Machine learning based (likelihood-free) statistical inference

Tom Charnock
Guilhem Lavaux & Ben Wandelt

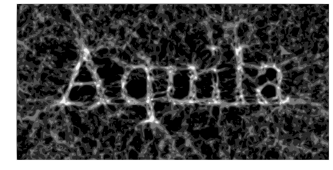
Alessandro Manzotti
Justin Alsing

Institut d'Astrophysique de Paris

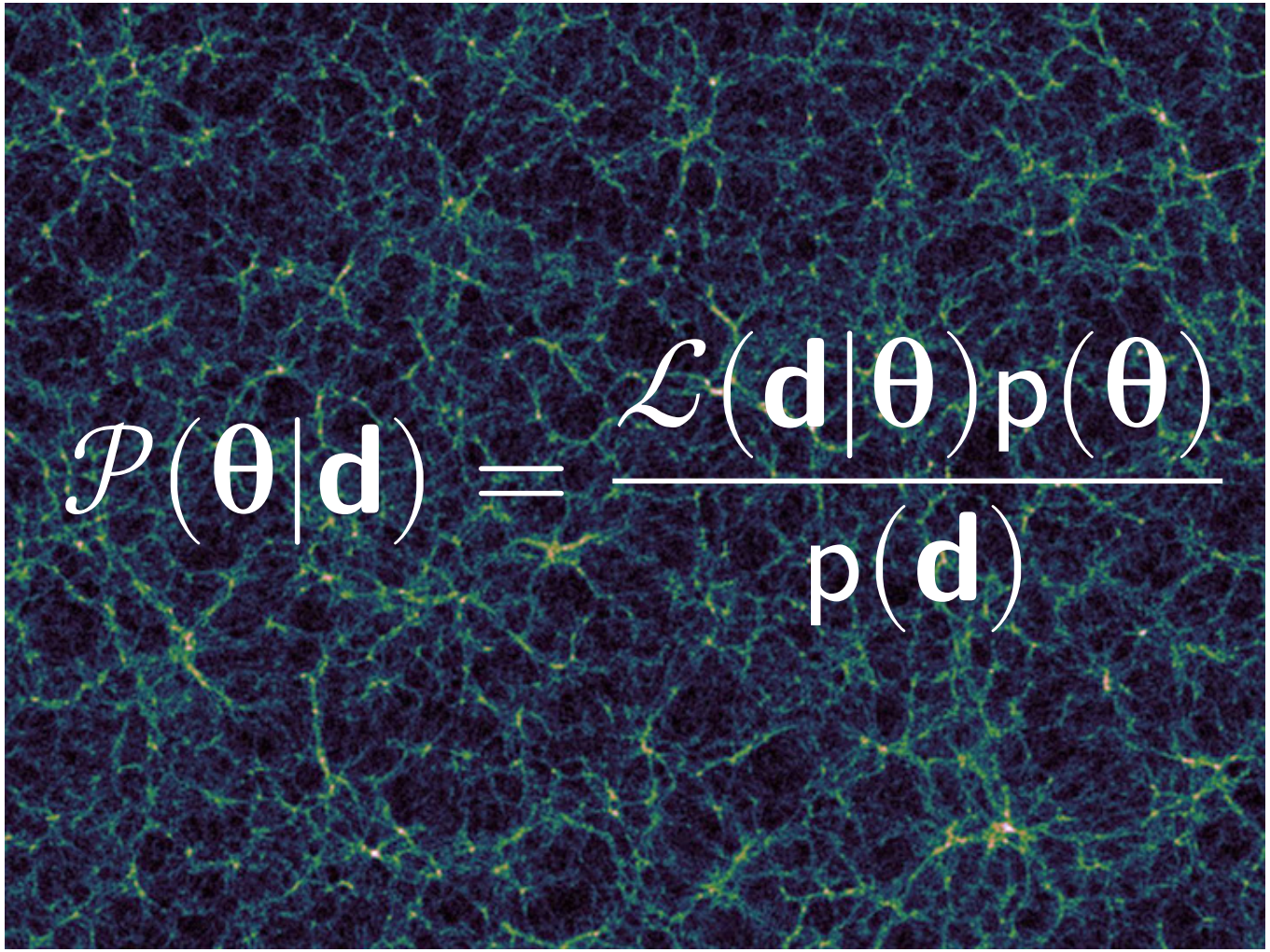
Physical Review D **97**, 083004

DOI 10.5281/zenodo.1175196

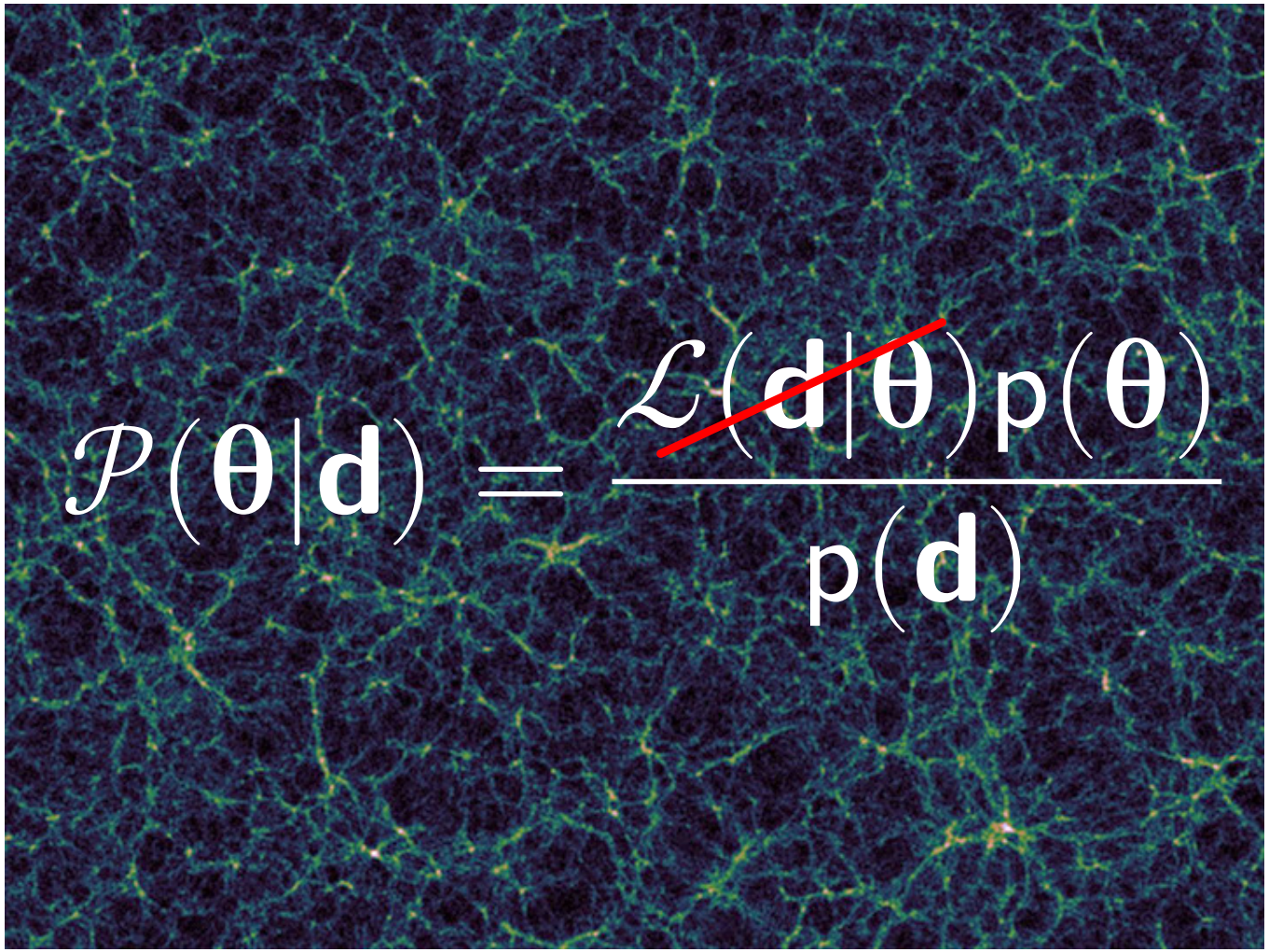
github:information_maximiser



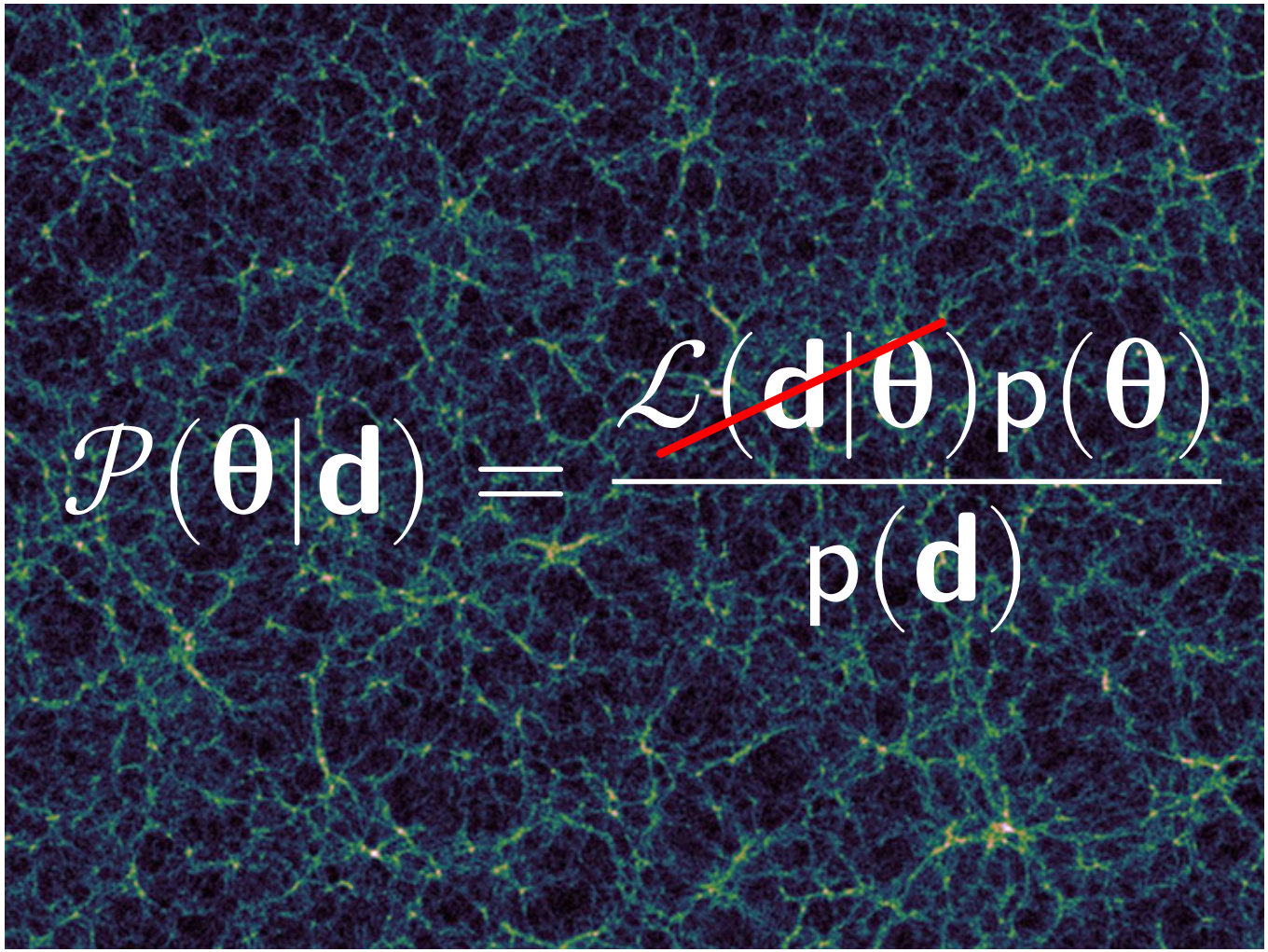
Statistical inference



Approximate Bayesian computation

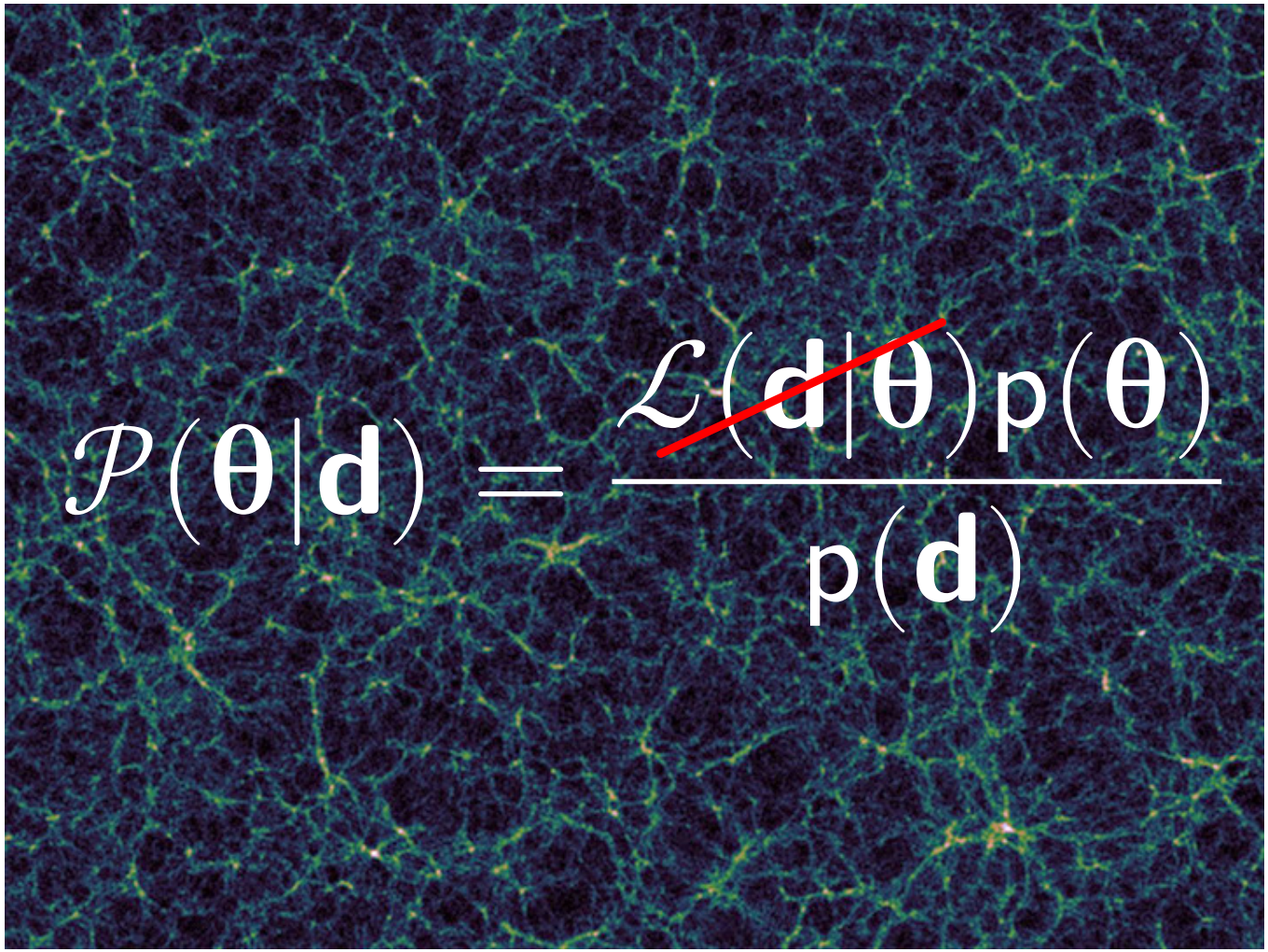


Approximate Bayesian computation



Simulate data at different $\boldsymbol{\theta}$ drawn from $\mathrm{p}(\boldsymbol{\theta})$

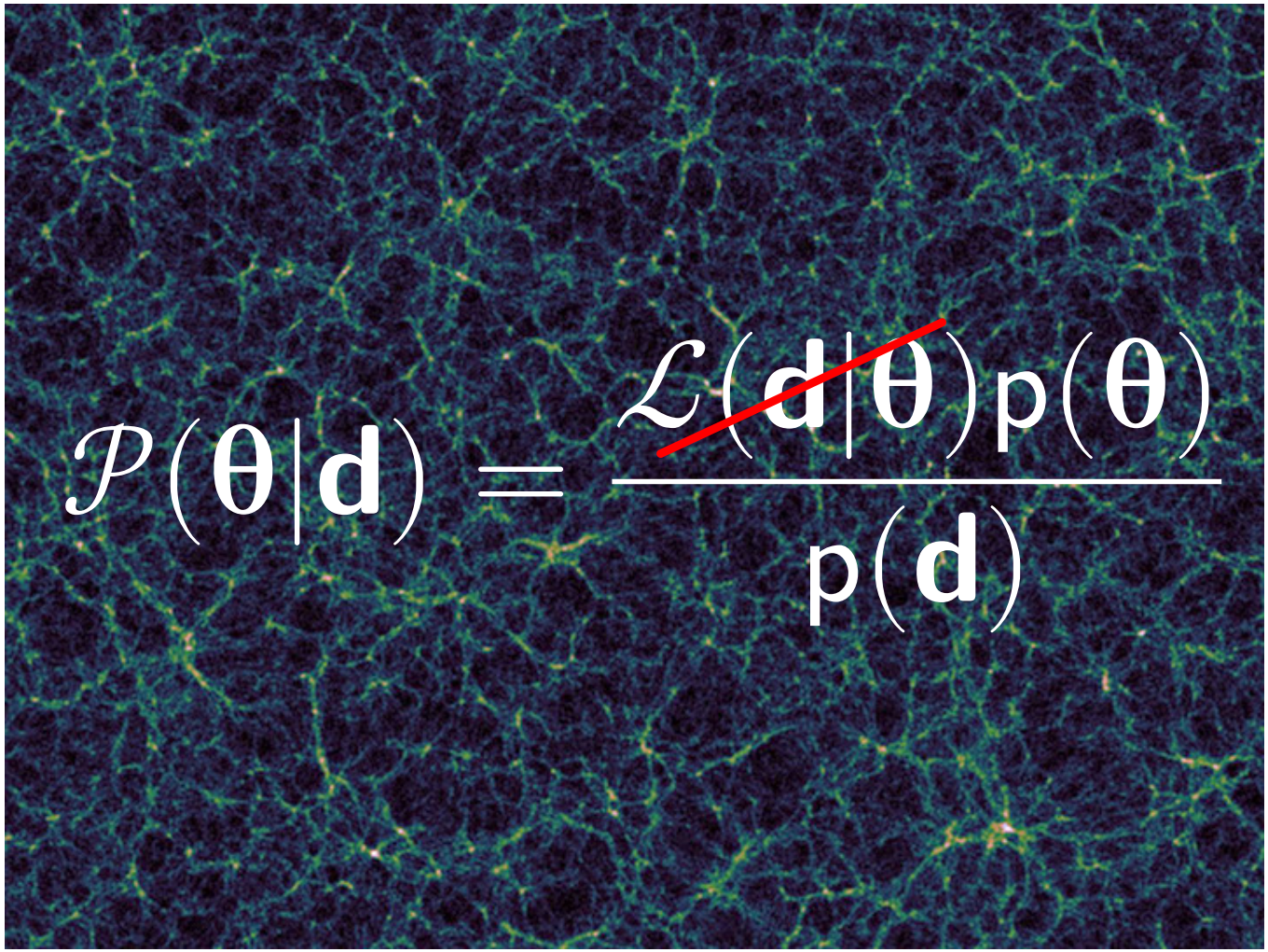
Approximate Bayesian computation



Simulate data at different $\boldsymbol{\theta}$ drawn from $\mathrm{p}(\boldsymbol{\theta})$

Accept or reject $\boldsymbol{\theta}$ using similarity of simulation and real data

Likelihood-free rejection sampling



Simulate data at different $\boldsymbol{\theta}$ drawn from $\mathrm{p}(\boldsymbol{\theta})$

Accept or reject $\boldsymbol{\theta}$ using similarity of simulation and real data

Why we need compressed summary statistics

Reduce dimensionality of data

Unlikely for simulation to hit real data in high dimensions

Use summary statistics such as power spectrum, C_ℓ, ect.

Why we need compressed summary statistics

Reduce dimensionality of data

Unlikely for simulation to hit real data in high dimensions

Use summary statistics such as power spectrum, C_ℓ, ect.

Inadequately compressed summary statistics

Upcoming surveys will observe huge amounts of raw data

Even the summary statistics will be $O(10^4)$ (LSST, EUCLID, etc.)

Huge amount of computing power needed

Still nearly impossible to do ABC

Compression for general likelihoods

Optimal compression for general likelihood functions

Fisher information

- Alsing & Wandelt, 2017

Amount of information data $\mathbf{d}$ contains about parameters $\boldsymbol{\theta}$

$$\mathbf{F}_{\alpha\beta} = -\left\langle \frac{\partial^2 \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{fid}}}$$

$$= \left\langle \frac{\partial \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha} \frac{\partial \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{fid}}}$$

Optimal compression for general likelihood functions

Fisher information

- Alsing & Wandelt, 2017

Amount of information data $\mathbf{d}$ contains about parameters $\boldsymbol{\theta}$

