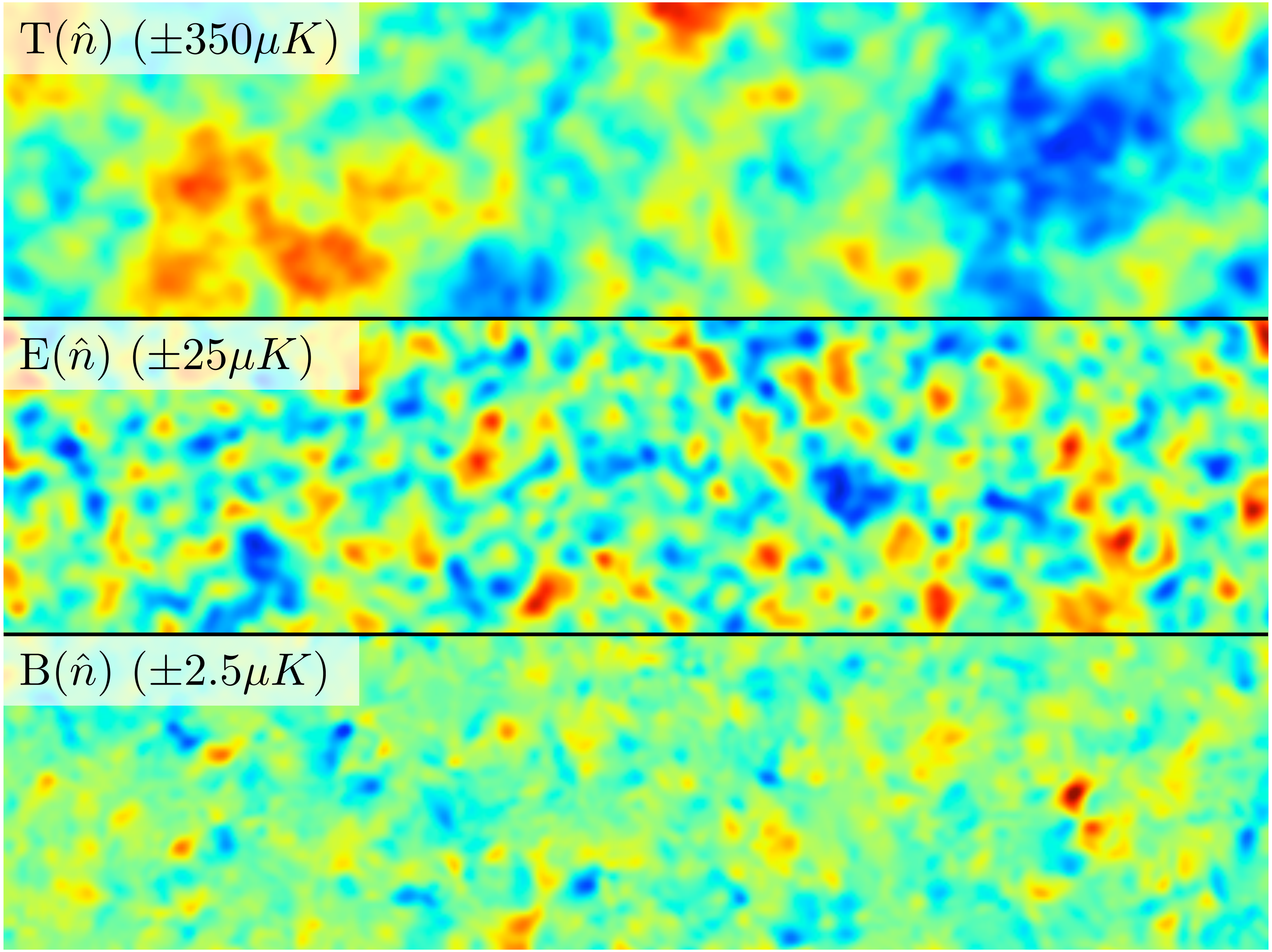
B-mode Inference and De-Lensing

with Marius Millea and Ethan Anderes

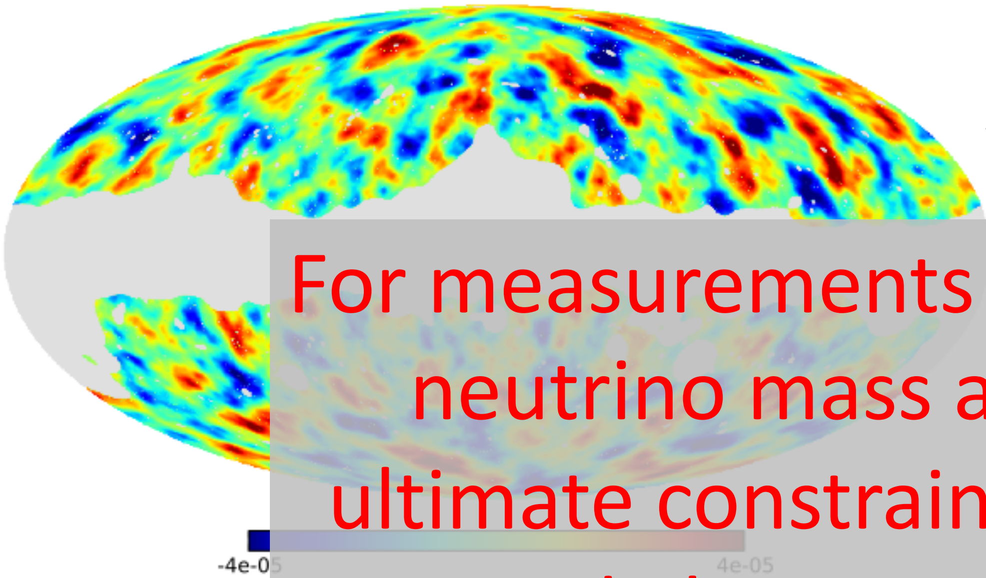




$\mathrm{T}(\hat{n})$ $(\pm 350\mu K)$
$\mathrm{E}(\hat{n})$ $(\pm 25\mu K)$
$\mathrm{B}(\hat{n})$ $(\pm 2.5\mu K)$

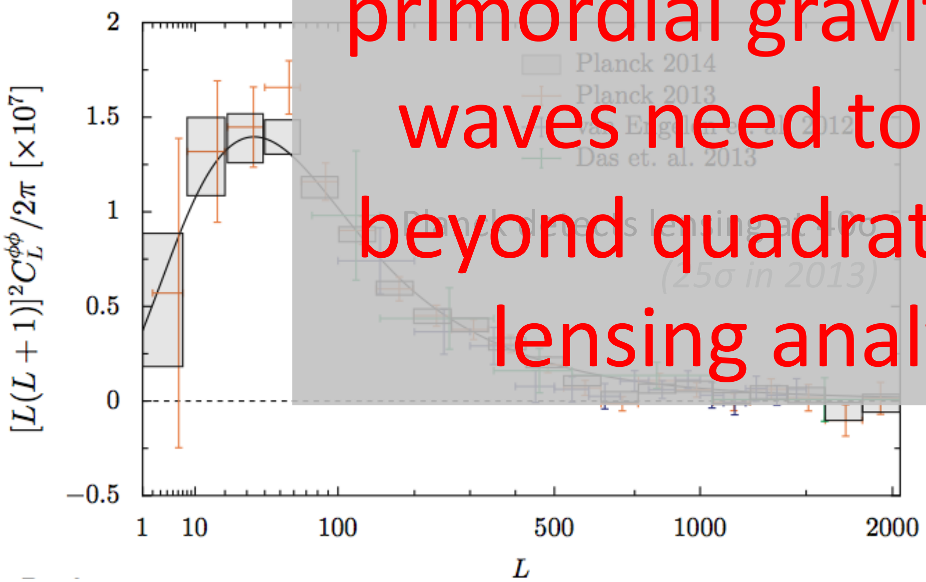


T($\hat{n}$) ($\pm 350\mu K$)
E($\hat{n}$) ($\pm 25\mu K$)
B($\hat{n}$) ($\pm 2.5\mu K$)



All clumping mass in the Universe, as it lenses the CMB

seen using the quadratic estimator



For measurements of the neutrino mass and ultimate constraints on primordial gravitational waves need to do go beyond quadratic CMB lensing analysis

Bayesian lensing potential reconstruction

- A fully Bayesian approach for non-Gaussian fields would be **optimal**. But studying the posterior pdf for the lensing potential given the data is a "doubly intractable" problem that has remained unsolved for ~20 years.
- All approaches to full Bayesian inference approaches fail in practice
- The key problem is the *lensing determinant*

Hirata & Seljak (2003); Anderes, Wandelt & Lavaux (2015); Carron & Lewis (2017)

