LLM-Era Compute for 21-cm Cosmology:

Accelerating Bayesian Inference for the SKA Era

Jacob Tutt[1,2]

[1]Cavendish Astrophysics, University of Cambridge, UK
[2]Kavli Institute for Cosmology, Cambridge, UK

Bayesian Inference

$$\mathcal{P}(\theta \mid D, M) = \frac{p(D \mid \theta, M)\, p(\theta \mid M)}{\int p(D \mid \theta, M)\, p(\theta \mid M)\, d\theta} = \frac{\mathcal{L}(D \mid \theta, M)\, \Pi(\theta \mid M)}{Z(M)}$$

Prior, $\Pi(\theta \mid M)$

Describes our knowledge/ assumptions of the parameters θ *prior* to any data.

Bayesian Inference

$$\mathcal{P}(\theta \mid D, M) = \frac{p(D \mid \theta, M)\, p(\theta \mid M)}{\int p(D \mid \theta, M)\, p(\theta \mid M)\, d\theta} = \frac{\mathcal{L}(D \mid \theta, M)\, \Pi(\theta \mid M)}{Z(M)}$$

Prior, $\Pi(\theta \mid M)$

Describes our knowledge/ assumptions of the parameters θ *prior* to any data.

Likelihood, $\mathcal{L}(D \mid \theta, M)$

Quantifies how well a parameter choice θ explains the observed data D.

Bayesian Inference

$$\mathcal{P}(\theta \mid D, M) = \frac{p(D \mid \theta, M)\, p(\theta \mid M)}{\int p(D \mid \theta, M)\, p(\theta \mid M)\, d\theta} = \frac{\mathcal{L}(D \mid \theta, M)\, \Pi(\theta \mid M)}{Z(M)}$$

Prior, $\Pi(\theta \mid M)$

Describes our knowledge/ assumptions of the parameters θ *prior* to any data.

Likelihood, $\mathcal{L}(D \mid \theta, M)$

Quantifies how well a parameter choice θ explains the observed data D.

Posterior, $\mathcal{P}(\theta \mid D, M)$

An updated state of belief about the parameters θ after incorporating the data.

Bayesian Inference

$$\mathcal{P}(\theta \mid D, M) = \frac{p(D \mid \theta, M)\, p(\theta \mid M)}{\int p(D \mid \theta, M)\, p(\theta \mid M)\, d\theta} = \frac{\mathcal{L}(D \mid \theta, M)\, \Pi(\theta \mid M)}{Z(M)}$$

Prior, $\Pi(\theta \mid M)$

Describes our knowledge/ assumptions of the parameters θ *prior* to any data.

Likelihood, $\mathcal{L}(D \mid \theta, M)$

Quantifies how well a parameter choice θ explains the observed data D.

Posterior, $\mathcal{P}(\theta \mid D, M)$

An updated state of belief about the parameters θ after incorporating the data.

Evidence, $Z(M)$

The total support the data provides for a model. Crucial for model comparison and validation.

The Challenge in Radio Astronomy

SKA-Era Inference Problem

- ▷ Petabyte-scale datasets:

$$\boldsymbol{D}_{b,\nu,t,p}$$

- ▷ High-dimensional Models:
  - ▷ Foregrounds
  - ▷ Beam
  - ▷ Calibration
  - ▷ 21-cm Signal

$$\boldsymbol{M}(\theta) \qquad \theta \in \mathbb{R}^{10^5-10^6}$$

- ▷ Run-time accumulation:

$$\log \mathcal{L}(\boldsymbol{D} \mid \theta) = \sum_{b,\nu,t,p} \log \mathcal{L}\left(D_{b,\nu,t,p} \mid \theta\right)$$

The Challenge in Radio Astronomy

SKA-Era Inference Problem

- Petabyte-scale datasets:

$$\boldsymbol{D}_{b,\nu,t,p}$$

- High-dimensional Models:
  - Foregrounds
  - Beam
  - Calibration
  - 21-cm Signal

$$\boldsymbol{M}(\theta) \qquad \theta \in \mathbb{R}^{10^5-10^6}$$

- Run-time accumulation:

$$\log \mathcal{L}(\boldsymbol{D} \mid \theta) = \sum_{b,\nu,t,p} \log \mathcal{L}\left(D_{b,\nu,t,p} \mid \theta\right)$$

Statistical Inference

- Traditional sampling: many likelihood evaluations
- SBI: many simulated datasets
- **Inference cost grows rapidly**

Radio Astronomy ↔ Accelerators

Radio Astronomy

- Bulk linear algebra operations:
  - Matrix multplications
  - Fourier transforms
  - Matrix inversions
- Intrinsic batch dimensions:
  - Baselines (b)
  - Time (t)
  - Frequency (ν)
  - Polarisation (p)

