Exploring Phase Space with Nested Sampling

David Yallup (dy297@cam.ac.uk)

Work based on [2205.02030] w. T. Janßen, S. Schumann, W. Handley

Bayesian Inference in High Energy Physics, Durham University

25-27th May 2022

UNIVERSITY OF CAMBRIDGE

Exploring phase space?

$$\sigma = \int_{\Omega} d\Phi |\mathcal{M}|^2(\Phi)$$

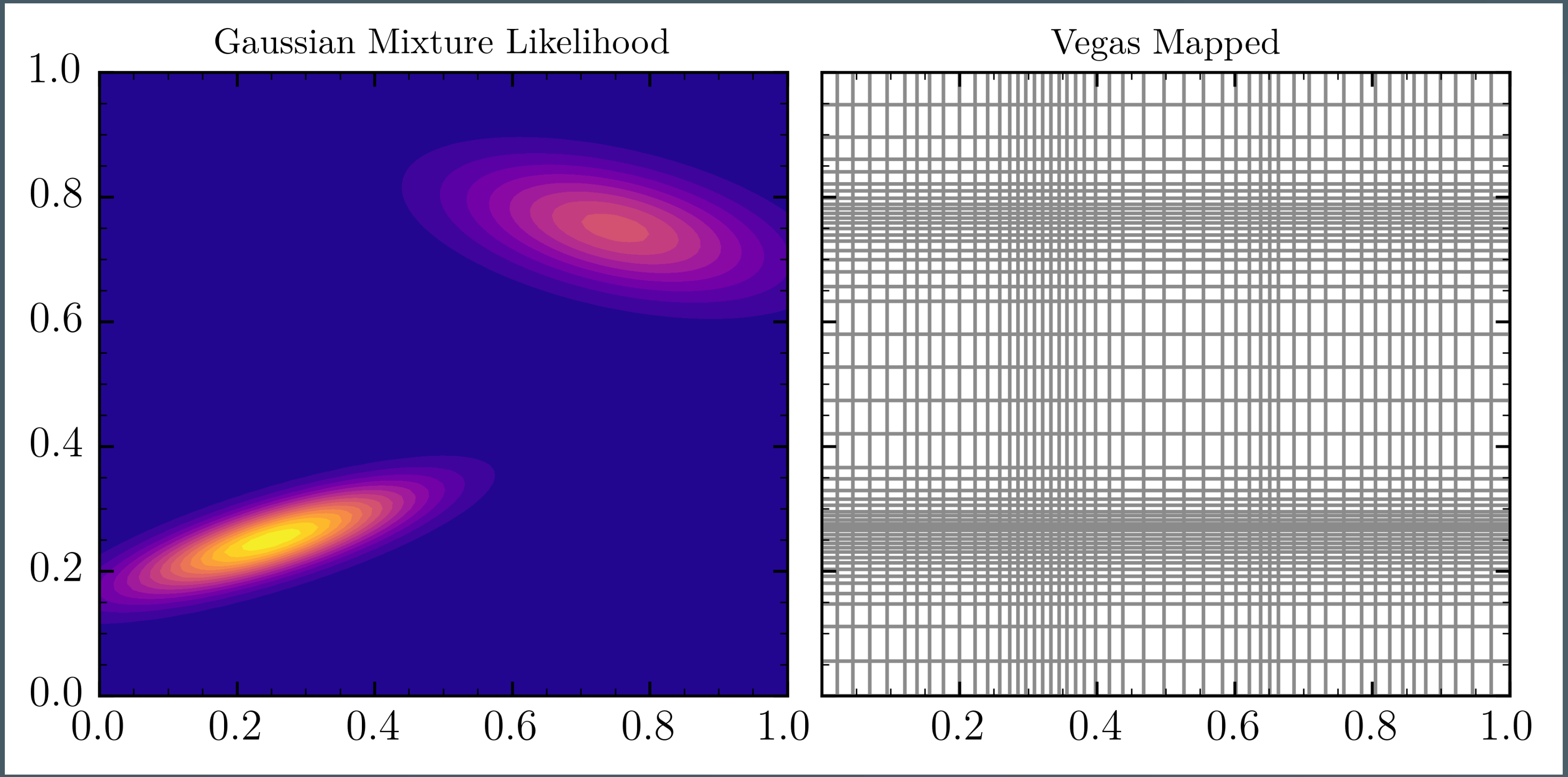
Cross section = integral$_{\text{(over kinematic variables)}}$ (Matrix Element)