$$\mathbf{F}_{\alpha\beta} = -\left\langle \frac{\partial^2 \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{fid}}}$$

$$= \left\langle \frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha} \frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{fid}}}$$

Information inequality

Cramér-Rao bound for unbiased estimators $\langle \mathcal{T}_\alpha \rangle = \theta_\alpha$

$$\left\langle \left(\mathcal{T}_\alpha - \langle \mathcal{T}_\alpha \rangle\right)\left(\mathcal{T}_\beta - \langle \mathcal{T}_\beta \rangle\right) \right\rangle \geq \mathbf{F}^{-1}_{\alpha\beta}$$

Lower bound on the variances of sufficient statistics, $\mathcal{T}_\alpha$.

$$\left\langle \left(\mathcal{T}_\alpha - \langle \mathcal{T}_\alpha \rangle\right)\left(\mathcal{T}_\beta - \langle \mathcal{T}_\beta \rangle\right) \right\rangle \geq \left\langle \frac{\partial \mathcal{T}_\mu}{\partial\theta_\alpha} \right\rangle \mathbf{F}^{-1}_{\mu\nu} \left\langle \frac{\partial \mathcal{T}_\nu}{\partial\theta_\beta} \right\rangle$$

Optimal compression for geneal likelihood functions

Sufficient statistics

Expand the log-likelihood about some fiducial parameters, $\boldsymbol{\theta}^{\rm fid}$

$$\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}^{\rm fid}) + \delta\theta_\alpha^{\rm T}\frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}^{\rm fid})}{\partial\theta_\alpha} + \frac{1}{2}\delta\theta_\alpha^{\rm T}\frac{\partial^2 \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}^{\rm fid})}{\partial\theta_\alpha\partial\theta_\beta}\delta\theta_\beta + \cdots$$

To linear order the data only couples through the score function

$$\mathcal{S}_\alpha = \frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha}$$

This is a set of sufficient statistics for the linearised log-likelihood

$$\mathcal{T}_\alpha = \mathcal{S}_\alpha$$

Optimal compression for general likelihood functions

Fisher information

$$\left\langle \left(\mathcal{T}_\alpha - \left\langle \mathcal{T}_\alpha \right\rangle\right)\left(\mathcal{T}_\beta - \left\langle \mathcal{T}_\beta \right\rangle\right)\right\rangle = \left\langle \frac{\partial \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_\alpha} \frac{\partial \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\rm fid}}$$
$$= \mathbf{F}_{\alpha\beta}$$

$$\left\langle \frac{\partial \mathcal{T}_\alpha}{\partial \theta_\beta} \right\rangle = \left\langle \frac{\partial^2 \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\rm fid}}$$
$$= -\mathbf{F}_{\alpha\beta}$$

