GPE

GPU-accelerated CBC Parameter Estimation

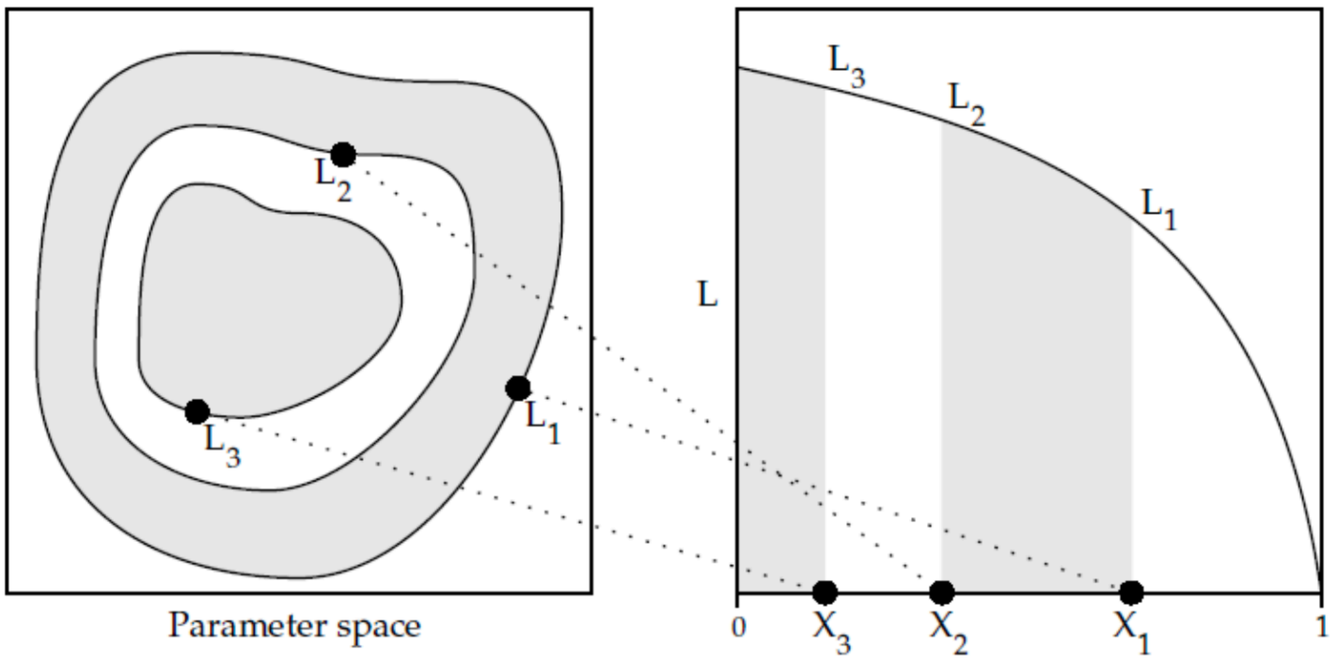
S. Haino

Introduction

- Parameter Estimation (PE) is a time-consuming process; we need to find a new way to accelerate it in view of many-GW-detection era towards 3G
- Freq.-domain CBC PE on
  15 parameters (3(M_c,q,φ) +6(spin) +6(ext.)) with
  - Nested sampling with MCMC sub-chains and
  - *IMRPhenomPv2* waveform model
  has been accelerated with GPU
- The codes are all newly written in C++ and CUDA and produce the same output for *cbcBayesPostProc*

Nested sampling

- Introduced by J. Skilling to compute Bayesian evidence(Z) with MCMC sub-chains by transforming the Multi-Dimensional integral into 1-dimension over the prior volume   Ref(e.g.): *arXiv:1409.7215*



Nested sampling in LAL

- Implemented in :
  LALInferenceNest, LALInferenceNestedSampler, LALInferenceProposal, LALInferencePrior, …
- 5 methods used for Jump proposals (LAL default) :
  - *CovarianceEigenvectorJump*
  - *DifferentialEvolution*
  - *EnsembleStretch*
  - *EnsembleWalk*
  - *DistanceLikelihood*
- MCMC sub-chain length is determined at every (*Nlive*/10) iterations from autocorrelation length

GPU Approach

- Waveform and likelihood calculations are the dominant sources of time consumption for P.E.