Central challenge for many physics tasks:

- Total cross section (σ) - Probability of process occuring
- Differential cross section $(d\sigma)$ - chunk integral into $d\Phi$ pieces
- Events - (unweight) and use as pseudo data

Workhorse in HEP on this set of problems is Importance Sampling

- Replace problem of sampling from unknown $P(\Phi)$ with a known $Q(\Phi)$
- Adjust importance of sample drawn from Q by weighting, $w = \frac{P(\Phi)}{Q(\Phi)}$



Problem seemingly reduces to coming up with good mappings for target

However, Even in $D = \mathcal{O}(10)$ Dimensions this starts to break.

- Massless glue scattering, $D = 3n_g - 4$:
  - $gg \to 3g, D = 5$
  - $gg \to 4g, D = 8$

Even modern ML (normalising flows) won't save you [2001.05478]

Algorithm	Efficiency $gg \to 3g$	Efficiency $gg \to 4g$
HAAG	3.0%	2.7%
Vegas	27.7%	31.8%
Neural Network	64.3%	33.6%

A sampling problem? Anyone for Bayes?

Central problem:

- Convergent integral means you have good posterior samples
- **Reverse not true**, Samples from a convergent MCMC chain **not** guaranteed a good integral
- Multimodal targets well established failure mode.
  - Multichannel decompositions in MCMC HEP, $(\mathrm{MC})^3$ [1404.4328]

$$P(\Phi) = \frac{\mathcal{L}(\Phi)\Pi(\Phi)}{\mathcal{Z}} \propto \mathcal{L}(\Phi)\Pi(\Phi)$$

MCMC kicks in as we go to high dimensions, grey area between IS and MCMC, can ML help?

Where's the Evidence?

In neglecting the Evidence ($\mathcal{Z}$) we have neglected precisely the quantity we want,

