

# Modeling anomalous electron transport in Hall thrusters using surrogate methods

IEPC-2022-344

*Presented at the 37th International Electric Propulsion Conference  
Massachusetts Institute of Technology, Cambridge, MA, USA  
June 19-23, 2022*

Thomas A. Marks<sup>1</sup> and Benjamin A. Jorns<sup>2</sup>

*University of Michigan, Ann Arbor, MI, 48109*

A Bayesian surrogate optimization method is developed which seeks to automate and accelerate the discovery and calibration of data-driven models for the non-classical electron transport in Hall thrusters. The proposed method is then used to identify best-fit model coefficients for a three-parameter Bohm-like stationary anomalous collision frequency profile for the SPT-100 Hall thruster. It is shown that the proposed approach can automatically identify model coefficients which minimize the error between simulations and experimental data with minimal human intervention. In addition, it is discovered that incorporating low-fidelity one-dimensional simulations in addition to higher-fidelity two-dimensional axisymmetric simulations may not increase the rate of convergence of this process unless the lower-fidelity code correlates sufficiently well with the higher-fidelity code. The results are discussed in the context of other data-driven approaches to modeling non-classical electron transport in Hall thrusters.

## I. Introduction

All thrusters are in-space electric propulsion devices with moderate specific impulse and high thrust density relative to other state-of-the-art electric propulsion systems[1]. They have been widely employed on satellites as station-keeping thrusters and are becoming increasingly popular for deep space exploration[2]. Despite this relative maturity and widespread use, there remain poorly-understood aspects of their operation. The most consequential of these is the problem of enhanced non-classical electron transport across the thruster's magnetic field lines[1]. The inability to self-consistently model this so-called "anomalous" electron transport has prevented the development of simulations which could predict the performance and plasma properties of a device from its geometry and operating conditions alone. Such simulations would be invaluable as part of the Hall thruster design and qualification process, reducing the reliance on lengthy and expensive vacuum chamber testing. They would also accelerate the design and optimization of new types of Hall thrusters, such as those which employ propellants other than xenon, unconventional geometries, or which operate at high power densities.

To date, efforts to explain anomalous transport in Hall thrusters from first principles have not yet produced a model which can be self-consistently incorporated into a whole-device simulation and yield results which are consistent with experiment. This stems from the lack of understanding of the precise mechanism of the enhanced electron transport. While multiple theories have been proposed to date [3] [4][5], none of these has proven fully predictive when incorporated self-consistently into higher fidelity models. To address the need for predictive tools, it has become common practice to employ data-driven methods to approximate the electron dynamics and[6][7][8].

In these methods, either a base model is proposed which includes model coefficients which must be tuned to match experiment, or an algorithm is used to try to discover a model. In state-of-the-art Hall thruster simulations, the most common model is a spatially-varying, stationary anomalous collision frequency profile which is applied along the Hall thruster channel centerline[9] and then projected outward along the magnetic field lines with a user-defined scaling[10]. These models typically feature 6 to 12 parameters which govern how the anomalous collision frequency varies in space. They typically yield the best fit to experiment, but generalize poorly across thrusters and operating conditions.

---

<sup>1</sup>PhD. Candidate, Aerospace Engineering, AIAA Student Member

<sup>2</sup>Associate Professor, Aerospace Engineering, AIAA Associate Fellow

Other models include those derived from first-principles, which usually include a few numeric constants which can be adjusted to improve the model's fidelity[11][12][13]. Being rooted in physics, these are more likely to generalize well, but so far have not proven predictive. Traditionally, this sort of parameter calibration has been done by hand, but this is time-consuming and requires one or more human experts. We can also try and learn new models from existing experimental data[6], with the hope of finding models that are more predictive than existing first-principles models, but which have fewer model coefficients and generalize better than the spatially varying models.

Ideally, data-driven methods for model discovery and calibration should be performed with a high fidelity simulation. Typically, one begins with a model and a proposed set of model coefficients and then runs a simulation using these inputs. We then compare the results to experiment and adjust the coefficients accordingly. Methods for doing this range from hand-based, intuition-driven methods to rigorous gradient-based optimization. With that said, one of the largest obstacles for this approach is that the computational expense of each iteration for higher fidelity simulations can be prohibitive, with simulations in JPL's state of the art fluid code Hall2De taking on the order of tens of hours to complete. Hand-tuning is time-consuming, non-rigorous, and requires the time and attention of a human expert in the loop. Gradients are not typically available in Hall thruster simulations, making gradient-based optimization impractical.

In an effort to circumvent these limitations, it helps to employ information from lower-fidelity methods to accelerate the calibration process. The simplest way to do this is to try to tune a model to match measurements of electron transport without integrating the model into a full Hall thruster simulation. In 2018[6], we used a dataset of spatially-varying anomalous transport profiles combined with validated time-averaged simulation results in order to try and discover new models of anomalous transport. The models obtained in this way yielded good fits with the profiles they were trained on and with those not in their training dataset, but did not perform well when integrated into a Hall thruster simulation[14]. This may be due to the known non-uniqueness of such profiles[10] and the highly oscillatory nature of Hall thrusters. For such methods to work, we would likely need time-resolved measurements of the anomalous collision frequency and local plasma properties, which are not widely available.

In light of the limitations of this previous approach, we propose in this work a surrogate-based method for calibrating Hall thruster anomalous transport model parameters which makes use of information from both low-fidelity and high-fidelity codes. The overarching goal is to be able to improve model fidelity while increasing the speed of calibration. To this end, this paper is organized in the following way. In Sec. II, we describe data-driven modeling of anomalous transport and describe possible approaches for optimizing such models. In Sec. III, we outline our proposed methodology and describe how we map the results of the low-fidelity simulations onto a global model of the high-fidelity simulation output. In Sec. IV, we then illustrate the utility of our approach for Hall thruster applications by performing a proof-of-concept study in which we infer the parameters of a simple three-parameter Bohm-like anomalous transport model. simulations[10]. In Sec. V, we discuss the performance of our method relative to other calibration methods. Lastly, in Sec. VI, we summarize our work and its key conclusions.

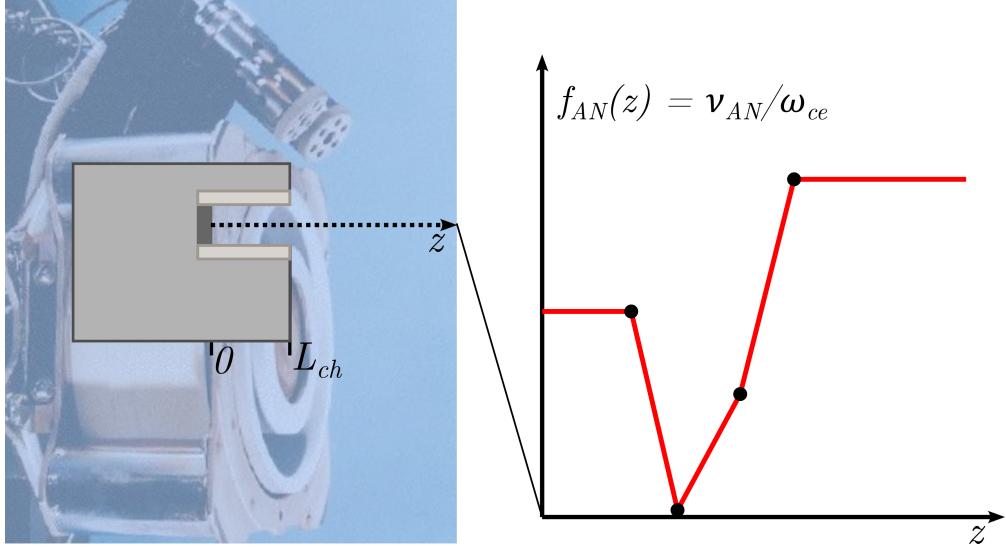
## II. Data-driven approach to modeling anomalous transport

