# Subpopulations in the Gravitational Wave Binary Black Hole Merger Population with UMAP

**Julius Gassert**[1,2] on behalf of A. J. Amsellem[1], I. Magana Hernandez[1] and A. Palmese[1]
Carnegie Mellon University[1], Ludwig-Maximilians-University[2]

Initial Results: Amsellem et. al
arxiv.org/abs/2603.06566

*graviflows* package soon:
github.com/jgassert

## Motivation & Background

- Black holes (BH) merge & emit *gravitational* waves (GW)
- Interferometers detect GWs and properties (such as masses, distance, ...) of the sources can be inferred
- Growing catalog (GWTC) of binary BH mergers since first detection in 2015

***How do these systems form? What does their population look like?***

- Clear features in the mass spectrum
- Evolution with time (redshift)
- Different formation channels proposed: how do they contribute to the overall population?

***Model-independent analysis***

- Many previous analysis were model-dependent
- New approach: data-driven, model-independent techniques to uncover trends and subpopulations

**Uniform Manifold Approximation and Projection (UMAP)**

## Data & Methods

- 69 BBH events (FAR < 1/yr), 400 samples each, reweighted for prior removal and detection probability (source population)
- UMAP dimensionality reduction input: masses, redshift and spin parameters
- HDBScan for clustering

***Ongoing Extensions***

- Catalog update
- Detector network sensitivity increase
- Stochasticity of UMAP
- 3D embeddings and different input parameters

## Results

- UMAP effectively partitions BBH population into subgroups
- Masses are the primary driver (little/no overlap between groups)
- Clear outlier event: GW190521 - mass gap progenitor?
- Mass spin correlation: mass ratio and effective spin correlation in overall catalog driven by low mass group(s)
- Astrophysical interpretation: groups trace distinct formation pathways (e.g. the ~10 solar masses BBH buildup appears as a separate, spin-aligned subpopulation)

***Data-driven clustering reveals distinct BBH subpopulations consistent with multiple formation channels, without assuming a model***

## New Data & Updated Methods

- 156 BBH events (FAR < 1/yr)
- New detection probability emulator for updated detector network

$$P(\mathbb{D}|\theta) = \frac{p(\theta|\mathbb{D})P(\mathbb{D})}{p(\theta)}$$ ***Normalizing Flow density estimator***

- UMAP, HDBScan and data sampling: seeds and hyperparameters

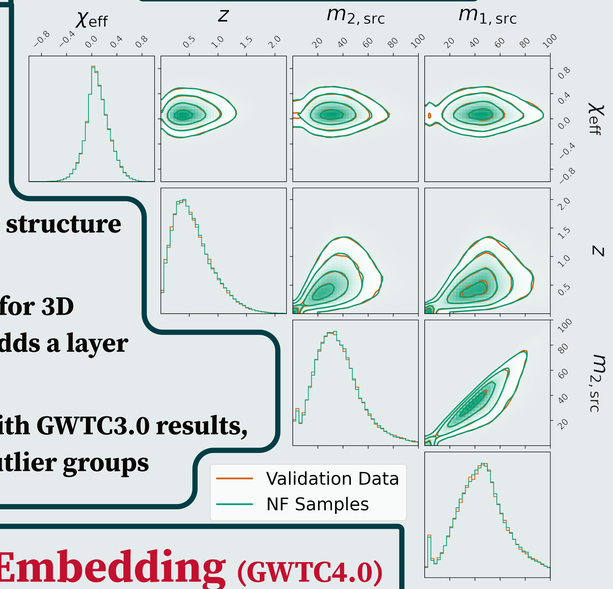
➡ Optuna hyperparameter optimization over seeds