Radio Astronomy $\leftrightarrow$ Accelerators

Radio Astronomy

- Bulk linear algebra operations:
  - Matrix multplications
  - Fourier transforms
  - Matrix inversions
- Intrinsic batch dimensions:
  - Baselines (b)
  - Frequency (ν)
  - Time (t)
  - Polarisation (p)

LLM-era Accelerators

- Transformer architectures:
  - Dense matrix multiplication
  - Specialised tensor cores
- High-bandwidth memory
- Native multi-device scaling

Radio Astronomy $\leftrightarrow$ Accelerators

Radio Astronomy

- Bulk linear algebra operations:
  - Matrix multplications
  - Fourier transforms
  - Matrix inversions
- Intrinsic batch dimensions:
  - Baselines (b)
  - Frequency (ν)
  - Time (t)
  - Polarisation (p)

LLM-era Accelerators

- Transformer architectures:
  - Dense matrix multiplication
  - Specialised tensor cores
- High-bandwidth memory
- Native multi-device scaling

Accelerator Software

- High-level array programming, low-level kernel execution
- Accelerator-level compilation without writing CUDA

Radio Astronomy ↔ Accelerators

Radio Astronomy

- Bulk linear algebra operations:
  - Matrix multplications
  - Fourier transforms
  - Matrix inversions
- Intrinsic batch dimensions:
  - Baselines (b)
  - Frequency (ν)
  - Time (t)
  - Polarisation (p)

LLM-era Accelerators

- Transformer architectures:
  - Dense matrix multiplication
  - Specialised tensor cores
- High-bandwidth memory
- Native multi-device scaling

Accelerator Software

- High-level array programming, low-level kernel execution
- Accelerator-level compilation without writing CUDA

Takeaway

- Radio-astronomy inference maps naturally to LLM-era compute
- Redefining the scale and complexity with which we can perform inference

Next-Generation AI Supercomputers

Isambard-AI

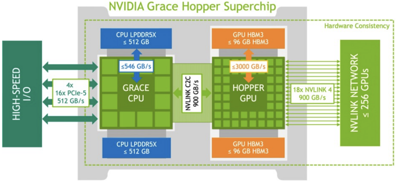
- 1,320 nodes overall
- 5,280 GH200 superchips

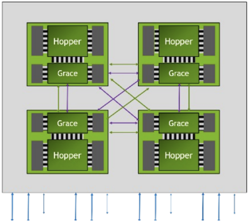
GH200 Node

- 4x Grace ARM CPUs
  - 288 CPU cores
  - 512 GB CPU memory
- 4x Hopper GPUs
  - 384 GB high-bandwidth memory
- 896 GB total memory

Future Directions

- Google TPUs
- Promising for SKA-scale inference





Source: NVIDIA Grace Hopper Superchip Architecture

An Introduction to JAX

High-performance numerical-computing and large-scale machine learning

High-level array code
⇓
JAX tracing
⇓
XLA compilation
⇓
Hardware-optimised kernels
⇓
CPU / GPU / TPU execution

High-performance numerical-computing and large-scale machine learning

High-level array code
⇓
JAX tracing
⇓
XLA compilation
⇓
Hardware-optimised kernels
⇓
CPU / GPU / TPU execution

XLA Compilation

- Compiles high-level array code into optimised machine code
- Combines Python productivity with compiled-performance execution $f(x) \implies \texttt{jax.jit}(f)(x)$

An Introduction to JAX

High-performance numerical-computing and large-scale machine learning

High-level array code
⇓
JAX tracing
⇓
XLA compilation
⇓
Hardware-optimised kernels
⇓
CPU / GPU / TPU execution

XLA Compilation

- Compiles high-level array code into optimised machine code
- Combines Python productivity with compiled-performance execution $f(x) \implies \texttt{jax.jit}(f)(x)$

Parallelisation/ Distribution

- **Vectorisation**: automatic parallelisation over batches of data
- **Sharding**: distributing arrays and workloads across multi-GPU / multi-TPU architectures

An Introduction to JAX

High-performance numerical-computing and large-scale machine learning

High-level array code
⇓
JAX tracing
⇓
XLA compilation
⇓
Hardware-optimised kernels
⇓
CPU / GPU / TPU execution

XLA Compilation

- Compiles high-level array code into optimised machine code
- Combines Python productivity with compiled-performance execution $f(x) \implies \texttt{jax.jit}(f)(x)$

Parallelisation/ Distribution

- **Vectorisation**: automatic parallelisation over batches of data
- **Sharding**: distributing arrays and workloads across multi-GPU / multi-TPU architectures