In Hall thrusters, the electric field which accelerates ions to produce thrust is established across the applied magnetic field. Neglecting electron inertia and assuming the plasma is strongly magnetized ( $\nu_e \ll \omega_{ce}$ , where  $\nu_e$  is the electron collision frequency) the cross-field electron momentum equation reduces to a generalized Ohm's law of the form:

$$j_{e\perp} = \frac{en_e\nu_e}{\omega_{ce}B} \left( E_\perp + \frac{1}{n_e} \nabla_\perp(n_e T_{eV}) \right). \quad (1)$$

In the above expression,  $j_{e\perp}$  is the cross-field electron current density in A/m<sup>2</sup>,  $e$  is the fundamental charge,  $1.6 \times 10^{-19}$  C,  $n_e$  is the plasma density in m<sup>-3</sup>,  $E_\perp$  is the cross-field electron current, and  $T_{eV}$  is the electron temperature in eV. We see that in the magnetized limit  $j_{e\perp}$  scales directly with  $\nu_e$ . Experimentally, we observe that the cross-field electron current is much higher than that predicted when all known classical momentum transfer collisions, such as electron-ion and electron-neutral, are accounted for in  $\nu_e$ [1]. We thus typically include an extra "anomalous" collision frequency (denoted  $\nu_{AN}$ ) in simulations in order to resolve this discrepancy. This frequency may dominate the combined classical collision frequency by an order of magnitude or more, depending on the location in the discharge. This introduces a closure problem, however, as to solve the plasma governing equations we need an expression that ties  $\nu_{AN}$  back to the state of the plasma.

In order to close the plasma fluid equations, it is necessary to find a form for this anomalous collision frequency, for which two primary paths have been pursued to date. The first is to assume a spatially dependent form:



**Fig. 1** Example of a spatially-resolved anomalous collision frequency profile, with a reference thruster geometry.

$$v_{AN} = f_{AN}(z) \omega_{ce}. \quad (2)$$

Here,  $f_{AN}(z)$  is some shape function of the axial coordinate  $z$ , and  $\omega_{ce}$  is the electron cyclotron frequency. Physically, this expression captures the intuition that the transport is non-classical and therefore may be Bohm-like, i.e. with collision frequency scaling with the cyclotron frequency to a position-dependent degree. Fig. 1 depicts an example of such a profile and how it is typically applied in simulations.

One common way to do this is to make  $f_{AN}$  a piecewise-continuous function whose shape is tuned by the user in order to make simulations match experiment. Once the simulation matches available experimental data to a satisfactory degree, we can then use it as a surrogate measurement of physical properties in regions of the discharge that are difficult to probe non-perturbatively. This is useful for understanding how the plasma behaves inside of the thruster channel or near the anode. However, the calibration procedure can be laborious and time-consuming. A recent effort by one of the authors to calibrate such a profile for a single operating condition of a magnetically shielded Hall thruster[15] took over three weeks. Additionally, these profiles can be non-unique[10], and in certain regions of the discharge, changing the anomalous collision frequency by an order of magnitude has little effect on the resulting simulation.

The other approach is to attempt to find physics-based models of the electron transport which can be self-consistently implemented into Hall thruster simulations. Such models have  $v_{AN}$  as a function of the plasma properties instead of the axial location. These too often have parameters which may be adjusted to improve the fit with experiment. For the ensuing discussion, we would like to unify such models into a single framework. To this end, we seek models of the anomalous collision frequency  $v_{AN}$  of the form:

$$m(v_{AN}, \mathbf{p}, \mathbf{r}, t, \mathbf{x}) = 0 \quad (3)$$

Here,  $m$  is the model, which may in general be an implicit function of  $v_{AN}$ ;  $\mathbf{p} = (T_e, \mathbf{u}_i, n_e, \dots)$  represents the state of the Hall thruster discharge plasma in terms of macroscopic fluid variables, such as the electron temperature  $T_e$ , the ion velocity vector  $\mathbf{u}_i$ , and plasma density  $n_e$ ;  $\mathbf{r}$  represents the spatial coordinate; and  $t$  is the time, and  $\mathbf{x}$  is a vector of constant model parameters. These parameters are model-dependent and there may be any number of them.

Equation 3 contains both the piecewise profiles discussed above (a special case where  $m$  does not depend on  $\mathbf{p}$  and  $t$ ), as well as physics-based models. Many authors have proposed algebraic models where  $m$  is an explicit function of  $\mathbf{p}$  and  $\mathbf{x}$  only[11][12][13], and some have proposed multi-equation models where  $m$  is a coupled system of partial differential equations[16]. In most cases, however,  $\mathbf{x}$  needs to be tuned to make simulations give decent agreement with experiment. For piecewise profiles, there can be as few as two constants, or more than twelve, while first-principles models in the literature typically feature three or less. In the former case, these constants are specific to a single device or operating regime, while in the latter case, it is hoped that if the model properly captures key aspects of the underlying

physics, then the parameters will generalize well across thrusters and conditions. In all cases, we need to optimize the  $\mathbf{x}$  such that the simulation(s) match experimental results as accurately as possible.

We can frame this more formally as an optimization problem. Let  $f(\mathbf{x} \mid m)$  be some metric of the difference between an experimental measurement and the output of a simulation performed using model  $m$  and parameters  $\mathbf{x}$ . For example,  $f$  may measure the difference in thrust or discharge current between the simulation and experiment, or the residual of the difference between the spatially-resolved ion velocity curve produced by the code and one obtained in the laboratory via laser-induced fluorescence. With  $f$  defined, our optimization problem is:

$$\mathbf{x} \in \mathbb{R}^d f(\mathbf{x} \mid m),$$

where  $d$  is the length of  $\mathbf{x}$ , the dimension of the optimization problem. For a general function  $f$  that we seek to minimize, we would normally employ some sort of gradient-based local optimization, such as gradient descent or Newton's method as our first approach. If we do not have access to the gradients of  $f$  with respect to the parameters  $\mathbf{x}$ , which is almost always the case in Hall thruster modeling, then we need to compute them via finite differences. If  $\mathbf{x} \in \mathbb{R}^d$ , then the cost of doing so is proportional to  $d$ . These local search methods also often require additional function evaluations in order to pick the step size at each optimization iteration. They are quick to converge in terms of the number of iterations, but they rapidly become impractical as the cost of each function evaluation grows. When  $f$  is a computer simulation that may take hours to complete, which is the case for 2D axisymmetric Hall thruster simulations, the cost of such methods may be prohibitive.

Additionally, local search methods tend to refine the solution to a high degree of accuracy, but in engineering optimization we are more interested in finding the approximate location of the minimum and seldom interested in accuracy beyond a couple of decimal digits. Lastly, these methods are *local* and can thus converge at local minima, whereas we are interested in finding the globally optimal values of  $\mathbf{x}$ , i.e. the model parameters that make the simulation most closely match experiment.

One common way to tackle such problems is via *surrogate optimization*. Given some initial set of parameters  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$  and the results of simulations at those parameters  $\mathcal{Y} = \{y_1, y_2, \dots\}$ , where  $y_1 = f(\mathbf{x}_1 \mid m)$ , we construct a *surrogate model*, which predicts the output of  $f$  at values of  $\mathbf{x}$  that we have not yet tried. In contrast to the original function  $f$ , the surrogate  $\hat{f}(\mathbf{x})$  is quick to evaluate and has gradient information available. We can then optimize the surrogate model instead of the original function. We can also use the surrogate to suggest new values of  $\mathbf{x}$  that we should run in our simulation code, and over time find an approximation of the global minimum value of  $f$  using as few function evaluations as possible.