  \- 4,096(*srate*) x8(*seglen*)/2 x2(*Nifo*) = 32,768 /call
  \- ~1,000 times called per iteration
  \- ~16,000 iterations per run
  => 5 x10^{11} calculations / run

- These particular parts are implemented in CUDA

* In *LALSimulation*, waveform calculation can run in OpenMP
but the performance didn't improve so significantly on Core™ i7

Performance test

- GW150914 data from LIGO Open Science Center
- LAL: *lalinferene_nest* with *IMRPhenomPv2*, *seglen*=8, *Nlive*=500 are used as a reference
  - running on single CPU (4 parallel jobs / machine)

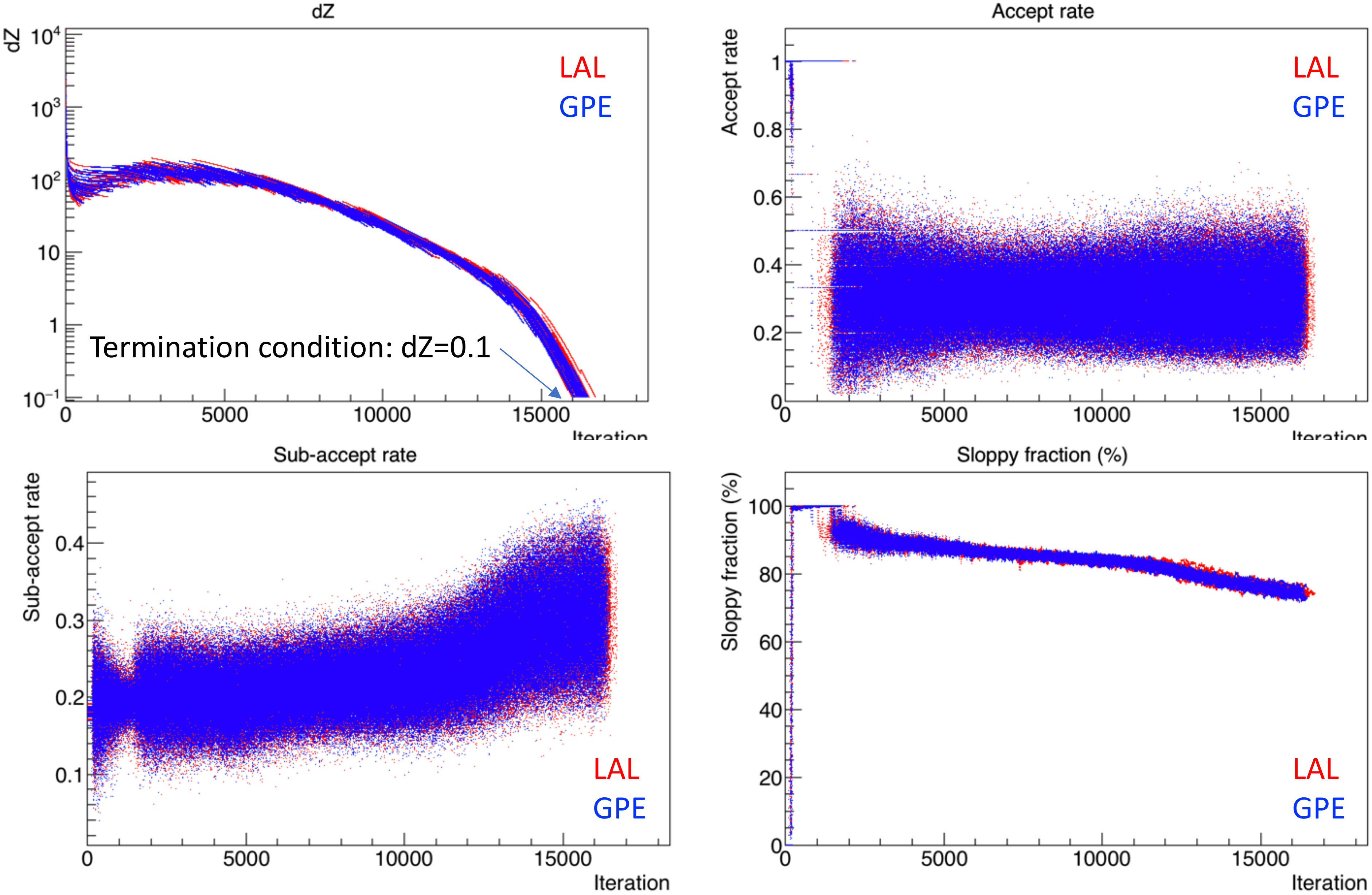
- GPE (GPU-accelerated P.E.) :
  - running on single CPU/machine and tested with 3 different GPU boards (NVIDIA™ GeForce™)
- 23 independent runs performed and time consumptions and output results are compared