Automatic Differentiation

$$(\nabla f)(x)_i = \frac{\partial f}{\partial x_i}(x) \implies \texttt{jax.grad(f)(x)}$$

JacobTutt/dual_autodiff_package

UNIVERSITY OF CAMBRIDGE
Cavendish Laboratory
Department of Physics
REACH
JacobTutt

JAX Ecosystem

JAX

Optax

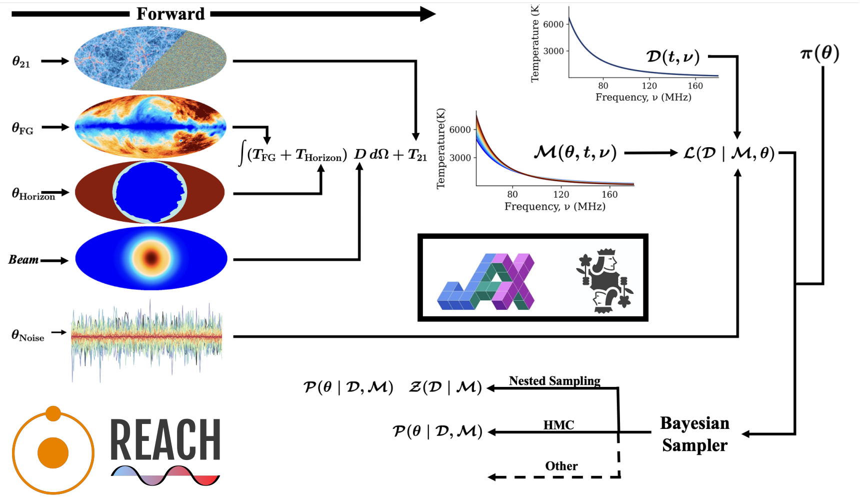
BlackJAX

Flax

XLA

JAXtronomy

Case Study I: Global 21-cm Cosmology

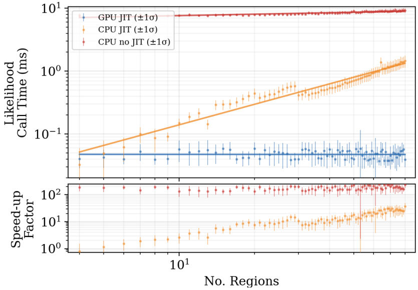


with P. Sims, J. Pattison, D. Anstey, S. Leeney and E. de Lera Acedo

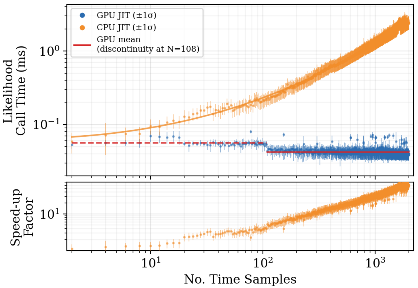
Source: arXiv:2603.13196

REACH: Scaling Behaviour

Model Complexity



Data Volume



*GPU: NVIDIA A100 (40GB) *CPU: Intel Cascade Lake CPU

What this means for REACH:

- Full first-year early dataset fits: ∼4000 spectra at near single-spectrum cost
- More complex foreground models for higher-resolution low-frequency sky maps

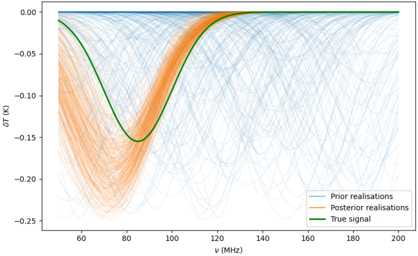
BaNTER Validation Framework

Motivation

▷ Two hypothesised models $M_{\mathrm{FG}}/M_{\mathrm{FG+21}}$:

$$\ln B_{\det} = \ln\left(\frac{\mathcal{Z}_{\mathrm{FG+21}}}{\mathcal{Z}_{\mathrm{FG}}}\right)$$

▷ A high detection Bayes factor alone is not enough



UKSRC Validation Team

with P. Sims, J. Pattison, D. Anstey, S. Leeney and E. de Lera Acedo

Sources: arXiv:2603.13196; arXiv:2502.14029

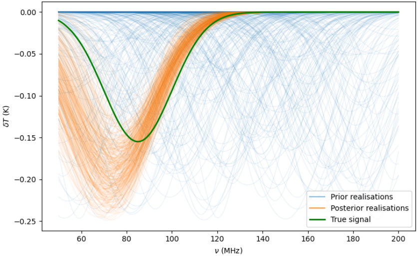
BaNTER Validation Framework

Motivation

▷ Two hypothesised models $M_{\mathrm{FG}}/M_{\mathrm{FG+21}}$:

$$\ln B_{\mathrm{det}} = \ln\left(\frac{\mathcal{Z}_{\mathrm{FG+21}}}{\mathcal{Z}_{\mathrm{FG}}}\right)$$