$$\sigma = \int_{\Omega} d\Phi |\mathcal{M}|^2(\Phi)$$

$$\mathcal{Z} = \int d\theta \mathcal{L}(\theta)\Pi(\theta)$$

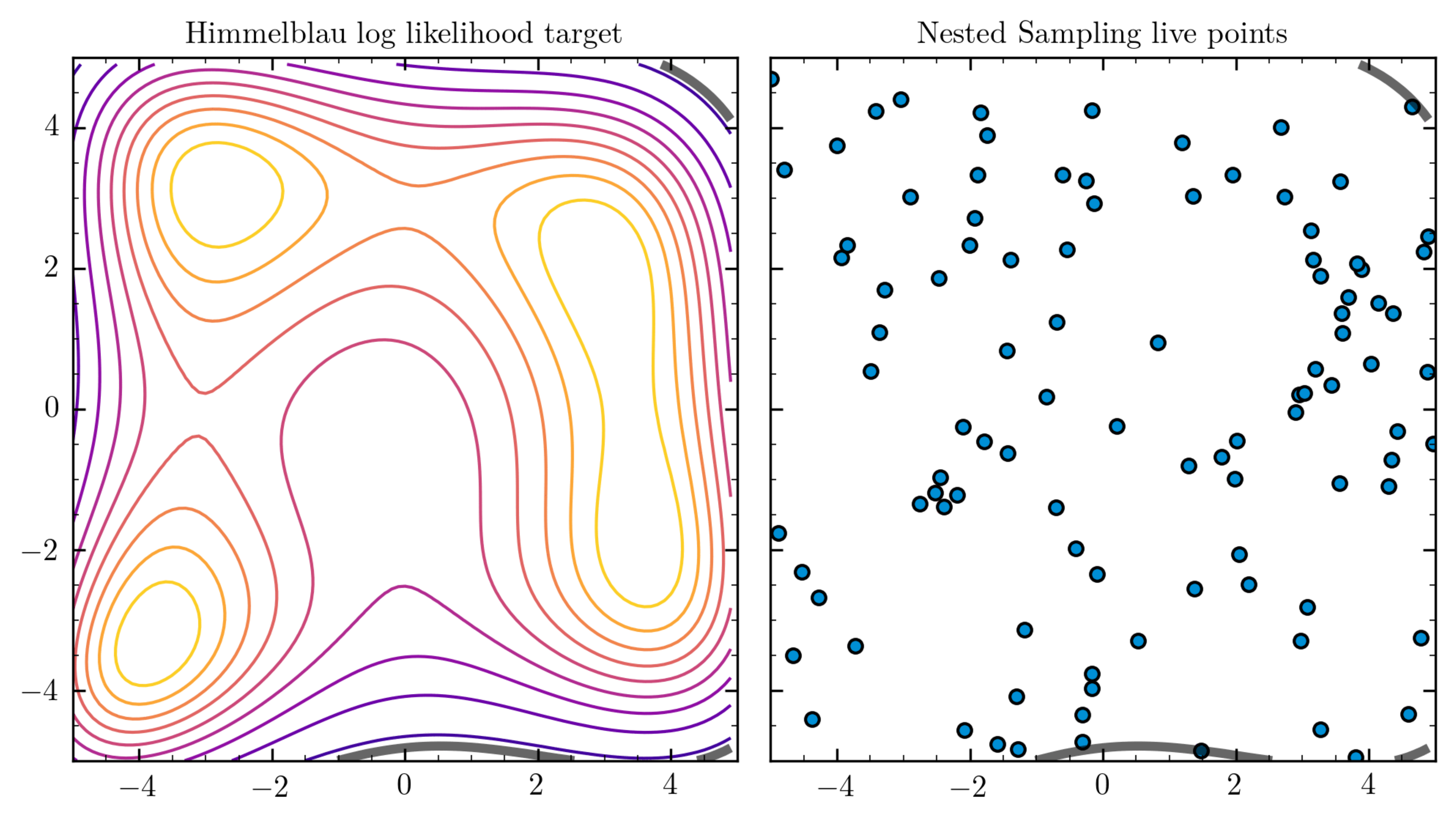
- Mapping $\rightarrow$ Prior
- Matrix element $\rightarrow$ Likelihood
- Cross section $\rightarrow$ Evidence

