# Binary mergers: a common format

November 4, 2023

**Abstract**

We propose a common output format for storing astrophysical predictions of compact binary coalescences.

## 1 Motivation

The quantity and quality of models of compact binary mergers experienced rapid growth in recent years. There is increasing interest in comparing these models against each other and against observations, combining models that explore different formation environments (e.g., isolated binary evolution, hierarchical triples, young clusters, globular clusters, AGN disks, ...), and creating astrophysically realistic injection populations to scope out data analysis challenges for future detectors. All of these goals would benefit from having predictions from a broad range of model-building tools available in a common format. Here we propose a simple format for storing astrophysical predictions of compact binary coalescences.

## 2 Main output

Since most population-synthesis codes carry out Monte Carlo sampling to produce binary populations, we propose to directly report the samples of merging binaries, rather than any smoothed distributions. Thus, outputs will contain the parameters of simulated binaries, each of which is representative of a population of systems with similar parameters. In practice, if the sample size is large enough, only a subset of these can be used (e.g., for ejection campaigns), avoiding repetitions and unwelcome statistical correlations.

We propose to store the main output in a simple ASCII (human-readable) comma-separated-value (CSV) file with four mandatory columns:

1. mass of the first compact object at the time of merger in solar masses
2. mass of the second compact object at the time of merger in solar masses
3. the merger redshift
4. the merger rate of binaries with these parameters, per unit comoving volume in Gpc$^3$ per unit of source time in years.

Additional optional columns can include other predictions, such as spins. An example is shown in table 1, following the file format described in section 5.

Although the general file format described in section 5 is flexible, it is strongly recommended to include as the first five rows the following meta-data: the model type (possible options include "isolated-binary", "triple", "globular-cluster", "open-cluster", "nuclear-cluster", "AGN"); the name of the code used to generate the data; the code version; the date (YYYY-MM-DD) on which the data were originally generated; and the name and e-mail of the content creator, to provide a point of contact. The next two rows contain the column headers and units. An arbitrary number of data rows follow.

In practice, many population synthesis codes "recycle" binaries: a binary is formed at a given metallicity, but could be assigned an arbitrary formation time (redshift), with the merger time (redshift) set equal to the sum of the formation time and the simulated delay time between the formation

| | | | | | |
|---|---|---|---|---|---|
| M, | model-type, | isolated-binary | | | |
| M, | code-name, | COMPAS | | | |
| M, | code-version, | 02.27.05 | | | |
| M, | date, | 2022-03-11 | | | |
| M, | creator, | Jeff Riley `<jeff.riley@awesome.edu>` | | | |
| H, | Mass1, | Mass2, | Redshift, | MergerRate, | Optional |
| U, | Msun, | Msun, | , | Gpc-3yr-1, | |
| D, | 12.3, | 1.4, | 0.05, | 0.61, | ... |
| D, | 12.3, | 1.4, | 0.05, | 0.71, | ... |
| D, | 36.3, | 11.4, | 0.05, | 0.04, | ... |
| D, | 36.3, | 11.4, | 0.15, | 0.05, | ... |

Table 1: An example main output format.

and merger for this binary. The merger rate at a given redshift is then determined by the convolution of the metallicity-dependent binary yield (with a distribution across delay times and component masses) and the metallicity-specific star formation history of the Universe. However, the proposed format allows the end user to avoid these details. Furthermore, it allows for results simulations that cannot be decomposed in this way (e.g., formation in nuclear clusters with AGN disks) to be presented in the same format.

Thus, the merger redshifts are determined by the content creator, but it may be the case that binaries with the same physical parameters appear in the same data set with different merger redshifts (and different merger rates at each redshift). Each merger redshift can be interpreted as the centre of a redshift bin (the bin width can be determined by the reader from the spacing of reported redshifts). In this case, the sum of merger rates for all binaries recorded in a given redshift bin is the total merger rate at that redshift.

It is recommended that the content creator choose a sufficiently small bin (say, no more than 0.1 in redshift) to ensure that the merger rate does not vary appreciably across the bin, so the bin centre is representative of the full bin. At the same time, care should be taken to avoid excessive file sizes: e.g., with 40 bytes per line (with no auxiliary data), 20 redshift bins and 50,000 binaries, the ASCII file will take up approximately 40 MB.

If a much larger number of binaries are available and storage is an issue, masses can also be pixellated into a 2-dimensional grid in $M_1$–$M_2$ space (with entries in each pixel corresponding to the sum of rates across all binaries falling into that pixel) at the contents creator's discretion.