▷ A high detection Bayes factor alone is not enough



Metric 1: Null Test

$$\ln B_{\mathrm{val}} = \ln\left(\frac{\mathcal{Z}^{\mathrm{v}}_{\mathrm{FG+21}}}{\mathcal{Z}^{\mathrm{v}}_{\mathrm{FG}}}\right)$$

Fit signal-free validation data

▷ Fail if $\ln B_{\mathrm{val}} \geq 0$

UKSRC Validation Team

with P. Sims, J. Pattison, D. Anstey, S. Leeney and E. de Lera Acedo

Sources: arXiv:2603.13196; arXiv:2502.14029

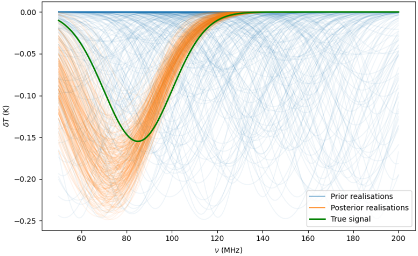
BaNTER Validation Framework

Motivation

▷ Two hypothesised models $M_{\mathrm{FG}}/M_{\mathrm{FG+21}}$:

$$\ln B_{\mathrm{det}} = \ln\left(\frac{\mathcal{Z}_{\mathrm{FG+21}}}{\mathcal{Z}_{\mathrm{FG}}}\right)$$

▷ A high detection Bayes factor alone is not enough



Metric 1: Null Test

$$\ln B_{\mathrm{val}} = \ln\left(\frac{\mathcal{Z}^{\mathrm{v}}_{\mathrm{FG+21}}}{\mathcal{Z}^{\mathrm{v}}_{\mathrm{FG}}}\right)$$

Fit signal-free validation data

▷ Fail if $\ln B_{\mathrm{val}} \geq 0$

Metric 2: Residual Structure

$$q_i = \mathbb{P}\left(\mathcal{L}_{\mathrm{noise}} \leq \overline{\mathcal{L}}_i\right)$$

Ask whether the residuals are noise-like

▷ Require $q_i \geq q_{\mathrm{threshold}}$

UKSRC Validation Team

with P. Sims, J. Pattison, D. Anstey, S. Leeney and E. de Lera Acedo

Sources: arXiv:2603.13196; arXiv:2502.14029

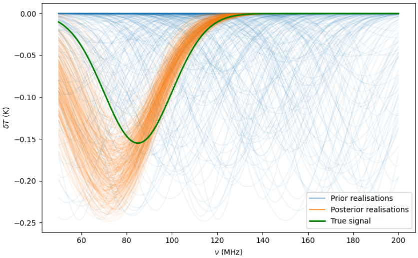
BaNTER Validation Framework

Motivation

▷ Two hypothesised models $M_{\mathrm{FG}}/M_{\mathrm{FG+21}}$:

$$\ln B_{\mathrm{det}} = \ln\left(\frac{\mathcal{Z}_{\mathrm{FG+21}}}{\mathcal{Z}_{\mathrm{FG}}}\right)$$

▷ A high detection Bayes factor alone is not enough



Metric 1: Null Test

$$\ln B_{\mathrm{val}} = \ln\left(\frac{\mathcal{Z}^{\mathrm{v}}_{\mathrm{FG+21}}}{\mathcal{Z}^{\mathrm{v}}_{\mathrm{FG}}}\right)$$

Fit signal-free validation data

▷ Fail if $\ln B_{\mathrm{val}} \geq 0$

Metric 2: Residual Structure

$$q_i = \mathbb{P}\left(\mathcal{L}_{\mathrm{noise}} \leq \overline{\mathcal{L}}_i\right)$$

