

Accessing the Data

The data for each model are labeled with ETAM and TASEP. All data files have the same naming scheme. For example, under TASEP, the folders are labeled with the structure "type1_alphavalue_results". To be specific *type1* is either pause or baseline (no pauses in the simulation). The 'alphavalue' is as written with the '-' taking the place of the decimal in the number, for example $\alpha = 0.01$ is labeled 'alpha0-01'.

1 The Available Data

Let's investigate a specific folder under TASEP, 'pause_alpha0-01_results'. You'll notice there are three folders each of which contains 50 csv files, one for each run. Here what each folder contains and how to read the data.

1.1 Collision

The Collision folder contains 50 csv files labeled collisions0, ..., collisions49. Each RNAP collision for that run was recorded in the csv file, with the start time in the second column and the end time in the third column. The first column is the position of the head of the RNAP.

1.2 Flux

The Flux folder contains 50 csv files labeled flux0, ..., flux49. Each RNAP that completed transcription during the simulation time in each run was recorded with the initiation time in the first column and termination time in the second column.

1.3 Pause

The Pause folder contains 50 csv files labeled pauses0, ..., pauses49. The RNAP pause for that run was recorded in the csv file, the start time is in the second column, the end time is in the third column. The first column is the position of the head of the RNAP. The baseline simulations do not have pauses in them so this folder is absent for those simulations.

2 TASEP Data

The TASEP data has two parts; the baseline runs (without pauses), and the pause runs. Each folder is labeled according to the value of α under the naming scheme described above.

3 ETAM Data

The ETAM data is divided into several parts that can be accessed individually. Now we will explain the data that is available in each ETAM folder.

- Main - This data set was used in the majority of analysis and is in Fig1, Fig2, and Fig3 labeled ETAM, Fig10 and Fig11 labeled Nonlinear, and in Fig14 labeled freq1.
- Linear - The piecewise linear fit for all functions: velocity, pause frequency, and pause duration. This data was used in the analysis that generated Fig10 and Fig11 labeled Linear.
- Frequency - The frequency is available in separate folders. For example, freq0-5 contains the data for the pause frequency value of 0.5 when torque is 11 pN·nm. This data was used in the analysis that generated Fig14.
- Duration - This folder contains that data for the nonlinear vs linear pause duration fits.
 - pause_duration_NL_right - The piecewise linear fit for velocity and pause frequency, linear fit for pause duration with torque values [-10,5] and the nonlinear fit for pause duration with torque values [5,11]. This data was used in the analysis that generated Fig11.
 - pause_duration_NL_left - The piecewise linear fit for velocity and pause frequency, non-linear fit for pause duration with torque values [-10,5] and the linear fit for pause duration with torque values [5,11]. This data was used in the analysis that generated Fig11.
- Count - This data kept a count of what interval of torque values the computed value belonged to. This data was used in the analysis that generated Fig15.

4 Downloading the Data

The data is stored on Montana State University's Hyalite supercomputing cluster. Links are available to transfer the data using Globus, which is free to all but does require registration. The TASEP and ETAM data is transferred separately. For the user's convenience, the ETAM data can be obtained by transferring each ETAM subfolder individually.