

# BBN Uncertainties with Gaussian Processes

Tim Launders<sup>1</sup>

<sup>1</sup>*Boston University*

(Dated: August 12, 2025)

## I. INTRODUCTION

With experimental constraints of primordial element abundances reaching percent-level precision<sup>1</sup>, theoretical predictions must be rigorously analyzed to achieve similar precision. Using a novel approach with Gaussian processes, this work directly incorporates the experimentally reported data and uncertainties (both statistical and systematic) for the nuclear reactions relevant to BBN into primordial element abundance calculations.

The complicated nature of the strong force makes theoretical calculations for these reactions very difficult even with a few interacting nucleons; therefore, experimental data must be incorporated into BBN calculations as directly as possible. Two leading BBN codes, PRIMAT [1] and PARthENoPE [2], use different techniques to fit experimental nuclear reaction data. For reactions in PRIMAT, theoretical calculations are scaled to fit data in a fully Bayesian framework. On the other hand, the reactions in PARthENoPE are fit with a low-degree polynomial. Both approaches assume specific functional forms to fit the data; for some reactions, they yield different central values for rates, which then propagate to a difference in the predicted D/H of  $\sim 2\sigma$ . Both estimate the uncertainties in primordial abundances<sup>2</sup> due to uncertainties in the reaction rates by shifting them by  $1\sigma$ .

The three deuterium burning reactions –  $d(d,n)^3\text{He}$ ,  $d(d,p)t$ , and  $d(p,\gamma)^3\text{He}$  – have the greatest influence on primordial deuterium. Since the methods employed for PRIMAT and PARthENoPE disagree on D/H while agreeing within experimental constraints on  $Y_P$ , these reactions are selected for this analysis.

Using Gaussian processes for these fits provides several benefits. (1) Passing function samples from the Gaussian process posterior through the BBN code (in this case, LINX) provides robust uncertainties on primordial abundances. (2) Assumptions regarding the functional form of reaction rates are relaxed from exact expressions to general properties that these functions are expected to follow, expanding the space of allowed functions allowed by the data. (3) More freedom is allowed in energy regions with few or no data points by not fixing an exact functional form, giving strong uncertainty estimates in these regions. (4) Systematic uncertainties are directly incorporated within the Gaussian process framework. (5) Gaussian process regression can be performed either with a purely data-driven approach or by using *ab initio* theoretical calculations as a prior.

This note outlines the current status of this work. Section 2 outlines the Gaussian process regression process for the deuterium burning reactions. The resulting abundance predictions are presented and analyzed in section 3. Section 4 discusses outstanding questions that are left for future work. The remaining sections provide further details that are omitted earlier.

## II. METHODS

Nuclear cross sections as a function of center of mass energy  $\sigma(E)$  are measured in labs, typically using a fixed-target collider. Due to the Coulomb repulsion between nuclei, cross sections typically span many orders of magnitude; it is convenient to express the cross section in terms of the astrophysical S-factor

$$S(E) = \sigma(E)Ee^{2\pi\eta}, \quad (1)$$

where  $\eta = \hbar^{-1}\sqrt{\frac{\mu}{2E}}Z_0Z_1e^2$  is the Sommerfeld parameter [3].  $\mu$  is the reduced mass of the two-body system, and  $Z_0$  and  $Z_1$  are the charges of the two nuclei. The exponential term as a whole (the Gamow factor) is the leading

---

<sup>1</sup> Need sources for this for both D/H and  $Y_P$ .

<sup>2</sup> Should look into exactly how these uncertainties are handled when factoring in other error sources.

order term for the  $\ell = 0$  transmission coefficient in a Coulomb potential. Thermally-averaging the S-factor using a Maxwell-Boltzmann distribution gives the reaction rate:

$$R(T) \equiv N_A \langle \sigma v \rangle = \sqrt{\frac{8}{\pi \mu}} \frac{N_A}{(k_B T)^{3/2}} \int_0^\infty dE e^{-2\pi\eta} S(E) e^{-E/k_B T}. \quad (2)$$

To use Gaussian processes here, certain assumptions must be made. All reported uncertainties in input datasets are assumed to follow Gaussian statistics<sup>3</sup> and to be accurate representations of the uncertainty in each measurement. In line with the literature, **systematic errors within each dataset are assumed to affect each data point in the same way**, shifting each point up or down by the same (percent) amount. To incorporate these assumptions within the Gaussian process framework, the covariance matrix of the input datasets is added to the upper-left block of the partitioned full covariance matrix, in the same spirit as the standard procedure of adding statistical errors to the diagonal.

Additionally, any kernel choice includes assumptions about the latent function. The kernel encodes correlations between S-factor values at different energies; a simple, common kernel choice is the squared exponential<sup>4</sup> [4]

$$k_{SE}(d) = e^{-\frac{d^2}{2\ell^2}}$$

where  $\ell$  is a hyperparameter representing the characteristic length scale of correlations in input space and  $d$  is the distance between two points in input space. When the latent function is expected to be smooth throughout input space with correlations that only depend on the distance between points in input space, the squared exponential is an appropriate kernel choice. The Matérn class of kernels somewhat relax this smoothness assumption while still including a characteristic length scale in a similar manner:

$$k_{M,\nu}(d) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2\nu}d}{\ell} \right)^\nu K_\nu \left( \frac{\sqrt{2\nu}d}{\ell} \right),$$

where  $K_\nu$  is the modified Bessel function of the second kind [4].  $\nu$  is a positive parameter that controls the smoothness of functions drawn from the Gaussian process posterior. A smaller  $\nu$  gives less smooth functions, and the Matérn kernel becomes the squared exponential in the limit  $\nu \rightarrow \infty$ .

Kernel choices for this application should yield S-factor draws should not deviate strongly from leading theoretical calculations with error bars that are reasonable given reported experimental uncertainties. Sampled S-factor functions are expected to be globally smooth, as deuterium burning reactions have no resonances in this energy range. This observation suggests that the squared exponential kernel would work well; however, a Matérn kernel must be added to the squared exponential to fit these criteria. With only the squared exponential kernel, the fit is driven to a small correlation amplitude and length scale since the Gaussian process assumes that all points are drawn from the same latent function. When multiple perfectly correlated datasets have data points at very similar energies that differ by  $> 1\sigma$ , there are two options for the Gaussian process: either some of the datasets are multiple  $\sigma$  off from the latent function, or the latent function itself has significant small-scale fluctuations. During hyperparameter optimization, a small characteristic length scale for the squared exponential kernel is chosen as that is more probable than datasets being off by multiple  $\sigma$ . The addition of a small- $\nu$  ( $\nu < 1$ ) Matérn kernel allows the model to fit these small-scale fluctuations while accommodating the global shape provided by the squared exponential kernel. Therefore, the following kernel form<sup>5</sup> is used for the deuterium burning reactions:

$$k(d) = \sigma_0 k_{SE} + \sigma_1 k_{M,1/4}$$

with four hyperparameters – two constants controlling the amplitude of correlations and two characteristic length scales. This kernel choice implicitly assumes that correlations only depend on differences in energies. Whenever any kernel is used in this work, **log energies are used**.

These hyperparameters are selected with a form of cross-validation in which each dataset is withheld from the rest, and the model is evaluated on its performance on the unseen dataset<sup>6</sup>. The Adam optimizer is used for gradient

<sup>3</sup> Perhaps we want to elaborate on this point and show why this is a reasonable assumption.

<sup>4</sup> The squared exponential kernel is often referred to as the RBF (radial basis function) kernel.

<sup>5</sup> This notation feels super clunky...

<sup>6</sup> Details for this method are in section 5.

descent to avoid convergence to a local minimum. Optimization is considered converged when the step size drops below  $10^{-5}$ . The standard marginal likelihood to observe all of the data points is included for comparison. Two different prior mean functions are considered for the S-factor: an uninformative prior that is zero everywhere and the theoretical calculations used by PRIMAT.

Once the model has been fit to the experimental data, sample functions are drawn from the posterior and integrated at a discretized interval of temperatures using (2) to get the rate distribution, and these rate samples are directly used in LINX [5]. Averaging the rate distribution at each temperature gives the  $1\sigma$  error bars<sup>7</sup> shown in subsequent plots.

### A. $d(d,n)^3\text{He}$

Figure 1 shows four plots to analyze the Gaussian process fit to  $d(d,n)^3\text{He}$  S-factor data. The model fitted using the a zero prior mean and the “Leave-Dataset-Out” hyperparameter selection method is shown in all plots. The posterior mean is somewhat jagged in some regions from the contribution of the Matérn kernel component, while the smoothness from the squared exponential kernel component dominates at higher energies. If the data had preferred this smooth behavior at all energies, the constant hyperparameter multiplying the Matérn component would have been driven to zero. Individual sample draws have more jagged small-scale fluctuations than the mean, but when integrated rate draws are smooth. These small-scale fluctuations are a product of needing the small- $\nu$  Matérn component in the kernel to capture correlations. S-factor uncertainties are comparable to reported experimental uncertainties and increase slightly in regions with without any data points. The overall shape is comparable with the theoretical prediction, notably between 0.3-0.8 MeV where there are no data. The Gaussian process  $1\sigma$  rate envelope is around 1.5% in the relevant temperature range for BBN, slightly larger than those of both PRIMAT and PArthENoPE.