Optimal compression for general likelihood functions

Fisher information

$$\left\langle (\mathcal{T}_\alpha - \langle \mathcal{T}_\alpha \rangle)(\mathcal{T}_\beta - \langle \mathcal{T}_\beta \rangle) \right\rangle = \left\langle \frac{\partial \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_\alpha} \frac{\partial \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\rm fid}}$$
$$= \mathbf{F}_{\alpha\beta}$$

$$\left\langle \frac{\partial \mathcal{T}_\alpha}{\partial \theta_\beta} \right\rangle = \left\langle \frac{\partial^2 \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\rm fid}}$$
$$= -\mathbf{F}_{\alpha\beta}$$

Information inequality saturated

$$\left\langle (\mathcal{T}_\alpha - \langle \mathcal{T}_\alpha \rangle)(\mathcal{T}_\beta - \langle \mathcal{T}_\beta \rangle) \right\rangle \geq \left\langle \frac{\partial \mathcal{T}_\mu}{\partial \theta_\alpha} \right\rangle \mathbf{F}^{-1}_{\mu\nu} \left\langle \frac{\partial \mathcal{T}_\nu}{\partial \theta_\beta} \right\rangle$$

Massively optimised parameter estimation and data (MOPED) compression

- Heavens et. al, 2000

Assume a Gaussian likelihood ($\mu(\boldsymbol{\theta}) \equiv \boldsymbol{\mu}$ and $\partial \mathbf{C}/\partial\theta_\alpha = 0$)

$$-2\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = (\mathbf{d}-\boldsymbol{\mu})^{\mathsf{T}}\,\mathbf{C}^{-1}\,(\mathbf{d}-\boldsymbol{\mu}) + \ln|2\pi\mathbf{C}|$$

Massively optimised parameter estimation and data (MOPED) compression

- Heavens et. al, 2000

Assume a Gaussian likelihood ($\mu(\boldsymbol{\theta}) \equiv \boldsymbol{\mu}$ and $\partial \mathbf{C}/\partial\theta_\alpha = 0$)

$$-2\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = (\mathbf{d}-\boldsymbol{\mu})^{\mathsf{T}}\,\mathbf{C}^{-1}\,(\mathbf{d}-\boldsymbol{\mu}) + \ln|2\pi\mathbf{C}|$$

Compression function

$$\mathcal{T}_\alpha = \left.\frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{fid}}}$$

$$= \left.\boldsymbol{\mu}_{,\alpha}^{\mathsf{T}}\,\mathbf{C}^{-1}\mathbf{d}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{fid}}} \quad \text{using} \quad \frac{\partial\boldsymbol{\mu}}{\partial\theta_\alpha} \equiv \boldsymbol{\mu}_{,\alpha}$$

Massively optimised parameter estimation and data (MOPED) compression

- Heavens et. al, 2000

Assume a Gaussian likelihood ($\mu(\boldsymbol{\theta}) \equiv \boldsymbol{\mu}$ and $\partial \mathbf{C}/\partial\theta_\alpha = 0$)

$$-2\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = (\mathbf{d}-\boldsymbol{\mu})^{\mathsf{T}}\,\mathbf{C}^{-1}\,(\mathbf{d}-\boldsymbol{\mu}) + \ln|2\pi\mathbf{C}|$$

Compression function

$$\mathcal{T}_\alpha = \left.\frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\rm fid}}$$

$$= \left.\boldsymbol{\mu}_{,\alpha}^{\mathsf{T}}\,\mathbf{C}^{-1}\mathbf{d}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\rm fid}} \quad \text{using} \quad \frac{\partial\boldsymbol{\mu}}{\partial\theta_\alpha} \equiv \boldsymbol{\mu}_{,\alpha}$$

Fisher information

$$\mathbf{F}_{\alpha\beta} = \left.\left\langle \frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha}\frac{\partial \ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\beta}\right\rangle\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\rm fid}}$$

$$= \mathrm{Tr}\left[\boldsymbol{\mu}_{,\alpha}^{\mathsf{T}}\mathbf{C}^{-1}\boldsymbol{\mu}_{,\beta}\right]$$

How can we use this compression?

Calculated summary statistics from raw data

Calculate summary statistics (such as power spectrum) from raw data

Use MOPED (or similar) on summary statistics (assuming Gaussianity)

How can we use this compression?

Calculated summary statistics from raw data

Calculate summary statistics (such as power spectrum) from raw data

Use MOPED (or similar) on summary statistics (assuming Gaussianity)

Likelihood-free compression

Forward simulate to get realisations of summary statistics

Use MOPED (or similar) on summary statistics (assuming Gaussianity)

How can we use this compression?

Calculated summary statistics from raw data

Calculate summary statistics (such as power spectrum) from raw data

Use MOPED (or similar) on summary statistics (assuming Gaussianity)

Likelihood-free compression

Forward simulate to get realisations of summary statistics

Use MOPED (or similar) on summary statistics (assuming Gaussianity)

Don't have to approximate your likelihoods!

Non-linear likelihood-free compression

Non-linear likelihood-free compression

Start with unknown likelihood $\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})$

Data $\mathbf{d}$ can be generated using a model $\mathcal{M}(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$

Non-linear likelihood-free compression

Start with unknown likelihood $\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})$

Data $\mathbf{d}$ can be generated using a model $\mathcal{M}(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$

Transform data

Find a function, $\mathscr{f} : \mathbf{d} \rightarrow \mathbf{x}$ which Gaussianizes the summaries, i.e.

$$-2\ln\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \left(\mathbf{x}-\boldsymbol{\mu}_{\mathscr{f}}\right)^{\mathsf{T}} \mathbf{C}_{\mathscr{f}}^{-1} \left(\mathbf{x}-\boldsymbol{\mu}_{\mathscr{f}}\right)$$

Non-linear likelihood-free compression

Start with unknown likelihood $\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})$

Data $\mathbf{d}$ can be generated using a model $\mathcal{M}(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$

Transform data

Find a function, $\mathscr{f} : \mathbf{d} \rightarrow \mathbf{x}$ which Gaussianizes the summaries, i.e.

$$-2\ln\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \left(\mathbf{x}-\boldsymbol{\mu}_{\mathscr{f}}\right)^{\mathsf{T}} \mathbf{C}_{\mathscr{f}}^{-1} \left(\mathbf{x}-\boldsymbol{\mu}_{\mathscr{f}}\right)$$

What is $\mathscr{f}$?

Non-linear likelihood-free compression

Start with unknown likelihood $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})$

Data $\mathbf{d}$ can be generated using a model $\mathcal{M}(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$

Transform data

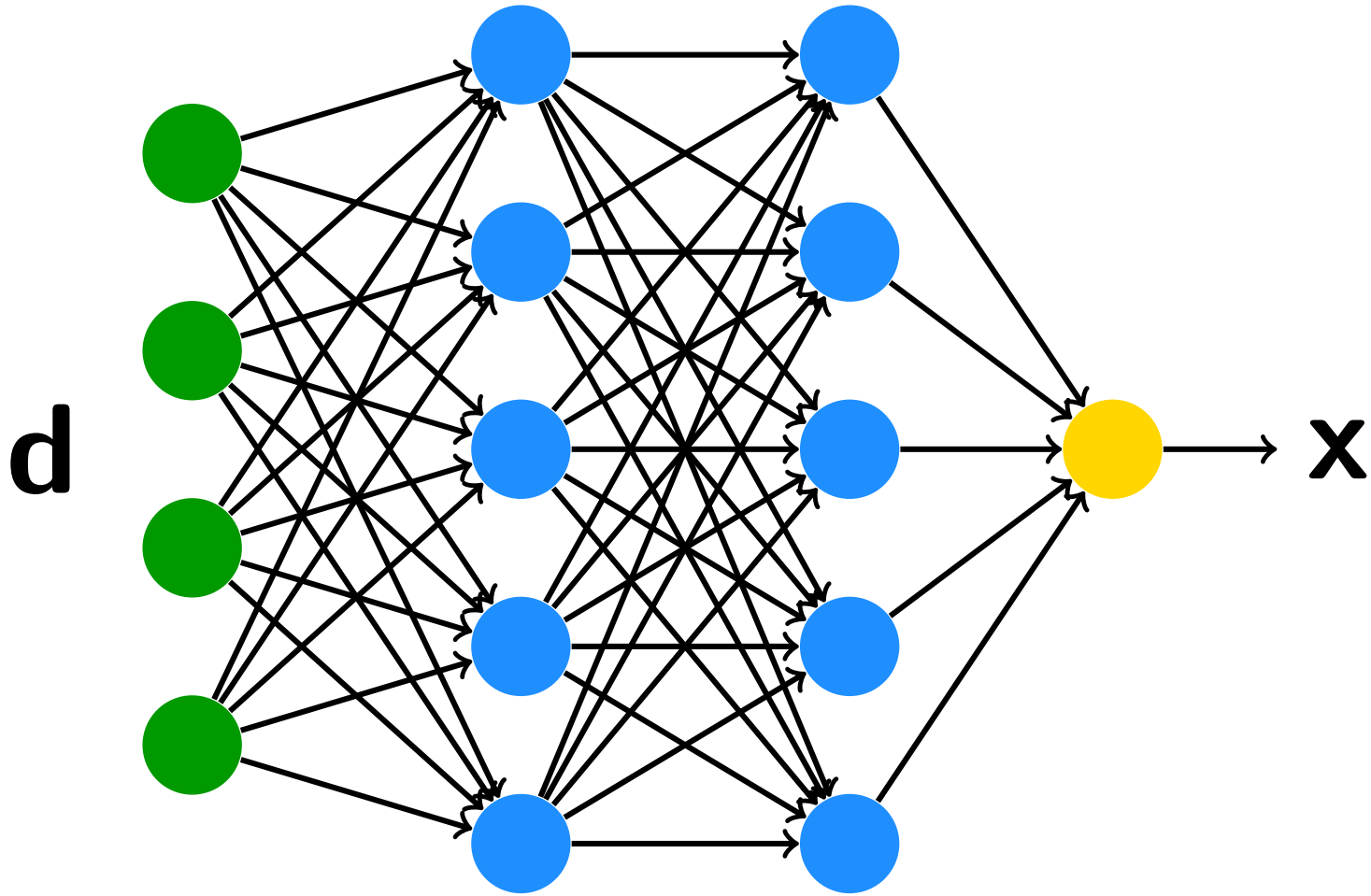
Find a function, $\mathscr{f} : \mathbf{d} \rightarrow \mathbf{x}$ which Gaussianizes the summaries, i.e.

$$-2\ln\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)^{\mathsf{T}} \mathbf{C}_{\mathscr{f}}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)$$

What is $\mathscr{f}$?

A neural network!