### III. Methodology

#### A. Gaussian process regression

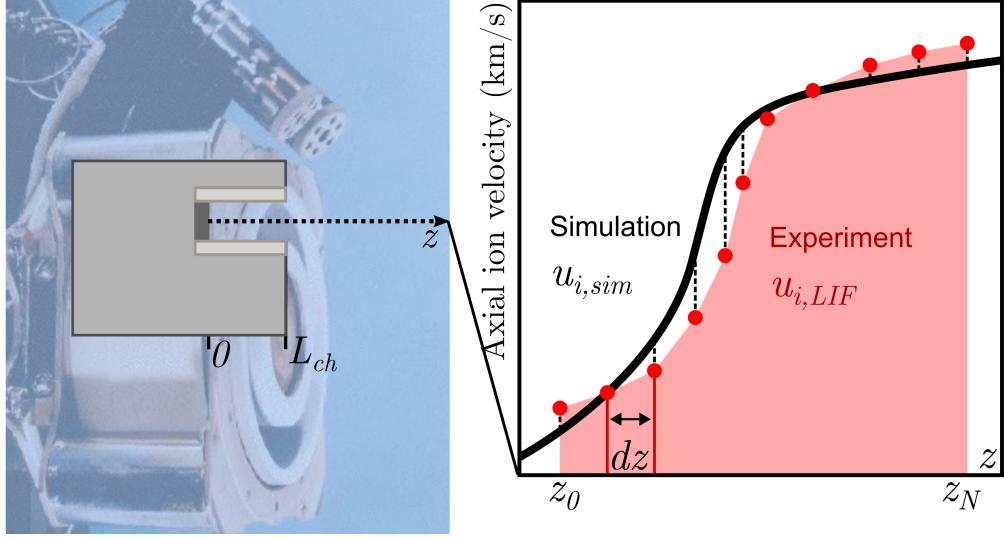
One popular approach to constructing a surrogate model is Gaussian process regression (GPR), also known as *Kriging*. Let  $\hat{f}(\mathbf{x})$  be an approximation of  $f(\mathbf{x})$ . If  $\hat{f}$  is a Gaussian process, that means that it is characterized by both a mean prediction  $\mu(\mathbf{x})$  and a covariance function or *kernel*  $k(\mathbf{x}, \mathbf{x}')$  for all values of  $\mathbf{x}$ . This is a key feature, as it not only allows us to predict the value of our computationally expensive function at an untested point, but also estimate our uncertainty in our prediction at that point. There are many choices of covariance kernel, but in general the covariance between two points falls off as the distance between them increases. One popular choice, and the one we employ in this work, is the squared-exponential kernel:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(\frac{1}{2\ell} \|\mathbf{x} - \mathbf{x}'\|_2^2\right).$$

Here,  $\sigma$  and  $\ell$  are hyperparameters which reflect how far the function strays from its mean and how smooth the function is, respectively. These are then optimized via cross-validation or maximum-likelihood estimation[17] in order to obtain the final surrogate model.

#### 1. Multi-fidelity surrogates

Suppose now that we have two codes available, a low-fidelity code and a high-fidelity code. We denote the result of the low-fidelity code as  $f_{LF}(\mathbf{x})$  and the result of the high-fidelity code as  $f_{HF}(\mathbf{x})$ . We wish to use the low-fidelity code to predict the response of the high-fidelity code to changes in  $\mathbf{x}$ , in order to reduce the number of times we need to run



**Fig. 2** Diagram of how the integrated velocity error is defined. The numerator of Eq. 5 is the integral of the squared difference (dashed black lines) between the experimental data (red markers) and the simulation (solid black line). The denominator of Eq. 5 is the area denoted in light red. The integral is performed over the differential length elements integrated by  $dz$  from  $z_0$  to  $z_N$ .

the high-fidelity code. We accomplish this via *Multi-Fidelity Kriging*[18] (MFK), where we model the high fidelity function as

$$\hat{f}_{HF}(\mathbf{x}) = \rho(\mathbf{x})\hat{f}_{LF}(\mathbf{x}) + \hat{\delta}(\mathbf{x}).$$

where  $\hat{\delta}(\mathbf{x})$  is a difference function and  $\rho(\mathbf{x})$  is a correlation function (generally a constant or a low-order polynomial). For the purposes of this work, we employ the Surrogate Modeling Toolbox[19], a Python package developed at the University of Michigan, to construct and train the multi-fidelity surrogate models.

## B. Optimization

Once we have obtained a multi-fidelity surrogate, we need to pick new points at which we can run our codes in order to improve the surrogate and find the parameter values which make the simulations give the best fit to data. We define the following objective functions for this work. The first is the relative error in the discharge current:

$$f_I(\mathbf{x}) = \frac{|I_D - I_{sim}(\mathbf{x})|}{I_D}. \quad (4)$$

The second is the *integrated velocity error*, which we define as:

$$f_V(\mathbf{x}) = \sqrt{\frac{\int_{z_0}^{z_N} (u_{i,LIF}(z) - u_{i,sim}(z, \mathbf{x}))^2 dz}{\int_{z_0}^{z_N} u_{i,LIF}^2(z) dz}} \quad (5)$$

Here,  $u_{i,LIF}(z)$  is the mean ion velocity as a function of axial location  $z$  as measured by laser-induced fluorescence,  $u_{i,sim}(z, \mathbf{x})$  is the simulated ion velocity at  $z$  given model parameters  $\mathbf{x}$ , and  $z_0$  and  $z_N$  are the axial locations of the first and last LIF measurements, respectively. In Fig. 2, we illustrate graphically how the IVE is computed.

The best fit to experiment is found when both of these functions are minimized. While optimizing, we need an *acquisition function* which tells us where we should run the high-fidelity code next. For a single-objective Bayesian optimization problem where we attempt to minimize a function  $f(\mathbf{x})$  and have constructed a kriging surrogate  $\hat{y}(\mathbf{x})$ , the *Expected Improvement* function is often used[17]:

$$EI(\mathbf{x}) = \mathbb{E}[y_{min} - \hat{f}(\mathbf{x})] = (y_{min} - \hat{f}(\mathbf{x}))\Phi\left(\frac{y_{min} - \hat{f}(\mathbf{x})}{\sigma(\mathbf{x})}\right) + \sigma(\mathbf{x})\phi\left(\frac{y_{min} - \hat{f}(\mathbf{x})}{\sigma(\mathbf{x})}\right) \quad (6)$$

Here,  $y_{min}$  is the smallest objective function evaluated,  $\sigma(\mathbf{x})$  is the predicted standard error at  $\mathbf{x}$  and  $\Phi$  and  $\phi$  are the cumulative distribution function and probability density functions, respectively, of the standard normal distribution. At each optimization iteration, we find the point  $\mathbf{x}$  which maximizes the expected improvement function and compute our expensive model at that point. However, we have a multi-objective optimization problem, and it is not guaranteed that the point which minimizes the integrated velocity error will also minimize the error in the discharge current. There are many approaches to such problems in literature, but we adopt a simple *no-preference* method in which we transform our multi-objective problem into a single-objective problem via a combined objective function:

$$f(\mathbf{x}) = f_V(\mathbf{x}) + f_I(\mathbf{x}), \quad (7)$$

The integrated velocity error and discharge current error are defined such that they should have similar orders of magnitude. If the ion velocity curve is off by 10% from the experimental profile, then  $f_V(\mathbf{x})$  will be 0.1. Similarly, if the discharge current is off by 10%, then  $f_I(\mathbf{x})$  will also be 0.1. This scaling captures our lack of a-priori reason to weigh one of these quantities more highly than the other. If a simulation fails for some reason, we would set  $f(\mathbf{x})$  to 20.0, although this did not happen in the present study. We compute this objective function at all low-fidelity and high-fidelity sample points, then build the multi-fidelity surrogate as described in Section III.A.1. We can then compute and maximize the expected improvement. We find this point by sampling randomly in the domain, finding the points with the highest expected improvement, and initializing a local optimizer at each of these points to find the one which maximizes EI. We then run both the low- and high-fidelity codes at the new point and use the output to update the multi-fidelity surrogate. We repeat this until a termination criterion (typically a fixed number of simulations) is reached.