The Gaussian process mean lies closest to the BR90 dataset (shown in blue) in the energy range with overlapping datasets, as one expects since this dataset has the smallest reported uncertainties. Additionally, each of the higher energy points in the KR87M dataset (shown in green) lie below the Gaussian process mean by the same percentage. This suggests that this technique handles correlations as desired since these points have negligible statistical uncertainty. The fit is mostly insensitive to prior mean choice and hyperparameter optimizer; using the theoretical prediction for the prior mean and optimizing the marginal likelihood make very little impact on the fit. The most impactful difference when using the marginal likelihood comes between 0.03 and 0.1 MeV, where the Gaussian process mean is consistently very slightly lower than the other two correlated Gaussian process means, resulting in a slightly lower rate<sup>8</sup>. However, excluding correlations from the model entirely makes a meaningful impact, aligning the fit more closely with that from PArthENoPE. Perhaps this suggests that fitting a higher degree polynomial with their method would indeed have little impact on the fit, and some aspect for their handling of systematic uncertainties is insufficient<sup>9</sup>.

Both the S-factor and rate agree better with those used in PRIMAT, despite Gaussian process regression seeming more similar to fitting a polynomial. However, the Gaussian process has much greater flexibility in fitting the data than a polynomial, and the data prefer a fit that is more similar to the theoretical calculation regardless of the prior mean choice.

### B. $d(d,p)t$

Figure 2 shows four plots to analyze the Gaussian process fit to  $d(d,p)t$  S-factor data. As with  $d(d,n)^3\text{He}$ , the model fitted using the a zero prior mean and the “Leave-Dataset-Out” hyperparameter selection method is shown in all plots. Many qualities of this reaction are similar to those of  $d(d,n)^3\text{He}$  as five of the six experiments used in each are the same. Without the highest energy dataset of  $d(d,n)^3\text{He}$ , the uncertainties grow much more significantly after  $\sim 0.3$  MeV. However, the addition of GR95C (shown in brown) greatly constrains the fit at low energies. The PRIMAT fit does not include this dataset, driving the Gaussian process fit above that of PRIMAT in this region.

<sup>7</sup> Could show a plot demonstrating that the rate samples also follow Gaussian distributions, but I am not sure that it is necessary.

<sup>8</sup> Seen in the next section this translates to a slightly higher predicted value of D/H.

<sup>9</sup> This is a big claim...

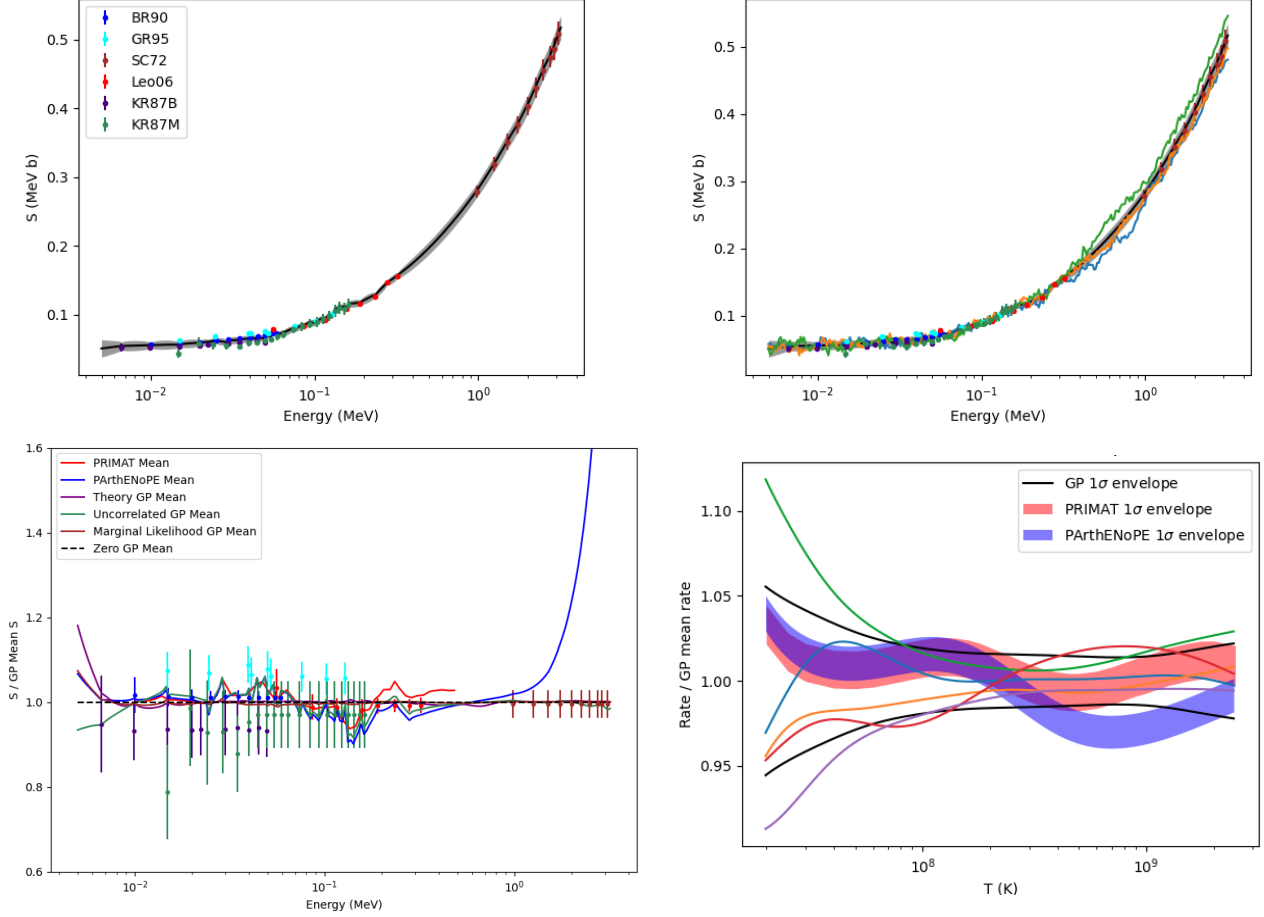


FIG. 1. Gaussian process regression fits to  $d(d,n)^3\text{He}$  S-factor data. *Top left*: Gaussian process mean and  $1\sigma$  envelope in the BBN energy range. An uninformative zero prior mean and “Leave-Dataset-Out” hyperparameter selection are used. The six datasets (citations here) used to fit the model are shown with  $1\sigma$  total error bars. *Top right*: Three samples drawn from the posterior overlayed on the mean and  $1\sigma$  envelope. *Bottom left*: Three additional Gaussian process means – one using the *ab initio* theoretical calculation from (citation here) as the prior mean, one excluding correlations within datasets, and one using the marginal likelihood for hyperparameter selection – as well as the fits used by PRIMAT and ParthENoPE, normalized to the fit above. *Bottom right*: Gaussian process S-factor fit in the top plots along with those of PRIMAT and ParthENoPE integrated to get the reaction rate. 5 samples drawn from the posterior are shown.

Again, the addition of the Matérn component to the kernel makes the mean and subsequent sample draws jagged with small-scale fluctuations which become smooth after integration. S-factor uncertainties are reasonable given the reported experimental uncertainties, and the shape of the Gaussian process mean is comparable with the theoretical prediction. It is somewhat interesting that after 0.3 MeV, where there are no longer experimental data, the Gaussian process mean does not collapse to the zero prior mean and continues to increase. The Gaussian process  $1\sigma$  rate envelope is around 1.1% in the relevant temperature range for BBN, comparable in size to those of PRIMAT and ParthENoPE.

The Gaussian process mean again lies closest to those datasets with smallest error bars, and correlations can be seen in the same way as with  $d(d,n)^3\text{He}$  in the normalized S-factor plot. The fit is slightly more sensitive to using the theoretical prediction for the prior mean but only at the lowest and highest energies where there are no data. The uncorrelated Gaussian process mean is not as similar to the fit from ParthENoPE<sup>10</sup>, but still aligns closely between 0.01-0.09 MeV. Between 0.02 and 0.3 MeV the Gaussian process means match the PRIMAT fit very well, which is

<sup>10</sup> Further suggesting that a higher degree polynomial would not fit the data better and that their method is not handling correlations well