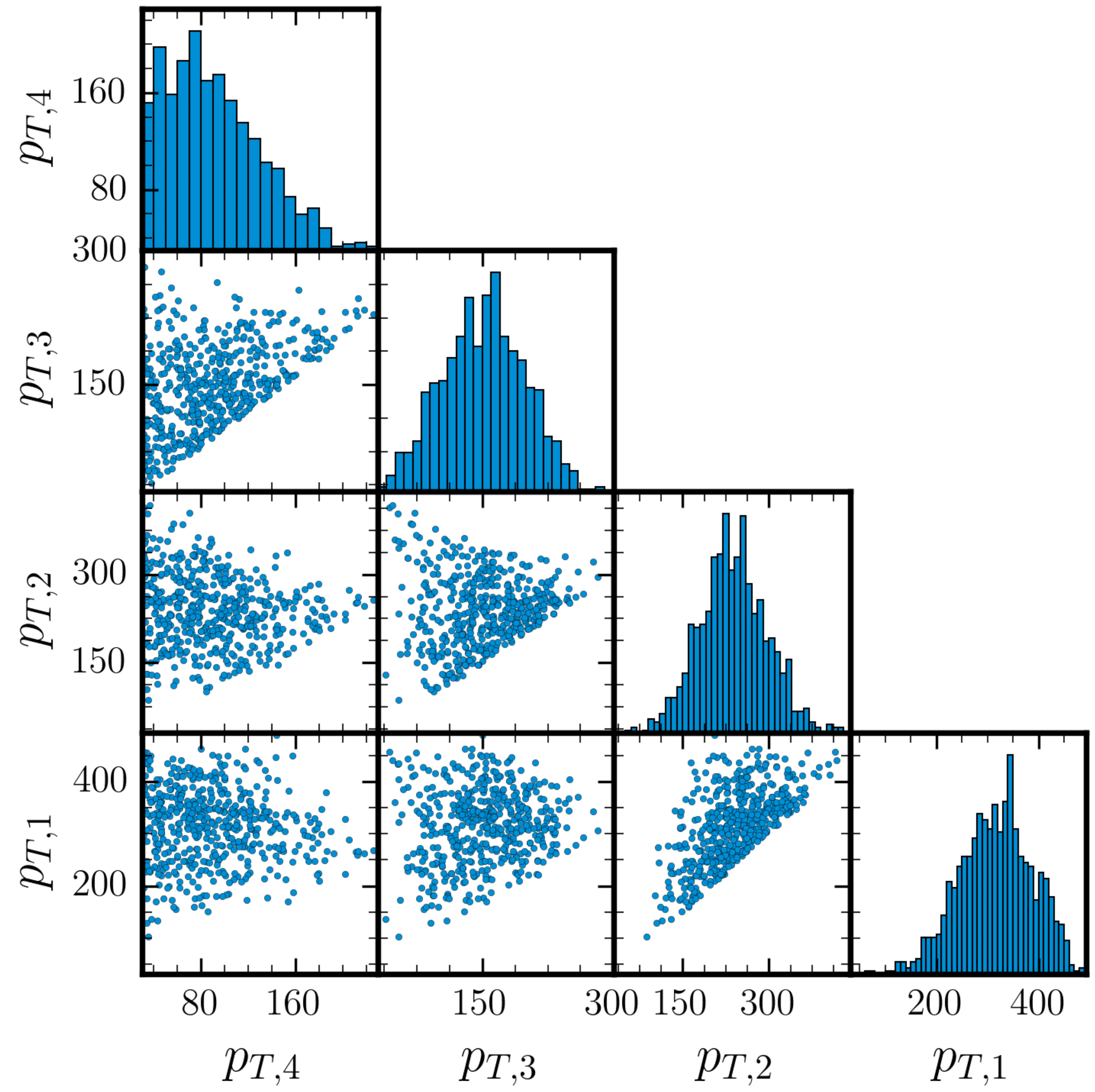
Nested Sampling

Nested Sampling [Skilling 2006], implemented for in PolyChord [1506.00171]. Is a good way to generically approach this problem for $\mathcal{O}(10) \rightarrow \mathcal{O}(100)$ dimensions