A new way to think about weak lensing

Millea, Anderes & Wandelt arXiv:1708.06753

Expansion:

$$T(x+\nabla\phi(x)) = \left[\sum_{i=0}^{N} \frac{1}{i!} \left[\nabla\phi(x)\right]^{i} \nabla^{i}\right] T(x)$$

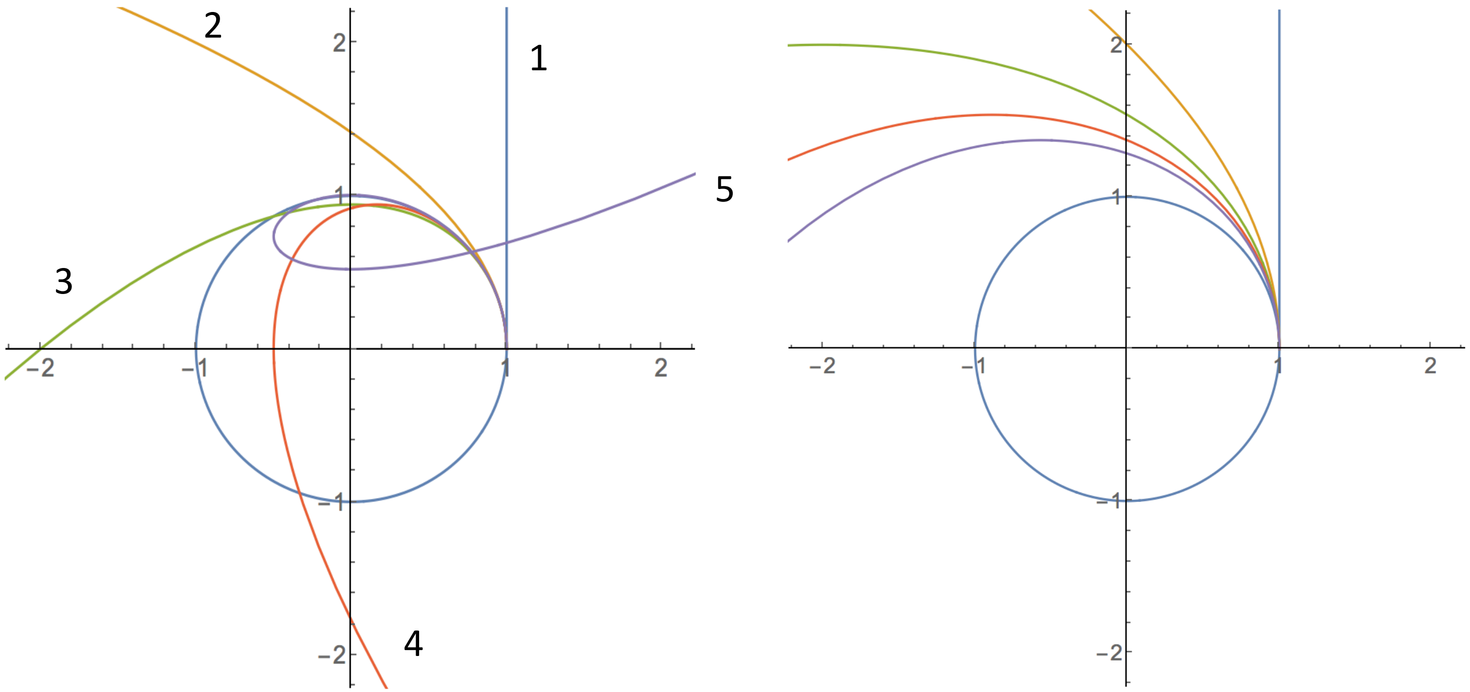
LensFlow:

$$T(x+\nabla\phi(x)) = \left[\prod_{i=0}^{N} \left(1+\frac{1}{N} p_i \nabla \ \right)\right] T(x)$$

Conserves the lensing determinant!

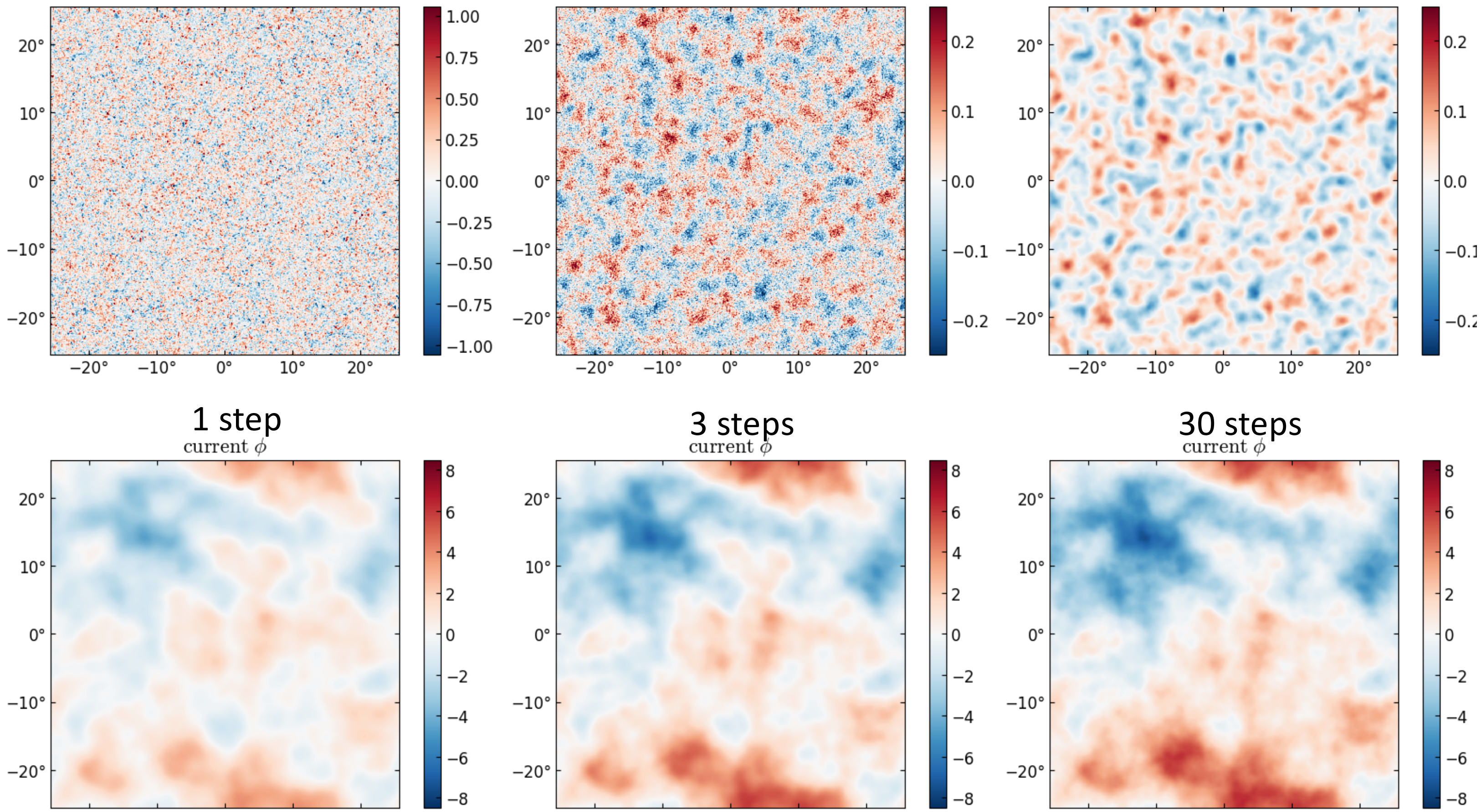
$$p_t^i = (\nabla^j \phi)(M_t^{-1})^{ji}$$