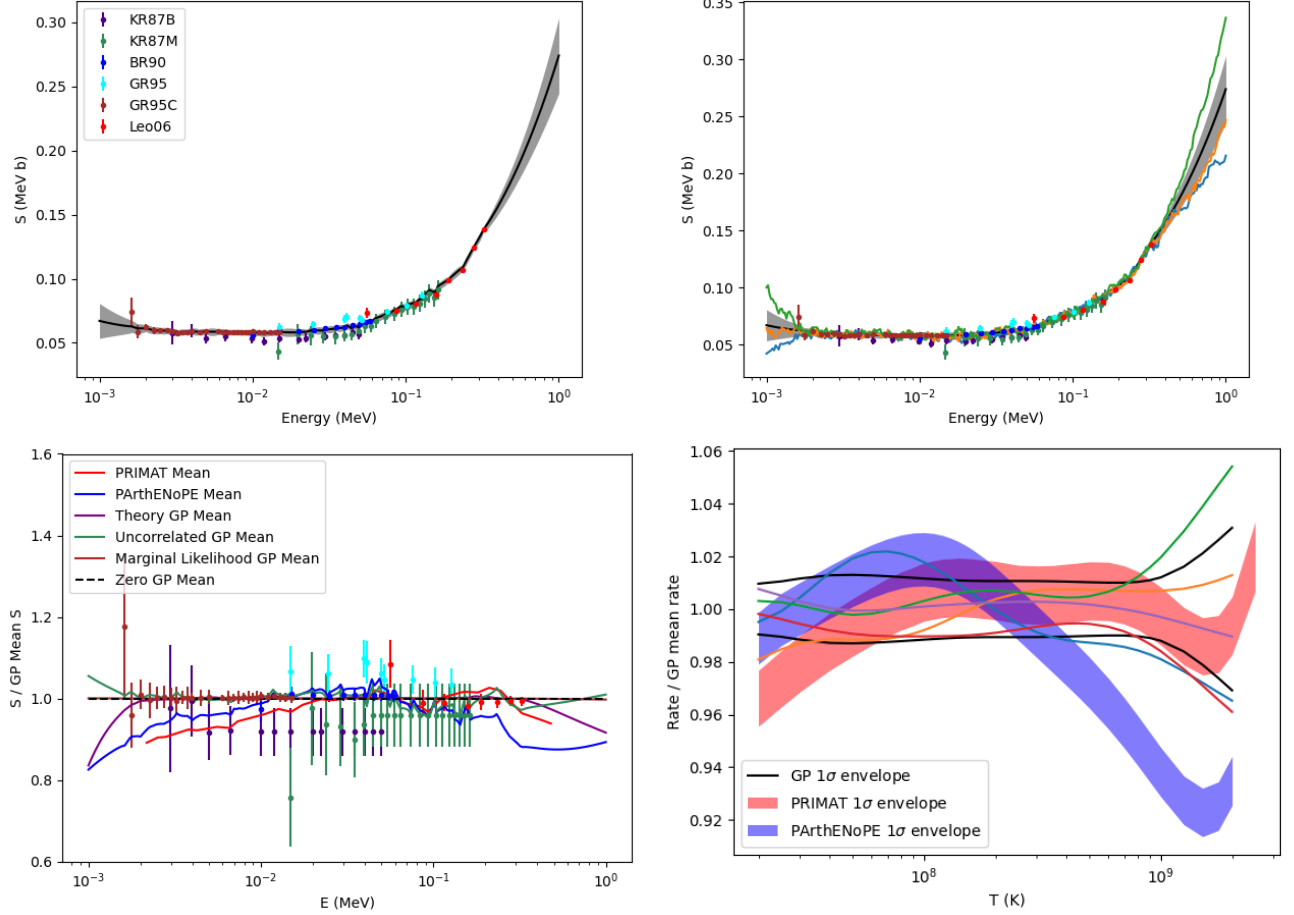


FIG. 2. Gaussian process regression fits to  $d(d,p)t$  S-factor data. *Top left*: Gaussian process mean and  $1\sigma$  envelope in the BBN energy range. An uninformative zero prior mean and “Leave-Dataset-Out” hyperparameter selection are used. The six datasets (citations here) used to fit the model are shown with  $1\sigma$  total error bars. *Top right*: Three samples drawn from the posterior overlayed on the mean and  $1\sigma$  envelope. *Bottom left*: Three additional Gaussian process means – one using the *ab initio* theoretical calculation from (citation here) as the prior mean, one excluding correlations within datasets, and one using the marginal likelihood for hyperparameter selection – as well as the fits used by PRIMAT and ParthENoPE, normalized to the fit above. *Bottom right*: Gaussian process S-factor fit in the top plots along with those of PRIMAT and ParthENoPE integrated to get the reaction rate. 5 samples drawn from the posterior are shown.

reflected in the rate envelopes as well. Again, the data prefer a fit that is more similar to the theoretical calculation regardless of prior mean choice.

### C. $d(p,\gamma)^3\text{He}$

Fitting a Gaussian process model to  $d(p,\gamma)^3\text{He}$  is much more difficult than with either of the  $dd$  reactions – in the relevant energy range, the S-factor spans approximately two orders of magnitude. Luckily, this reaction has been studied recently by the LUNA collaboration, and the rates from PRIMAT and ParthENoPE agree very well. Since the two fitting methods agree well, it is less important to use Gaussian processes for this reaction. Figure 3 shows four plots to analyze Gaussian process fits to  $d(p,\gamma)^3\text{He}$  S-factor data. Two fits using a zero prior mean are considered: one fitting to the S-factor data directly, and one fitting to  $\log S$ . The first of these fits is consistently below both of the two LUNA datasets (CA02 in yellow and MO20 in blue), which are fit more satisfactorily by PRIMAT and ParthENoPE. Using the theoretical calculation for the prior mean makes little difference up to 1 MeV. Different kernel choices do not help nearly enough in this regard. After integrating to get the reaction rate, the Gaussian process rate is over  $1\sigma$

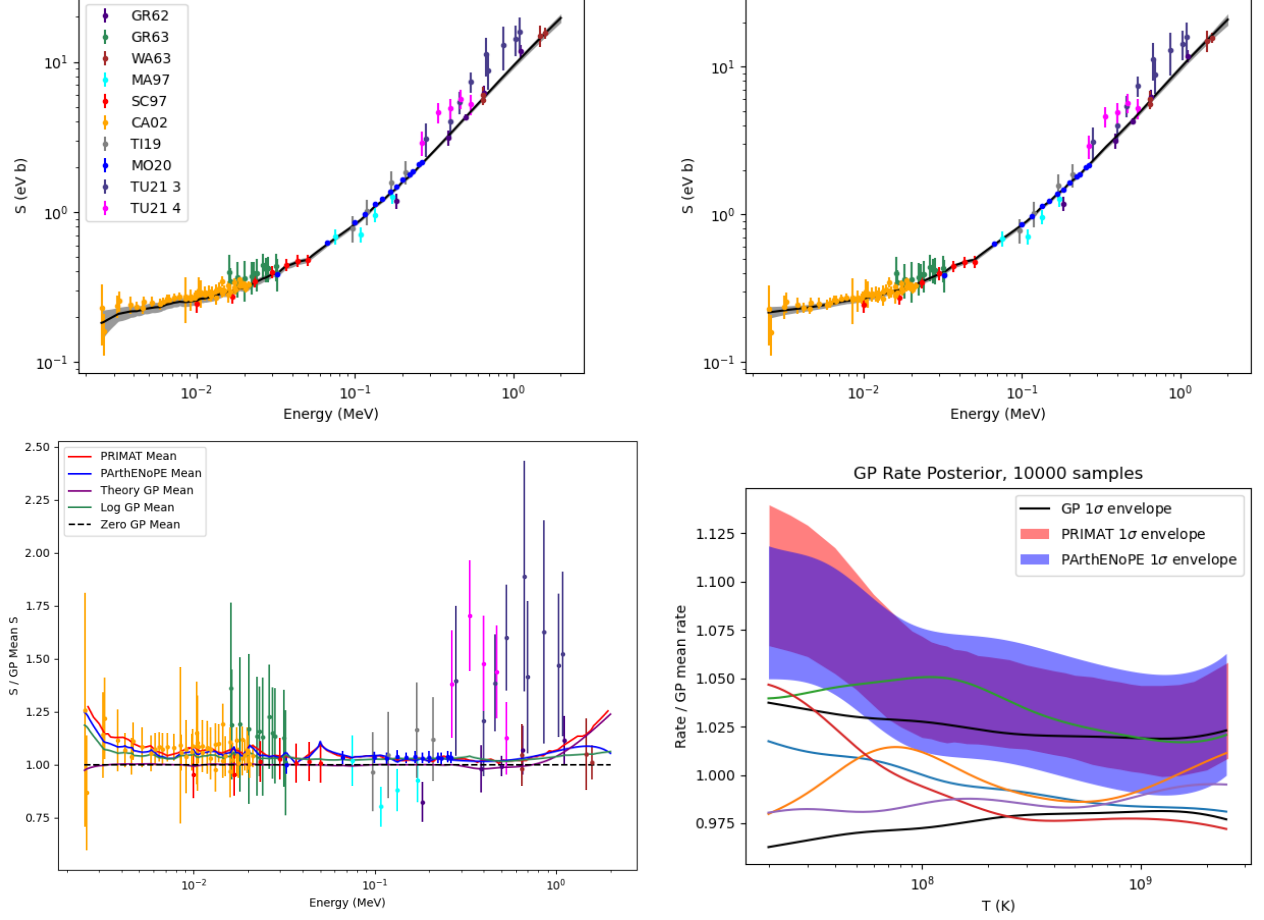


FIG. 3. Gaussian process regression fits to  $d(p,\gamma)^3\text{He}$  S-factor data. *Top left*: Gaussian process mean and  $1\sigma$  envelope in the BBN energy range. An uninformative zero prior mean and “Leave-Dataset-Out” hyperparameter selection are used. The ten datasets (citations here) used to fit the model are shown with  $1\sigma$  total error bars. *Top right*: Gaussian process mean and  $1\sigma$  envelope in the same energy range. This model is fit in  $\log S$ ; everything else in the fit process is the same as in the top left plot. *Bottom left*: In addition to the two models above, a Gaussian process fit to S-factor data (not  $\log S$ ) with the theoretical calculation as the prior mean. All three models, along with the best-fit S-factors from PRIMAT and ParthENoPE, are normalized to the Gaussian process mean in the top left plot. *Bottom right*: Gaussian process S-factor fit in the top left plot along with those of PRIMAT and ParthENoPE integrated to get the reaction rate. 5 samples drawn from the posterior are shown.

discrepant with both PRIMAT and ParthENoPE in the relevant BBN temperature range.