Information maximising neural network



Information maximising neural network

Compression network

Fewer outputs than inputs (outputs can be as few as the number of model parameters)

Should be built for your data structure for best performance

Information maximising neural network

Compression network

Fewer outputs than inputs (outputs can be as few as the number of model parameters)

Should be built for your data structure for best performance

Loss function

$$\begin{aligned} \text{Loss} &= -\frac{1}{2}|\mathbf{F}|^2 \\ &= -\frac{1}{2}|\boldsymbol{\mu}_{\ell,\alpha}^{\mathsf{T}}\,\mathbf{C}^{-1}\boldsymbol{\mu}_{\ell,\beta}\,|^2 \end{aligned}$$

Information maximising neural network

Compression network

Fewer outputs than inputs (outputs can be as few as the number of model parameters)

Should be built for your data structure for best performance

Loss function

$$\begin{aligned} \text{Loss} &= -\frac{1}{2}|\mathbf{F}|^2 \\ &= -\frac{1}{2}|\boldsymbol{\mu}_{\ell,\alpha}^{\mathsf{T}}\,\mathbf{C}^{-1}\boldsymbol{\mu}_{\ell,\beta}\,|^2 \end{aligned}$$

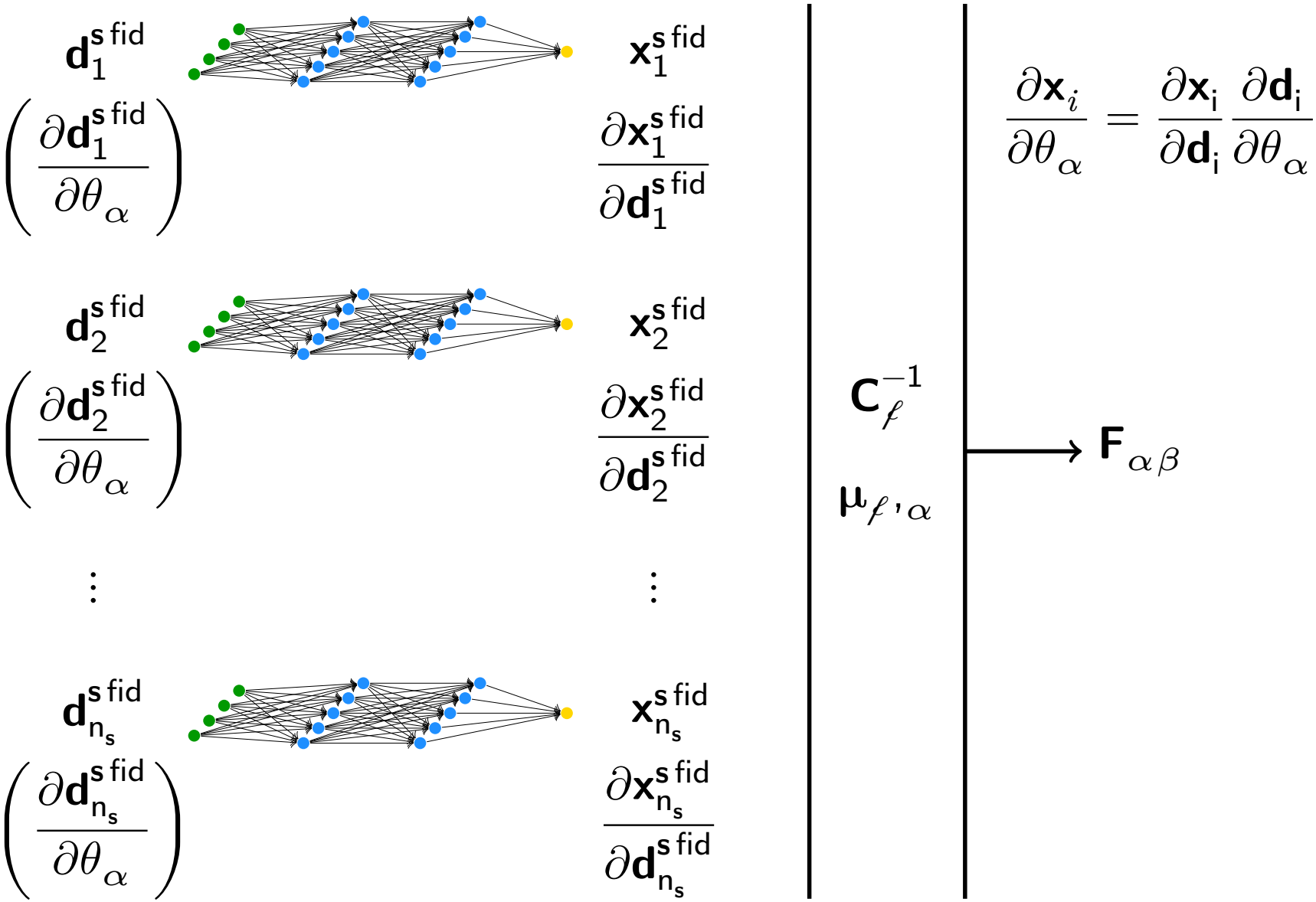
Simulations

Need high quality simulations of the real data

Only needed at the fiducial parameter values for $\mathbf{C}^{-1}$

Need the derivative of the simulation (analytic or numerical (surpressed sample variance) for $\boldsymbol{\mu}_{\ell,\alpha}$

Information maximising neural network



Automatic physical inference

Automatic physical inference

Make simulations of data, $\left\{ \mathbf{d}_{\mathsf{i}}^{\mathbf{s}\,\mathsf{fid}}, \dfrac{\partial \mathbf{d}_{\mathsf{i}}^{\mathbf{s}\,\mathsf{fid}}}{\partial \theta_\alpha} \right\}$ at $\boldsymbol{\theta}^{\mathsf{fid}}$

Automatic physical inference

Make simulations of data, $\left\{ \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}, \frac{\partial \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}}{\partial \theta_\alpha} \right\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

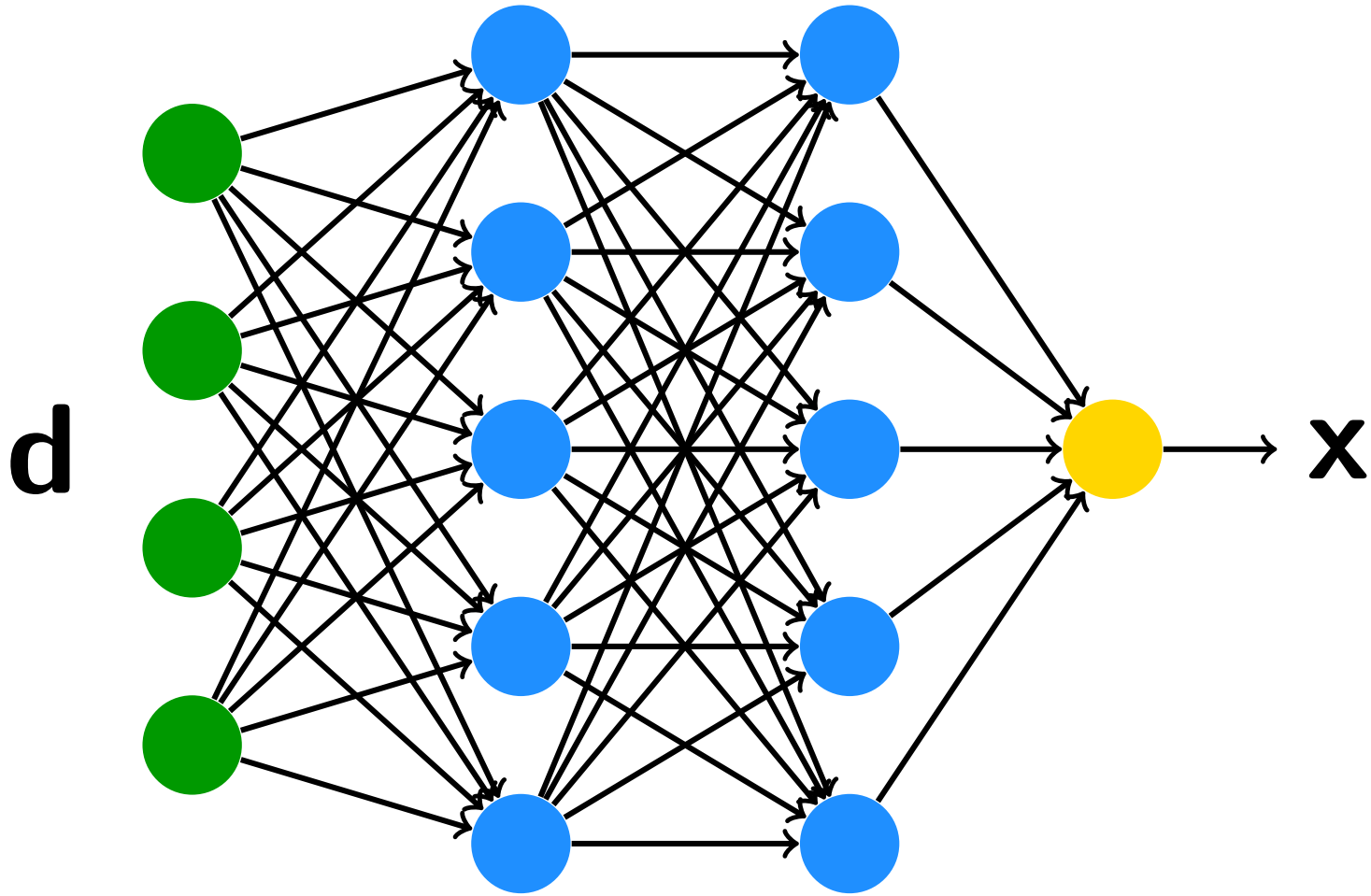
Automatic physical inference

Make simulations of data, $\left\{ \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}, \frac{\partial \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}}{\partial \theta_\alpha} \right\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

Compress real data

Automatic physical inference



Automatic physical inference

Make simulations of data, $\left\{ \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}, \frac{\partial \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}}{\partial \theta_\alpha} \right\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

Compress real data

Draw $\boldsymbol{\theta}$ from $\mathrm{p}(\boldsymbol{\theta})$ and create simulation

Automatic physical inference

Make simulations of data, $\left\{ \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}, \frac{\partial \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}}{\partial \theta_\alpha} \right\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

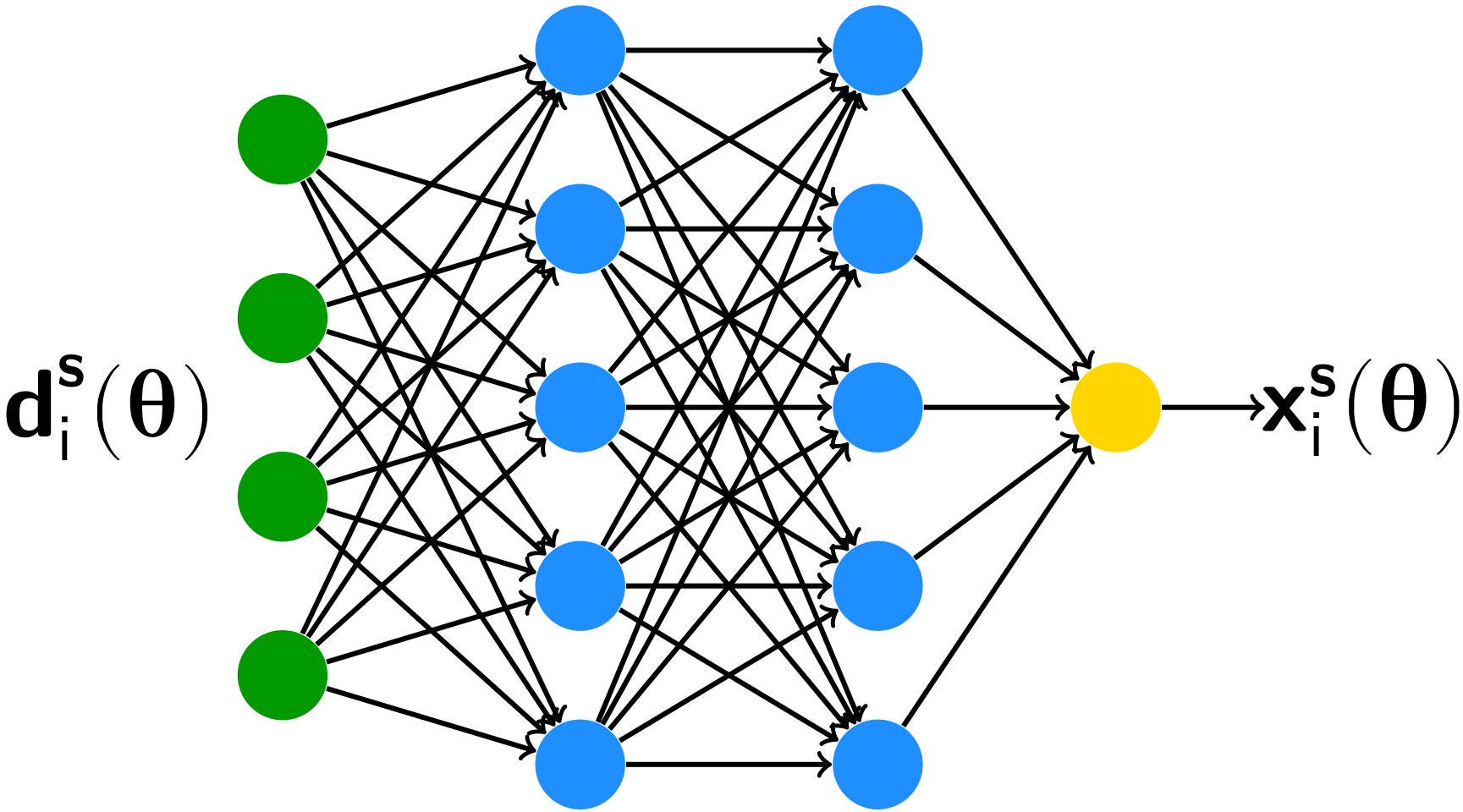
Train IMNN to maximise Fisher information