Ask whether the residuals are noise-like

▷ Require $q_i \geq q_{\mathrm{threshold}}$

Evidence Evaluation

BlackJAX Nested Slice Sampling

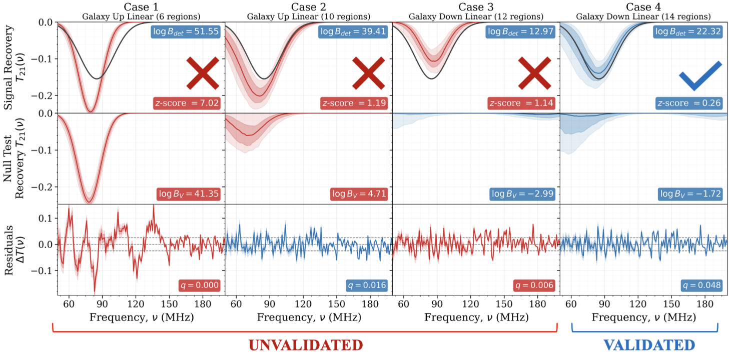
BlackjaxNSS

UKSRC Validation Team

with P. Sims, J. Pattison, D. Anstey, S. Leeney and E. de Lera Acedo

Sources: arXiv:2603.13196; arXiv:2502.14029

Validation Across Model Configurations



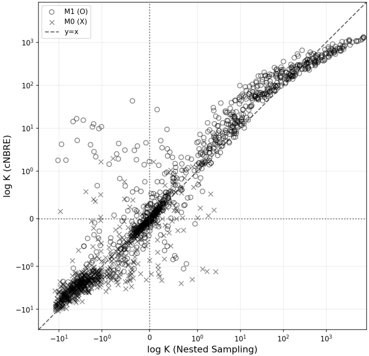
Key Takeaways

- Validated recovery rejects cases that ln $B_{\rm det}$ alone would accept
- 1052 nested-sampling runs:
  - $\sim$ 100 CPU years $\rightarrow$ under 2 GPU days ($\sim$ 100$\times$ lower cost)

with P. Sims, J. Pattison, D. Anstey, S. Leeney and E. de Lera Acedo

Sources: arXiv:2603.13196; arXiv:2502.14029

Opening the Door to SBI



Conditional Bayes Neural Ratio Estimation

- ▷ **Sub-ms forward model**
- ▷ **What this enables for SBI**
  - ○ Dynamic simulations during training
  - ○ End-to-end gradients
  - ○ No fixed simulation set:
    - ⇒ Better coverage of the prior
    - ⇒ Less overfitting

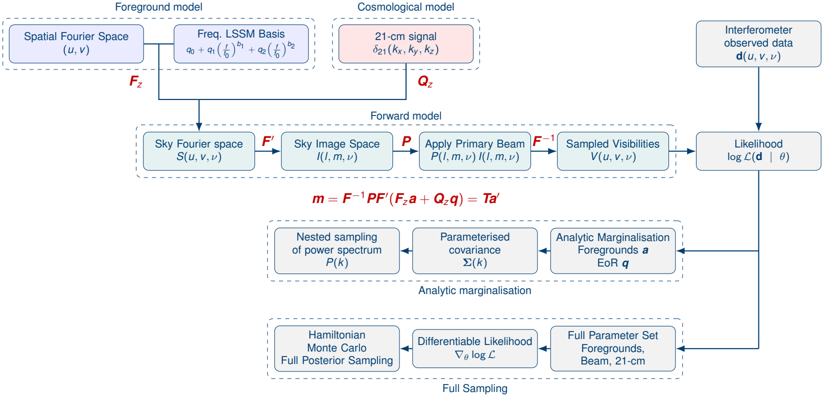
with **S. Leeney**, T. Gessey-Jones, W. Handley, E. de Lera Acedo, H. Bevins

Source: Leeney et al. (in Prep)

Case Study II: BayesEoR

BayesEoR Approach

- ▷ Foreground Reconstruction
- ▷ 21-cm Power Spectrum Analysis



with P. Sims, D. Anstey and E. de Lera Acedo

Takeaways

Hardware

Radio-astronomy inference workloads map naturally onto accelerator systems developed for AI: massively parallel, tensor-core, multi-device hardware.

Takeaways

Hardware

Radio-astronomy inference workloads map naturally onto accelerator systems developed for AI: massively parallel, tensor-core, multi-device hardware.

Software

Modern software libraries, such as JAX, significantly reduce the barrier to using these architectures effectively.

Takeaways

Hardware

Radio-astronomy inference workloads map naturally onto accelerator systems developed for AI: massively parallel, tensor-core, multi-device hardware.

Software

Modern software libraries, such as JAX, significantly reduce the barrier to using these architectures effectively.

Impact

Acceleration of the REACH and BayesEoR Pipelines illustrate how these methods are changing the speed, scale, complexity and rigor of radio-astronomy inference.

Questions

Thank you for your attention!