Taylor and “Flow” approximations to rotations



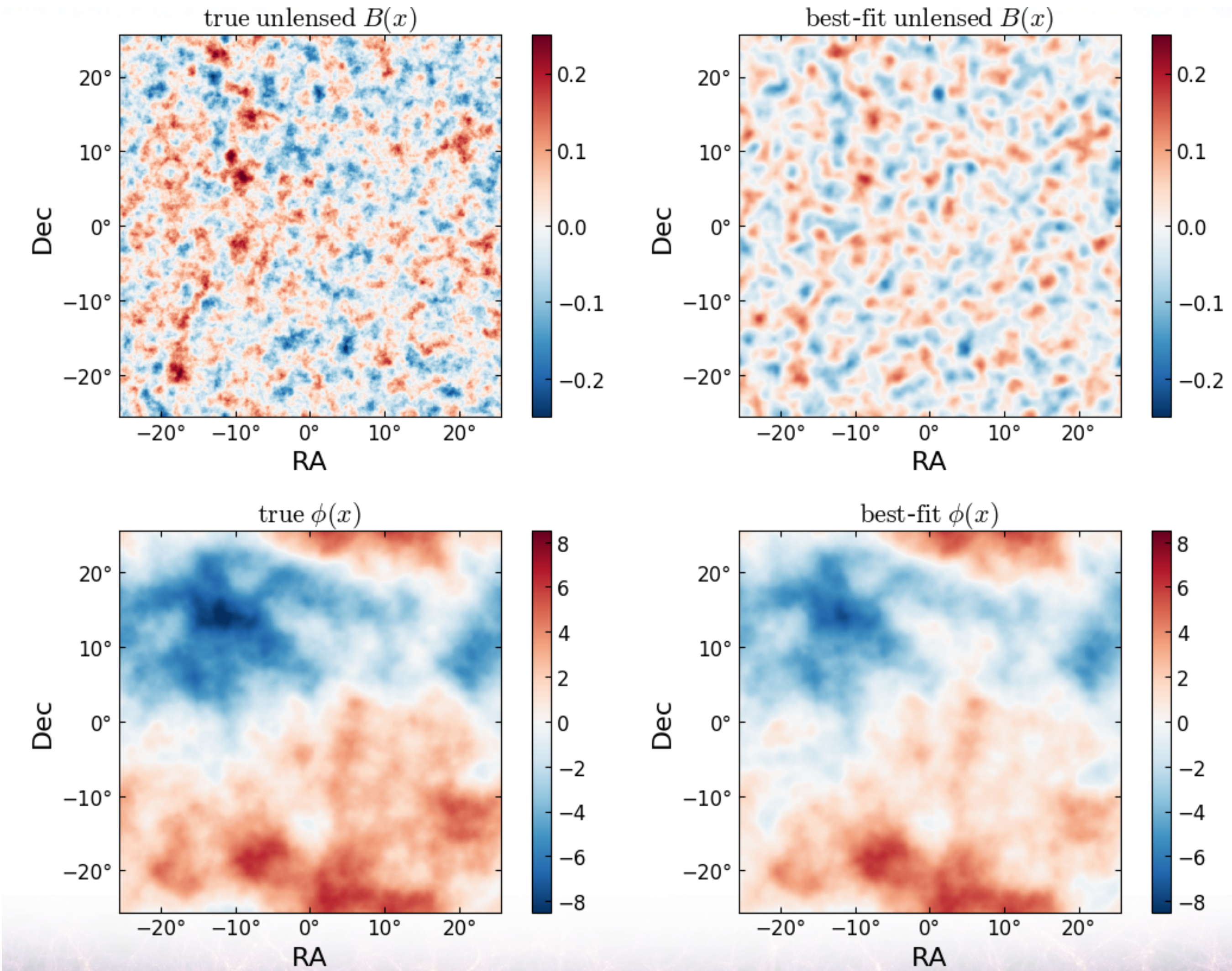
Benjamin Wandelt

Example de-lensing



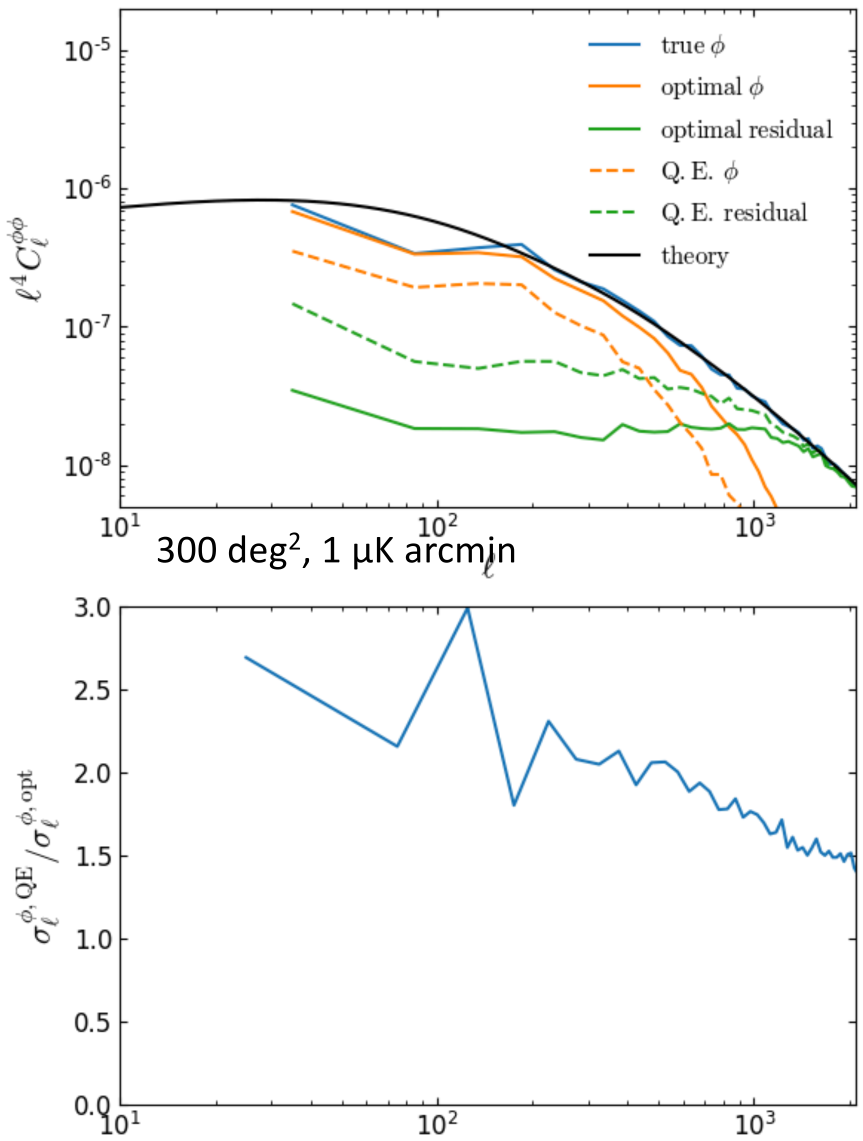
Millea, Anderes & Wandelt arXiv:1708.06753

Example de-lensing



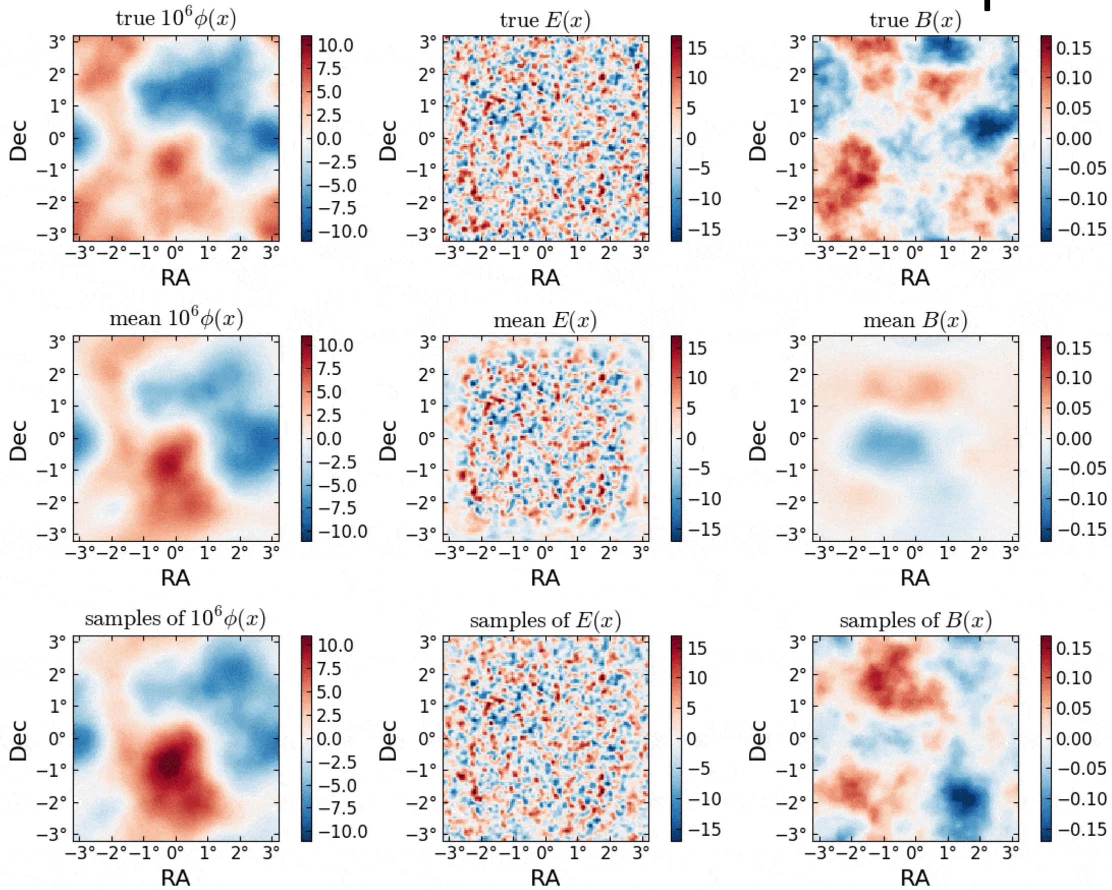
Millea, Anderes & Wandelt arXiv:1708.06753