For a multi-fidelity problem such as ours with an expensive high-fidelity simulator, it is advantageous to be able to select a batch of several points to run in parallel instead of a single point. One simple approach to this problem (and the one we employ in this work) is the *Constant Liar* (CL) method[20]. Let  $\mathbf{x}^{n+1}$  be the point which maximizes the expected improvement as described above. Instead of evaluating the objective function to obtain  $y^{n+1}$  and adding  $(\mathbf{x}^{n+1}, y^{n+1})$  to the surrogate, we use a constant pretend value of  $y^{n+1} = L$ . We then find the point which maximizes the expected improvement of the updated surrogate and call it  $\mathbf{x}^{n+2}$ . We again add  $(\mathbf{x}^{n+2}, L)$  to the surrogate, and repeat this process until we have added however many new points we desire.  $L$  is often chosen to be the maximum, minimum, or mean of all  $y$ -values in the surrogate. Using  $L = \min(Y)$  will tend to find points that are close to the current best point (exploitation), while using  $L = \max(Y)$  will tend to generate points that are spread out (exploration).  $L = \text{mean}(Y)$  will be somewhere in-between these extremes, and it is the value we will use in this work. We summarize the surrogate construction and optimization procedure in Fig. 3.

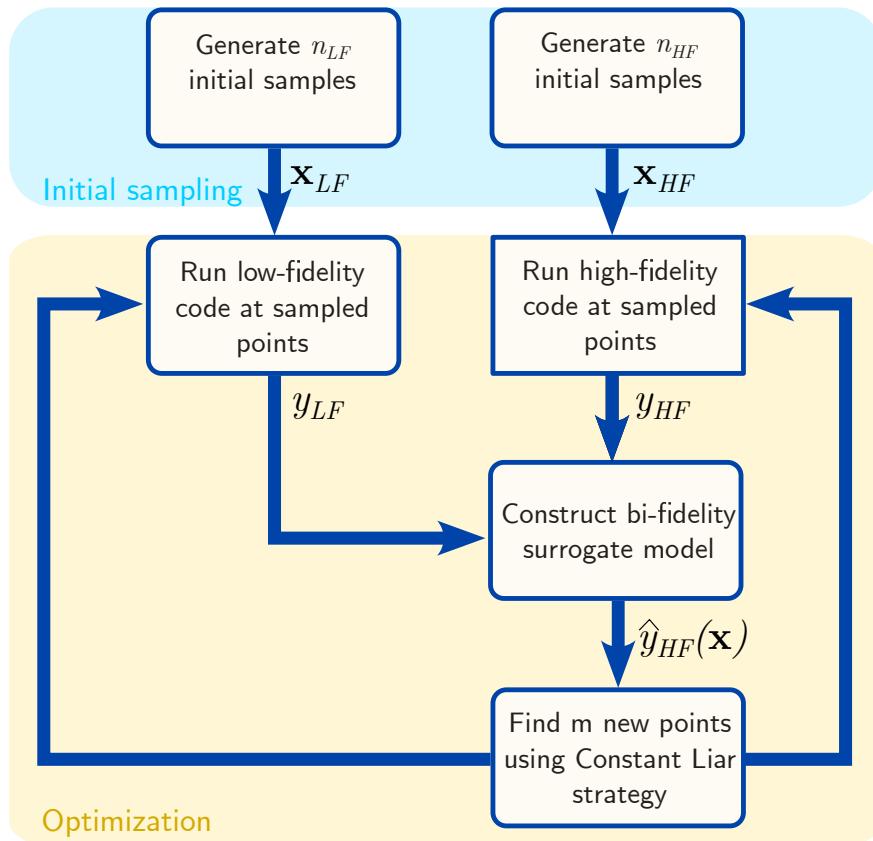
### C. Thruster

In this work, we simulate the SPT-100 Hall thruster (depicted in Fig. 4). This 1.35 kW unshielded thruster was developed in the Soviet Union and is one of the oldest and best-studied thrusters still in use today[21]. It is extensively-studied by the electric propulsion community and often serves as a model thruster for Hall thruster research.

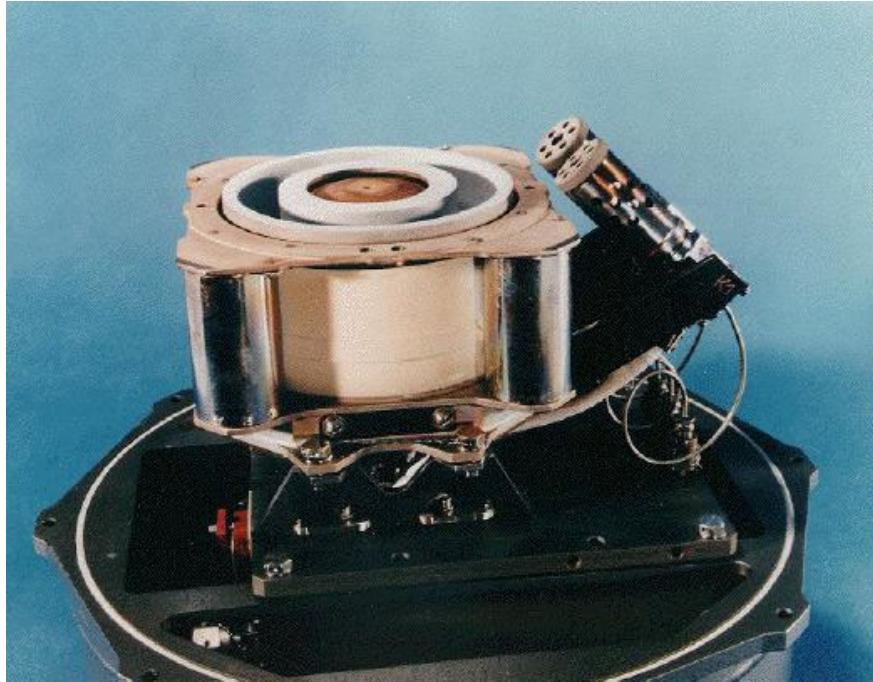
The data used for comparison in this section comes from Macdonald-Tenenbaum et al[22], who measured ion velocity curves using laser-induced fluorescence at multiple background pressures at a discharge voltage of 300 V and a discharge current of 4.24 A. We attempt to calibrate our model to match their measured discharge current as well as the ion velocity curve measured at a background pressure of  $3.5 \times 10^{-5}$  Torr.

### D. Codes

The last things we need in order to apply our proposed multi-fidelity surrogate modeling method are two Hall thruster codes – a low-fidelity code and a high-fidelity code. Our high-fidelity code of choice is Hall2De, a two-dimensional axisymmetric multi-fluid code developed at the Jet Propulsion Laboratory[9]. Our low-fidelity code is HallThruster.jl[23], an open source one-dimensional fluid code recently developed at the University of Michigan. HallThruster.jl can simulate 2 ms of Hall thruster operation in seconds to minutes, depending on the grid resolution, making it ideal for sampling densely from the parameter space. In contrast, Hall2De takes several hours to run a single simulation of the same duration, but features much greater fidelity compared to HallThruster.jl.