## GWTC4.0 Results

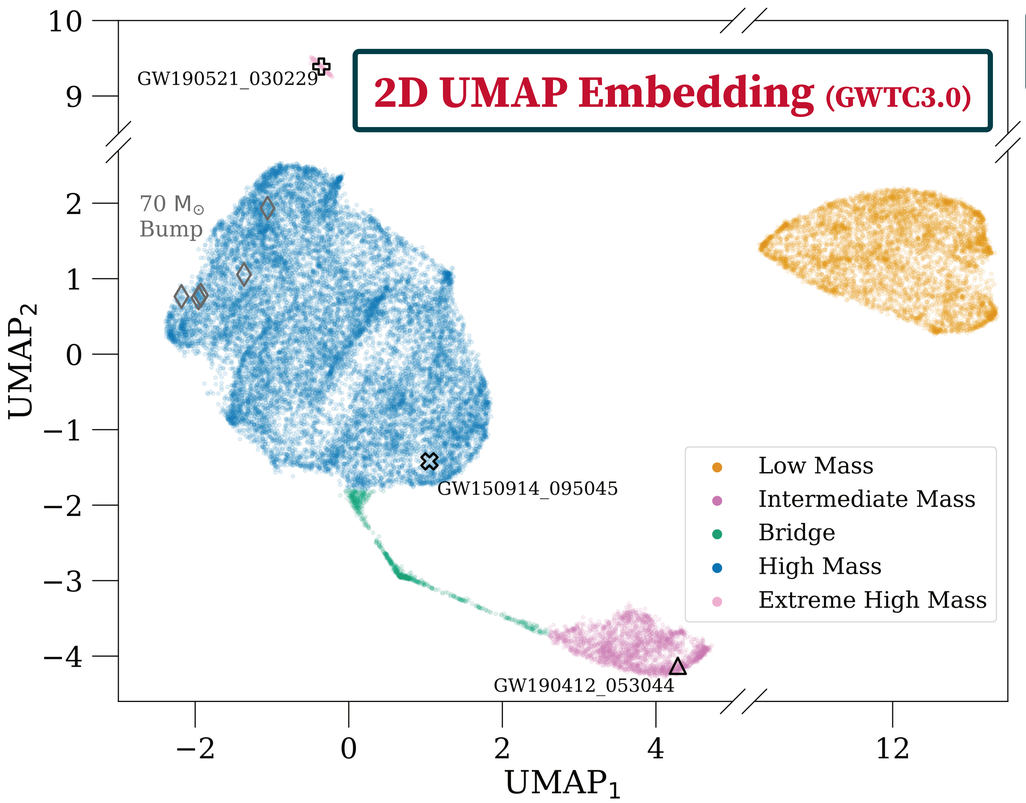
*preliminary*

- Stable embeddings over different optimization runs
- Adding more input parameters smears out structure (as expected)
- No significant changes for 3D embeddings, but this adds a layer of interpretability
- Generally consistent with GWTC3.0 results, yet identifying more outlier groups

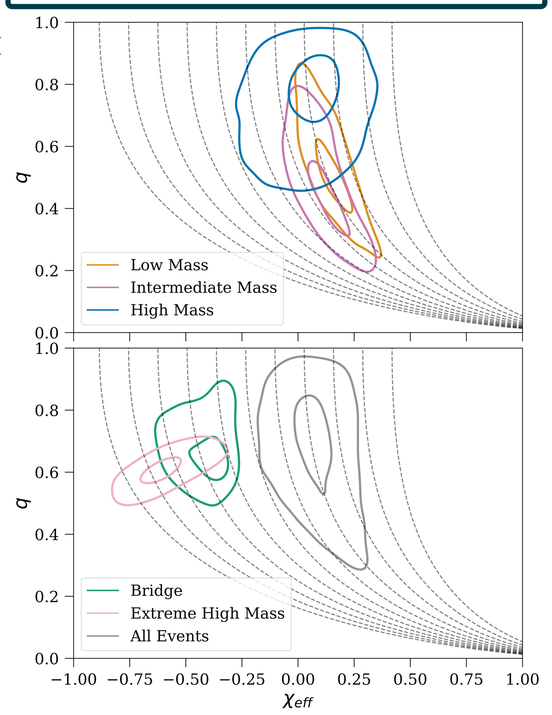
## NF Density Fit




## 2D UMAP Embedding (GWTC3.0)




## Parameter Degeneracy




## 2D UMAP Embedding (GWTC4.0)