Millea, Anderes & Wandelt, arXiv:1708.06753



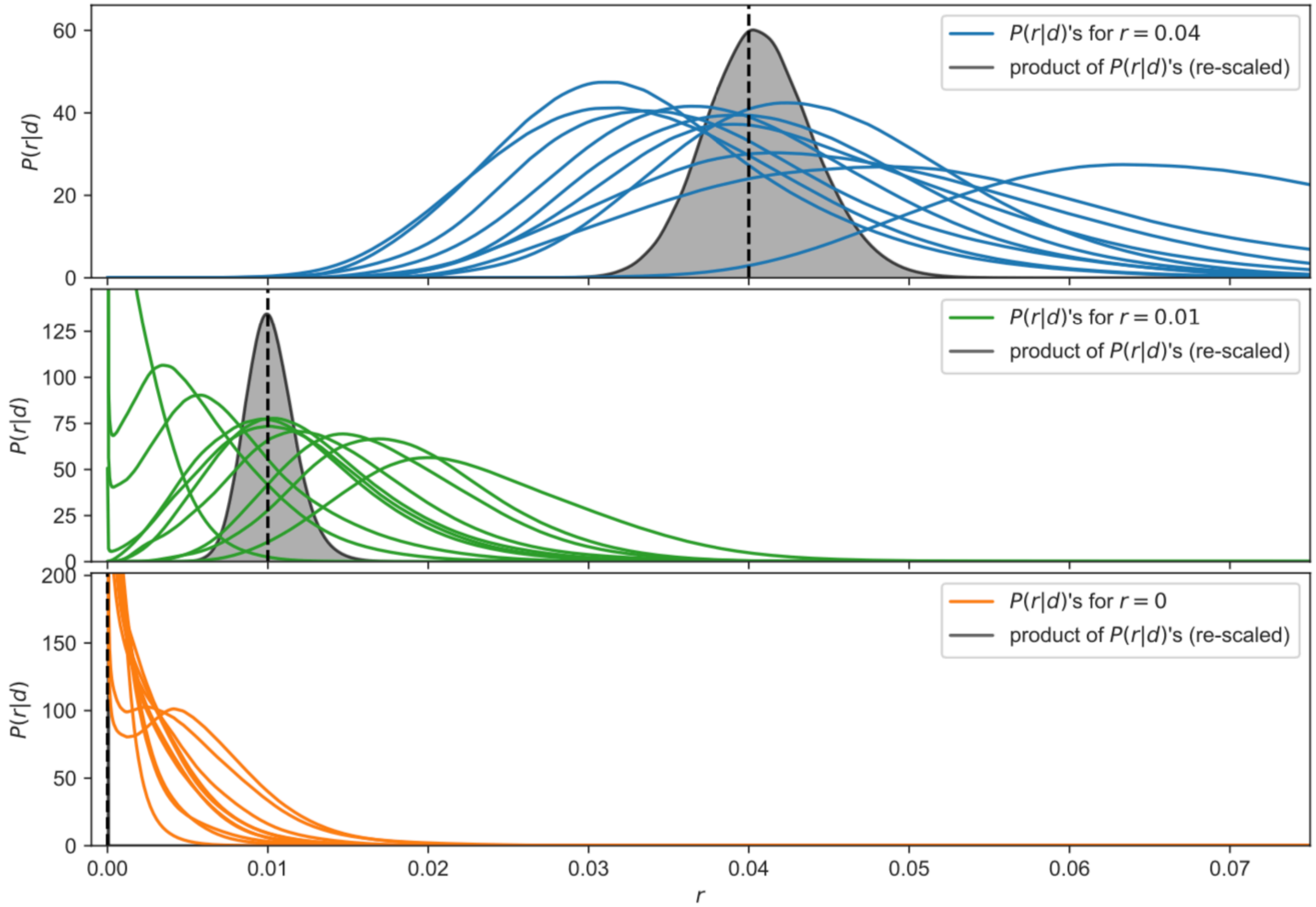
Bayesian lensing inference gives 50-200% improvement over standard quadratic estimator

Full delensing B-mode reconstruction and r inference with masks and anisotropic noise



Millea, Anderes & Wandelt, in prep.

Bayesian approach gives consistent inferences; and reveals strong realization dependence

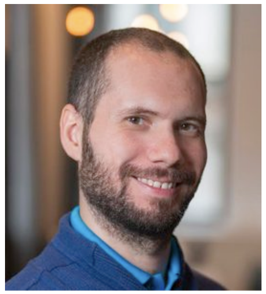


Millea, Anderes & Wandelt, in prep.

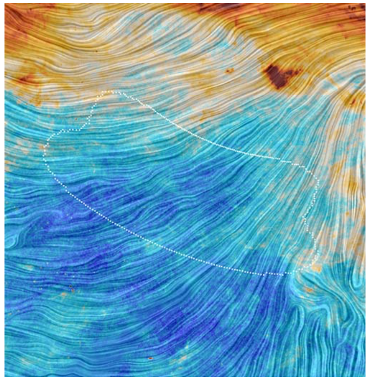
Simulation based (likelihood-free) inference for non-Gaussian data models

with Justin Alsing, Tom Charnock, Stephen Feeney, Francisco Villaescusa, Guilhem Lavaux



How to science, Bayesianly

What if d =  ?

1. Get data.
2. Write down full physical and stochastic model of data given parameter.

   ➔ **Likelihood**

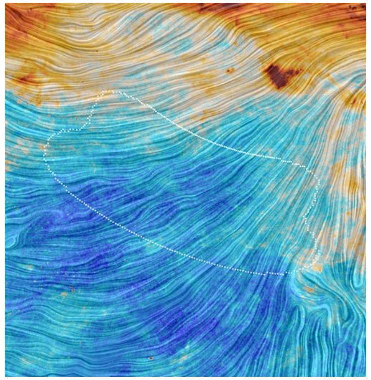
3. Specify prior
4. Write down posterior

$$P(\boldsymbol{\theta}|\mathrm{d}) = \frac{\boxed{P(\mathrm{d}|\boldsymbol{\theta})}\,P(\boldsymbol{\theta})}{P(\mathrm{d})}$$

5. Explore posterior for fixed data as a function of parameters

How to science, Bayesianly

What if d =



?

1. Get data.
2. Write down full physical and stochastic model of data given parameter.

   ➔ **Likelihood**

3. Specify prior
4. Write down posterior

$$P(\boldsymbol{\theta}|\mathrm{d}) = \frac{\cancel{P(\mathrm{d}|\boldsymbol{\theta})}\,P(\boldsymbol{\theta})}{P(\mathrm{d})}$$

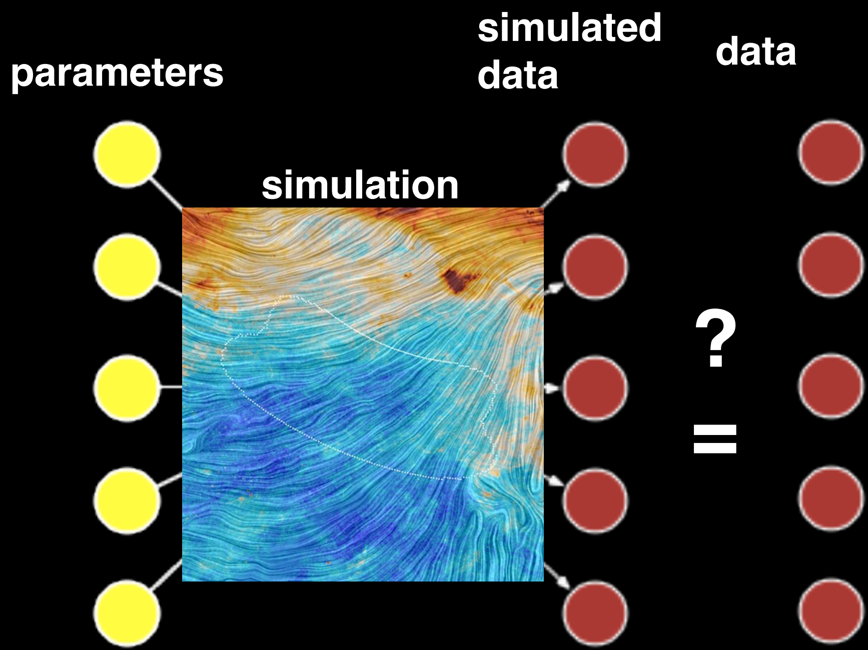
5. Explore posterior for fixed data as a function of parameters

What if we can only do simulations?