**Fig. 3** Summary of the multi-fidelity surrogate modeling and optimization procedure used in this work.



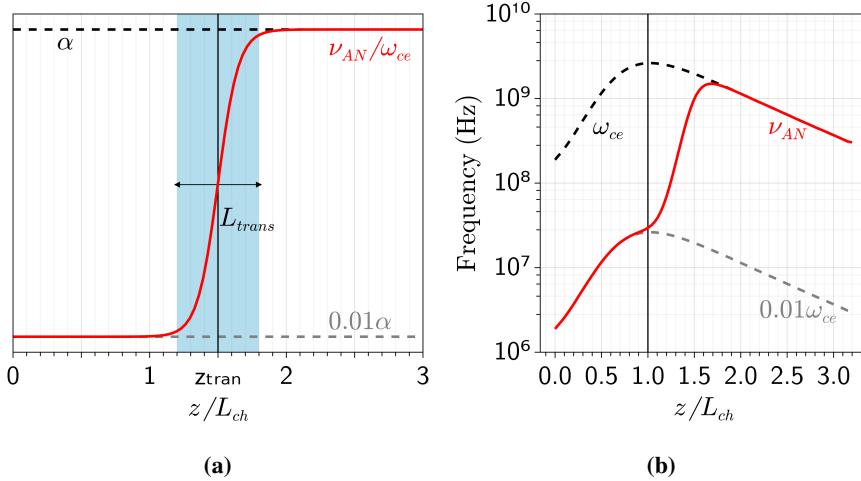
**Fig. 4** The SPT-100 Hall thruster.

#### IV. Study: Three-parameter model

In order to assess the utility of the bi-fidelity surrogate modeling approach, we first attempted to use it to calibrate a static anomalous transport model. We employ a simple two-zone Bohm-like mobility model similar to that used by Giannetti et al[24]. In this model, the anomalous collision frequency is described by three model parameters:  $\alpha$ ,  $z_{trans}$ , and  $L_{trans}$ . These parameters represent the scale of the anomalous transport, the axial location of the transition from low to high anomalous collision frequency, and the width of the transition, respectively. The latter two parameters are normalized by the discharge channel length  $L_{ch}$  which for the SPT-100 is 2.5 cm. The model is given by:

$$\nu_{AN} = \omega_{ce} \begin{cases} \alpha & z/L_{ch} > z_{trans} + 0.5 L_{trans} \\ 0.01 \alpha & z/L_{ch} < z_{trans} - 0.5 L_{trans}, \end{cases} \quad (8)$$

When  $-0.5 L_{trans} < (z/L_{ch} - z_{trans}) < 0.5 L_{trans}$ , the anomalous collision frequency transitions smoothly between the two zones using a hyperbolic tangent function, as depicted in Fig. 5a. In Fig. 5b, we show an example anomalous collision frequency profile for  $\alpha = 1.0$ ,  $z_{trans} = 1.5$ ,  $L_{trans} = 0.5$ .



**Fig. 5** (a) Illustration of the anomalous transport model, given parameters  $(\alpha, z_{trans}, L_{trans})$ , showing transition region in blue. (b) Anomalous collision frequency profile for  $\alpha = 1.0, z_{trans} = 1.5, L_{trans} = 0.6$ .

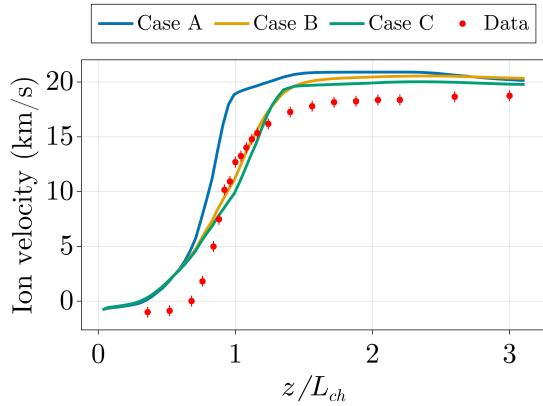
We compare three cases. In the first (case A), we use only the one-dimensional code to perform the parameter optimization. In the second (case B), we only run the two-dimensional code. In the last (case C), we employ a bi-fidelity approach, using the two-dimensional code to augment the response of the one-dimensional code. For cases A and C, we initialize the optimizer by sampling 250 points from the region of  $\alpha \in (0.01, 1.0)$ ,  $z_{trans} \in (1.0, 2.0)$  and  $L_{trans} \in (0.1, 1.0)$  and running HallThruster.jl at these points. For case B and C, we then subsample 20 points from the initial 500 and run Hall2De at this subset of points. We use an optimal Latin Hypercube sampling plan[25] to pick the initial sample locations, ensuring the points are well-distributed and not clumped. In Fig. ??.

With the initial samples generated, we then run the optimization procedure described in the preceding section. Since we sampled the parameter space comparatively densely using the 1D code and since both metrics vary smoothly, we simply select the best point from the initial samples as the final solution. For the other two cases, we optimize the surrogate as described in the previous section. At each iteration of case B and C, we obtain five new points using the Expected Improvement - Constant Liar strategy, then run Hall2De at each of the sample points. For case C, we also run HallThruster.jl at the sampled points, as the surrogate construction process requires that high-fidelity samples be co-located with low-fidelity samples. We present the results of this study in Tab. 1, while in Fig. 6a, we show how the simulated ion velocity curves compare across the three cases.

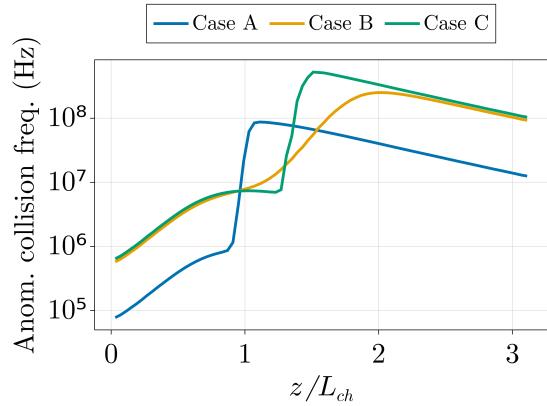
As Tab. 1 demonstrates, both methods are able to converge to points that match the discharge current well and which minimize the objective function. The fits to the ion velocity curves so obtained are not exact, but this is a limitation of the chosen model rather than the optimization procedure. Unsurprisingly, both cases B and C outperform case A, as they have access to information from the high-fidelity code and case A does not. Unexpectedly, case B seems to moderately outperform case C, despite the former's lack of low-fidelity code information. The results obtained

Case	Description	# 1D evals	# 2D evals	$\mathbf{x}^* = (\alpha^*, z_{trans}^*, L_{trans}^*)$	$I_D(\mathbf{x}^*)$	$f_V(\mathbf{x}^*)$	$f(\mathbf{x}^*)$
A	1D code only	250	1	(0.034, 1.012, 0.1325)	4.064 A	0.228	0.270
B	2D code only	0	80	(0.251, 1.785, 1.000)	4.212 A	0.139	0.146
C	Bi-fidelity	330	80	(0.281, 1.424, 0.159)	4.082 A	0.125	0.162

**Table 1** Results of the bi-fidelity surrogate optimization procedure for the three-parameter anomalous transport model. “#1D evals” and “#2D evals” refer to the number of times HallThruster.jl and Hall2De were run, respectively. The optimal parameter values found in each case are denoted as  $\mathbf{x}^*$ , and the discharge current, integrated velocity error, and combined objective function at the found optimum are  $I_D(\mathbf{x}^*)$ ,  $f_V(\mathbf{x}^*)$ , and  $f(\mathbf{x}^*)$ , respectively.