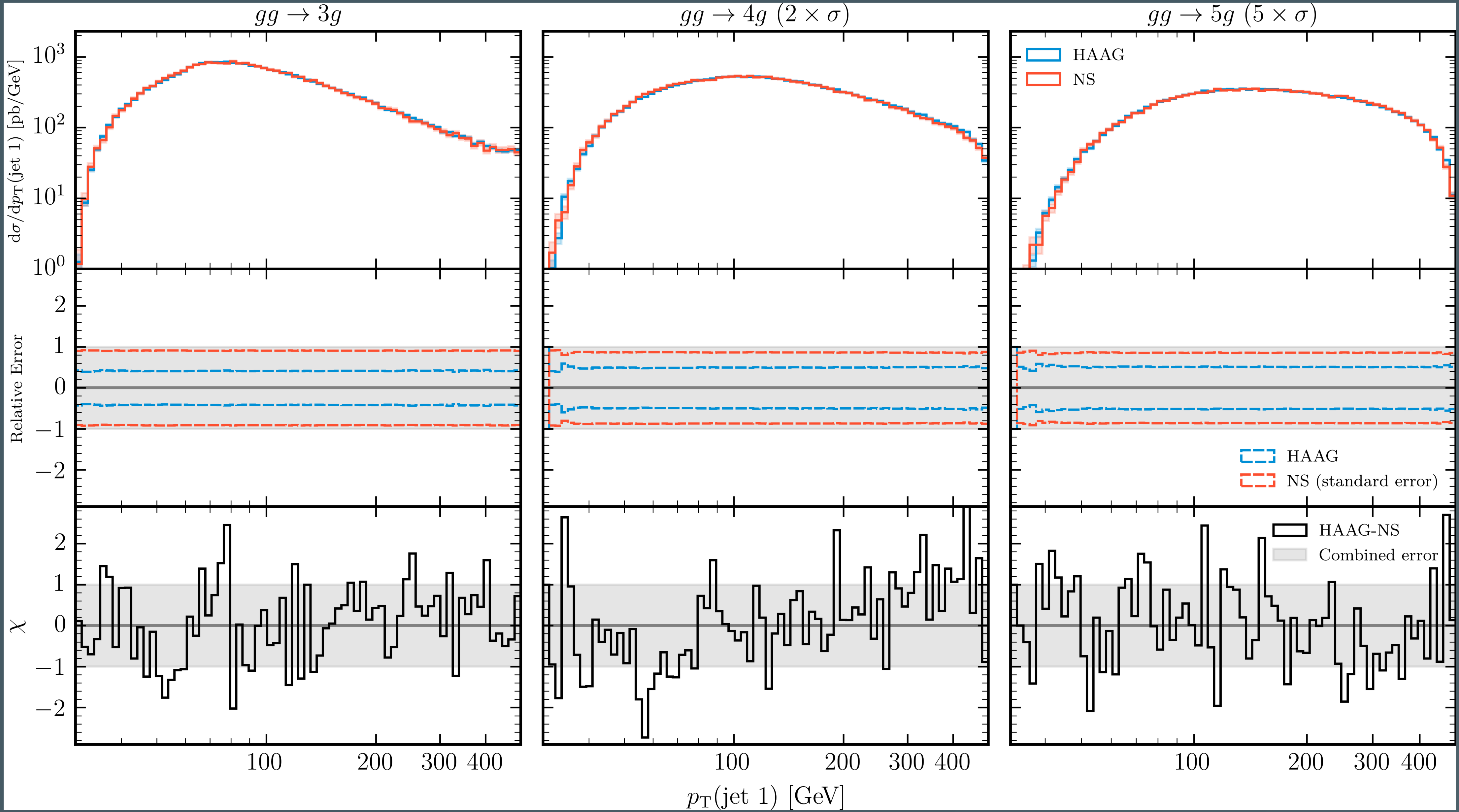
- Primarily an integral algorithm (largely unique vs other MCMC approaches)
- Designed for multimodal problems from inception
- Requires construction that can sample under hard likelihood constraint
- Largely self tuning
  - Little interal hyperparameterization
  - More importantly, tunes any reasonable prior to posterior

