The UK effort for 3G computing

Gareth Davies, Steve Fairhurst, Ian Harry, Qian Hu, Duncan Macleod, Geraint Pratten, Vivien Raymond, Patricia Schmidt, John Veitch, Christopher Whittall.

Digital infrastructure

Develop new computational algorithms and infrastructure designed to process the hundreds of thousands to millions of black holes and neutron star mergers each year observed by the next-generation GW observatories.

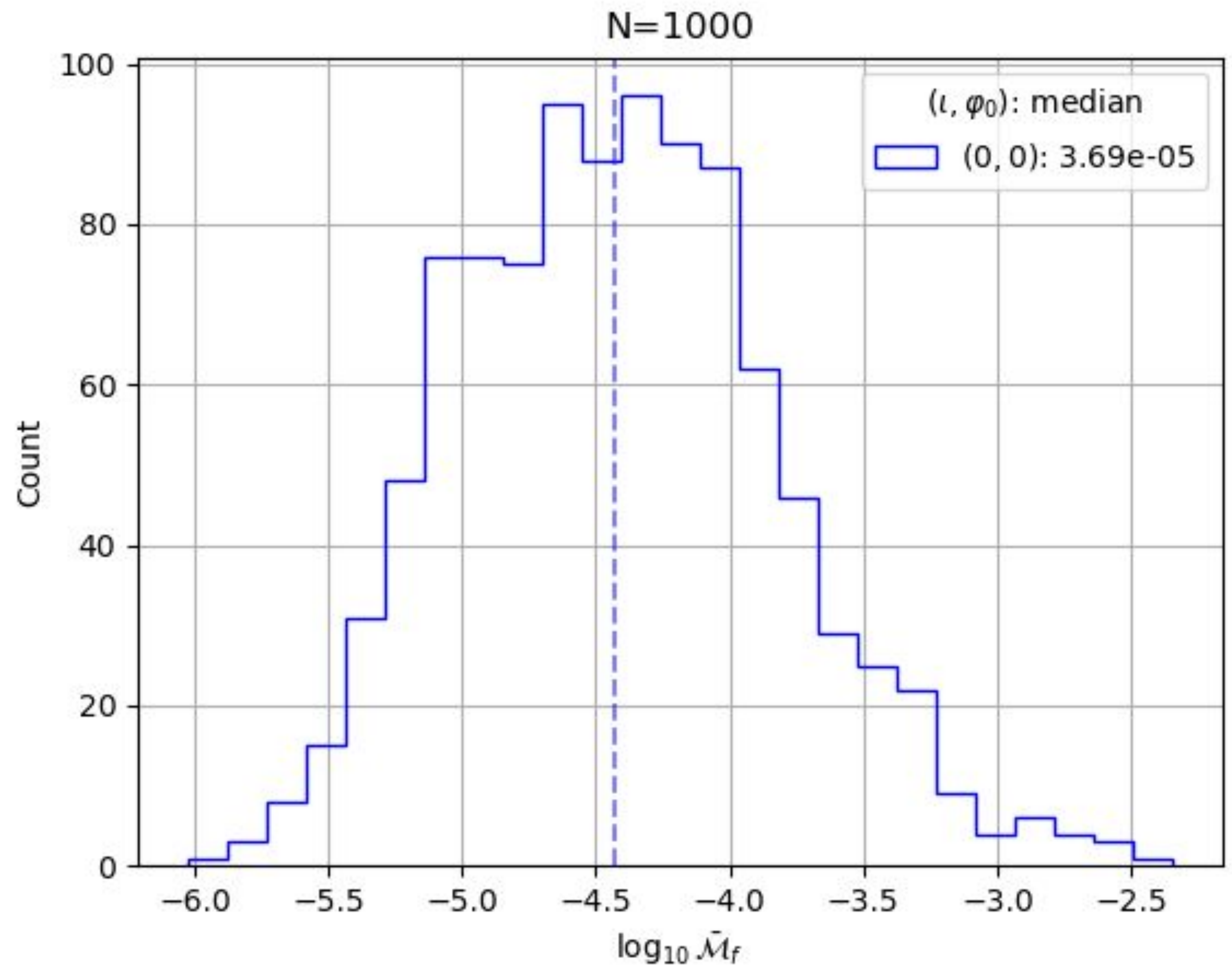
- Computational infrastructure and algorithms
  - **Waveform generation**, led by the University of Birmingham
  - **Real time searches**, led by the University of Portsmouth
  - **Signal and population inference**, led by the University of Glasgow
- **Prototype event database**, led by Cardiff University

Waveform acceleration (C. Whittall & G. Pratten)

UNIVERSITY OF BIRMINGHAM

Developed semi-automated pipeline to train reduced-order-models with artificial neural networks for parameter space fitting. Agnostic to base waveform model.

- Proof of concept: SEOBNRv5PHM with arbitrary precessing spins, mass ratio 1 <= q <= 2 and 5000M signal duration:
    - Neural networks interpolate across the 7d parameter space accurately and efficiently.
    - Suitable for GPU acceleration and waveform batching.
- Now investigating how these techniques scale up to longer durations and higher mass ratios, critical for next-gen detectors.



Mismatches in the surrogate compared to the base EOB model.

How will searches deal with 3G GW detector data?