Compress real data

Draw $\boldsymbol{\theta}$ from $\mathrm{p}(\boldsymbol{\theta})$ and create simulation

Compress simulation

Automatic physical inference



Automatic physical inference

Make simulations of data, $\left\{ \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}, \frac{\partial \mathbf{d}_{\mathrm{i}}^{\mathbf{s}\,\mathrm{fid}}}{\partial \theta_\alpha} \right\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

Compress real data

Draw $\boldsymbol{\theta}$ from $\mathrm{p}(\boldsymbol{\theta})$ and create simulation

Compress simulation

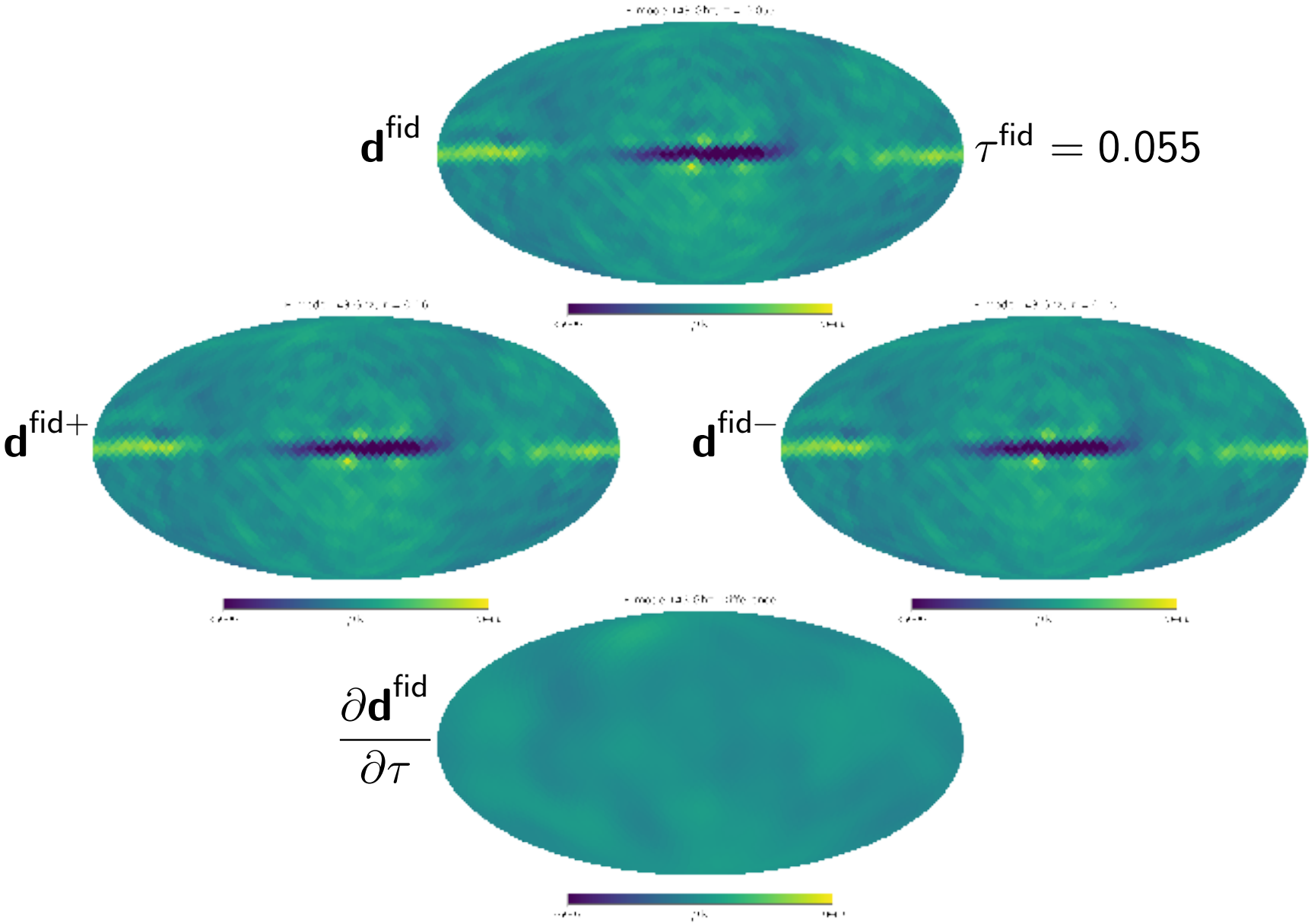
Accept or reject proposed sample to build approximate posterior distribution

Example time!

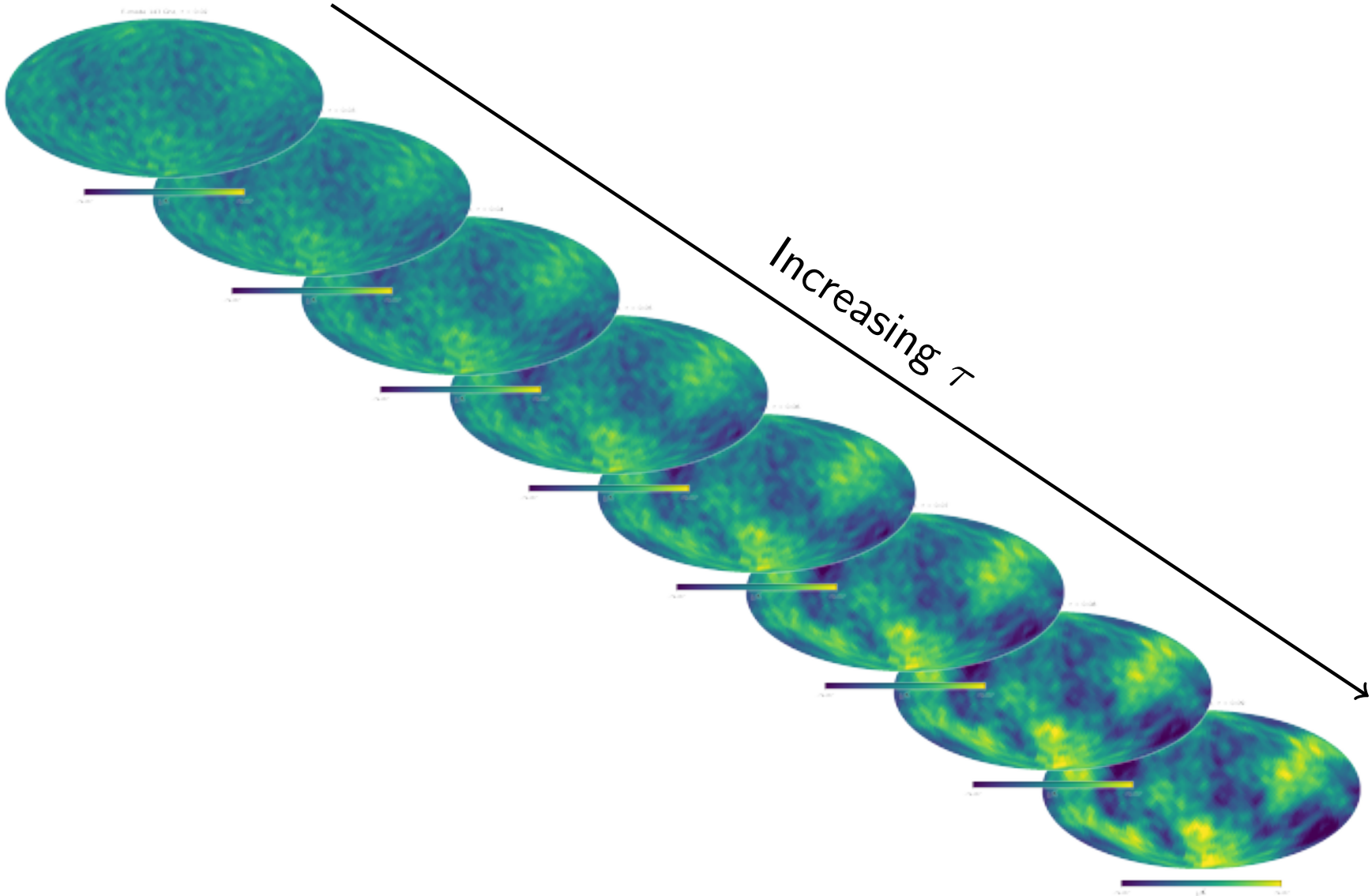
Can we infer τ from CMB polarisation maps?

Make simulations of data, $\{\mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}-}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}+}\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

CMB polarisation maps



Effect of τ on the maps

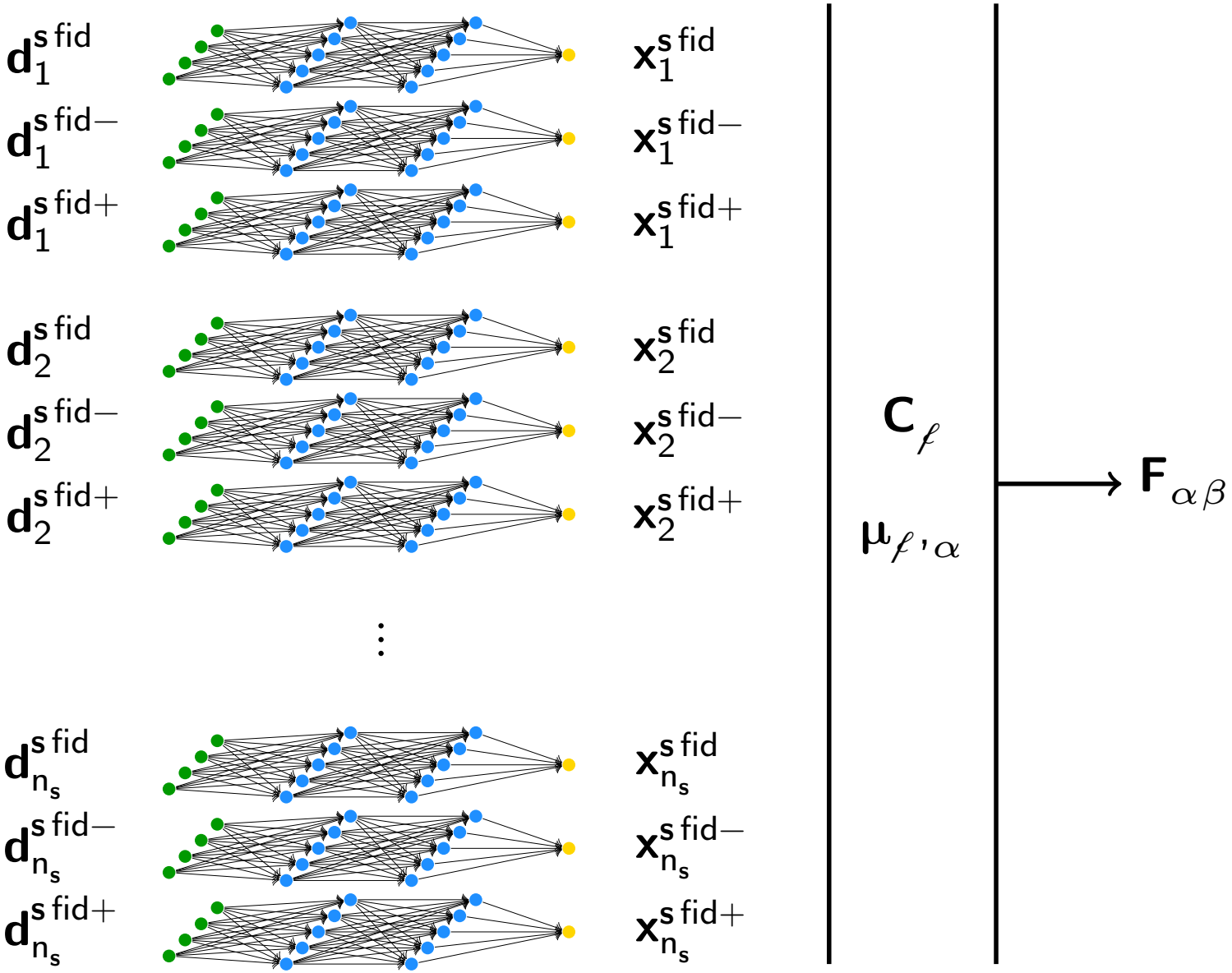


Can we infer τ from maps?