$$P(\boldsymbol{\theta}|\mathbf{d}) = \frac{P(\mathbf{d}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{d})}$$

$$\mathbf{d}^* \leftarrow \mathrm{simulation}(\mathbf{d}^*|\boldsymbol{\theta})$$

Likelihood-free inference 101



Draw from prior:

$$\boldsymbol{\theta} \leftarrow P(\boldsymbol{\theta})$$

Simulate data:

$$\mathbf{d}^* \leftarrow P(\mathbf{d}^*|\boldsymbol{\theta})$$

If $\rho(\mathbf{d}^*, \mathbf{d}) < \epsilon$

accept;

else:

reject;

In the limit $\epsilon \rightarrow 0, \ \{\boldsymbol{\theta}\} \leftarrow P(\boldsymbol{\theta}|\mathbf{d})$

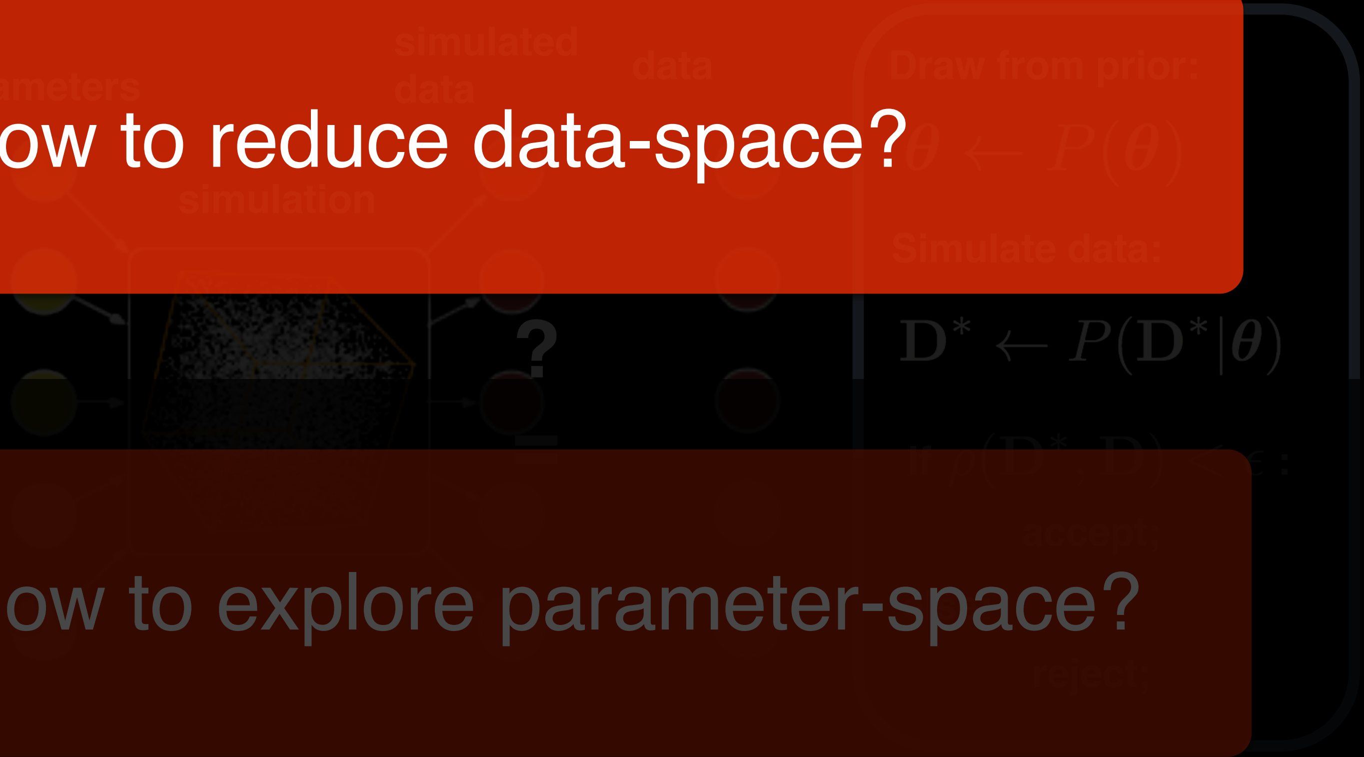
Likelihood-free inference 101

How to reduce data-space?

How to explore parameter-space?

Likelihood-free inference 101

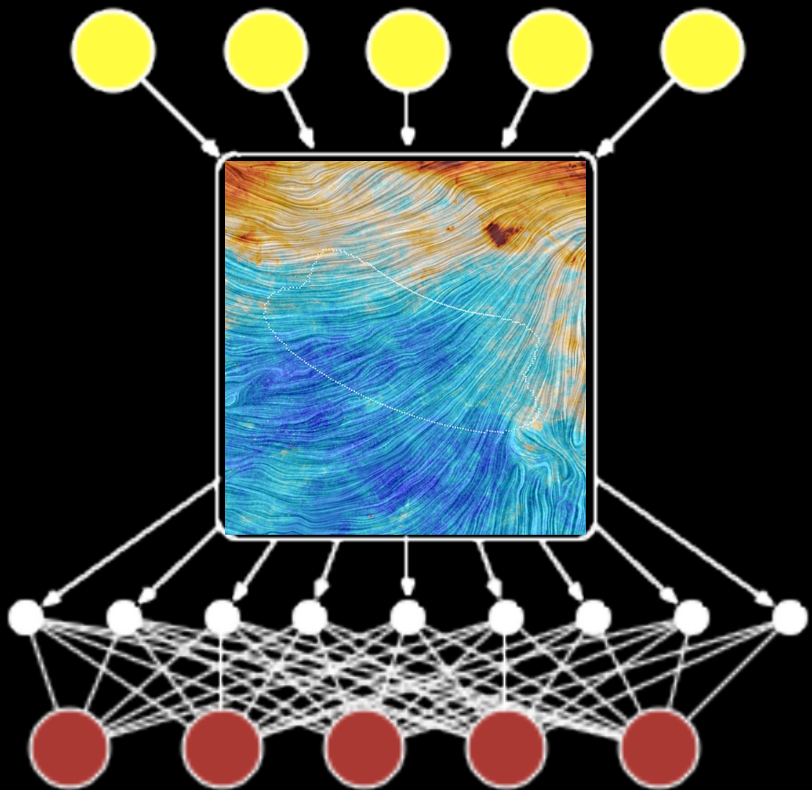
How to reduce data-space?



How to explore parameter-space?

In the limit $\epsilon \rightarrow 0, \quad \{\boldsymbol{\theta}\} \leftarrow P(\boldsymbol{\theta}|\mathbf{D})$

Reducing data space: massive data compression



n parameters

N data

M summary statistics

n compressed statistics

Massive data compression

Can we derive *n* compressed quantities that contain all the Fisher information?

Fisher information

$$\mathbf{F} \equiv -\mathbb{E}_{\boldsymbol{\theta}}(\nabla\nabla^T \mathcal{L})$$

Information inequality

$$\mathbb{V}_{\boldsymbol{\theta}}(t_\alpha) \geq [\nabla\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{t})^T \mathbf{F}^{-1} \nabla\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{t})]_{\alpha\alpha}$$

YES! Answer: $\mathbf{t} = \nabla_\theta \mathcal{L}|_{\theta_*}$

Alsing & Wandelt
arXiv:1712.00012

Likelihood-free inference 101

How to reduce data-space?

How to explore parameter-space?