Performance comparison

Code	Hardware	Spec.	Wall Time Mean ± RMS	Acceleration w.r.t. LAL	Improvement
LAL	Core™ i7	4 cores (x2 HT) 3.6 GHz	24:27:24 ± 47:42		
GPE	GeForce™ GTX 1060	1152 cores 1.76 GHz 192 bit Bus	17:21 ± 0:28	× 84.5	
GPE	GeForce™ GTX 1070	1920 cores 1.68 GHz 256 bit Bus	13:58 ± 0:17	× 105.0	24% to 1060
GPE	GeForce™ GTX 1080	2560 cores 1.85 GHz 256 bit Bus	12:25 ± 0:15	× 118.1	40% to 1060 13% to 1070

Log Bayes factors $\ln(B_{s/n})$: 254.7±0.3 and 254.5±0.3 in LAL and GPE, respectively

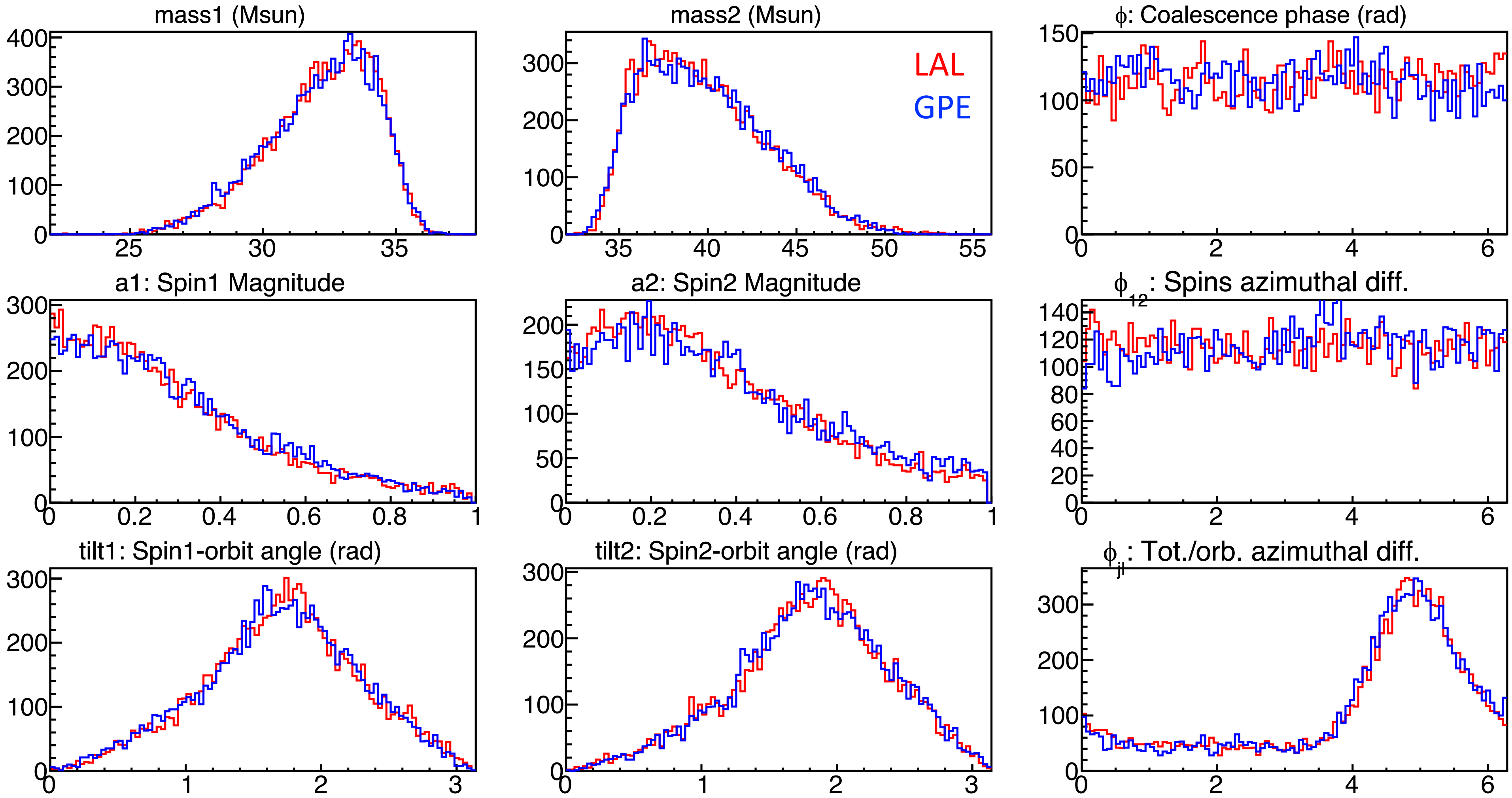
Comparisons (Nested sampling parameters)