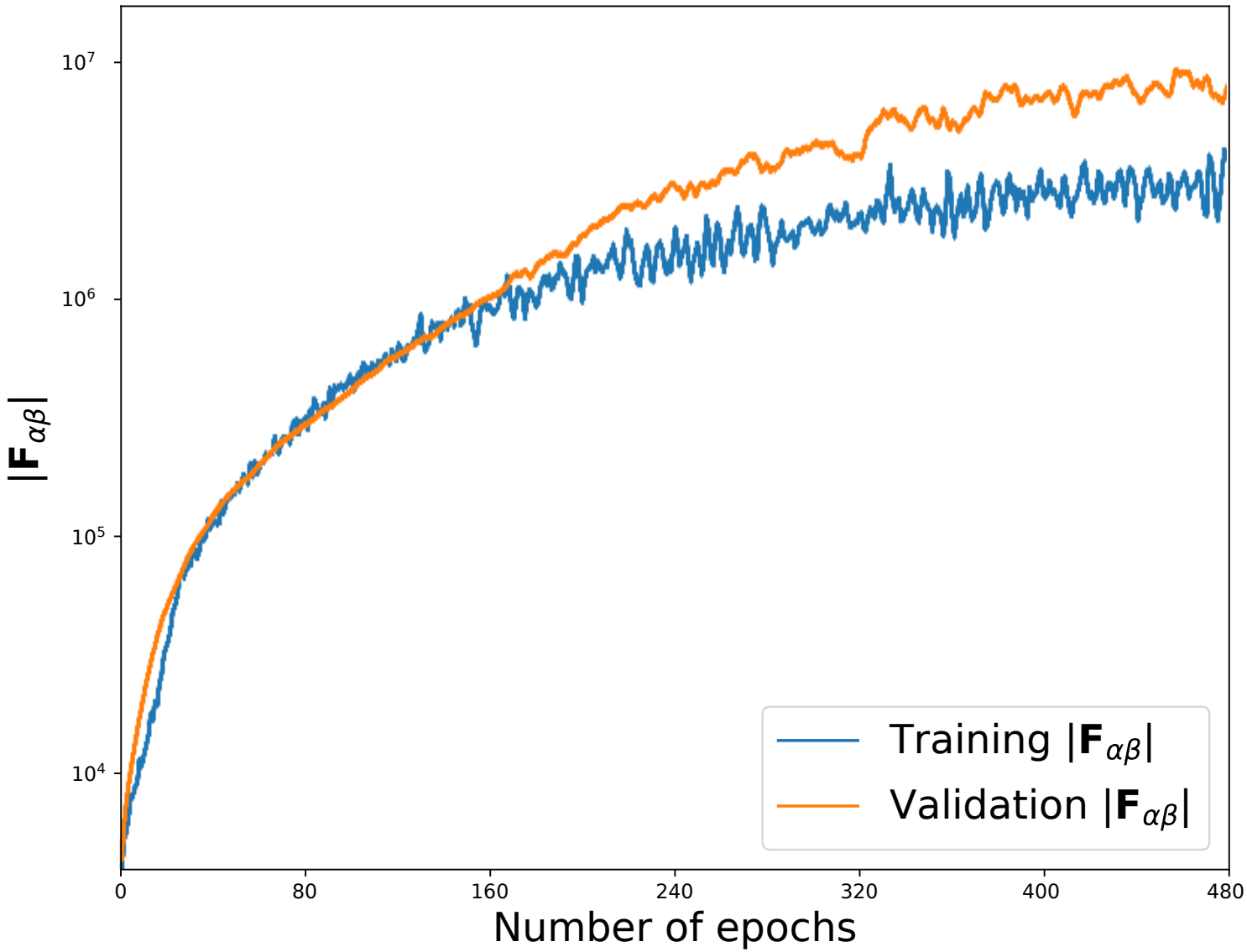
Make simulations of data, $\{\mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}-}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}+}\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

Information maximising neural network



How much does the Fisher grow?



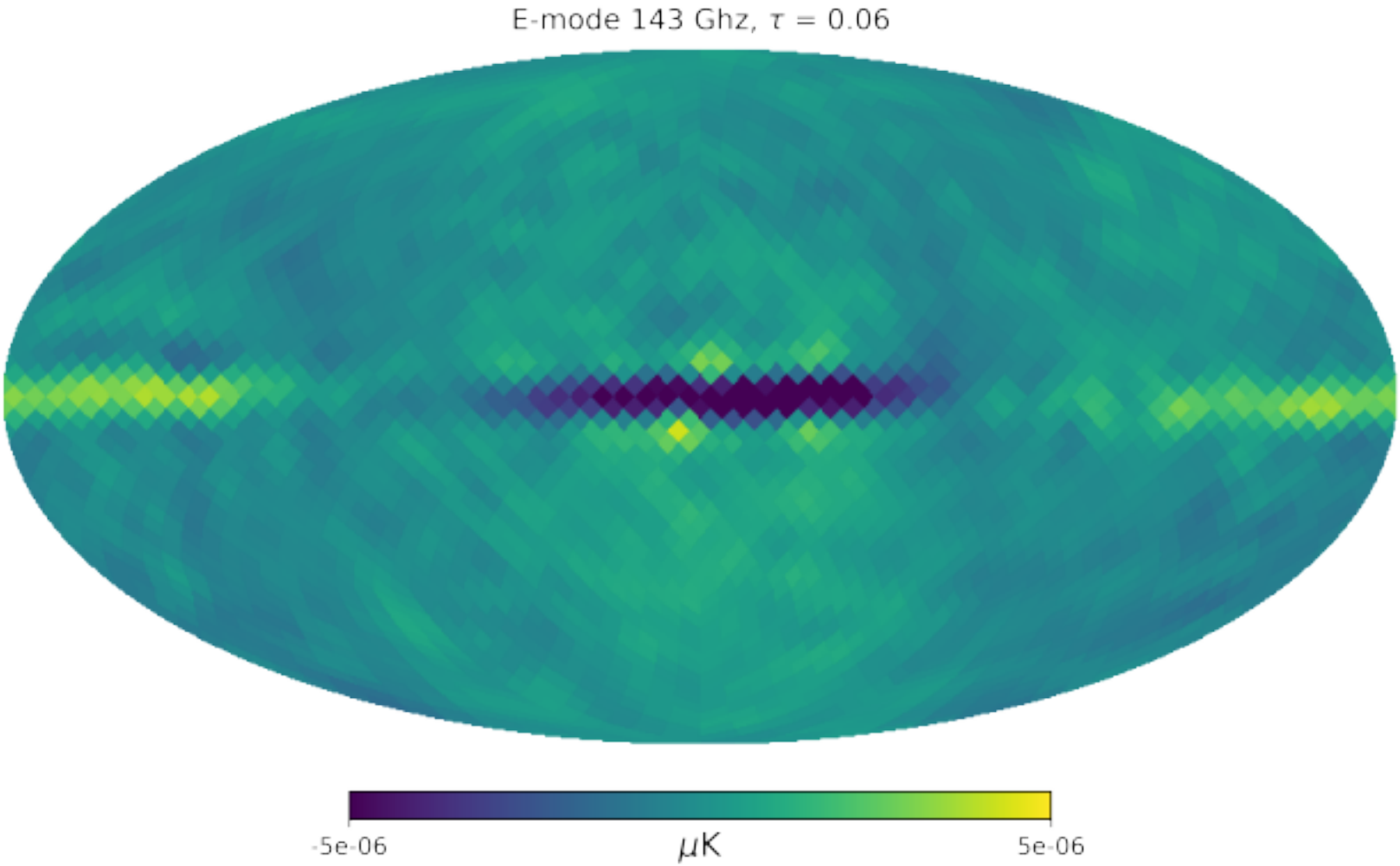
Can we infer τ from maps?

Make simulations of data, $\{\mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}-}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}+}\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

Compress real data

Our *real* data at $\tau = 0.06$



Can we infer τ from maps?

Make simulations of data, $\{\mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}-}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}+}\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

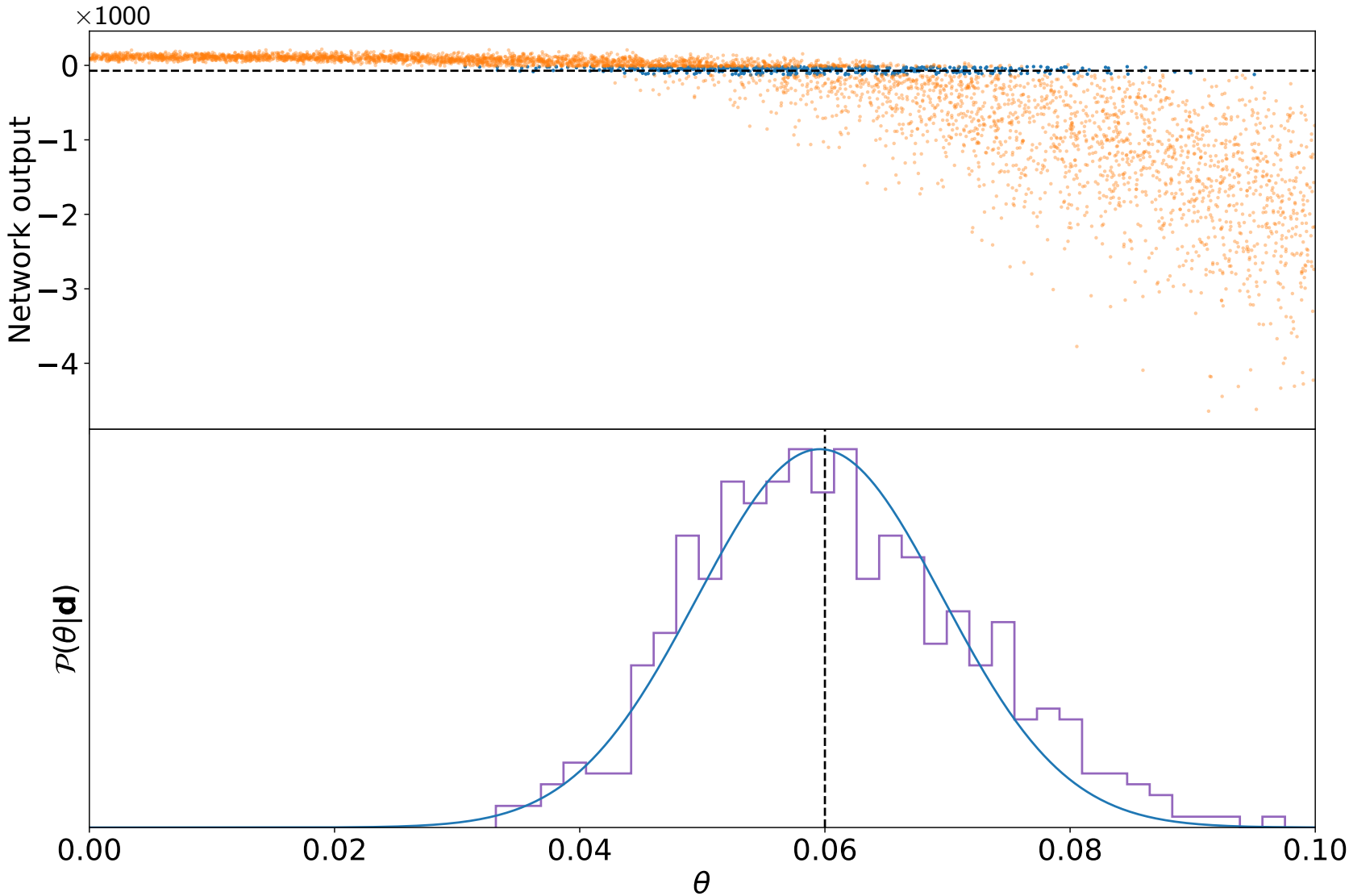
Compress real data

Draw $\boldsymbol{\theta}$ from $\mathrm{p}(\boldsymbol{\theta})$ and create simulation