(a) Comparison of best-fit velocity profiles for the three cases



(b) Comparison of optimized anomalous collision frequency profiles for the three cases

in cases B and C are very similar, with nearly identical ion velocity curves and only small differences in discharge current. However, the predicted optimal coefficients are quite different. The optimum found by case B is  $\alpha = 0.0251$ ,  $z_{trans} = 1.785$ , and  $L_{trans} = 1.00$ , while for case C we find  $\alpha = 0.281$ ,  $z_{trans} = 1.424$ , and  $L_{trans} = 0.159$ . As is evident, the parameter governing the scaling of the anomalous transport,  $\alpha$ , matches quite closely, but both  $z_{trans}$  and  $L_{trans}$  are very dissimilar. In Fig. 6b, we compare the best profiles discovered by the three cases. As indicated by the best-fit coefficients, the optimized profiles for the two cases differ strongly in the transition region, but match well both upstream and downstream of this point.

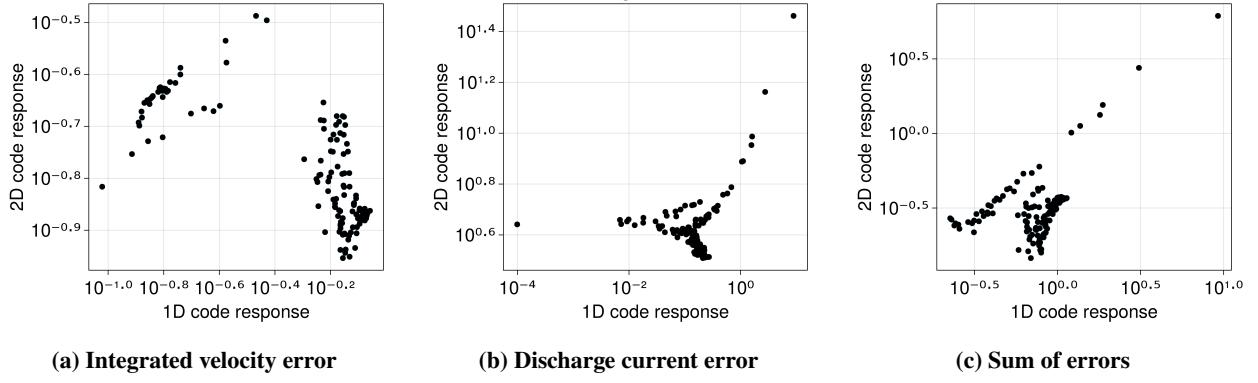
These results seems to suggest that there may not be much utility in our multi-fidelity approach. With that said, the model in question is simple and it is not necessarily surprising that the bi-fidelity approach will require as many numerical evaluations as the single-fidelity model to converge. Our method may be more useful for calibrating higher-dimensional models like those more commonly employed. We next turn to a more detailed discussion of the bi-fidelity surrogate’s poor performance and the implications of solution non-uniqueness on future calibration algorithms.

## V. Discussion

In both cases, our optimization procedure was able to minimize the combined objective function and find fits more optimal than those in the initial simulation batch. The rate of convergence was somewhat slow. It took four batches of five simulations each before the optimizer found a point which improved upon one in its initial batch. The high-fidelity optimizer performed slightly better than the bi-fidelity optimizer. This was unexpected, as the latter had more information than the former. Additionally, the

### A. Effect of conflicting minima

One possible reason for the lackluster performance of the bi-fidelity optimization routine over using the high-fidelity code by itself is that, for some anomalous collision frequency models, the optimal parameters are significantly different



**Fig. 7 Correlations between HallThruster.jl (x-axis) and Hall2De (y-axis) for (a) integrated velocity error, (b) error in discharge current, (c) sum of objectives for the three-parameter model evaluated in Section IV**

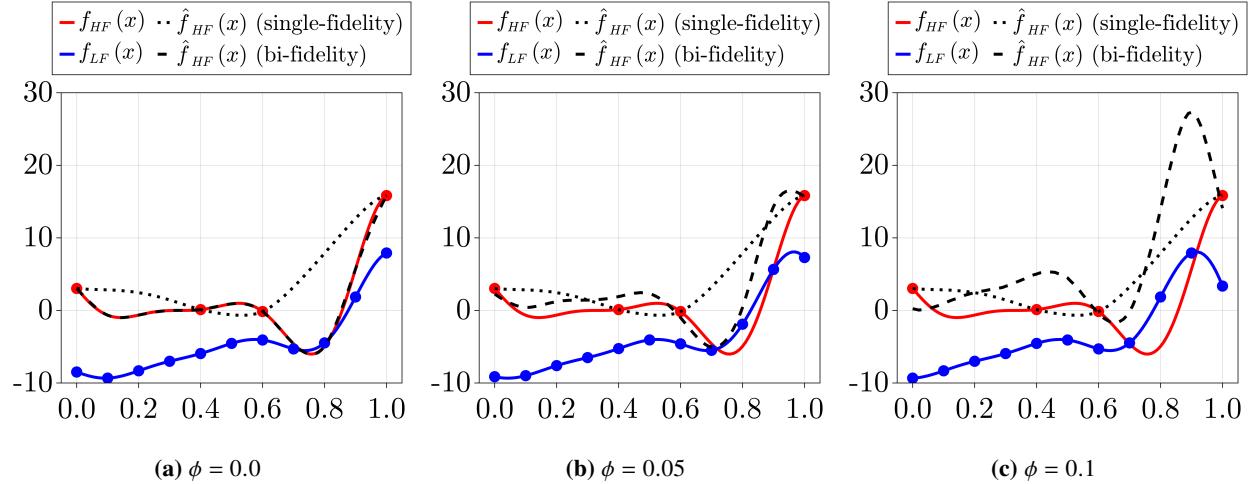
in the 1D code than in the 2D code. In Fig. 7, we show how the HallThruster.jl and Hall2De trend against each other. In the integrated velocity error plot (Fig. 7a), we see that there are two distinct populations of solutions. In the top left, there is a population that correlates very well between Hall2De and HallThruster.jl, while in the bottom right, there is a population which does not correlate at all. This is evident, but significantly less so, in the plot of discharge current error (Fig. 7b), as is it also in the plot of the combined objective function(Fig. 7c). This is problematic for the bi-fidelity surrogate (case C), as this second population contains the points that are lowest on the y-axis, i.e. which perform the best in Hall2De. This delays the convergence of the optimization routine to the point that its performance is very similar to the high-fidelity-only optimizer (case B). In fact, the minimum found in case B lies in the region of parameter space that correlates poorly between the two codes, reducing the probability that the bi-fidelity optimizer would find the same minimum. The reason for the poor correlation of the second population is related to the onset of very high-amplitude limit-cycle oscillations in HallThruster.jl which do not occur in Hall2De. These do not significantly affect the average discharge current but dramatically change the time-averaged velocity profile, increasing the integrated velocity error as measured by the 1D code.