Jacob Tutt

Cavendish Astrophysics, University of Cambridge

Kavli Institute for Cosmology, Cambridge,

`jlt67@cam.ac.uk`

Acceleration of BayesEoR

Forward-Operator Construction

- ▷ Build the dense operator:

$$\mathbf{T} = \mathbf{F}^{-1}\mathbf{P}\mathbf{F}'[\mathbf{F}_z\ \mathbf{Q}_z]$$

  - ▷ Discrete Fourier Transforms
  - ▷ Beam Response
  - ▷ LSSM Basis Terms
  - ▷ Reindexing
- ▷ Construct **T** through compositions of dense linear operators
  - ▷ Sharded across multiple GPUs (e.g. frequency, baselines)

Inference-Time Acceleration

- ▷ Just-in-time compilation
- ▷ Reduce large overheads of Analytical marginalisation
- ▷ JAX autodiff → Gradients for free → Enables HMC

Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_{\Theta} \mathcal{L}(\theta)\,\pi(\theta)\,d\theta = \int_0^1 \mathcal{L}(\xi)\,d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)\,d\theta \implies \text{inverse relation: } \mathcal{L}(\xi)$$

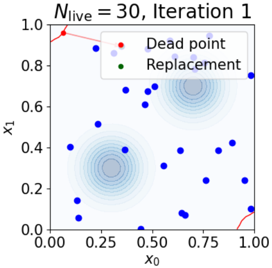
Shrinkage of the prior mass:

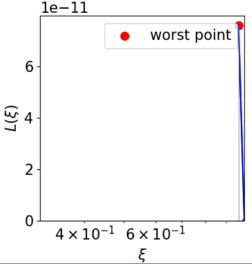
$$\xi_i = t\ \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\text{live}}} \implies \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i\left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right)\right]$$





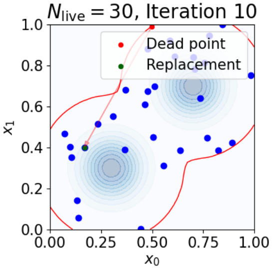
Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_\Theta \mathcal{L}(\theta)\,\pi(\theta)\,d\theta = \int_0^1 \mathcal{L}(\xi)\,d\xi.$$



Prior volume (mass):

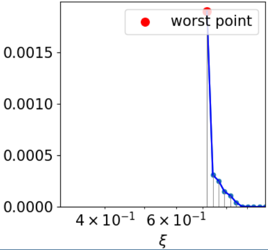
$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)\,d\theta \implies \text{inverse relation: } \mathcal{L}(\xi)$$

Shrinkage of the prior mass:

$$\xi_i = t\ \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\rm live}} \implies \xi_i \approx \exp\left(-\frac{i}{N_{\rm live}}\right)$$



$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i\left[\exp\left(-\frac{i-1}{N_{\rm live}}\right) - \exp\left(-\frac{i+1}{N_{\rm live}}\right)\right]$$

Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_\Theta \mathcal{L}(\theta)\,\pi(\theta)\,d\theta = \int_0^1 \mathcal{L}(\xi)\,d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)\,d\theta \implies \text{inverse relation: } \mathcal{L}(\xi)$$

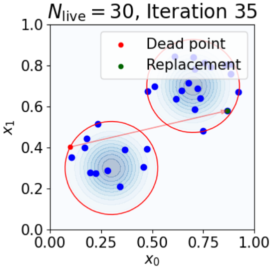
Shrinkage of the prior mass:

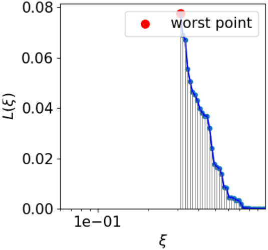
$$\xi_i = t\ \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\text{live}}} \implies \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i\left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right)\right]$$

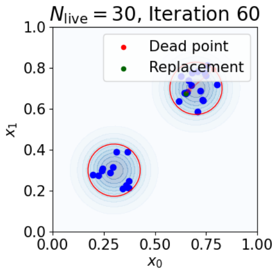




Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

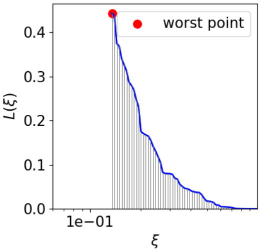


Evidence integral:

$$Z = \int_{\Theta} \mathcal{L}(\theta)\, \pi(\theta)\, d\theta = \int_0^1 \mathcal{L}(\xi)\, d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta) > \mathcal{L}} \pi(\theta)\, d\theta \implies \text{inverse relation: } \mathcal{L}(\xi)$$



Shrinkage of the prior mass:

$$\xi_i = t \;\; \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log \;\; t] = -\frac{1}{N_{live}} \implies \xi_i \approx \exp\left(-\frac{i}{N_{live}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i \left[\exp\left(-\frac{i-1}{N_{live}}\right) - \exp\left(-\frac{i+1}{N_{live}}\right) \right]$$

Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_{\Theta} \mathcal{L}(\theta)\, \pi(\theta)\, d\theta = \int_0^1 \mathcal{L}(\xi)\, d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta) > \mathcal{L}} \pi(\theta)\, d\theta \quad \Longrightarrow \quad \text{inverse relation: } \mathcal{L}(\xi)$$