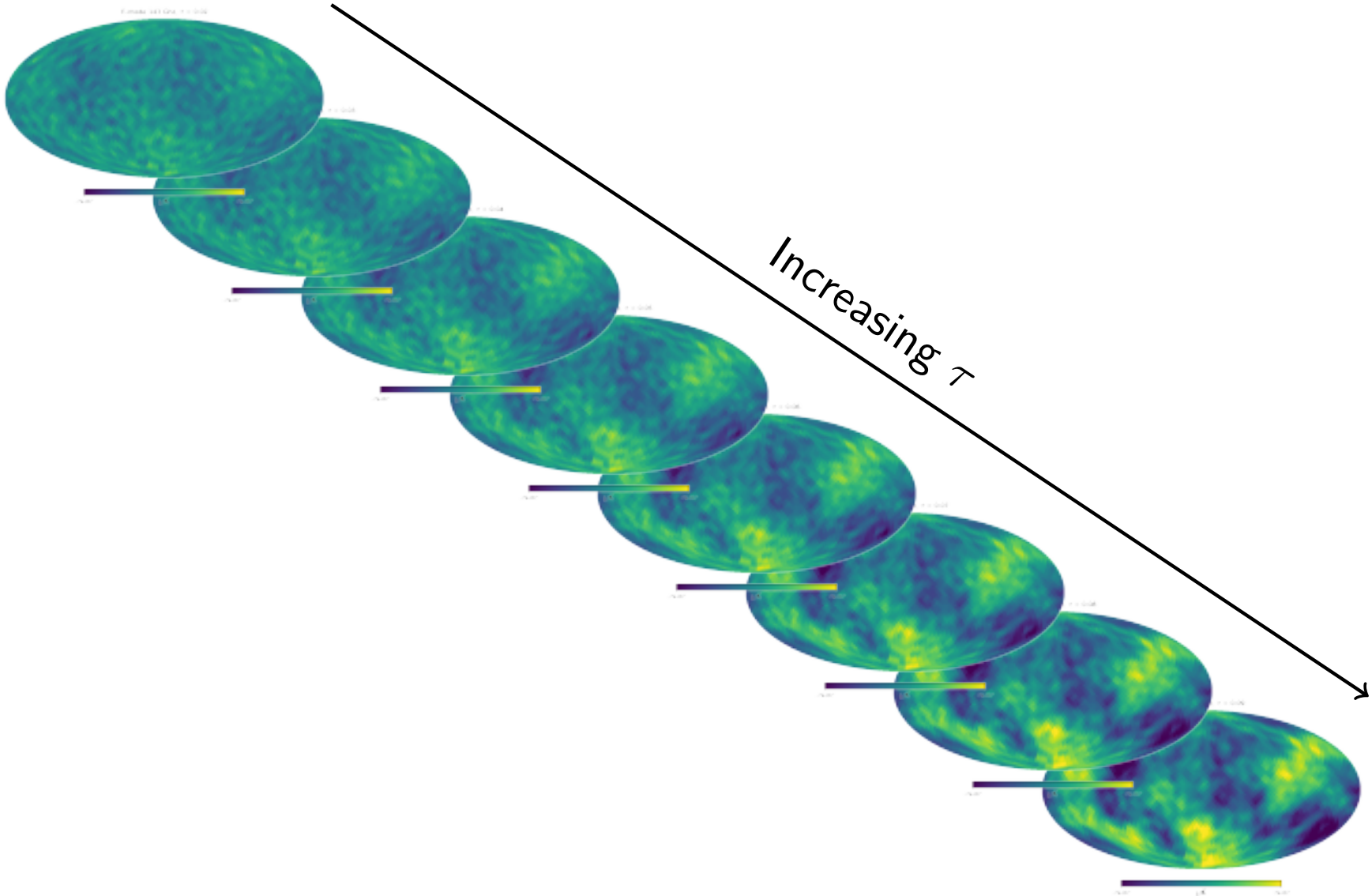
Compress simulation

Accept or reject proposed sample to build approximate posterior distribution

Approximate posterior distribution

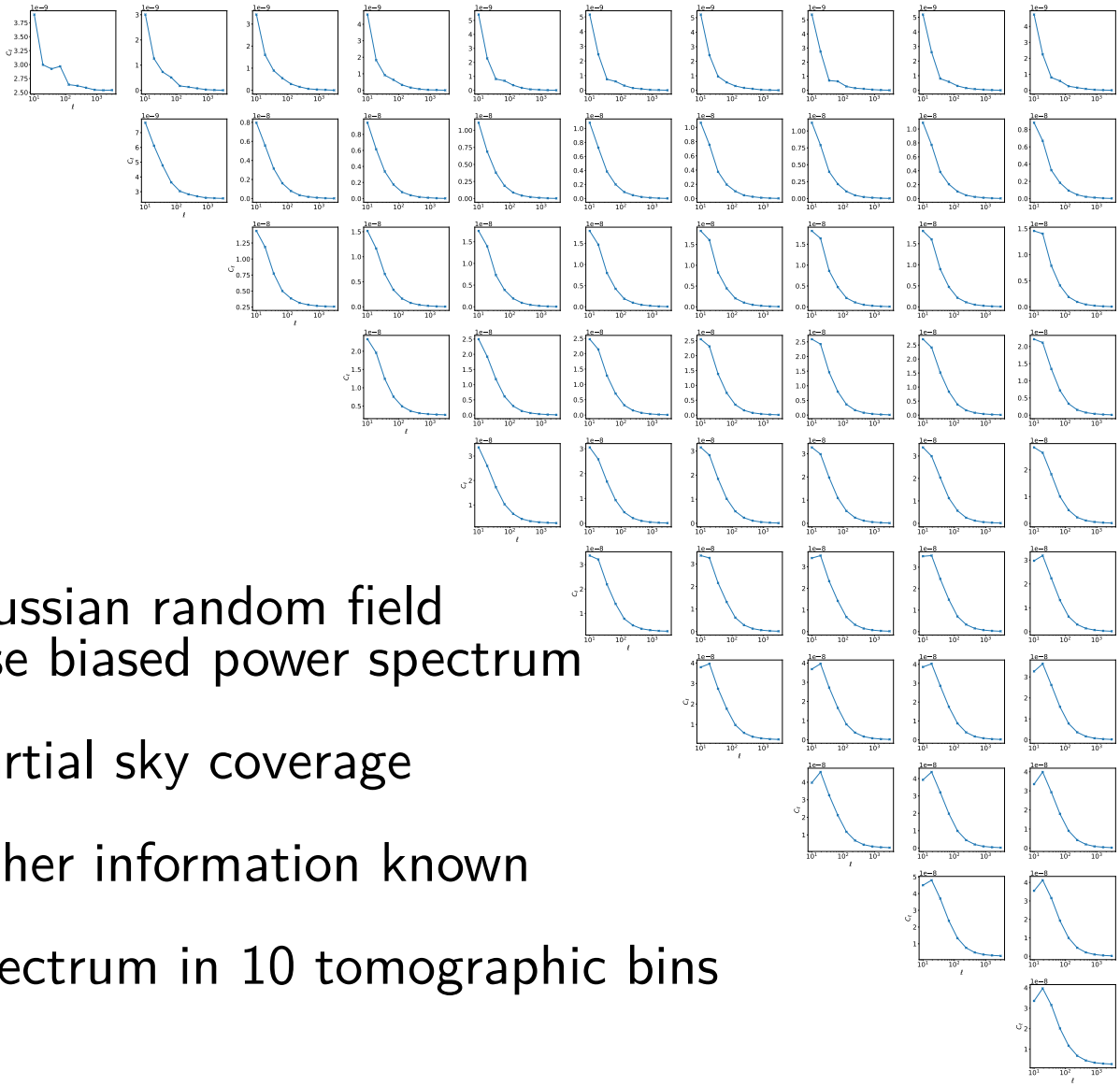


Effect of τ on the maps



Another quick example!

Cosmic shear



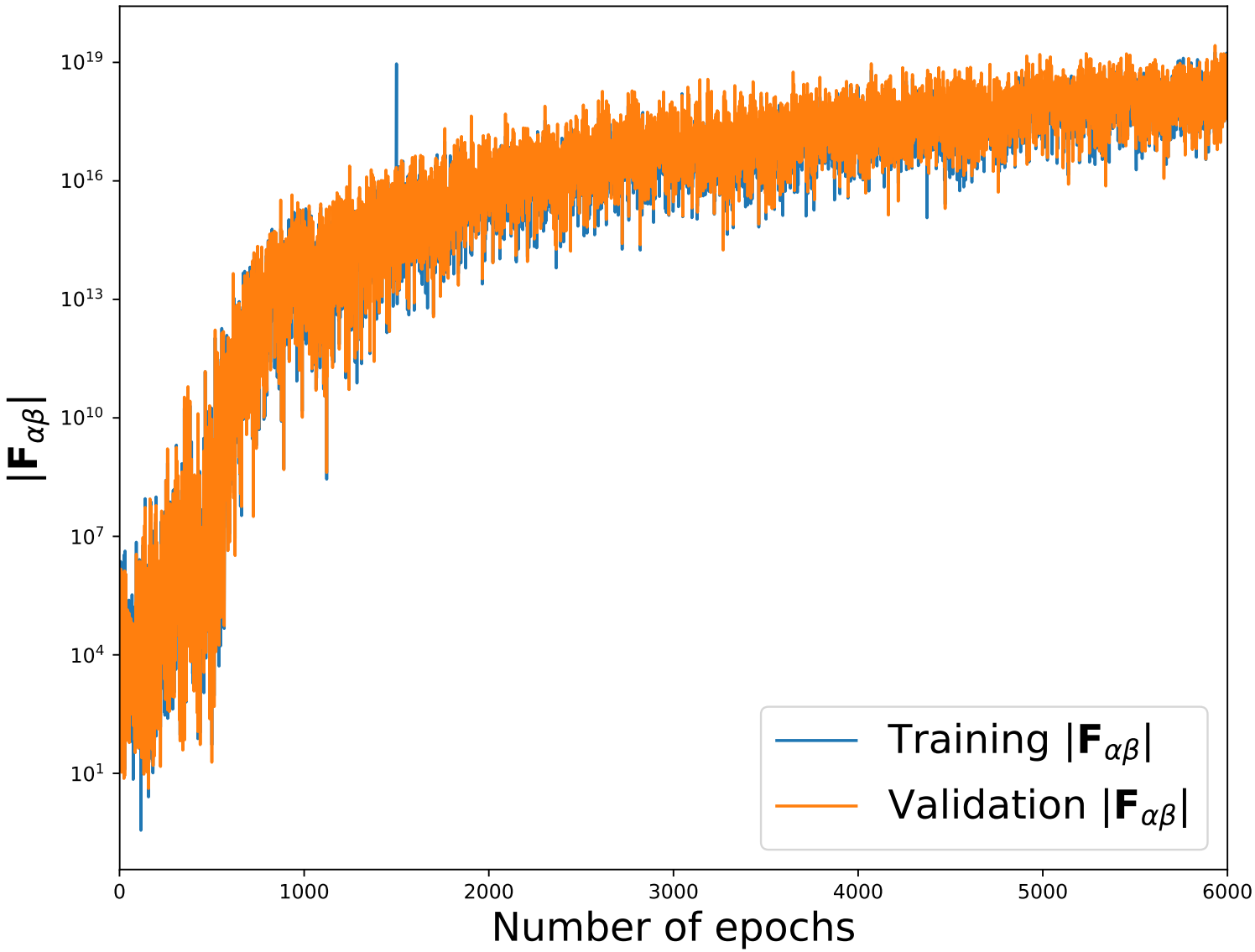
Draw Gaussian random field
from noise biased power spectrum