Will template bank searches even be applicable?:

- How big would a template bank be? Does this make this kind of search infeasible?
    - Template banks will be dominated by small changes at low frequency, how can we remove this?
    - Will Earth rotation be a problem for bank sizes?
- What kind of disk space will we need to store triggers?
- What sample rate do we use for the SNR series?

Additional questions:

Where can optimised approximants be used?

Do we need multi-banded approaches? Hierarchical searches? Machine learning?

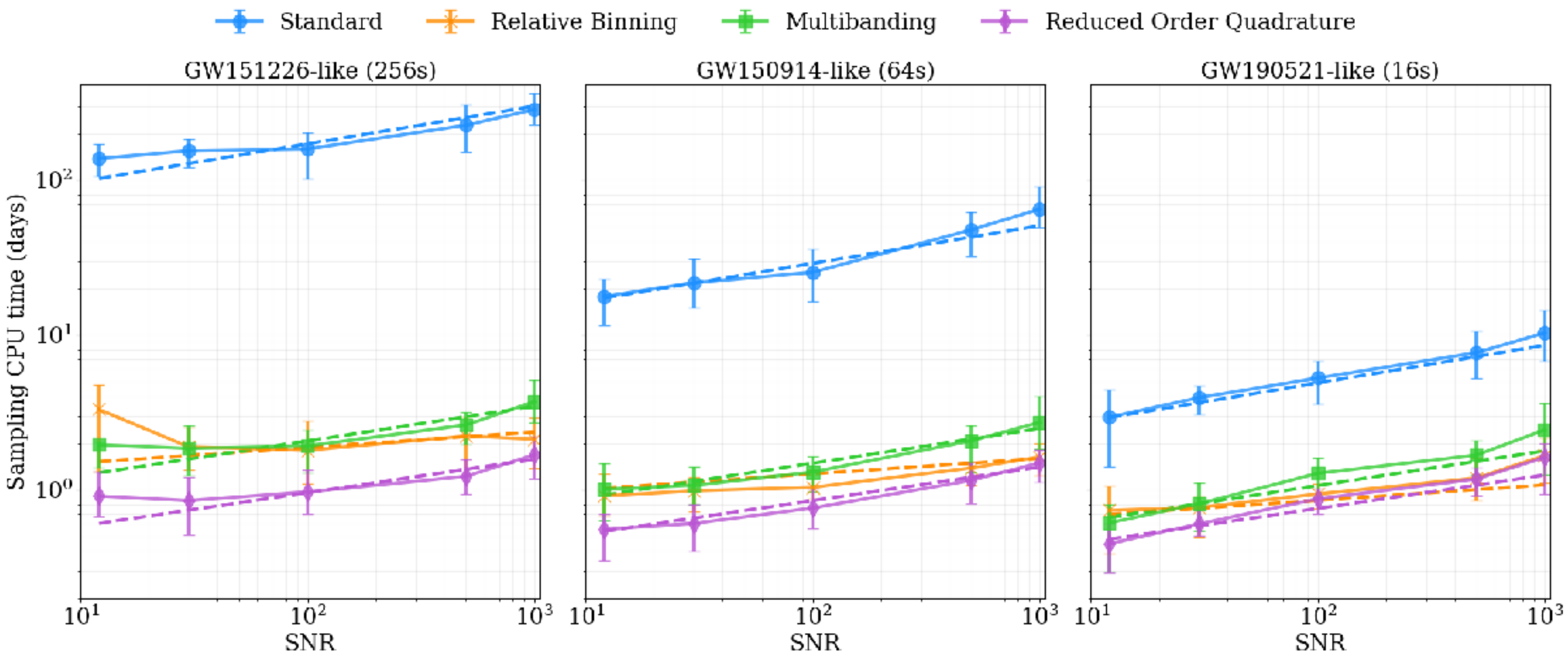
Are overlapping signals an issue?

Does it even matter that waveform approximants are not 100% accurate at this SNR?

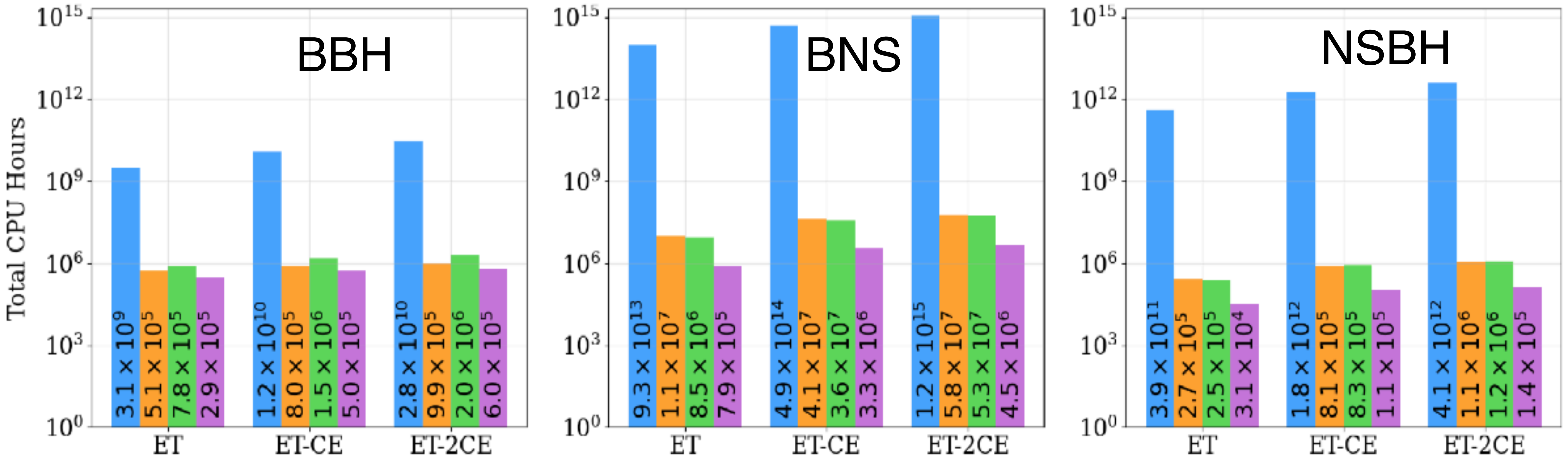
Will we need to include exotic physics in order to not miss signals?