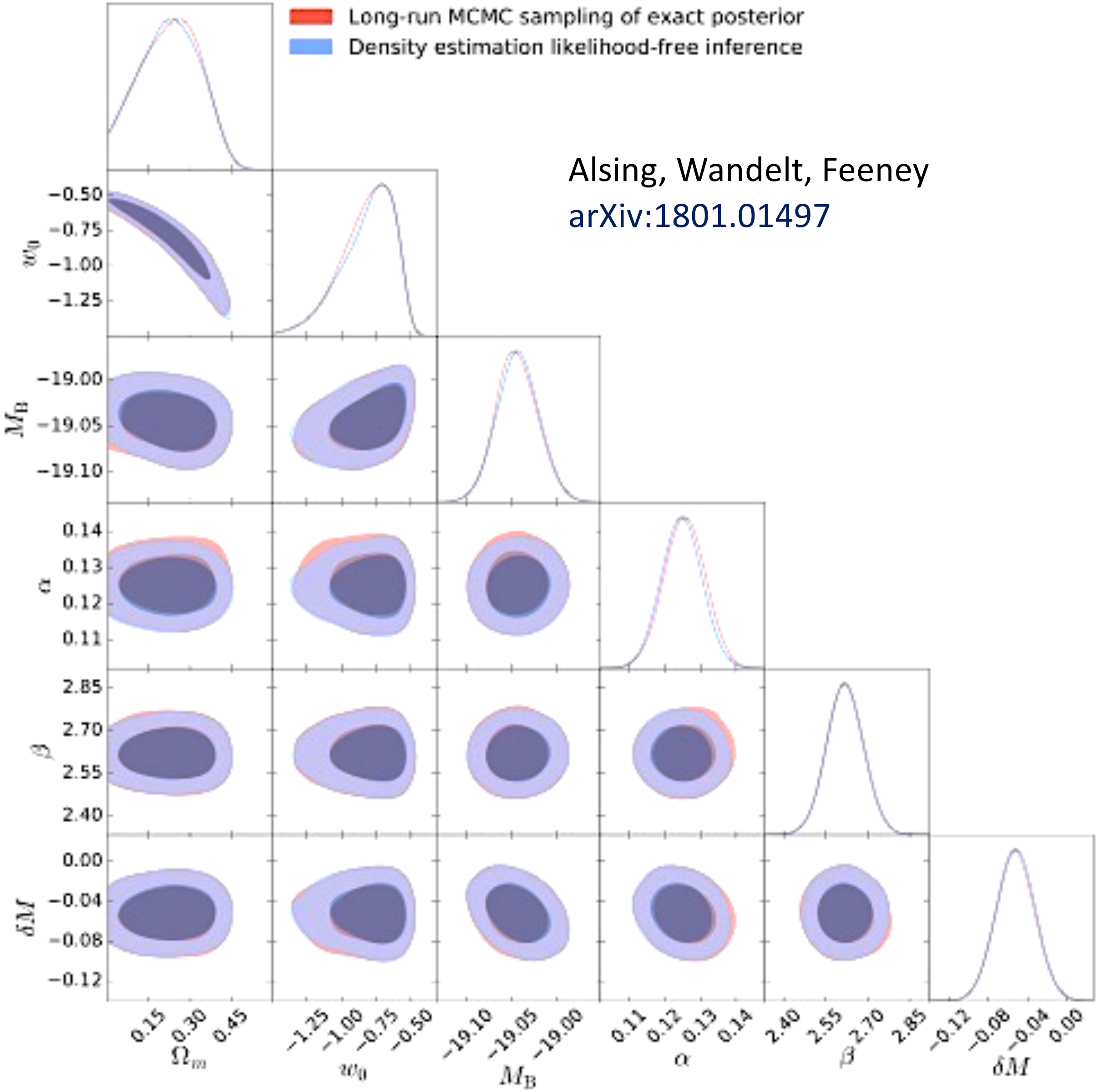
Solution

Density estimation Likelihood free inference (DELFI)

Learn *joint* probability density of parameters and compressed data

Alsing, Feeney &Wandelt arXiv: 1801.01497

DELFI
Posterior
inference
works…
faster than
MCMC!



Alsing, Wandelt, Feeney
arXiv:1801.01497

(O(1000) simulations)

The latest

- New *nuisance-hardened* compression greatly reduces required number of simulations and allows many more parameters (Alsing & Wandelt arXiv:1903.01473).

- New version of DELFI now released including neural density estimators to fit the likelihood (Alsing, Charnock, Feeney, Wandelt arXiv:1903.00007)
    - Also includes active learning for deciding where to run simulations

But what if you don’t know how to compute informative summaries of your data?

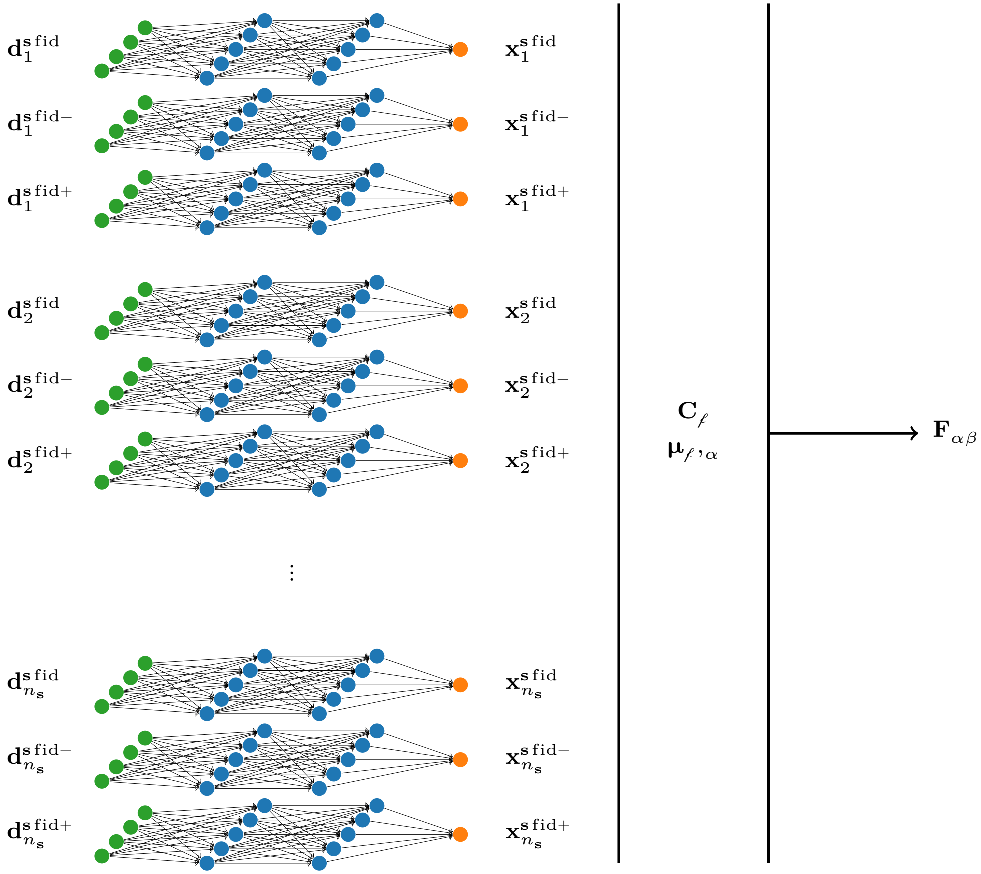
Automatic Physical Inference with Information Maximizing Neural Networks

- Goal: remove the need to “guess” heuristic, informative summaries of the data
- Setup: design a neural network that maps data into a small set of informative *summaries*.
- The training loss is (– the information)

- Use reinforcement learning on physical simulations to maximize the information in the network outputs about the model parameters
- The achieved loss on a test set is meaningful – it’s the information content of the data.

Charnock, Lavaux, Wandelt (arXiv:1802:03537)

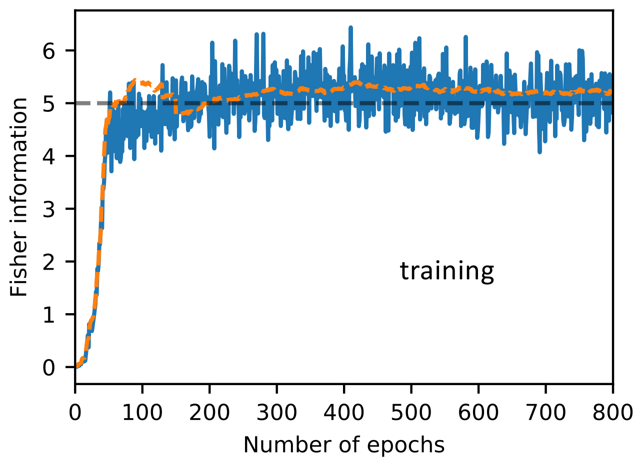
Information maximizing neural network

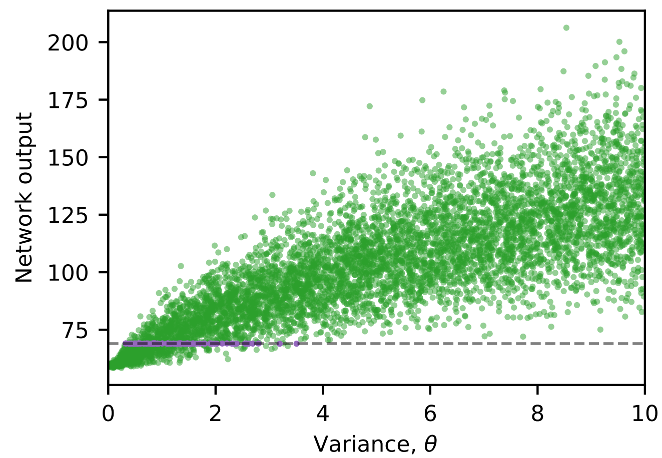


Charnock, Lavaux, Wandelt (arXiv:1802:03537)