Mimic partial sky coverage

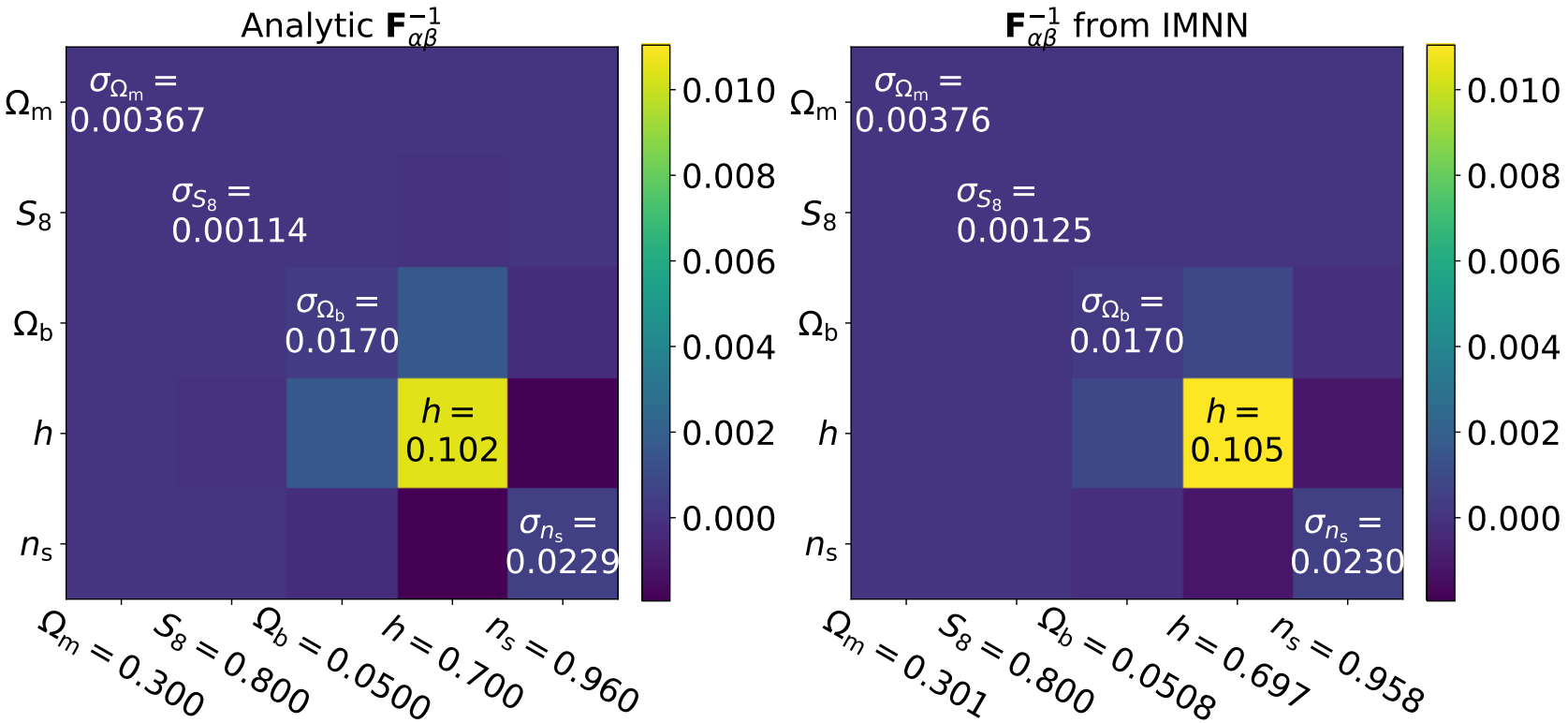
Exact Fisher information known

Power spectrum in 10 tomographic bins

How much does the Fisher grow?



Fisher comparison

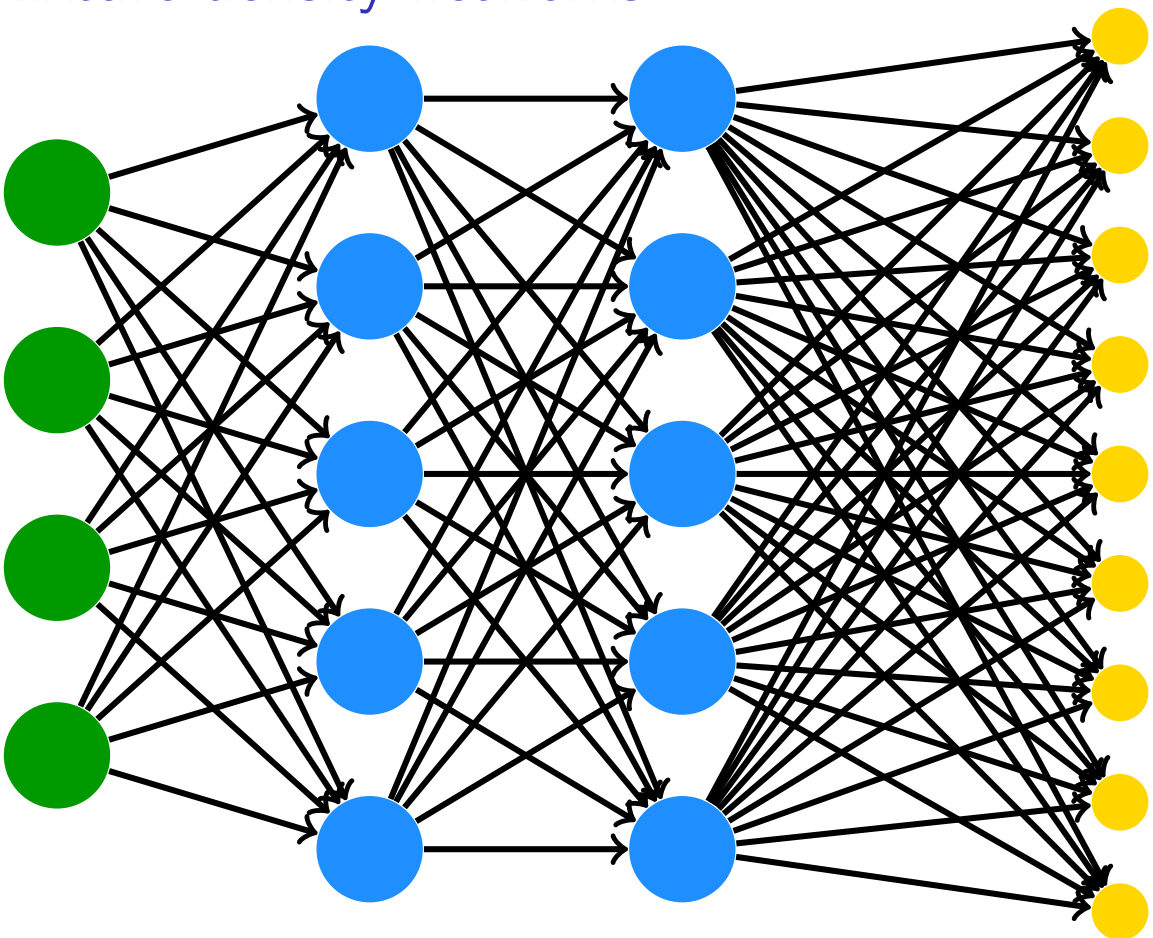


Density estimation
likelihood-free estimatation

Density estimation likelihood-free inference (DELFI)

- Alsing, Feeney & Wandelt, 2018

Mixture density networks



α_i - softmax

μ_i

σ_i - exponential

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta},\mathbf{w}) = \prod_{i=1}^{n_{\mathcal{N}}} \alpha_i(\mathbf{w})\mathcal{N}(\mathbf{x}|\boldsymbol{\theta},\mu_i(\mathbf{w}),\sigma_i(\mathbf{w}))$$

Super-efficient inference

Pretrained on the Fisher information

Really quick to produce samples to get MDN into right region

Taught actively

Geometric mean of each round of updates can be used as an updated proposal distribution

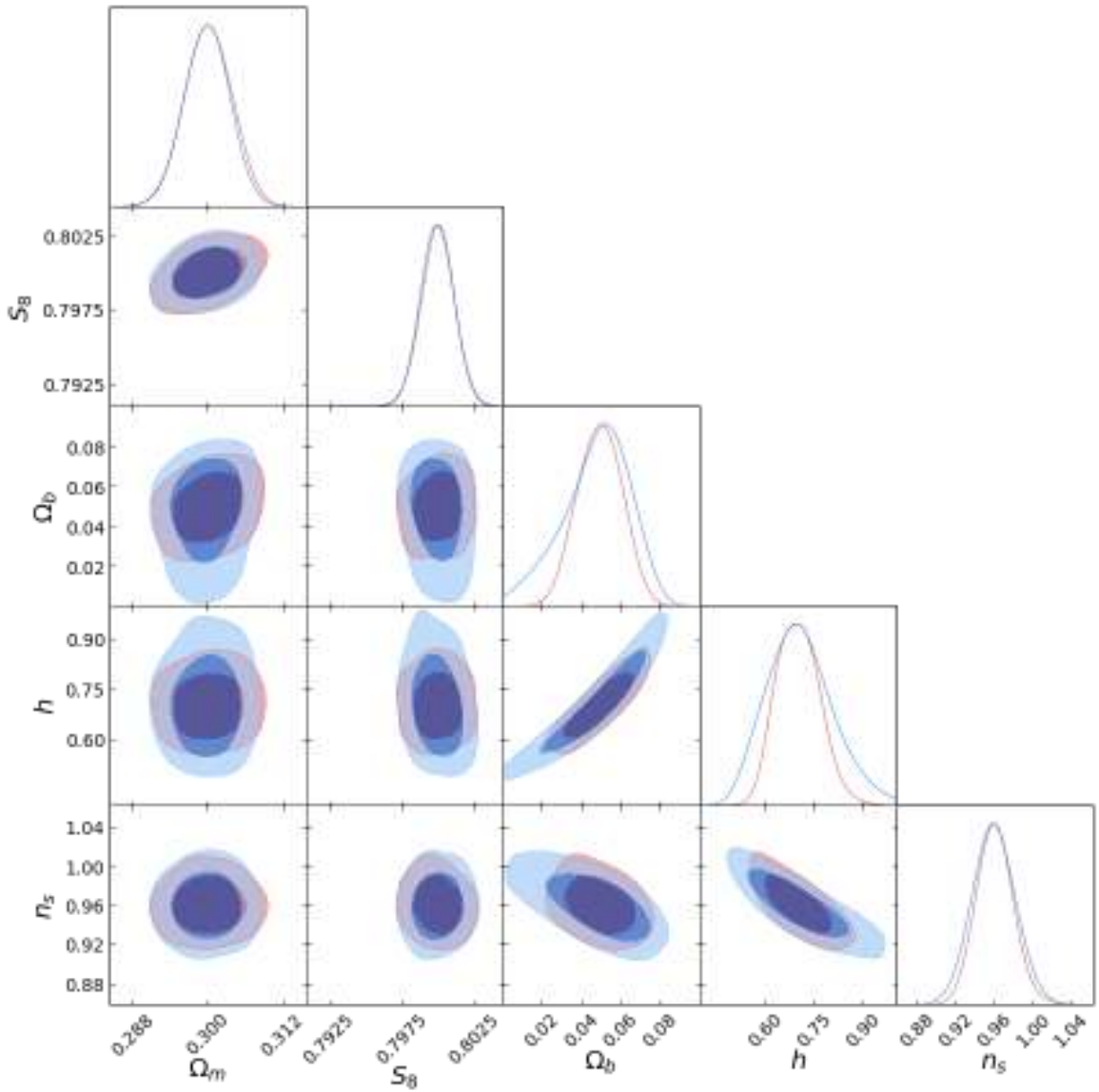
Reduce necessary simulations

Full MCMC $\sim$ 10^6 simulations

DELFI $\sim$ few $\times$ 10^3 simulations

MCMC vs DELFI constraints

Note these results are MOPED compression (not IMNN)



Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Totally ignorant inference

What I've just presented

An optimal compression scheme for general likelihoods

A non-linear compression network trained with the Fisher information

A couple of amazing (albeit toy) examples of how well the IMNN works

A way to drastically reduce the number of simulations needed to do inference using mixture density networks

A framework with which to do totally ignorant statistical inference

Machine learning based (likelihood-free) statistical inference

Tom Charnock
Guilhem Lavaux & Ben Wandelt

Alessandro Manzotti
Justin Alsing

Institut d'Astrophysique de Paris

Physical Review D **97**, 083004

DOI 10.5281/zenodo.1175196

github:information_maximiser