Cost of Bayesian Parameter Estimation

- Bayesian sampling time cost scales with signal duration(T) and SNR.
- CPU days (D) for PE can be fitted as logD = a logT + b logSNR +c
- Figure: Total CPU hours required to analyze ET-MDC-1 catalog (one-month observation)
- Current “standard” method used by LVK method is unfeasible
- Acceleration methods like ROQ: O(1-10) million CPU hours per month per round of analysis

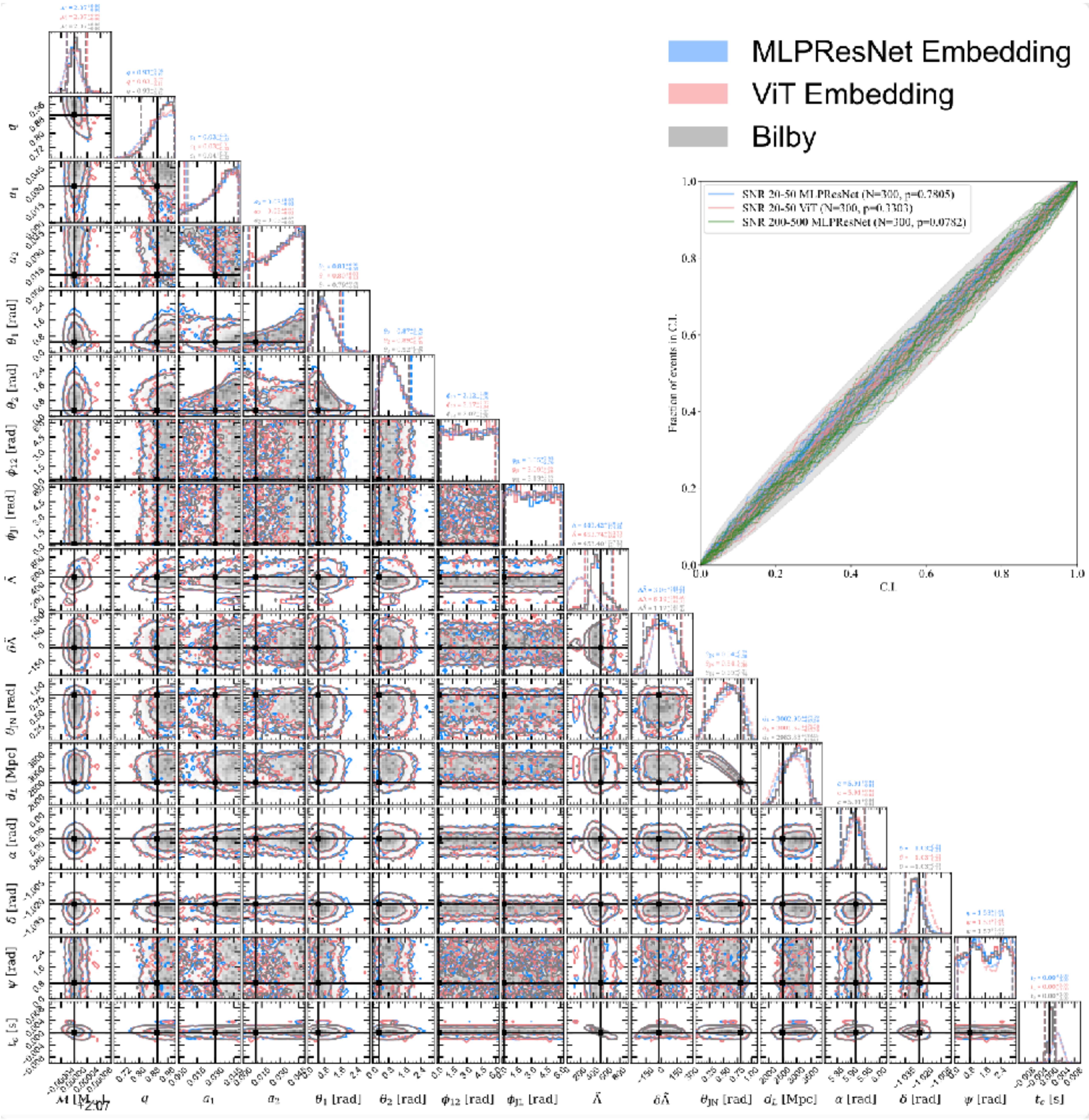


Q. Hu & J. Veitch, 2412.02651



Machine Learning Solution to Tough Problems

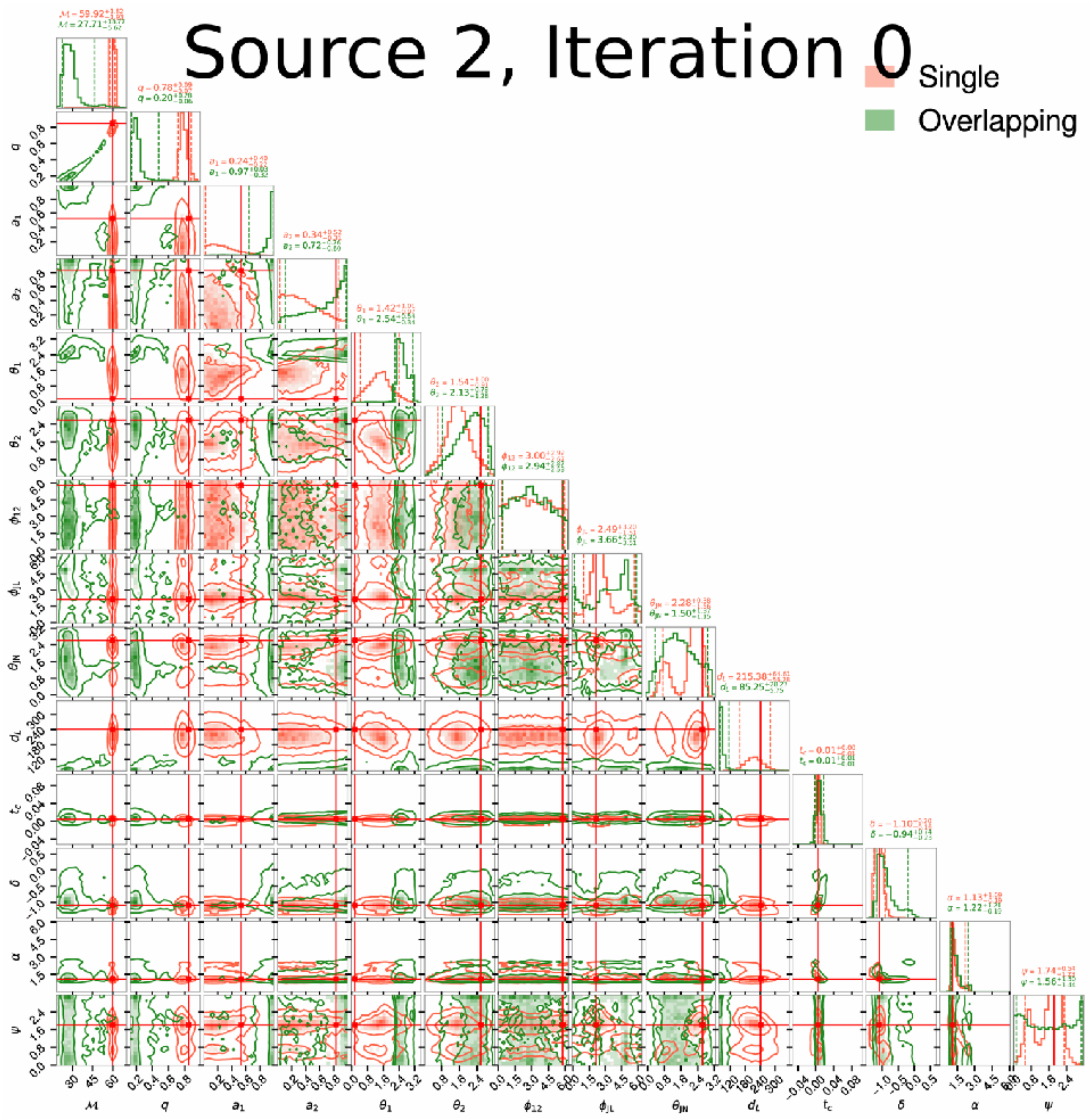
- Hours-long BNS Signals solved by ML within 1s
- 0.1% computational cost of traditional sampling methods (including training cost)
- Crucial for catalog-level analysis in the future