# 3 Optional: Meta-data

The content creator may optionally include an additional meta-data file in .yaml format. This file should contain at least four fields describing the model type, the code name, the code version, and the date (YYYY-MM-DD) which would match the main output file described above. Optional column names should be included. Other optional fields could include additional code parameters. An example is:

```
model-type: isolated-binary
code-name: COMPAS
code-version: 02.27.05
date: 2022-03-11
creator: Jeff Riley <jeff.riley@awesome.edu>
common-envelope-alpha: 1.0
wolf-rayet-multiplier: 0.2
```

# 4 Optional: Compact binary yields

Producing the main output file as above generally requires some assumptions to be made about the cosmic history of the Universe, particularly the metallicity and star formation rate. For isolated binary evolution predictions, it is often convenient to decouple these from binary evolution modelling per se, so that a range of cosmic history models can be considered independently. In this case, we recommend output in the CSV format for the full list of binaries consisting of the masses of the first and second compact objects at merger, the delay time between star formation and merger (in Myr), the binary metallicity (mass fraction), and the yield (number of binaries with these parameters per million solar masses of star formation). If the simulation proceeded at a fixed metallicity, and individual binaries are recorded, the yield would be 1 over the total simulated mass (or the total star formation mass represented by the simulation if the simulation omitted part of the parameter space for sampling efficiency). If the simulation stochastically sampled a distribution of metallicities, the yield would be 1 over the star forming mass represented by the simulation in a unit (natural) logarithmic metallicity bin centred on the metallicity of the reported binary, after resampling to a log-uniform distribution in metallicity if required. Additional columns containing further information, such as spin, may be added. An example is given in table 2.

| | | | | | | |
|---|---|---|---|---|---|---|
| M, | model-type, | isolated-binary | | | | |
| M, | code-name, | COMPAS | | | | |
| M, | code-version, | 02.27.05 | | | | |
| M, | date, | 2022-03-11 | | | | |
| M, | creator, | Jeff Riley `<jeff.riley@awesome.edu>` | | | | |
| H, | Mass1, | Mass2, | DelayTime, | Metallicity, | Yield, | Optional |
| U, | Msun, | Msun, | Myr, | , | MMsun-1, | |
| D, | 12.3, | 1.4, | 453.1, | 0.005, | 0.0032, | ... |
| D, | 36.3, | 11.4, | 23.4, | 0.011, | 0.0002, | ... |

Table 2: An example compact binary yields output format.

Note that this output format may not be suitable for dynamical formation models; a different format for storing direct model outputs may be appropriate in that case.

# 5 File format

The first column of each CSV file contains a label that describes the content of the rest of the row. The options are (the case is not significant):

- C: Comment row
- M: Meta-data row
- H: Header row
- U: Units row
- D: Data row

All row types can appear zero or more times in file. The cell contents for all cells must be zero or more ASCII (printable+extended) characters. Empty rows are ignored.

## 5.1 Comment row

Column 1: C Columns 2..unlimited: Comment (may be empty)

## 5.2 Meta-data row

Column 1: M Column 2: Meta-data field name (must not be empty) Columns 3..unlimited: Meta-data field values (may be empty)

## 5.3 Header row

Column 1: H Columns 2..unlimited: Column header (may be empty)

### 5.4 Units row

Column 1: U Columns 2..unlimited: Column units (may be empty)

### 5.5 Data row

Column 1: D Columns 2..unlimited: Column data values (may be empty)

It is up to the reader to interpret the:

- content of any column of any row (except column 1)
- meaning of empty cells in any row
- meaning of any new/repeated Meta-data row
- meaning of any new/repeated Header row
- meaning of any new/repeated Units row

For ease of tracing and association between different files, we recommend the following naming scheme: `CODE_Mergers_YYYY_MM_DD.csv` for the main output mergers file as described in section 2, where CODE is the code name (e.g., COMPAS, BPASS); `CODE_Meta_YYYY_MM_DD.yaml` for the meta data file as described in section 3; and `CODE_Yields_YYYY_MM_DD.csv` for the optional yields output as described in section 4. These filenames can be supplemented with additional tags (e.g., `***_A.csv`, `***_B.csv`) if needed to distinguish multiple outputs produced on the same date (for example, corresponding to different sets of assumptions).

## 6 Future enhancements

Our goal is to provide a list of common tools for parsing appropriately formatted outputs. We can also provide tools for creating gravitational-wave frame files with mock injections based on model data, or cosmic integration scripts to convolve merging binary yields with cosmic history models.