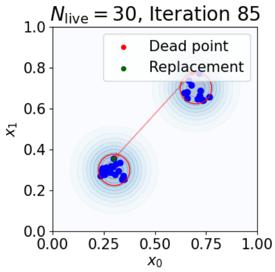
Shrinkage of the prior mass:

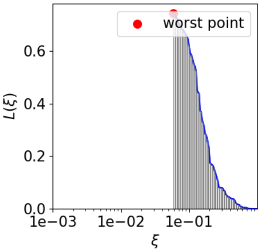
$$\xi_i = t \;\, \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\text{live}}} \quad \Longrightarrow \quad \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i \left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right) \right]$$





Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_{\Theta} \mathcal{L}(\theta)\, \pi(\theta)\, d\theta = \int_0^1 \mathcal{L}(\xi)\, d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta) > \mathcal{L}} \pi(\theta)\, d\theta \quad \Longrightarrow \quad \text{inverse relation: } \mathcal{L}(\xi)$$

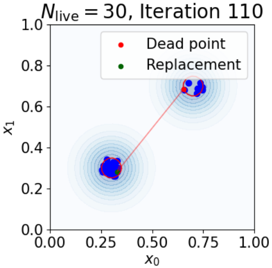
Shrinkage of the prior mass:

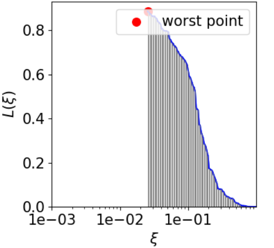
$$\xi_i = t \;\; \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\;\; t] = -\frac{1}{N_{\text{live}}} \quad \Longrightarrow \quad \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i \left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right) \right]$$





Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_{\Theta} \mathcal{L}(\theta)\,\pi(\theta)\,d\theta = \int_0^1 \mathcal{L}(\xi)\,d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)\,d\theta \quad \Longrightarrow \quad \text{inverse relation: } \mathcal{L}(\xi)$$

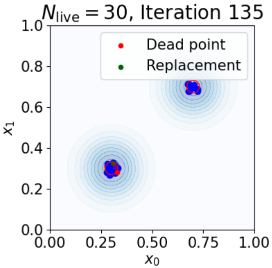
Shrinkage of the prior mass:

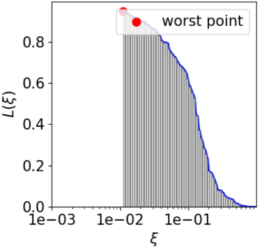
$$\xi_i = t\ \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\text{live}}} \quad \Longrightarrow \quad \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$\boldsymbol{Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i\left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right)\right]}$$





Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_{\Theta} \mathcal{L}(\theta)\, \pi(\theta)\, d\theta = \int_0^1 \mathcal{L}(\xi)\, d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta) > \mathcal{L}} \pi(\theta)\, d\theta \implies \text{inverse relation: } \mathcal{L}(\xi)$$

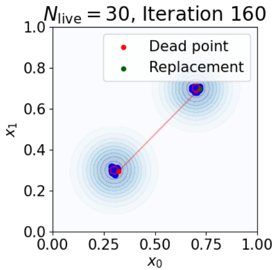
Shrinkage of the prior mass:

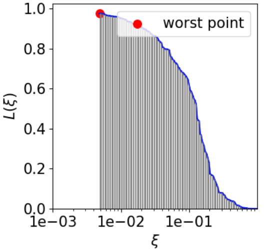
$$\xi_i = t\ \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\text{live}}} \implies \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i \left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right) \right]$$





Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_{\Theta} \mathcal{L}(\theta)\,\pi(\theta)\,d\theta = \int_0^1 \mathcal{L}(\xi)\,d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)\,d\theta \implies \text{inverse relation: } \mathcal{L}(\xi)$$

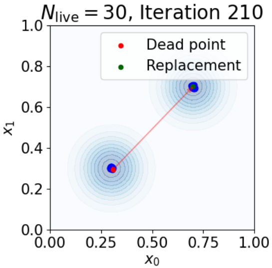
Shrinkage of the prior mass:

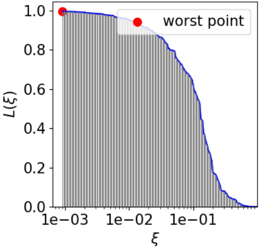
$$\xi_i = t\ \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\text{live}}} \implies \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i\left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right)\right]$$





Traditional Nested Sampling

Nested Sampling provides:

- Posterior samples
- **The Bayesian Evidence**

Evidence integral:

$$Z = \int_\Theta \mathcal{L}(\theta)\,\pi(\theta)\,d\theta = \int_0^1 \mathcal{L}(\xi)\,d\xi.$$