Q. Hu et al, 2412.03454

Machine Learning Solution to Tough Problems

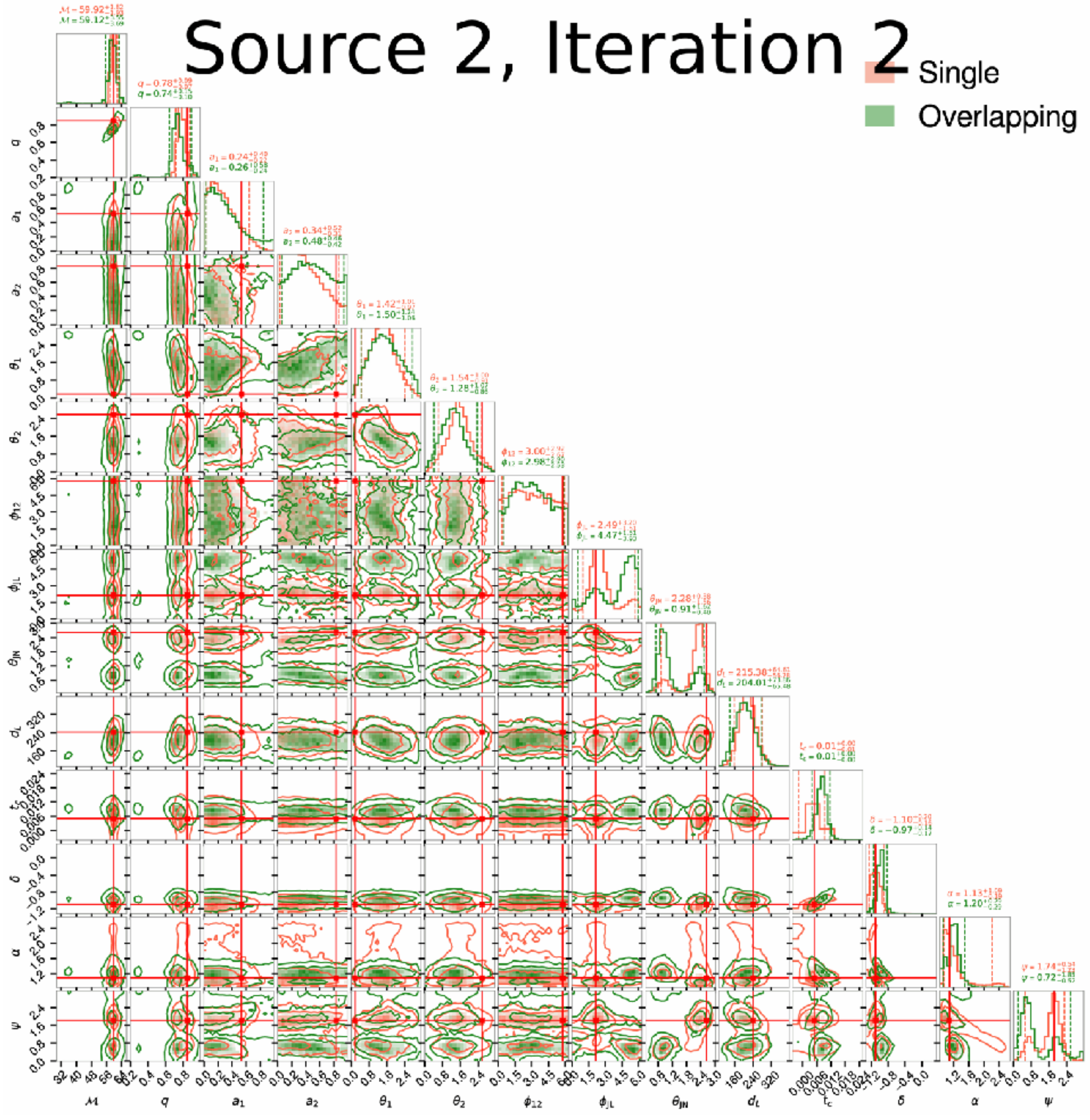
- A general solution to overlapping signal inference problem: hierarchical subtractions using ML PE models
- Only need to train ML PE models for single signals!
- Can adapt to different source types, time difference, and number of sources



Q. Hu & J. Veitch, in prep, 2025

Machine Learning Solution to Tough Problems

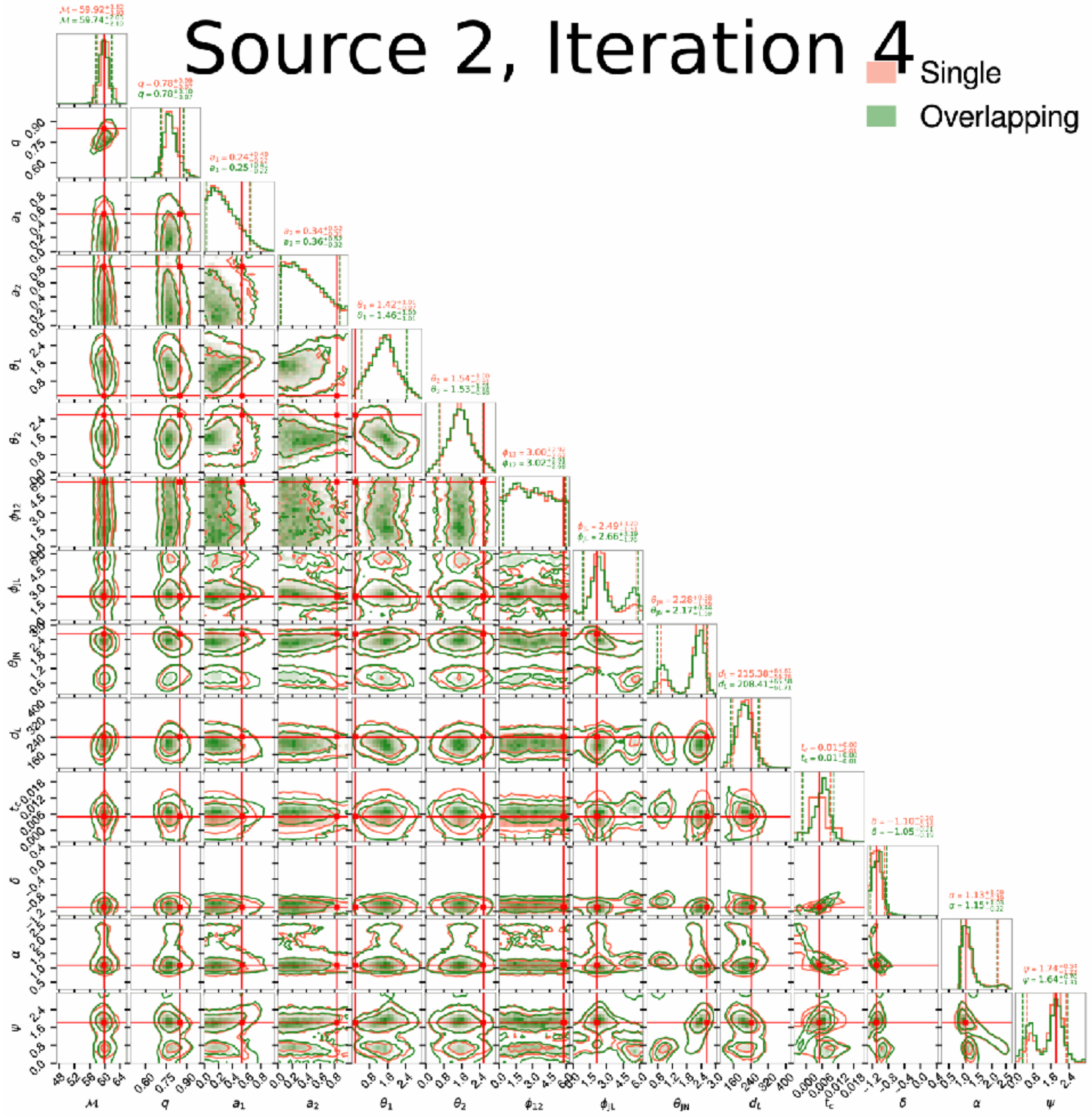
- A general solution to overlapping signal inference problem: hierarchical subtractions using ML PE models
- Only need to train ML PE models for single signals!
- Can adapt to different source types, time difference, and number of sources