Comparisons (CBC intrinsic parameters)

Parameters in the final 500 live points (from 23 independent runs)

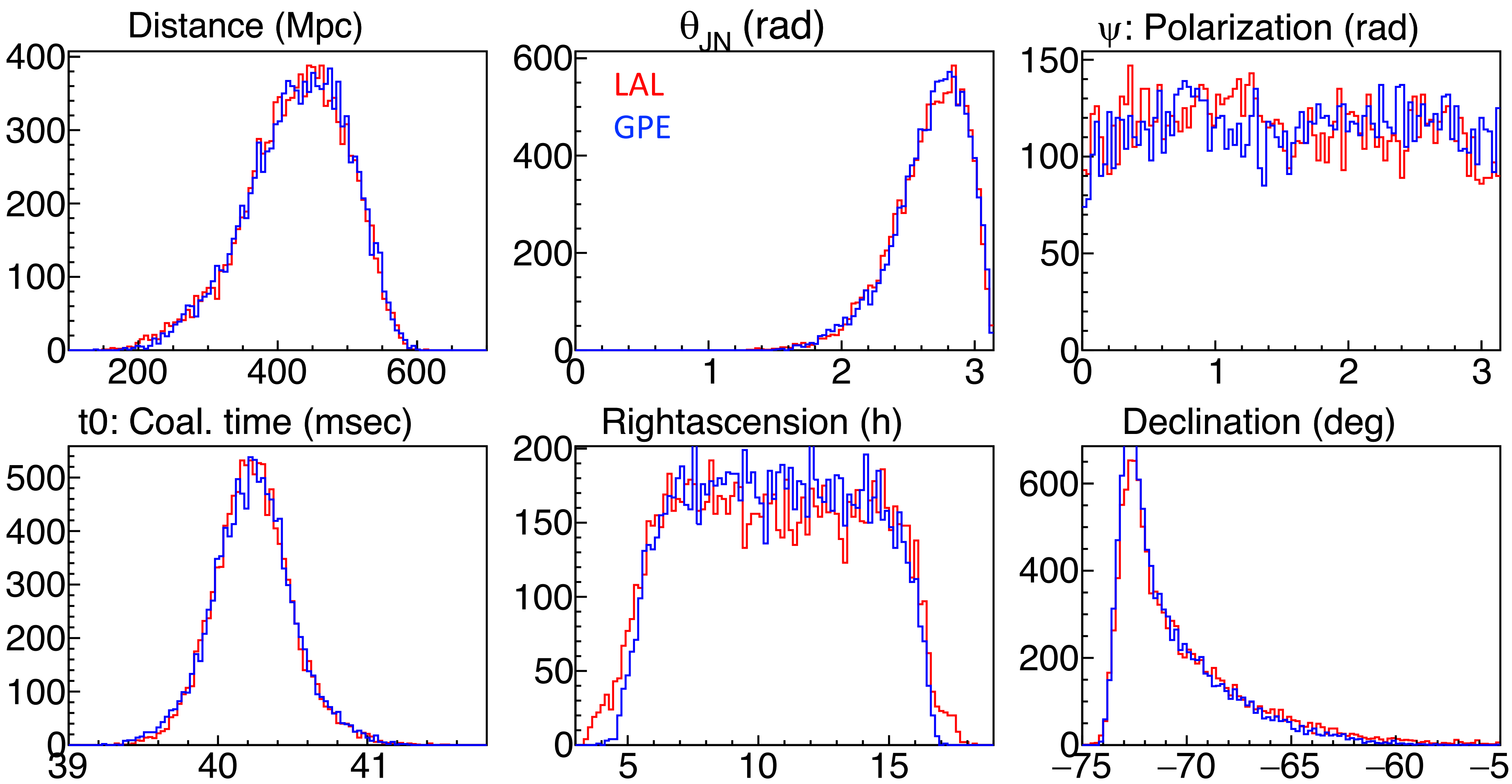
Note: they are not the posterior distributions



Comparisons (CBC extrinsic parameters)

Parameters in the final 500 live points (from 23 independent runs)