Example 1: inference of variance

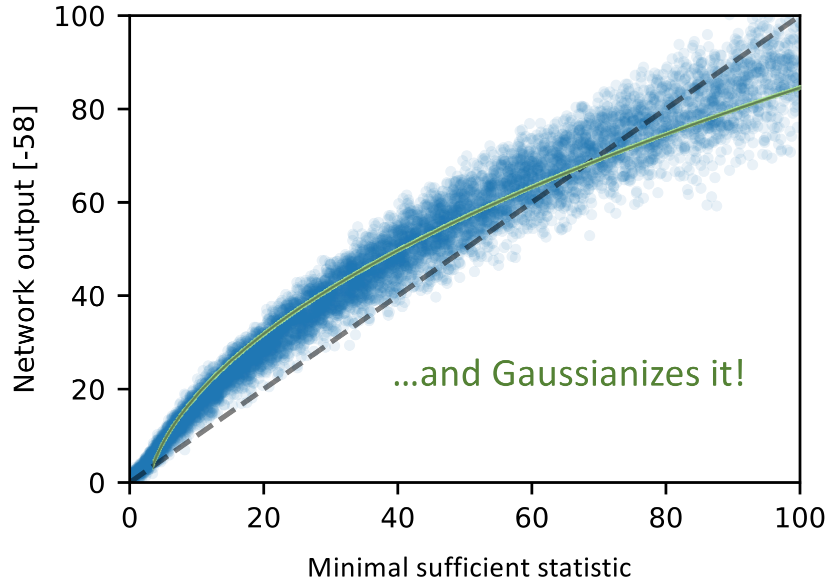
- Perfect information gives $|F| = 5$ in this problem
- Any linear summary gives $|\mathrm{F}| = 0.5$





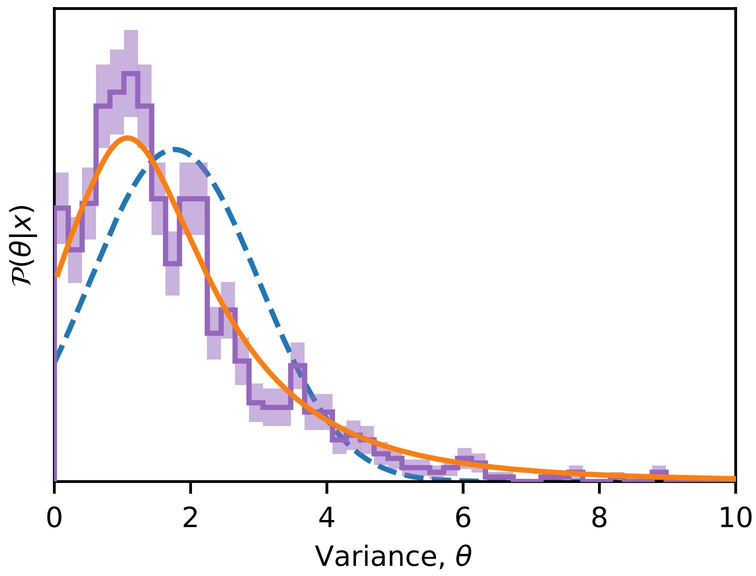
Charnock, Lavaux, Wandelt (arXiv:1802:03537)

The IMNN finds
a minimal
sufficient
statistic for this
inference
problem



Charnock, Lavaux, Wandelt (arXiv:1802:03537)

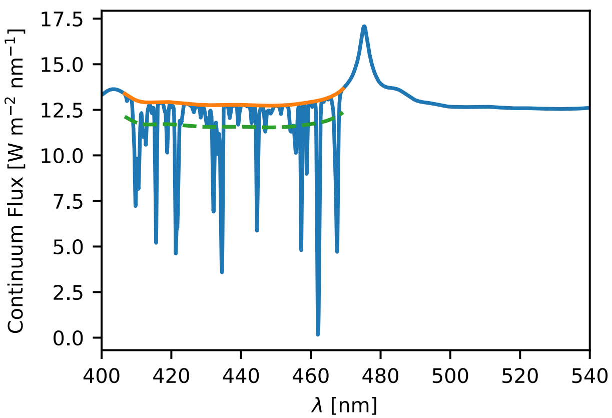
Example 2: Automatic physical inference with unknown noise

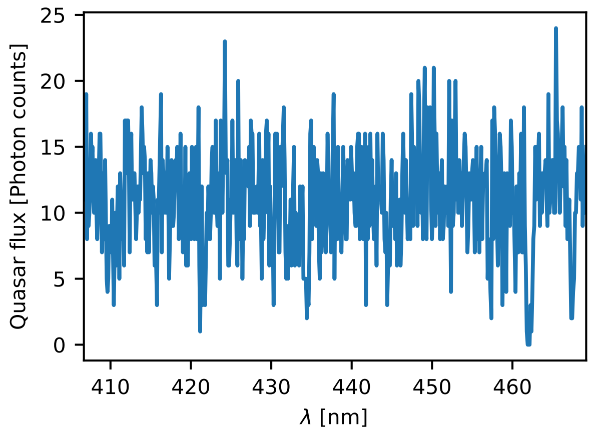


Charnock, Lavaux, Wandelt (arXiv:1802:03537)

Example 3: Lyman-α forest inference

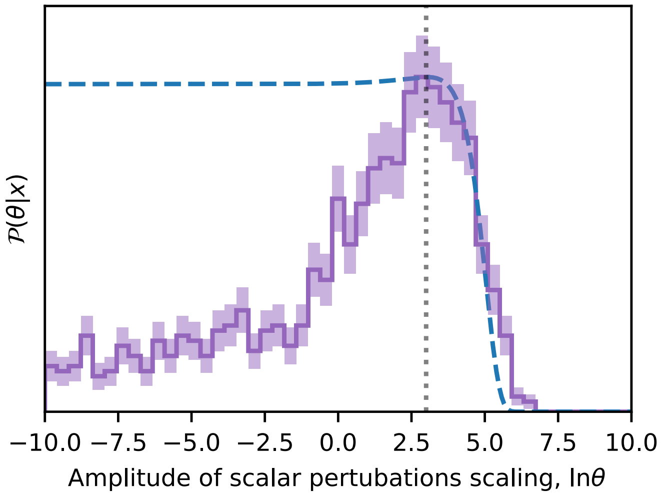
How to infer the variance of the underlying linear density field from a non-linearly transformed, photon-noise dominated Lyman-α forest spectrum?





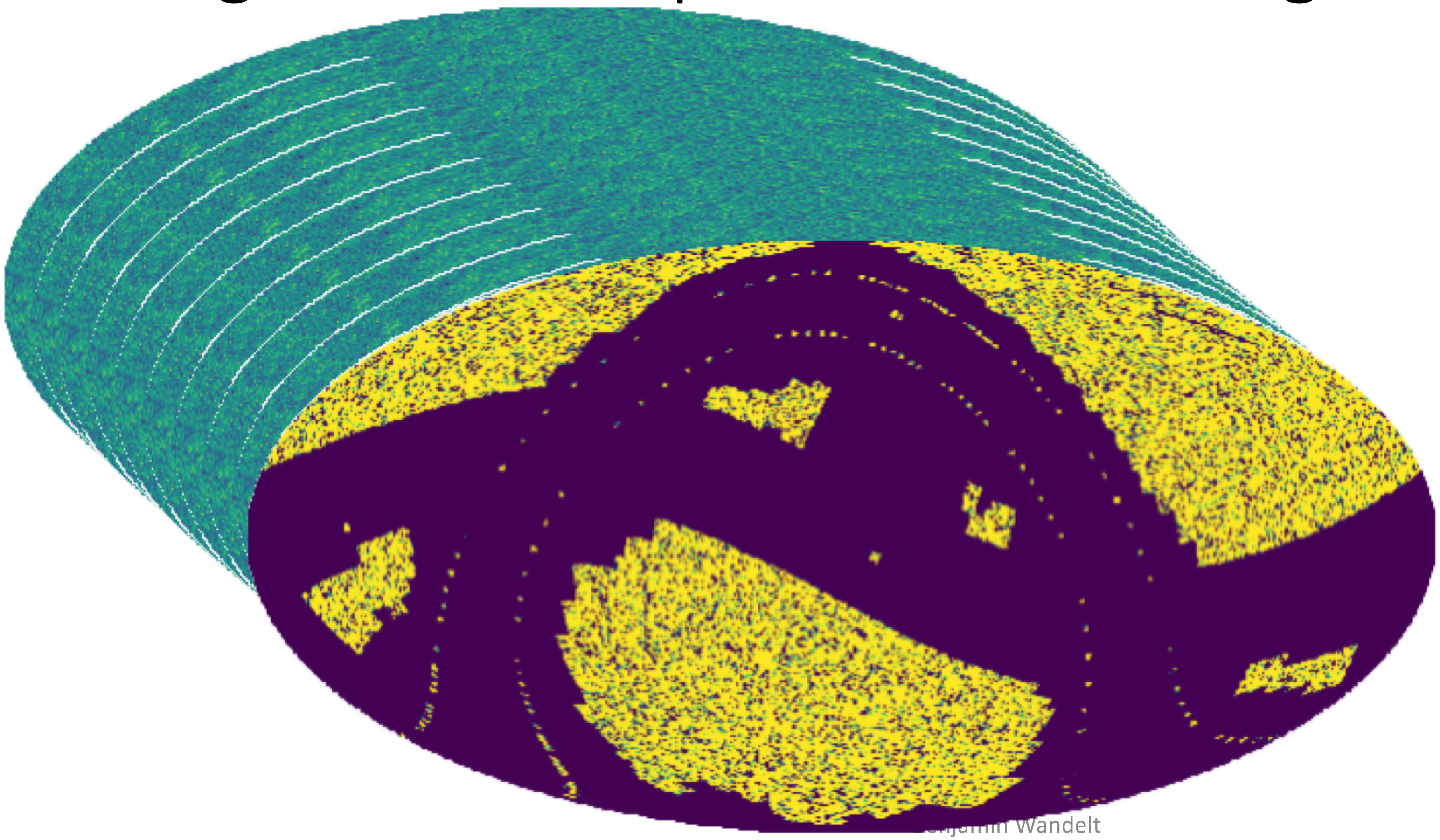
Charnock, Lavaux, Wandelt (arXiv:1802:03537)