Q. Hu & J. Veitch, in prep, 2025

Machine Learning Solution to Tough Problems

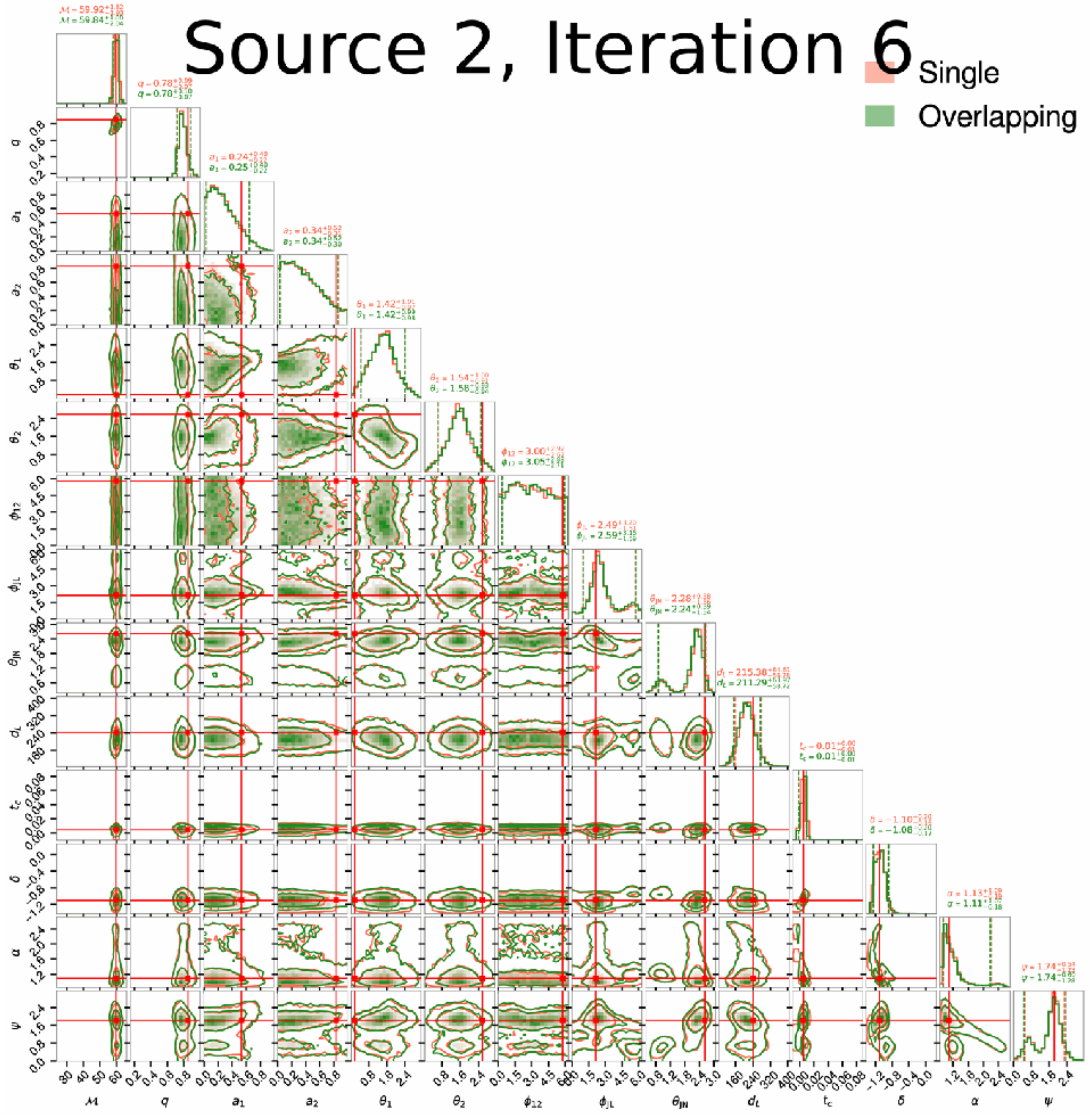
- A general solution to overlapping signal inference problem: hierarchical subtractions using ML PE models
- Only need to train ML PE models for single signals!
- Can adapt to different source types, time difference, and number of sources