To ameliorate this effect, it would be beneficial to employ a multi-fidelity optimization algorithm which is less sensitive to the differences in the location of the minimum between the low-fidelity and high-fidelity codes. To more clearly illustrate this difficulty, let us consider the Forrester function[26], a simple one-dimensional function commonly employed to demonstrate the utility of bi-fidelity optimization. Let  $f_{LF}$  be the low-fidelity version of the function and  $f_{HF}$  be the high-fidelity version, then the Forrester functions are given by:

$$f_{HF}(x) = (6x - 2)^2 \sin(12x - 4) \quad (9)$$

$$f_{LF}(x) = 0.5(6x - 2)^2 \sin(12x - 4) + 10(x - 0.5) - 5 \quad (10)$$

To examine the impacts of shifting the location of the minimum of the low-fidelity function compared to the high-fidelity function, we can shift the entire low fidelity function by  $\phi$  in the x direction. As in Forrester 2008, let us sample the low-fidelity function at 11 points in the interval  $[0, 1]$  (increments of 0.1), while we sample the high-fidelity function only at  $x = \{0.0, 0.4, 0.6, 1.0\}$ . We will show how the single-fidelity kriging and multi-fidelity kriging surrogates respond to different values of  $\phi$ . We will test  $\phi \in \{0.0, 0.05, 0.01\}$ . The results are shown in Fig. 8. In Fig. 8a, when  $\phi = 0$ , we see that adding the low-fidelity information dramatically improves the approximation over the only including high-fidelity information. However, when we shift the low-fidelity function only slightly, by 1/20 the width of the sample space (Fig. 8b), the approximation begins to degrade, and when we shift the low-fidelity function by twice as much (Fig. 8c), the approximation is only somewhat better than when we include only the high-fidelity sample points. Thus, when the low-fidelity function and high-fidelity function have minima that significantly conflict, the multi-fidelity kriging model might have misleading ideas about where the true minimum lies, and its performance may be only as good, or even worse than, a single-fidelity model. For calibrating a specific model of anomalous transport model, researchers should thus determine how well the 1D code correlates with the 2D code before committing to a bi-fidelity optimization method.



**Fig. 8 Multi-fidelity approximations of the Forrester function (Eqs. 9-10) with the low-fidelity shifted by  $\phi$  compared to Eq. 10. The sample locations of the low-fidelity and high-fidelity functions are indicated by the blue and red circular markers, respectively. The surrogate constructed using only the high-fidelity information is denoted by a dotted line, while the bi-fidelity surrogate is indicated by a dashed line.**

## B. Solution non-uniqueness

Non-uniqueness poses a significant challenge to verifying and validating anomalous transport models. Mikellides and Lopez Ortega showed in 2019[10] that one can significantly alter the anomalous collision frequency in parts of the discharge without altering the predicted electrostatic potential profile and discharge current. We encountered a similar result here, where two quite different coefficient sets and anomalous collision frequency profiles (c.f. Tab. 1 and Fig. 6b) gave very similar ion velocity curves (Fig. 6a) and predicted discharge currents. Without non-invasive measurements of plasma properties deep inside the channel, for example, it is hard to validate that the anomalous collision frequency near the anode is correctly calibrated. This is likely to pose more of a challenge for the spatially-varying profiles than for first-principles models, but may be mitigated by including additional data in the calibration procedure. For example, component efficiencies, such as mass and beam utilization, may give information about conditions in hard-to-measure regions of the discharge that are not apparent from the ion velocity profile and discharge current alone.

## C. Comparison to other calibration methods

We now turn to discussing our results in the context of other calibration procedures. For ad-hoc piecewise anomalous transport models, our proposed approach is likely preferable to hand-tuning, provided enough computing resources are available to run the needed simulations in parallel. Due to the model's simplicity, a human expert would likely have been able to calibrate the three-parameter model (Sec. IV) in fewer simulations than our procedure required. However, our procedure can be fully automated, and it is ultimately objective.

The Expected Improvement-based optimization algorithm we employ in this work may not be the ideal choice for our problem it is likely that other global optimization algorithms may provide significant speedups. In head-to-head comparisons[27] with a suite of other global optimization routines, those based on the Dividing Rectangles (DIRECT) [17] algorithm were able to solve the most number of problems in a fixed number of iterations. Some combination of this algorithm with a surrogate modeling effort may address the shortcomings of both methods[28] and lead to improved convergence behavior.

Although we did not directly investigate it in this work, our procedure would be useful for discovering and calibrating physics-based self-consistent models of anomalous transport. While our previous approach[6] produced a suite of new models which gave promising fits to experiment when trained on spatially-varying anomalous collision frequency profiles, we have found that this does not correlate strongly with how well the discovered model performs when integrated into a high-fidelity Hall thruster simulation[14]. Therefore, we need to include the high-fidelity simulation in the calibration procedure, and our approach provides a systematic way to do that. With these models, there may be less human intuition to guide what the appropriate model parameters should be, as the plasma and anomalous collision

frequency are strongly coupled and interact in unexpected ways. For example, we might want to evaluate a model of the following form:

$$\nu_{AN} = \alpha \omega_{ce} \left( \frac{u_i}{v_{de}} \right)^\beta \left( \frac{\nu_e}{\omega_{ce}} \right)^\gamma \left( \frac{T_e}{T_i} \right)^\delta \quad (11)$$

In this model, we incorporate several terms that have appeared in both prior model discovery efforts[6] and in first-principles models[12][13]. Obtaining a good fit to experiment with such a model would give us information about which plasma parameters the anomalous collision frequency is likely to scale with, which could then inform future first-principles modeling efforts. With such a model, though, there may be less intuition about how to tune the parameters (here  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ) in order to make simulations consistent with experiment, and so an automated procedure like ours might be useful.

## VI. Conclusion

Self-consistent predictive models of Hall thruster anomalous transport have proven elusive in the past twenty years. Even the ad-hoc piecewise profiles commonly employed to make simulations match experiment are problematic due to their potential non-uniqueness and the large amount of human time required to calibrate individual operating conditions. In this work, we have presented a multi-fidelity surrogate-based calibration procedure to attempt to automate the calibration of anomalous transport models, whether they be empirical, physics-based, or data-driven. We found that including information from reduced-fidelity codes may not speed up the calibration procedure, but only when the two codes are sufficiently well-correlated given the model being calibrated. However, this depends strongly on the model, and researchers ought to determine how well the codes correlate prior to choosing a single-fidelity or multi-fidelity calibration procedure. Despite these caveats, we have shown that a significant amount of the calibration work can be automated and parallelized, paving the way for more rapid design, testing, and iteration on more predictive engineering models of anomalous transport.

## VII. Acknowledgements

This work was funded by Air Force Office of Scientific Research Grant FA9550-19-1-0022 under the Space Propulsion and Program portfolio. We would like to thank Ioannis Mikellides and Alejandro Lopez-Ortega at the Jet Propulsion Laboratory for allowing us to use Hall2De, and Paul Schedler for his extensive work developing HallThruster.jl

## References

- [1] Boeuf, J. P., “Tutorial: Physics and modeling of Hall thrusters,” *Journal of Applied Physics*, Vol. 121, No. 1, 2017. <https://doi.org/10.1063/1.4972269>.
- [2] Snyder, J. S., Goebel, D. M., Chaplin, V., Lopez Ortega, A., Mikellides, I. G., Aghazadeh, F., Johnson, I., Kerl, T., and Lenguito, G., “Electric Propulsion for the Psyche Mission,” *36th International Electric Propulsion Conference*, 2019.
- [3] Morozov, A. I., and Savelyev, V. V., “Fundamentals of Stationary Plasma Thruster Theory,” Springer, Boston, MA, 2000, pp. 203–391. [https://doi.org/10.1007/978-1-4615-4309-1\\_2](https://doi.org/10.1007/978-1-4615-4309-1_2), URL [https://link.springer.com/chapter/10.1007/978-1-4615-4309-1\\_2](https://link.springer.com/chapter/10.1007/978-1-4615-4309-1_2).
- [4] Tsikata, S., Lemoine, N., Pisarev, V., and Gréillon, D. M., “Dispersion relations of electron density fluctuations in a Hall thruster plasma, observed by collective light scattering,” *Physics of Plasmas*, Vol. 16, No. 3, 2009, p. 33506. <https://doi.org/10.1063/1.3093261>, URL <https://doi.org/10.1063/1.3093261>.
- [5] Brown, Z. A., and Jorns, B. A., “Spatial evolution of small wavelength fluctuations in a Hall Thruster,” *Physics of Plasmas*, Vol. 26, No. 11, 2019, p. 113504. <https://doi.org/10.1063/1.5116708>, URL <https://doi.org/10.1063/1.5116708>.
- [6] Jorns, B., “Predictive, data-driven model for the anomalous electron collision frequency in a Hall effect thruster,” *Plasma Sources Science and Technology*, Vol. 27, No. 10, 2018. <https://doi.org/10.1088/1361-6595/aae472>.
- [7] Greve, C. M., Majji, M., and Hara, K., “Real-time state estimation of low-frequency plasma oscillations in Hall effect thrusters,” *Physics of Plasmas*, Vol. 28, No. 9, 2021, p. 93509. <https://doi.org/10.1063/5.0057751>, URL <https://doi.org/10.1063/5.0057751>.