When fitting to  $\log S$ , one must assume that the uncertainties on S-factor measurements follow a log-normal distribution as opposed to a Gaussian distribution. This assumption is less sound for statistical errors, but may be appropriate for systematic errors<sup>11</sup>. For uncertainties less than 10%, there is little difference between normal and log-normal distributions. From the reported mean and variance in an S-factor measurement  $\mu_S$  and  $\sigma_S^2$ , the standard deviation of the lognormal distribution  $\sigma$  is

$$\sigma = \sqrt{\log \left( \frac{\sigma_S^2}{\mu_S^2} + 1 \right)}.$$

The Gaussian process fit to  $\log S$  uses this expression when constructing the covariance matrix, yielding a much better

<sup>11</sup> Since systematic errors apply multiplicatively and the S-factor must be positive

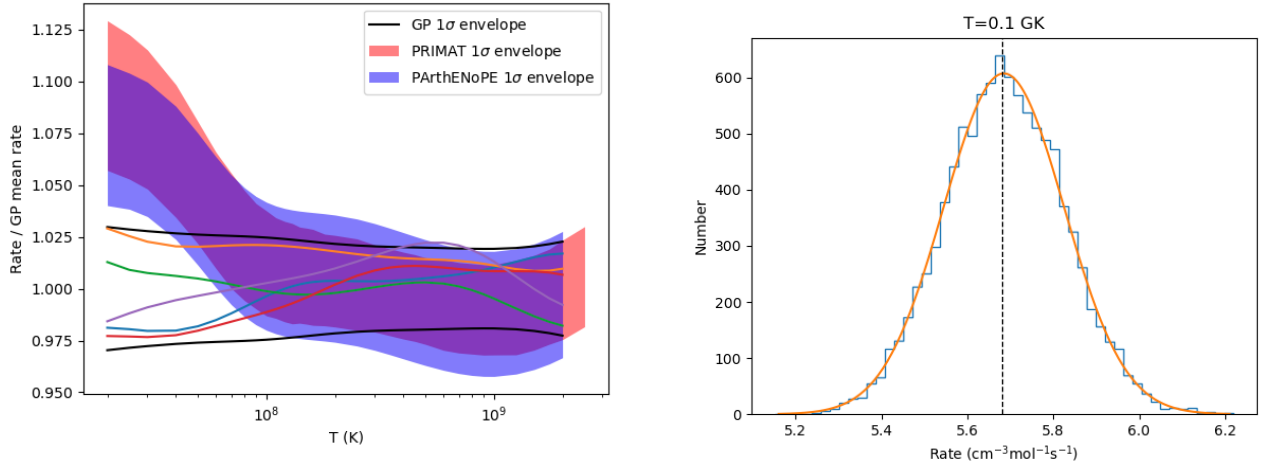


FIG. 4.  $d(p,\gamma)^3\text{He}$  reaction rates for a Gaussian process fit to  $\log S$  with a zero prior mean. *Left*: The  $1\sigma$  rate envelope, along with those of PRIMAT and PArthENoPE. 5 sample draws are shown. *Right*: Histogram of the reaction rate at  $10^8$  K, overlaid with a Gaussian fit. The black dashed line is the rate for the Gaussian process mean.

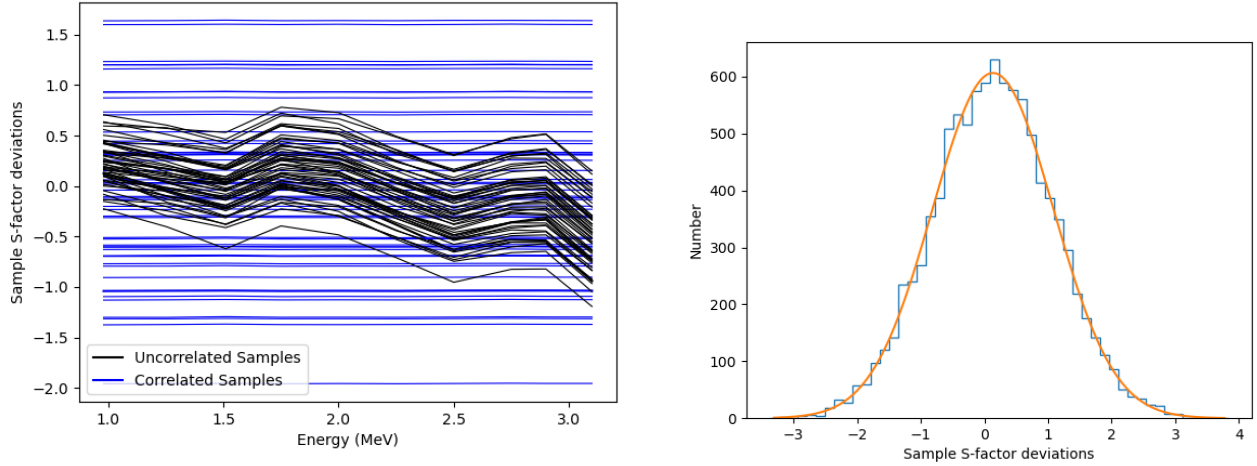


FIG. 5. Sample draws from Gaussian process posteriors both including and excluding correlations. A zero prior mean is used for both. *Left*: 50 sample draws from both posteriors at the energies measured by SC72 normalized to their reported values. *Right*: Histogram of 10,000 samples from the correlated posterior showing the number of SC72 standard deviations at the lowest energy measured, overlaid with a Gaussian fit.

fit to both LUNA datasets. Unfortunately, fitting to  $\log S$  for the  $dd$  reactions does not yield satisfactory fits as those S-factors do not span multiple orders of magnitude<sup>12</sup>.

Figure 4 shows the reaction rate posterior for the Gaussian process model fit to  $\log S$ . Above  $10^8$  K the rate agrees significantly better with those of PRIMAT and PArthENoPE when compared to models fit directly to  $S$ . The rate remains  $\gtrsim 1\sigma$  discrepant at lower temperatures as the S-factor fit to the lowest energy LUNA data is still below the other fits (see figure 4). At each temperature, the rate posterior still follows a Gaussian distribution closely despite treating the S-factor as following a log-normal distribution, shown in the right plot.

<sup>12</sup> It is not clear what to do about this issue. Either one fits to just the two  $dd$  reactions or includes  $d(p,\gamma)^3\text{He}$  using a fit to  $\log S$ .



## D. Correlations

The S-factor plots normalized to the Gaussian process means shown in figures 1 and 2 suggest that correlations are handled well by the Gaussian process model, since perfectly correlated data points lie the same amount above or below. To further confirm that correlations are handled in a similar manner in individual S-factor samples, consider the high energy regime for  $d(d,n)^3\text{He}$ . SC72 collected data in this energy range with negligible statistical uncertainty, and so each point in this set is considered perfectly correlated with every other. If correlations are handled as expected, shifting each data point up or down by the same amount, then each sample draw should lie the same amount above or below the Gaussian process mean at the energies of those data points.

Figure 5 shows sample draws at the SC72 energies from two Gaussian processes with zero prior mean: one including correlations and one without. Every sample from the correlated posterior is the same distance from the reported values from SC72, while this is not true for the uncorrelated samples. At each energy, the mean of the samples is  $0.13\sigma$  above the experimental mean, and the standard deviation of the samples is 95% of the experimental standard deviation. The small differences here from 0 and 1 respectively are from the influence of lower energy data on the Gaussian process fit. This result confirms that the correlations are handled robustly by Gaussian processes.

## III. PRIMORDIAL ABUNDANCES

To directly incorporate the reported uncertainties in experimental S-factor data into uncertainties in primordial abundance predictions, samples drawn from Gaussian process posteriors are passed into a slightly modified version of LINX, which allows for custom rates to be used for as many of the 12 key reactions as desired. For each sample, the final abundances of hydrogen, deuterium, helium-3, and helium-4 are recorded. Reported central values and error bars are the mean and standard deviation for each abundance, always closely following a Gaussian distribution. 10,000 samples are used for very combination of rates, hyperparameter optimization, and kernel choice considered. When multiple reactions are varied simultaneously, they are considered uncorrelated<sup>13</sup>.

Each prediction is compared with those using exclusively the rates from PRIMAT and PArthENoPE. Error bars for these depend on which Gaussian process rates are varied. Without the Gaussian process framework for drawing samples, uncertainties when using these rates are estimated by modeling the rate as following a log-normal distribution

$$\log R(T) = \log \bar{R}(T) + q \log \Delta R(T),$$

where  $\bar{R}(T)$  and  $\Delta R(T)$  are the mean and  $1\sigma$  rates at temperature  $T$  and  $q \sim \mathcal{N}(0, 1)$ . When only a single rate is varied, the error bars reported below for the PRIMAT and PArthENoPE rates are simply obtained by calculating the abundances for  $q = 0$  and  $q = 1$ . When multiple rates are varied,  $q$  is sampled for each rate 10,000 times and used in LINX. The error bars for PRIMAT and PArthENoPE rates are therefore larger when more rates are considered.

The subsections below walk through the impact on predicted abundances when Gaussian process regression is used for each reaction. Generally, using theory instead of zero mean priors makes very little difference, and using marginal likelihood instead of LDO hyperparameter optimization has a larger, yet still small, impact. Each reaction agrees quite well with the PRIMAT result, and when combined the predicted deuterium abundances are within  $0.15\sigma$ . Uncertainties are roughly the same when combined.

### A. $d(d,n)^3\text{He}$

Figure 6 shows D/H,  $^3\text{He}/\text{H}$ , and  $Y_{\text{P}}$  predictions when using four different Gaussian process fits to  $d(d,n)^3\text{He}$ : a zero prior mean and theory prior mean using LDO optimization, a zero prior mean using marginal likelihood optimization, and a zero prior mean using LDO optimization without the inclusion of correlations. For each Gaussian process fit, both PRIMAT and PArthENoPE rates are used for the other rates, giving eight sets of predictions in

<sup>13</sup> Acknowledgement of this assumption is important as theoretical predictions suggest that the two  $d+d$  reactions are not independent of one another, but handling this in a more sophisticated manner is well beyond the scope of this work.