Prior volume (mass):

$$\xi(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)\,d\theta \implies \text{inverse relation: } \mathcal{L}(\xi)$$

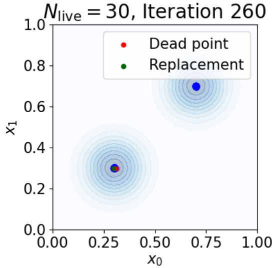
Shrinkage of the prior mass:

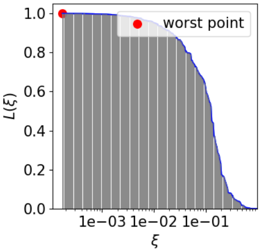
$$\xi_i = t\ \xi_{i-1},$$

Expected shrinkage:

$$\mathbb{E}[log\ t] = -\frac{1}{N_{\text{live}}} \implies \xi_i \approx \exp\left(-\frac{i}{N_{\text{live}}}\right)$$

$$Z \approx \sum_i \mathcal{L}_i(\xi_{i-1} - \xi_i) = \sum_i \mathcal{L}_i\left[\exp\left(-\frac{i-1}{N_{\text{live}}}\right) - \exp\left(-\frac{i+1}{N_{\text{live}}}\right)\right]$$





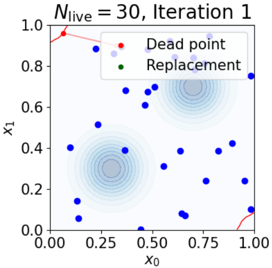
Accelerated Nested Sampling

Traditional nested sampling (serial):

- Remove one 'worst' live point each iteration.
- New point conditioned on:

$$\mathcal{L} > \mathcal{L}_{\min}$$

- Inherently sequential MCMC slice samples
- `PolyChord - Handley et al 2025`

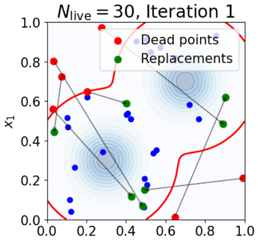


Accelerated nested sampling (parallel):

- Discard a batch of n_{del} 'worst' points.
- All replacements conditioned on:

$$\mathcal{L}(\theta) > \mathcal{L}_{\min} \quad \mathcal{L}_{\min} = \max\{\mathcal{L}_1, \ldots, \mathcal{L}_{n_{\text{del}}}\}$$

- Sample each replacement independently
  $\Rightarrow$ vectorised (vmap) across GPU
- `Blackjax`
  - `Cabezas at al 2024, Yallup et al 2025`



Accelerated Nested Sampling

Traditional nested sampling (serial):

- Remove one ‘worst’ live point each iteration.
- New point conditioned on:

$$\mathcal{L} > \mathcal{L}_{\text{min}}$$

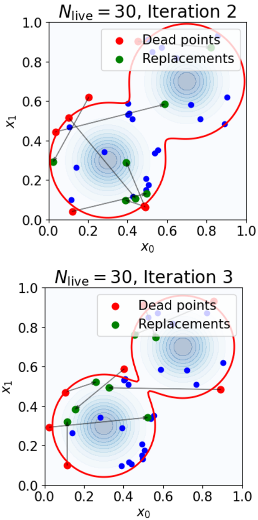
- Inherently sequential MCMC slice samples
- `PolyChord - Handley et al 2025`

Accelerated nested sampling (parallel):

- Discard a batch of n_{del} ‘worst’ points.
- All replacements conditioned on:

$$\mathcal{L}(\theta) > \mathcal{L}_{\text{min}} \quad \mathcal{L}_{\text{min}} = \max\{\mathcal{L}_1, \ldots, \mathcal{L}_{n_{\text{del}}}\}$$

- Sample each replacement independently
  $\Rightarrow$ vectorised (vmap) across GPU
- `Blackjax`
  - `Cabezas at al 2024, Yallup et al 2025`



Workshop Summary

Scientific Drivers

- 21-cm cosmology demands high-dimensional inference over foreground, instrumental, and cosmological models across large datasets.
- LLM-era hardware and software investment is changing what is computationally feasible in science.
- In both REACH and BayesEoR Pipelines, GPU conversion is changing what is feasible in model complexity, data volume, validation, and financial cost.
- This applies both to simulation-based inference and to accelerating traditional Bayesian sampling.

Technical Enablers

- Core stack: `JAX`, `XLA`, `BlackJAX`, `Optax`, `Flax`, `Distrax`
- Infrastructure: Google Compute Engine (`NVIDIA A100`), `AIRR` / `Isambard-AI` (`NVIDIA GH200`)

Policy and Best Practice

- Inference efficiency matters not only scientifically, but also financially and environmentally.