- [8] Shashkov, A., Tyushev, M., Lovtsov, A., Tomlin, D., Kravchenko, and Dmitrii, “Machine learning-based method to adjust electron anomalous conductivity profile to experimentally measured operating parameters of Hall thruster,” *Plasma Science and Technology*, Vol. 24, No. 6, 2022, p. 65502. <https://doi.org/10.1088/2058-6272/ac59e1>, URL <https://doi.org/10.1088/2058-6272/ac59e1>.
- [9] Mikellides, I. G., and Katz, I., “Numerical simulations of Hall-effect plasma accelerators on a magnetic-field-aligned mesh,” *Physical Review E*, Vol. 86, No. 4, 2012. <https://doi.org/10.1103/PhysRevE.86.046703>.
- [10] Mikellides, I. G., and Lopez Ortega, A., “Challenges in the development and verification of first-principles models in Hall-effect thruster simulations that are based on anomalous resistivity and generalized Ohm’s law,” *Plasma Sources Science and Technology*, Vol. 28, No. 1, 2019, p. 48. <https://doi.org/10.1088/1361-6595/aae63b>, URL <https://doi.org/10.1088/1361-6595/aae63b>.
- [11] Scharfe, M. K., Thomas, C. A., Scharfe, D. B., Gascon, N., Cappelli, M. A., and Fernandez, E., “Shear-based model for electron transport in hybrid Hall thruster simulations,” *IEEE Transactions on Plasma Science*, Vol. 36, No. 5 PART 1, 2008, pp. 2058–2068. <https://doi.org/10.1109/TPS.2008.2004364>.
- [12] Cappelli, M. A., Young, C. V., Cha, E., and Fernandez, E., “A zero-equation turbulence model for two-dimensional hybrid Hall thruster simulations,” *Physics of Plasmas*, Vol. 22, No. 11, 2015. <https://doi.org/10.1063/1.4935891>.
- [13] Lafleur, T., Baalrud, S. D., and Chabert, P., “Theory for the anomalous electron transport in Hall effect thrusters. I. Insights from particle-in-cell simulations,” *Physics of Plasmas*, Vol. 23, No. 5, 2016. <https://doi.org/10.1063/1.4948495>.
- [14] Marks, T. A., Ortega, A. L., Mikellides, I. G., and Jorns, B., *Self-consistent implementation of a zero-equation transport model into a predictive model for a Hall effect thruster*, ????, <https://doi.org/10.2514/6.2021-3424>, URL <https://arc.aiaa.org/doi/abs/10.2514/6.2021-3424>.
- [15] Su, L. L., Marks, T. A., and Jorns, “Investigation into the efficiency gap between krypton and xenon operation on a magnetically shielded Hall thruster.” *International Electric Propulsion Conference*, 2022.
- [16] Jorns, B. A., “Two Equation Closure Model for Plasma Turbulence in a Hall Effect Thruster,” *36th International Electric Propulsion Conference*, 2019, pp. 1–12.
- [17] Jones, D. R., Schonlau, M., and Welch, W. J., “Efficient Global Optimization of Expensive Black-Box Functions,” *Journal of Global Optimization*, Vol. 13, 1998, pp. 455–492.
- [18] Kennedy, M. C., and O’Hagan, A., “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 3, 2001, pp. 425–464. <https://doi.org/10.1111/1467-9868.00294>, URL <https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9868.00294><https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00294><https://rss.onlinelibrary.wiley.com/doi/10.1111/1467-9868.00294>.
- [19] Bouhlel, M. A., Hwang, J. T., Bartoli, N., Lafage, R., Morlier, J., and Martins, J. R., “A Python surrogate modeling framework with derivatives,” *Advances in Engineering Software*, Vol. 135, 2019, p. 102662. <https://doi.org/10.1016/j.advengsoft.2019.03.005>.
- [20] Ginsbourger, D., Le Riche, R., and Carraro, L., *Kriging Is Well-Suited to Parallelize Optimization*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 131–162. [https://doi.org/10.1007/978-3-642-10701-6\\_6](https://doi.org/10.1007/978-3-642-10701-6_6), URL [https://doi.org/10.1007/978-3-642-10701-6\\_6](https://doi.org/10.1007/978-3-642-10701-6_6).
- [21] Sankovic, J. M., Hamley, J. A., Haaa, T. W., and Haagt, T. W., “Performance Evaluation of the Russian SPT-100 Thruster at NASA LeRC,” *International Electric Propulsion Conference*, Seattle, Washington, 1993, p. 094.
- [22] MacDonald-Tenenbaum, N., Pratt, Q., Nakles, M., Pilgram, N., Holmes, M., and Hargus, W., “Background Pressure Effects on Ion Velocity Distributions in an SPT-100 Hall Thruster,” *Journal of Propulsion and Power*, Vol. 35, No. 2, 2019, pp. 403–412. <https://doi.org/10.2514/1.B37133>, URL <https://doi.org/10.2514/1.B37133>.
- [23] Marks, T. A., Schedler, P., and Jorns, B. A., “HallThruster.jl - An open source 1D fluid Hall thruster code,” , 2022. <https://doi.org/10.5281/zenodo.6614275>, URL <https://github.com/UM-PEPL/HallThruster.jl>.
- [24] Giannetti, V., Saravia, M. M., Leporini, L., Camarri, S., and Andreussi, T., “Numerical and Experimental Investigation of Longitudinal Oscillations in Hall Thrusters,” *Aerospace*, Vol. 8, No. 6, 2021. <https://doi.org/10.3390/aerospace8060148>, URL <https://www.mdpi.com/2226-4310/8/6/148>.
- [25] Urquhart, M., Ljungskog, E., and Sebben, S., “Surrogate-based optimisation using adaptively scaled radial basis functions,” *Applied Soft Computing*, Vol. 88, 2020. <https://doi.org/10.1016/j.asoc.2019.106050>.

- [26] Forrester, A. I. J., Sóbester, A., and Keane, A. J., “Engineering Design via Surrogate Modelling,” *Engineering Design via Surrogate Modelling*, 2008. <https://doi.org/10.1002/9780470770801>, URL <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470770801>.
- [27] Rios, L. M., and Sahinidis, N. V., “Derivative-free optimization: A review of algorithms and comparison of software implementations,” *Journal of Global Optimization*, Vol. 56, No. 3, 2013, pp. 1247–1293. <https://doi.org/10.1007/S10898-012-9951-Y>.
- [28] Jones, D. R., and Martins, J. R., “The DIRECT algorithm: 25 years Later,” *Journal of Global Optimization*, Vol. 79, No. 3, 2021, pp. 521–566. <https://doi.org/10.1007/S10898-020-00952-6>, URL <https://link.springer.com/article/10.1007/s10898-020-00952-6>.