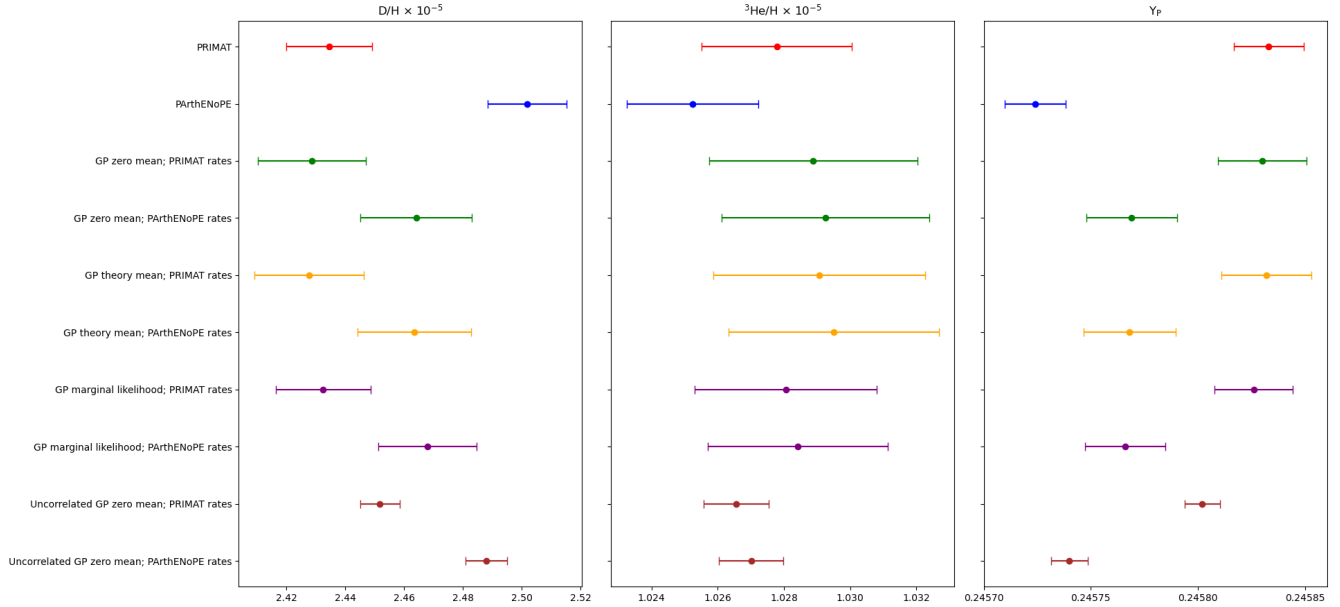


FIG. 6. Comparison of primordial abundance predictions using Gaussian process regression to fit  $d(d,n)^3\text{He}$  data. Unless otherwise specified, LDO optimization is used for hyperparameter selection and correlations are fully included during regression. Results from using exclusively rates from PRIMAT and PArthENoPE are also shown.

total. The uncorrelated Gaussian process produces error bars much smaller than the other scenarios, as expected with more degrees of freedom when performing regression. In general, a theory mean prior makes minimal difference on abundance predictions. Error bars for the two LDO cases are around 25% larger than PRIMAT and PArthENoPE<sup>14</sup>. Changing to marginal likelihood optimization has a slight effect, changing abundances by  $\sim 0.25\sigma$  while also shrinking error bars by around 10%. These observations apply when both PRIMAT and PArthENoPE rates are used for the other reactions.

The predicted D/H with PRIMAT rates<sup>15</sup> decreases slightly with the Gaussian process, while the decrease from PArthENoPE rates is much more significant, with the predicted value lying between the predictions from using only PRIMAT and PArthENoPE rates. One would expect predicted  $^3\text{He}/\text{H}$  values to therefore increase, since more deuterium is converted to helium-3 with the Gaussian process rate, and this is reflected in the abundance results. The predicted  $^3\text{He}/\text{H}$  is higher when the PArthENoPE network is used from other relevant reactions.  $Y_P$  remains relatively similar using the PRIMAT network and increases when using the PArthENoPE network.

## B. $d(d,p)t$

Figure 7 shows D/H,  $^3\text{He}/\text{H}$ , and  $Y_P$  predictions when using different Gaussian process fits to  $d(d,p)t$ . Again, error bars when the uncorrelated Gaussian process fit is used are smaller than all other cases, and a theory mean prior makes minimal difference in abundances. Using marginal likelihood hyperparameter optimization yields nearly identical results as LDO optimization for this reaction. Unlike  $d(d,n)^3\text{He}$ , error bars shrink by around 10% when the Gaussian process rates are used.

The predicted D/H with PRIMAT rates increases slightly with the Gaussian process. Again, D/H with PArthENoPE decreases significantly with the Gaussian process, with the predicted value lying between the predictions from using only PRIMAT and PArthENoPE rates. The predicted  $^3\text{He}/\text{H}$  now is lower when the PArthENoPE network is used from other relevant reactions.  $Y_P$  remains relatively similar using the PRIMAT network and increases when using the PArthENoPE network, although this increase is greater than with  $d(d,n)^3\text{He}$ .

<sup>14</sup> Include tables for these numbers in the appendix?

<sup>15</sup> This whole comparison feels super clunky, but I am not sure how best to explain otherwise.

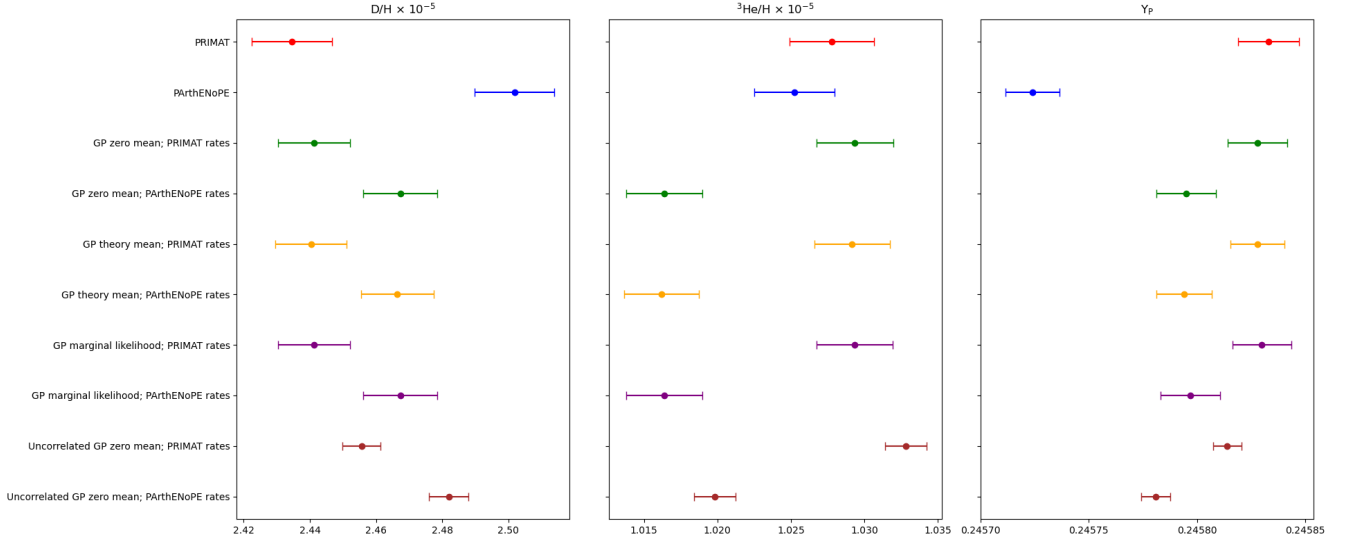


FIG. 7. Comparison of primordial abundance predictions using Gaussian process regression to fit  $d(d,p)t$  data. Unless otherwise specified, LDO optimization is used for hyperparameter selection and correlations are fully included during regression. Results from using exclusively rates from PRIMAT and PArthENoPE are also shown.

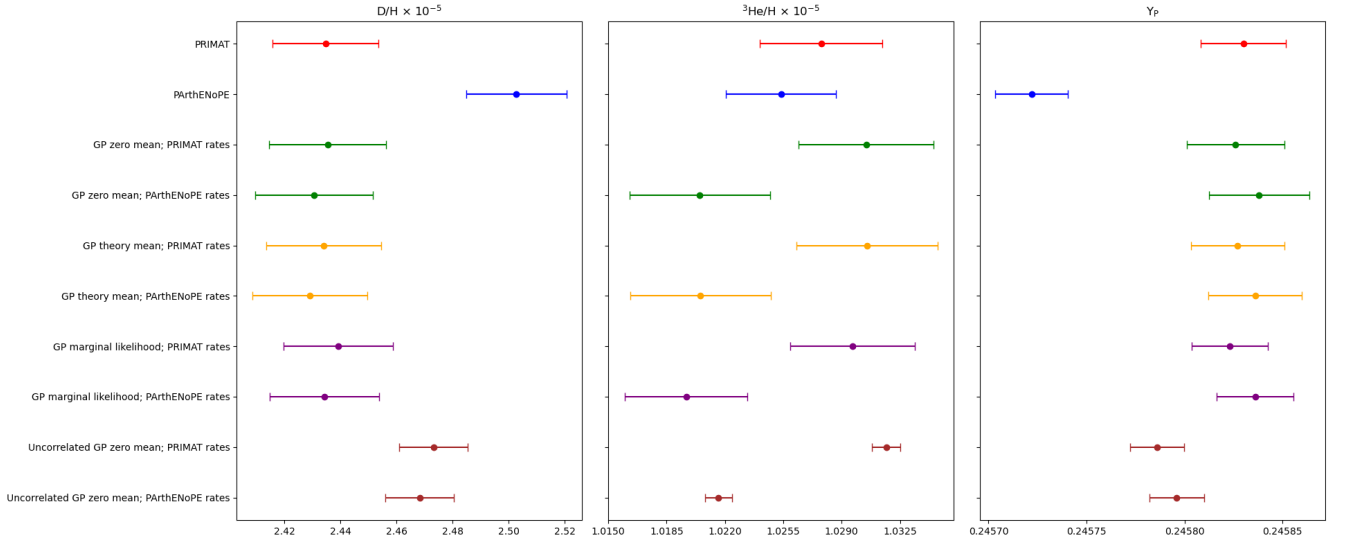


FIG. 8. Comparison of primordial abundance predictions using Gaussian process regression to fit both  $d(d,n)^3\text{He}$  and  $d(d,p)t$  data. Unless otherwise specified, LDO hyperparameter optimization is used and correlations are fully included during regression. Results from using exclusively rates from PRIMAT and PArthENoPE are also shown.