Q. Hu & J. Veitch, in prep, 2025

Machine Learning Solution to Tough Problems

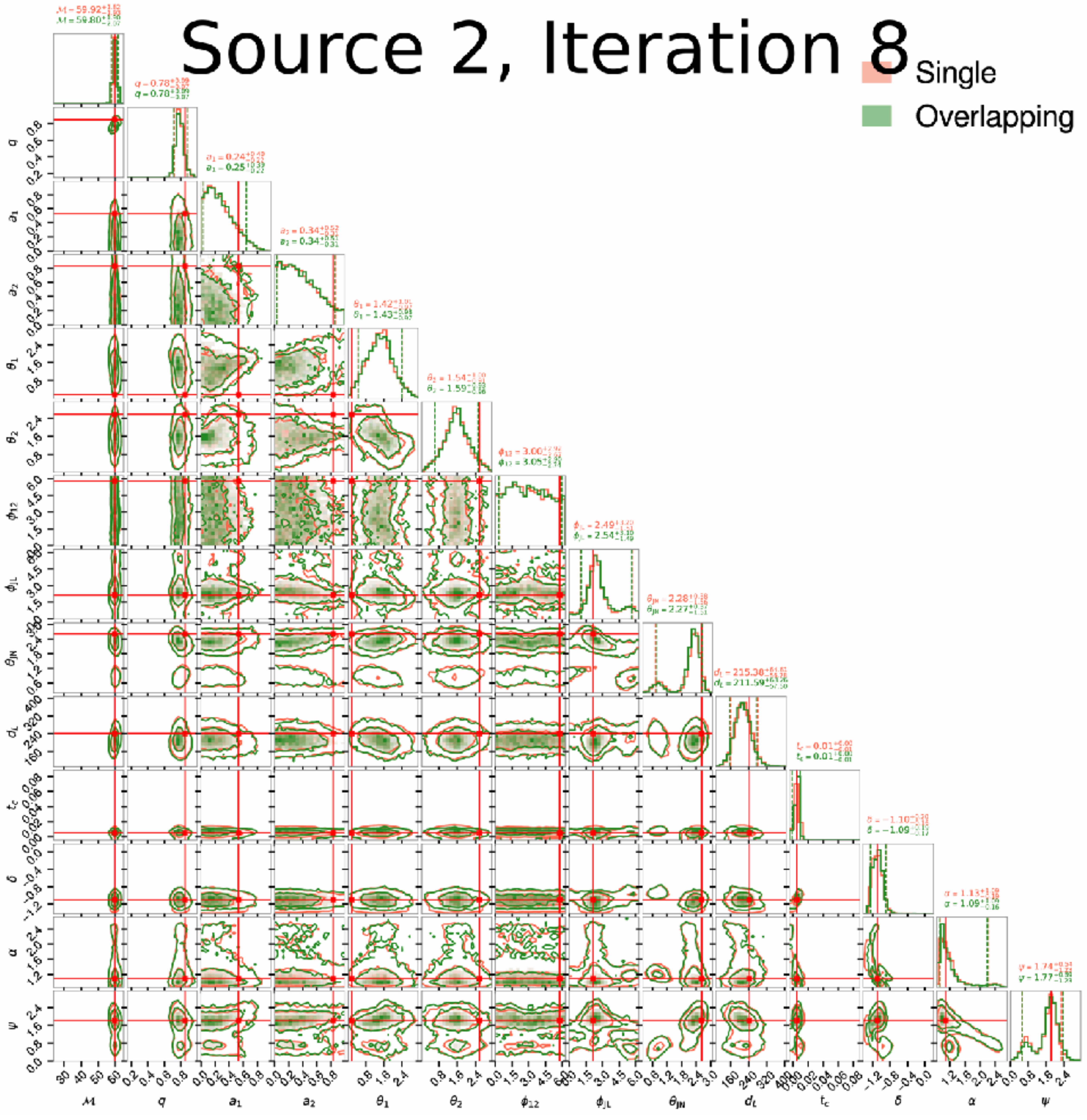
- A general solution to overlapping signal inference problem: hierarchical subtractions using ML PE models
- Only need to train ML PE models for single signals!
- Can adapt to different source types, time difference, and number of sources



Q. Hu & J. Veitch, in prep, 2025

Machine Learning Solution to Tough Problems

- A general solution to overlapping signal inference problem: hierarchical subtractions using ML PE models
- Only need to train ML PE models for single signals!
- Can adapt to different source types, time difference, and number of sources