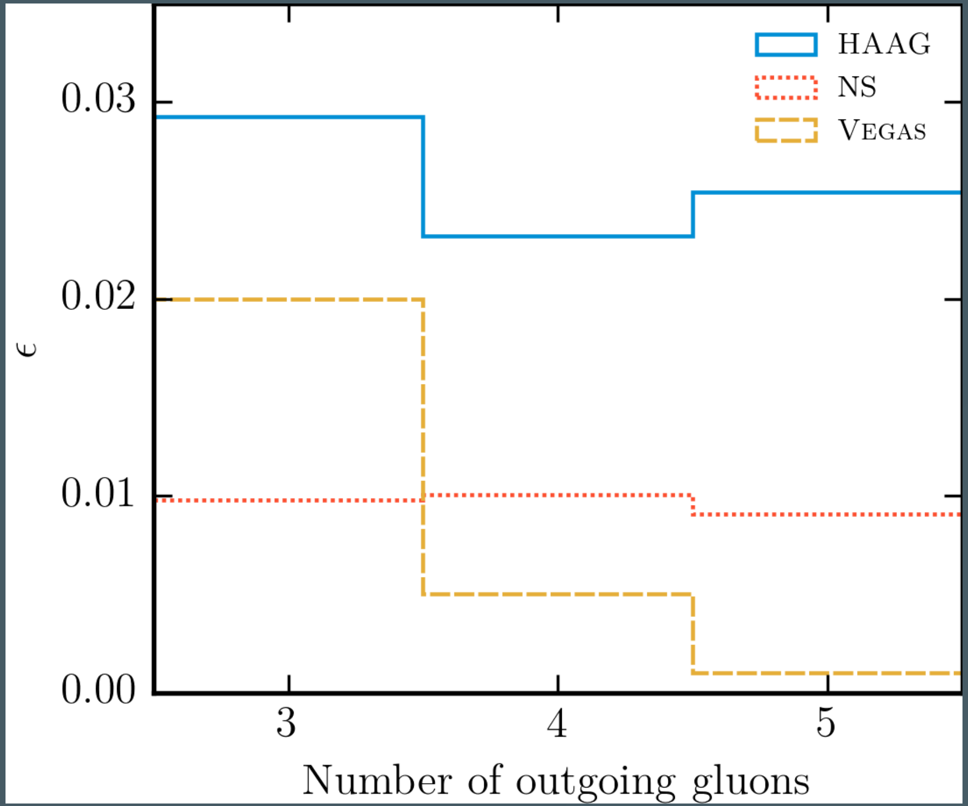
[yallup.github.io/bayeshep_durham] for animated versions




$p_{T,4}$
160
80
300
$p_{T,3}$
150
$p_{T,2}$
300
150
$p_{T,1}$
400
200
80
160
150
300
150
300
200
400
$p_{T,4}$
$p_{T,3}$
$p_{T,2}$
$p_{T,1}$

Unweighted Events





Algorithm	$gg \to 3g$	$gg \to 4g$	$gg \to 5g$
HAAG	3.0%	2.7%	2.8%
Vegas (cold start)	2.0%	0.05%	0.01%
NS	1.0%	1.0%	1.0%

Where do we go from here?

End to end stylised version of the problem demonstrated.

This is deeper than coming up with a new way of mapping phase space

Where do we go from here?

(dedicated section in paper)

- *Physics challenges*
- Variants of NS algorithm
- Prior information
- Fitting this together with modern ML

Physics challenges

The fundamental motivation for this work came from recognising not just an ML challenge but a physics challenge [2004.13687]

LO dijet isn't hard, NNNLO is. If your method isn't robust in these limits it doesn't solve the right problem. Unique features of NS open up interesting physics:

- **No mapping required:** NLO proposals generically harder, NNLO more so
- **No channel decomposition:** can we be *really* clever when it comes to counter events, negative events etc. with this?
- **Computation scaling guaranteed** to $\sim$ polynomial with D, other methods exponential: We can do *genuinely* high dimensional problems, $gg \rightarrow 10g$ anyone?

Conclusion

In my opinion (your milage may vary)

- The fundamental problem for LHC event generation trying to do Importance Sampling in high dimension.
- Machine learning can and will be useful, but this is not **just** a machine learning mapping problem.
- This **is** a Bayesian inference problem, precisely calculating Evidences or Posterior sampling.
- Nested Sampling is a high dimensional integration method, primarily from Bayesian Inference, that is an excellent choice for particle physics integrals