### C. $d+d$ Rates

Figure 8 shows  $D/H$ ,  $^3\text{He}/H$ , and  $Y_p$  predictions when using different Gaussian process fits to  $d(d,n)^3\text{He}$  and  $d(d,p)t$  simultaneously. As expected from varying these rates individually, error bars when the uncorrelated Gaussian process fit is used are smaller than all other cases, and a theory mean prior makes minimal difference in abundances. Since marginal likelihood optimization shifts central values by around 25% for  $d(d,n)^3\text{He}$  alone and has negligible impact for  $d(d,p)t$ , the 25% shift remains for this combination and error bars again slightly decrease.

Predicted  $D/H$  values now agree well with the base PRIMAT calculation when either PRIMAT or PArthENoPE networks are used along with the Gaussian processes; this makes sense as two of the three most relevant reactions for deuterium production are the same, and the third is very similar in both networks. Helium-3 becomes more

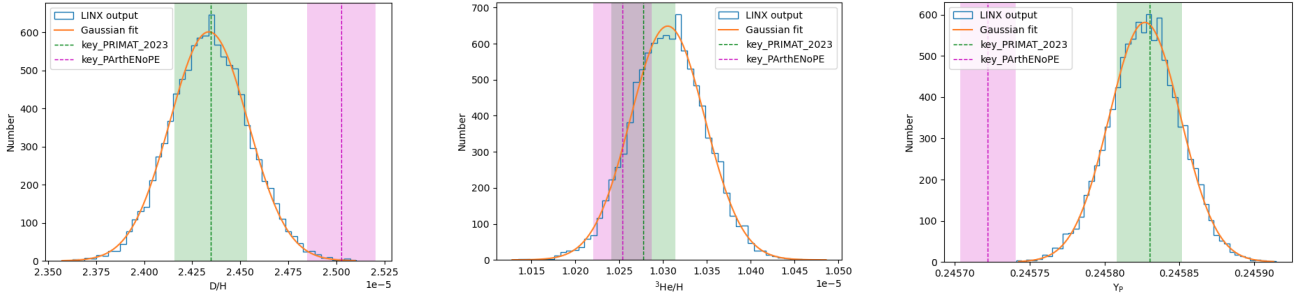


FIG. 9. Histograms of primordial abundance predictions using Gaussian process regression to fit both  $d(d,n)^3\text{He}$  and  $d(d,p)t$  data. 10,000 samples are shown. A Gaussian fit is overlayed on each, and  $1\sigma$  envelopes for PRIMAT and ParthENoPE are shown as vertical bands.

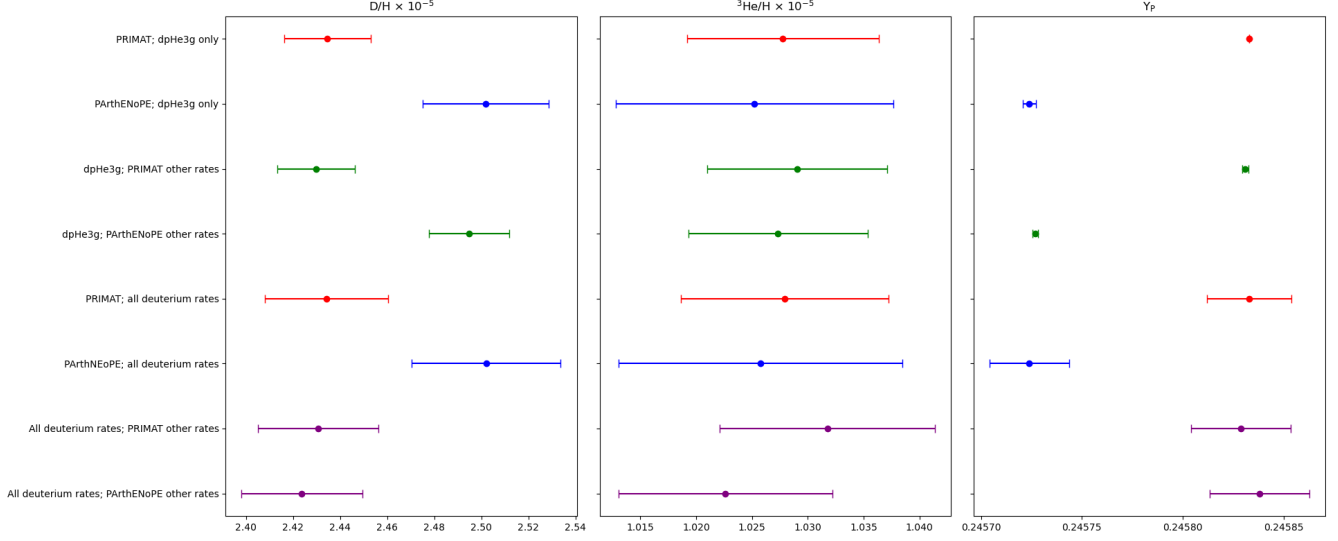


FIG. 10. Comparison of primordial abundance predictions using Gaussian process regression to fit  $d(p,\gamma)^3\text{He}$  data in  $\log S$ . The top four points vary only this reaction, while the bottom four also vary  $d(d,n)^3\text{He}$  and  $d(d,p)t$ . For each reaction, LDO hyperparameter optimization and a zero mean prior are used for regression.

discrepant due to the other relevant reactions, with abundances using the PRIMAT network slightly higher than the base PRIMAT calculation (from  $d(d,n)^3\text{He}$ ) and those using the ParthENoPE network lower.  $Y_p$  all remain close to the PRIMAT calculation even when using the ParthENoPE network.

Figure 9 shows histograms for  $D/H$ ,  $^3\text{He}/H$ , and  $Y_p$  predictions using Gaussian processes fit with theory priors for  $d(d,n)^3\text{He}$  and  $d(d,p)t$  and the PRIMAT network for other reactions. A Gaussian fit is overlayed on each; all three distributions closely follow Gaussian distributions. The  $1\sigma$  envelopes of PRIMAT and ParthENoPE are shown as vertical bands on each plot.

#### D. $d(p,\gamma)^3\text{He}$

Figure 10 shows  $D/H$ ,  $^3\text{He}/H$ , and  $Y_p$  predictions using the Gaussian process fit to  $d(p,\gamma)^3\text{He}$  in  $\log S$ , with a zero mean prior and using LDO hyperparameter optimization. The top four points use only this reaction, and the top four use the zero mean prior, LDO hyperparameter optimization Gaussian process fits to  $d(d,n)^3\text{He}$  and  $d(d,p)t$  as well. When varying only  $d(p,\gamma)^3\text{He}$ , error bars for  $D/H$  and  $^3\text{He}/H$  are smaller than for PRIMAT and ParthENoPE.  $D/H$  central values with both networks decrease slightly, and  $^3\text{He}/H$  central values increase; the Gaussian process rate is slightly higher between 0.3 and 2 GK, so this result is expected. It is interesting to note that the much lower rate at temperatures below 0.1 GK predicted by Gaussian process regression does not matter for these calculations. Small

shifts in  $Y_P$  are due to changes in deuterium and helium-3 propagating to other reactions, as this reaction has little direct impact on helium-4.

When all three deuterium burning reactions are sampled from Gaussian processes, there is remarkable agreement with PRIMAT for D/H and  $Y_P$  both in central values and uncertainties. Uncertainties on helium-3 are dominated by  $d(p,\gamma)^3\text{He}$ ; therefore, all predictions roughly agree.

#### IV. FUTURE WORK

There are a few outstanding questions for this method, but they are (for now) left for future work. It would be interesting to apply Gaussian process regression to the other nine key reactions to analyze impacts on lithium and helium-3, particularly  $^3\text{He}(d,p)\alpha$ ; the rates from PRIMAT and PARthENoPE disagree by multiple  $\sigma$  toward the low temperature range of BBN. Additionally, this reaction contains a broad resonance, and fitting a Gaussian process model would check if this method still works with resonances.

Cosmological parameter estimation becomes more difficult: with Gaussian processes, there are now infinitely many nuisance parameters to marginalize over instead of just one per reaction. The trained Gaussian process models provide priors for estimation; one could imagine performing sampling without treating these as nuisance parameters, drawing sample functions at each step, or choosing some pivot energies to marginalize over.

It would be fantastic to inject mock datasets into the regression to estimate to what extent uncertainties can be reduced from future experiments. If a new dataset could be generated with error bars that are reasonable given current experimental capabilities, it would be simple to add to the Gaussian process pipeline. It is still unclear how such a mock dataset should be created.

Kernel hyperparameter marginalization is, and will most likely always, be left for future work. However, considering marginal likelihood optimization in addition to LDO optimization helps to provide some sense of the sensitivity of results to hyperparameter choice. In both cases, the general results are the same with some details differing.