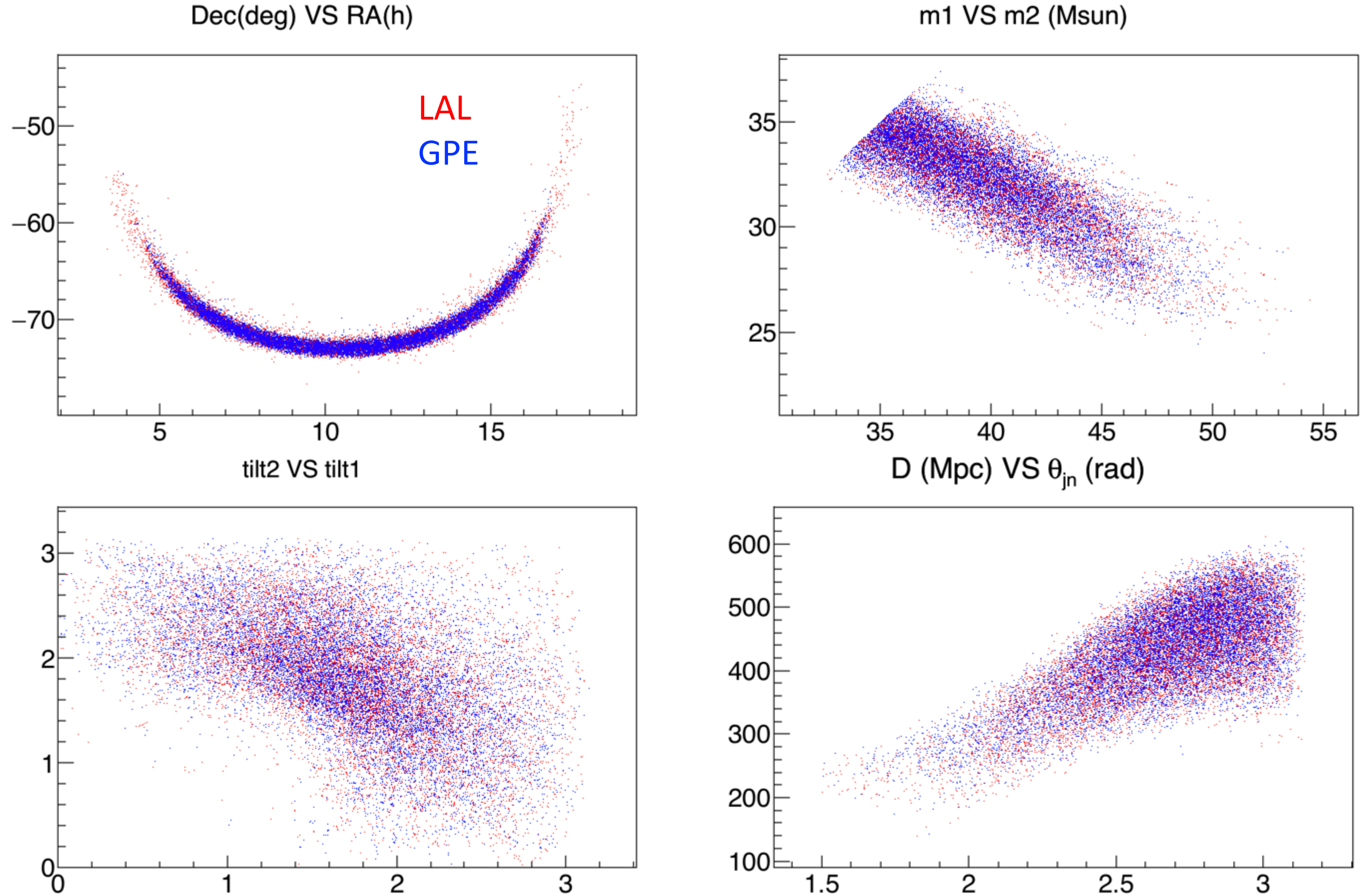
Note: they are not the posterior distributions



Comparisons (CBC parameter correlations)

Parameters in the final 500 live points (from 23 independent runs)

Note: they are not the posterior distributions



Summary and next steps

- Freq.-domain CBC PE on 15 parameters with Nested sampling and *IMRPhenomPv2* has been accelerated with GPU by factor of ~100 w.r.t. LAL

- Collaboration works are planned in Taiwan
  - ASGC and NCHC are encouraging GPU computing
- Possible applications:
  - Science studies for KAGRA science contributions
  - PE with realistic calibration parameter priors
  - Suggestions and collaborations a welcome !