Q. Hu & J. Veitch, in prep, 2025

Next-Gen event database

- 3G event rate 1000x higher than 2G
- High-SNR negative-latency detection allows early warning alerts
- Challenge to provide a robust, scalable, event database and alerting infrastructure to serve the GW observatory network and EM partners
- Cardiff leading UK effort to scope requirements and conceptual design
  - Requirements link
- Needs detailed understanding of access patterns
- Introductory study of prospective database technologies and data placement algorithms
  - Link

NGDB Software Requirements Specification

Development version 6dbb5bac

2024-12-13

Contents

Next-Gen event database

- 3G event rate 1000x higher than 2G
- High-SNR negative-latency detection allows early warning alerts
- Challenge to provide a robust, scalable, event database and alerting infrastructure to serve the GW observatory network and EM partners
- Cardiff leading UK effort to scope requirements and conceptual design
  - Requirements link
- Needs detailed understanding of access patterns
- Introductory study of prospective database technologies and data placement algorithms
  - Link

TABLE 1. Summary of relevant scalable databases or storage technologies

Database	Type	Use case	Open Source?
Apache Cassandra	NoSQL	High volume & throughput, column oriented, query-first, no joins (client-side implementation), no ad hoc queries	Yes
MongoDB	NoSQL	High volume & throughput, document-oriented, customisable sharding, vector search	Yes
CockroachDB	NewSQL	Distributed transactions, PostgreSQL compatibility, obscures partitioning	No
Google Spanner	NewSQL	PostgresSQL interface, strong consistency	No
ScyllaDB	NoSQL	Claimed improved performance over Cassan-	Yes
TiD			
Vites			
Apac			

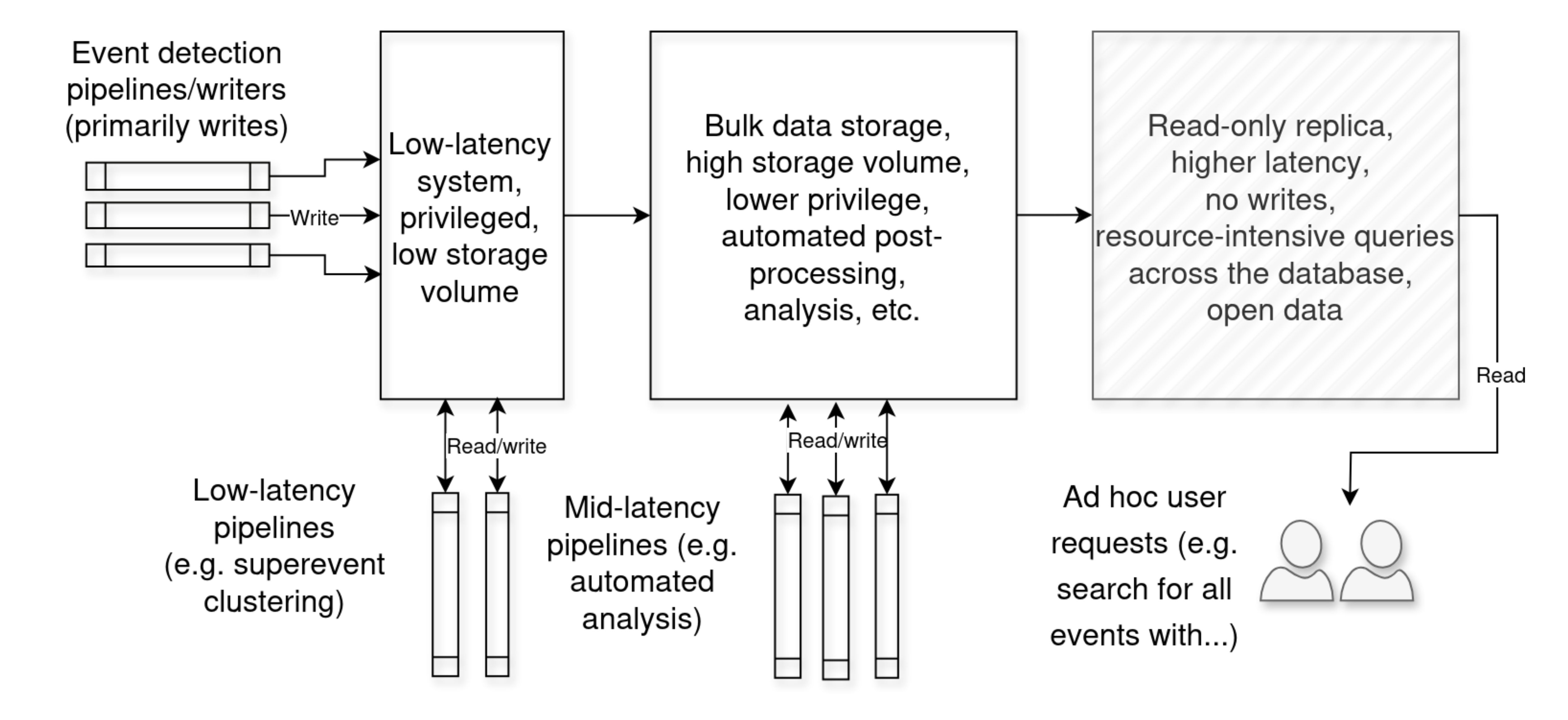


FIGURE 2. Logical decomposition of NGDB system into components by use case. In terms of concrete realisation, the system could be decomposed into streaming, OLTP operational databases, and OLAP systems for bulk analyses.

A. Southgate, D. Macleod, et al., in prep, 2025