#### V. GAUSSIAN PROCESSES

Gaussian processes enable the sampling of functions themselves from a fit to experimental data without assuming any particular functional form. This can be done either with a purely data-driven approach or by including the theoretical calculations (those used in PRIMAT) as a prior. Draws from the Gaussian process posterior are directly used downstream in LINX while varying as many reactions simultaneously as desired.

A Gaussian process is a stochastic process in which a collection of random variables jointly follow a multivariate normal distribution [4]. To use a Gaussian process for this application, the experimentally measured S-factor values are assumed to be drawn from the same latent function (the “true” S-factor) with Gaussian noise. By discretizing the relevant energy range, the probability distribution of the latent function can be modeled as following a multivariate normal distribution with some mean and covariance matrix.

For a given reaction, let  $S$  and  $S_*$  denote the measured S-factor values and latent function values (to be inferred) respectively. In subsequent equations, these quantities are vectors. The partitioned multivariate normal distribution is

$$\begin{bmatrix} S \\ S_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \right), \quad (3)$$

where  $\mu$  is the function mean prior and  $K$  describes the correlation between points in input space. The predictive mean and covariance for the underlying function are [4]

$$\mu_{S_*|S} = \mu_* + K_{21}K_{11}^{-1}(S - \mu), \quad \Sigma_{S_*|S} = K_{22} - K_{21}K_{11}^{-1}K_{12}. \quad (4)$$

$\mu$  is typically set to 0 for a purely data-driven approach since the predictive mean is non-zero; however, prior information (in this case, the theory prediction) can be included in  $\mu$ . The kernel is a function of two points in input

space that encodes the correlation between the latent function values at those points. Kernels also include other hyperparameters  $\theta$  that must be chosen. Selecting the functional form of the kernel is a key part of using Gaussian processes for regression and should be chosen by considering the properties that the underlying function is expected to follow. Kernels can be added or multiplied (including by an overall constant) and remain a valid kernel.

In (1),  $K_{12}$ ,  $K_{21}$  and  $K_{22}$  are entirely set by the kernel. For the selected kernel  $k$ , the  $ij^{th}$  component of  $K_{12}$  is given by

$$K_{12,ij} = k(E_i, E_{*j}),$$

where  $E_i$  is the energy for the  $i^{th}$  entry in the vector of S-factor measurements  $S$  and  $E_{*j}$  is the  $j^{th}$  entry in the vector of energies at which to make predictions. Similarly, the  $ij^{th}$  component of  $K_{22}$  is given by

$$K_{22,ij} = k(E_{*i}, E_{*j}).$$

$K_{11}$  can also be entirely set by the kernel, or known variance in the data can be included in addition. This allows both statistical and systematic errors as reported for each dataset to be included as a full covariance matrix  $\Sigma$  that is added to the kernel component, modeling both correlations from experiments and correlations from the latent function. The  $ij^{th}$  component of  $\Sigma$ , where point  $i$  is from dataset  $k$  and point  $j$  is from dataset  $l$  is given by

$$\Sigma_{ikjl} = \delta_{ij}\sigma_{stat,i}\sigma_{stat,j} + \delta_{kl}\sigma_{sys,i}\sigma_{sys,j}.$$

After selecting a particular kernel, the hyperparameters must be chosen in a principled manner before using (2) to draw S-factor samples from the Gaussian process posterior. One of the most common ways to do this is to maximize the marginal likelihood of observing the known data given a particular choice of hyperparameters, or equivalently minimizing the negative log likelihood:

$$-\log p(S|\vec{\theta}) = \frac{1}{2} (S^T K_{11}^{-1} S + \log |K_{11}| + n \log 2\pi). \quad (5)$$

where  $n$  is the number of data points [4]. Cross-validation, where a subset of the training data is left out from the rest of the set and the probability of observing those points given the remaining data is maximized, directly evaluates the model's performance on energy regions without known data. For this application, an individual dataset is left out from the rest. The model makes predictions using (2) at those energies and compares them to the experimental values. For a set of hyperparameters  $\theta$ , each dataset is left out one at a time, and the sum of the log likelihoods is used to evaluate the model's performance before updating the hyperparameters:

$$\mathcal{L}_{LDO} = \sum_{i \in \text{sets}} \log p(S_i | S_{-i}), \quad (6)$$

where  $S_i$  and  $S_{-i}$  denote the S-factor measurements in dataset  $i$  and all other datasets respectively. The log likelihood to observe dataset  $i$  given the other datasets is

$$\log p(S_i | S_{-i}) = -\frac{1}{2} \left( (S_i - \mu_{S_i|S_{-i}})^T \Sigma_{S_i|S_{-i}}^{-1} (S_i - \mu_{S_i|S_{-i}}) + \log \det \Sigma_{S_i|S_{-i}} + n_i \log 2\pi \right),$$

where  $\mu_{S_i|S_{-i}}$  is the conditional mean from (2),  $\Sigma_{S_i|S_{-i}}$  is the conditional covariance matrix from (2), and  $n_i$  is the number of datapoints in dataset  $i$ . This ‘‘Leave-Dataset-Out’’ (LDO) likelihood is that which is maximized for most hyperparameter selections in this work. Reasonable kernel choices considered have the flexibility (with very small correlation lengths) to perfectly fit through every data point. Using the ‘‘Leave-One-Out’’ cross-validation method, a common choice, leads to extreme overfitting during hyperparameter selection. This problem can either be solved by placing a highly-informative, somewhat arbitrarily-chosen prior for the correlation lengths, or by utilizing the LDO likelihood. Uninformative uniform priors are placed on the hyperparameters when either (3) or (4) are used.

With at most four hyperparameters for any given kernel used here, methods such as an MCMC or simulated annealing to optimize the likelihood are not necessary, and gradient descent is preferred. To avoid convergence to local minima in hyperparameter space, using a gradient descent algorithm that incorporates momentum when updating hyperparameters, in this case the Adam optimizer, is very helpful, as either likelihood choice is not generally convex. The code<sup>16</sup> used for all Gaussian process regression tasks in this project, fully implemented in JAX, can be found [here](#).

<sup>16</sup> It is simple to modify the default code provided by scikit-learn to add off-diagonal terms to  $\Sigma$ , but very challenging to implement hyperpriors and custom gradient descent algorithms. On the other hand, it is straightforward to have custom hyperpriors and gradient descent algorithms using the default code from PyTorch, but very challenging to add off-diagonal terms to  $\Sigma$  within their framework.

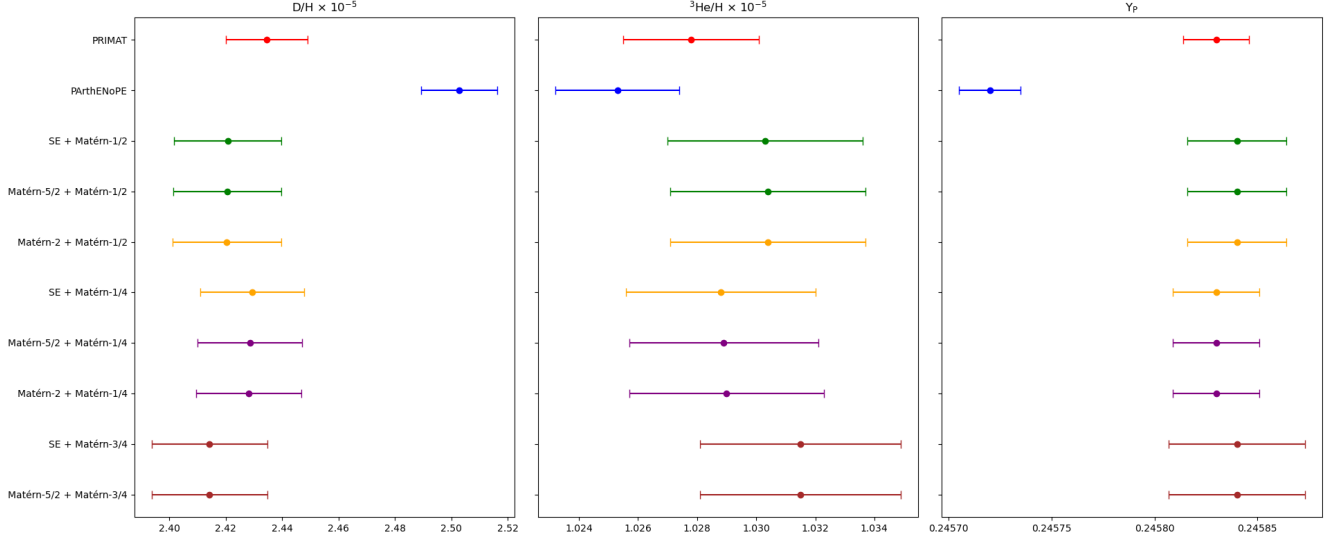


FIG. 11. Comparison primordial abundance predictions for different kernels used to perform Gaussian process regression on  $d(d,n)^3\text{He}$  S-factor data. For all kernel choices, a zero prior mean and LDO hyperparameter selection are used. The key\_PRIMAT\_2023 network is used for all other reactions. All error bars represent  $1\sigma$  uncertainties from 10,000 samples drawn from the posteriors. *Left*: D/H predictions for different kernels. *Middle*:  $^3\text{He}/\text{H}$  predictions for different kernels. *Right*:  $Y_P$  predictions for different kernels.