Example 3: Lyman-α forest inference



Charnock, Lavaux, Wandelt (arXiv:1802:03537)

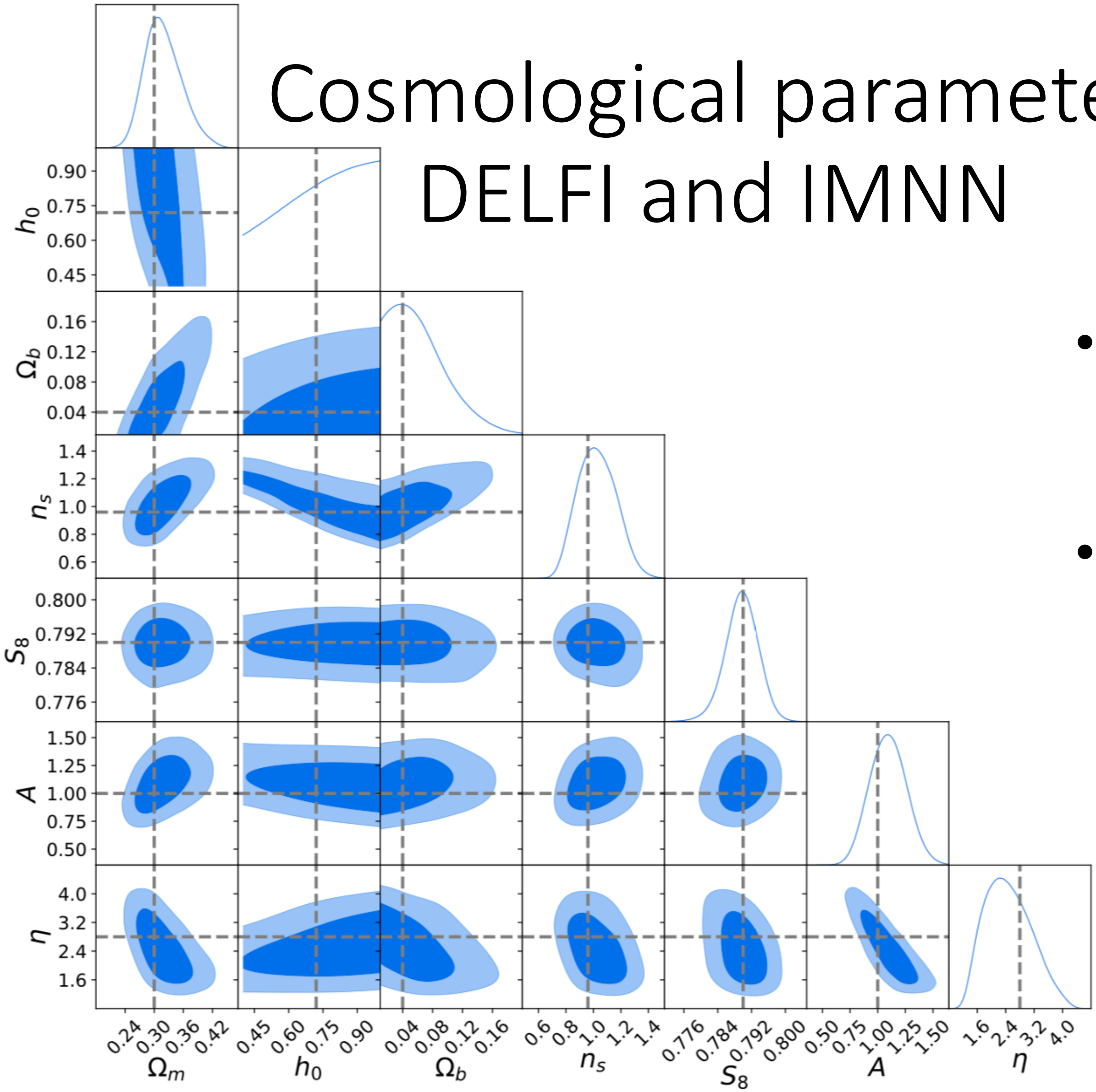
Big Data example: Weak lensing tomography



Mask and noise

10 spherical shells of correlated signal simulations

Cosmological parameter inferences using DELFI and IMNN



- First full **weak lensing analysis with Non-Gaussian lensing potential**
- Enabled by DELFI and IMNN

Taylor, et al, arXiv: 1904.05364

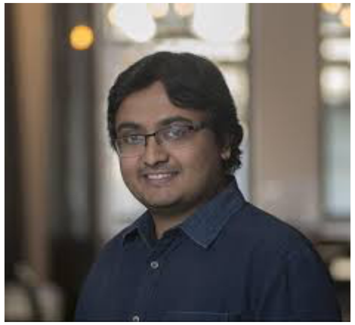
Conclusions

- The hunt for B-modes is on
  - Even with perfect instruments, data models will be complicated, non-linear and non-Gaussian
  - “Low-noise” means “systematics-dominated!”
- I showed advances in three areas
  - Optimal filtering to remove systematics
  - Full modeling and sampling to explore non-Gaussian physical models (lensing)
  - Highly flexible, simulation-based inference fueled by advances in machine-learning – many applications to B-mode science!

+ bonus science target for next-gen CMB polarization experiments…

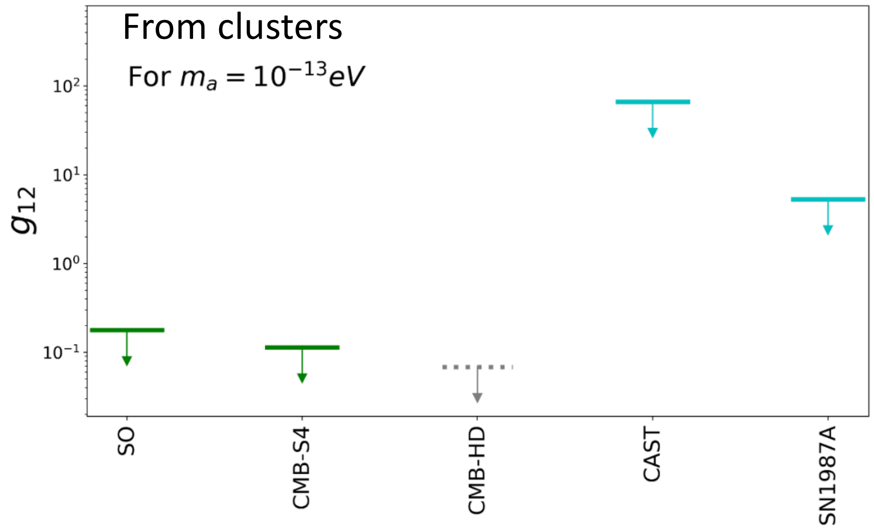
New science target: Axion physics with CMB polarization experiments

with Suvodip Mukherjee, Rishi Khatri, David Spergel



Photons convert to axions in magnetic field

- Resonance conversion of CMB photons to axions in magnetic fields produce a distinctive polarized spectral distortion
- Highly efficient process in magnetic fields of our galaxy or of clusters
- Promises to be a world-leading probe of light axion-like particles!



Mukherjee, Khatri & Wandelt (2018), arXiv:1801.09701

Mukherjee, Spergel, Khatri & Wandelt, arXiv: 1908.07534

To reproduce the results in the IMNN paper the code version used is archived on zenodo

https://doi.org/10.5281/zenodo.1175196

The most current development version is on github:

IMNN:

https://github.com/tomcharnock/IMNN

DELFI:

https://github.com/justinalsing/pydelfi