## VI. KERNEL SELECTION

An important test for this method is to consider sensitivity to kernel choice and determine which kernels are “valid” for this analysis. As mentioned above, kernel choices must include components to capture large scale correlations and small scale fluctuations. The first of these can be done with a large- $\nu$  Matérn or squared exponential; the second with a small- $\nu$  Matérn. Figure 9 shows D/H,  $^3\text{He}/\text{H}$ , and  $Y_P$  predictions when using various kernels that fit these criteria to fit  $d(d,n)^3\text{He}$  using a zero prior mean. Additionally, a simpler case using only a Matérn-1/2 kernel is considered, which leads to a “worse” fit overall. Decreasing the large correlation  $\nu$  slightly lowers final deuterium abundances. Raising the smaller  $\nu$  has a more significant effect, decreasing predicted deuterium abundances while increasing the error bars. The kernel combination of squared exponential with Matérn-1/4 is selected for further analyses, giving fit shapes with the least deviation from other fits<sup>17</sup>. Overall, this decision<sup>18</sup> has an effect on predicted abundances below  $1\sigma$ .

Interaction of helium abundance with Parth rates (see later figure, but can generally summarize here) – parth he3d rate is much higher at lower temperatures. at high temperatures – gp rate higher than parth rate, so get less deuterium and more helium. then at low temperatures he3d runs faster with parth than primat, so we get more helium 3 than with primat rates.

## VII. DATASET SELECTION

This section will discuss which experimental data sets are chosen for each deuterium burning reaction and why.

<sup>17</sup> We know, for example, that in the energy range without data points, samples should not dip very low before reaching the high energy dataset

<sup>18</sup> May want to select different kernels for this comparison, or only show fewer choices

A.  $d(d,n)^3\text{He}$

B.  $d(d,p)t$

C.  $d(p,\gamma)^3\text{He}$

## VIII. SUMMARY OF EXISTING METHODS

The techniques used in the PArthENoPE and PRIMAT codes are summarized as leading examples of the current state of reaction rate calculations and uncertainty propagation. In PArthENoPE (see [6] for more details on their procedure), the authors postulate that the S-factor can be well-approximated by a low-degree polynomial since the S-factor should be a smooth function of energy without resonances<sup>19</sup>. They use fourth degree polynomial of the form

$$S_{th}(E) = \sum_{\ell=0}^4 a_{\ell} E^{\ell}$$

where  $a_{\ell}$  are coefficients set by minimizing the  $\chi^2$  and the subscript  $th$  denotes the “theoretical” S-factor. They modify the expression for  $\chi^2$  to account for systematics in different experiments. The systematics for any experiment are expected to move each data point either up or down from the true value by the same amount. To account for this effect, each experiment is given some normalization factor that multiplies each measured value for that experiment. For the  $i^{th}$  datapoint of the  $k^{th}$  experiment, the measured S-factor  $S_{ik}$  is shifted to  $\omega_k S_{ik}$ , where  $\omega_k$  is the overall normalization for the  $k^{th}$  experiment. The statistical uncertainty is also normalized, shifting  $\sigma_{ik}$  to  $\omega_k \sigma_{ik}$ . In the case where only a total error is reported for an experiment, they use that value for  $\sigma_{ik}$ . From this procedure, the expression for the statistical  $\chi^2$  becomes

$$\chi_{stat}^2 = \sum_k \sum_i \frac{(S_{th}(E_{ik}, a_{\ell}) - \omega_k S_{ik})^2}{\omega_k^2 \sigma_{ik}^2}.$$

To account for normalization error, an additional term is added to the overall  $\chi^2$ :

$$\chi_{norm}^2 = \sum_k \frac{(\omega_k - 1)^2}{\epsilon_k^2},$$

where  $\epsilon_k$  is the normalization (systematic) error of the  $k^{th}$  experiment. The overall  $\chi^2$  is the sum of these terms:

$$\chi^2 = \chi_{stat}^2 + \chi_{norm}^2.$$

The free parameters  $\omega_k$  and  $a_{\ell}$  are obtained by minimizing this  $\chi^2$ . They use each dataset shown in Figure 1, including another (Tumino 2014) that is not publicly available.

Their estimation of the uncertainty in the reaction rate is computed from the  $a_{\ell}$  covariance matrix:

$$\Delta R^2(T) = \int_0^{\infty} dE' K(E', T) \int_0^{\infty} dE K(E, T) \sum_{i,j} \left. \frac{\partial S_{th}(E', a)}{\partial a_i} \right|_{\hat{a}} \left. \frac{\partial S_{th}(E, a)}{\partial a_j} \right|_{\hat{a}} \text{cov}(a_i, a_j),$$

where  $K(E, T)$  is the function in equation (2). The overall scale error for the data is estimated to be that of the dataset with the smallest normalization error. Their value for the total variance of the reaction rate is  $\Delta R^2$  inflated by  $\chi^2$  and summed in quadrature with the overall scale error:

$$\sigma_R^2(T) = \chi^2 \Delta R^2(T) + \min \epsilon_k^2.$$

---

<sup>19</sup> A smooth S-factor does not necessarily imply that a polynomial accurately models the true function.



For PRIMAT (see [7] for more details), the authors instead take a theory-based approach to calculate the S-factor for  $d(d,n)^3\text{He}$ . They adopt the *ab initio* calculation from [8] as the overall shape of the S-factor. In this calculation, interactions between nucleons are modeled with the following Hamiltonian:

$$H = \sum_{i=1}^4 T_i - T_{cm} + \sum_{i<j}^4 V_{ij} + \sum_{i<j<k}^4 V_{ijk}.$$

This form takes into account both two-body and three-body interactions between nucleons, accounting for spacial and spin distortions in the incident nuclei. The two-body terms include central, tensor, and spin-orbit contributions (modeling strong nuclear interactions), as well as Coulomb repulsion for protons. The tensor term is necessary to account for the non-spherical nature of nuclei and allows for mixing of orbital angular momentum states, here the *s*-wave and *d*-wave partial waves. A key result from [8] is the importance of the resulting *d*-wave components at low energies to align with experiment. The three-body term is phenomenological. Both the intrinsic nuclei wavefunctions and the relative motion component of the joint wavefunction for each relevant  $\ell$  are expressed in terms of Gaussian basis functions; for example, the relative motion wavefunction becomes

$$\chi_{\ell m} = \sum_i C_i r^\ell e^{-\lambda_i r^2} Y_{\ell m}(\hat{\mathbf{r}}).$$

$\{\lambda_i\}$  for the intrinsic nuclei wavefunctions are estimated with the stochastic variational method. The R-matrix method is used to calculate the matrix element for the relevant reaction, giving the theoretical prediction for the S-factor. Since this calculation is focused at low energies, it does not reproduce experimental results beyond 600 MeV. SHOW UNSCALED?

PRIMAT adopt a Bayesian approach to scale the S-factor result from [8] by some value of order unity to fit the experimental data, since the theory curve lies above nearly all of the experimental data points (seen in Figure 2). To determine this value, each of the five datasets they use (the collection of datasets in Figure 1 excluding RG85 and SC72) are each scaled by an overall constant to account for the normalization error of each experiment, in addition to the overall scaling. This gives six total parameters to determine in the fit. They use a Metropolis-Hastings MCMC to optimize and sample these parameters. Each dataset scaling parameter has a log-normal prior centered at one with the reported systematic error as the variance. The overall scaling has a non-informative prior, constrained to be positive. The likelihood is the sum of log-normal likelihoods for each datapoint. After burn-in, each S-factor sample is integrated using (2) to get the corresponding rate at 60 temperatures in the range relevant for BBN. The mean rate and variance at each temperature reported are directly obtained from this posterior.

Both PARthENoPE and PRIMAT compute their mean and  $1\sigma$  rates entirely separately from the BBN calculation. Abundances are calculated by passing the mean rates into each respective BBN code, while the variances are estimated by using their calculated  $1\sigma$  rates. This does not allow for variations in the reaction rate shape allowed by the experimental data. Additionally, both approaches assume a particular functional form for the S-factor and do not vary rates together.

- 
- [1] C. Pitrou, A. Coc, J.-P. Uzan, and E. Vangioni, “Precision big bang nucleosynthesis with improved Helium-4 predictions”, *Phys. Rept.* **754**, 1 (2018), [arXiv:1801.08023 \[astro-ph.CO\]](#).
  - [2] S. Gariazzo, P. de Salas, O. Pisanti, and R. Consiglio, “PARthENoPE revolutions”, *Computer Physics Communications* **271**, 108205 (2022), [arXiv:2103.05027 \[astro-ph.IM\]](#).
  - [3] C. Iliadis, *Nuclear Physics of Stars*, 2nd ed. (Wiley-VCH, Weinheim, 2015).
  - [4] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).
  - [5] C. Giovanetti, M. Lisanti, H. Liu, S. Mishra-Sharma, and J. T. Ruderman, “LINX: A Fast, Differentiable, and Extensible Big Bang Nucleosynthesis Package”, (2024), [arXiv:2408.14538 \[astro-ph.CO\]](#).
  - [6] O. Pisanti, G. Mangano, G. Miele, and P. Mazzella, “Primordial Deuterium after LUNA: concordances and error budges”, *JCAP* **04**, 020, [arXiv:2011.11537 \[astro-ph.CO\]](#).
  - [7] Á. Iñesta Gómez, C. Iliadis, and A. Coc, “Bayesian estimation of thermonuclear reaction rates for deuterium+deuterium reactions”, *Astrophys. J.* **849**, 134 (2017), [arXiv:1710.01647 \[astro-ph.IM\]](#).
  - [8] K. Arai, S. Aoyama, Y. Suzuki, P. Descouvemont, and D. Baye, “Tensor force manifestations in *ab initio* study of the  $^2\text{H}(d,\gamma)^4\text{He}$ ,  $^2\text{H}(d,p)^3\text{H}$ , and  $^2\text{H}(d,n)^3\text{He}$  reactions”, *Phys. Rev. Lett.* **107**, 132502.