# CONDITIONING SPATIAL REPRESENTATIONS IN FOURIER NEURAL OPERATORS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Geometric and operating parameters can change both the magnitude and spatial organization of a physical field. For Fourier neural operators (FNOs), this raises a design question: when should parameters participate in constructing spatial features, in addition to entering the input or modulating Fourier layers? We study this question with a parameter-conditioned sinusoidal coordinate network that supplies features to Fourier processing. Controlled heat-transport problems with translating and expanding material rings distinguish the benefit of learning spatial features from the additional benefit of conditioning them. A shared spatial representation is effective in the expanding-ring setting; conditioning provides a larger gain in the translating-ring setting, improving whole-field accuracy, localized thermal structure, and responses to geometry changes. In radio-frequency hollow cathode discharge, an industrial plasma problem in semiconductor manufacturing, the model predicts four coupled fields across operating conditions in a fixed cathode geometry and improves on the evaluated baselines. These results highlight spatial feature construction as a consequential choice in conditioning physical-field predictors.

## 1 INTRODUCTION

Many engineering systems are specified by a small set of geometric and operating parameters, yet their behavior is expressed through spatially distributed physical fields. Changing a material interface or the operating conditions can alter where heat accumulates, how particles are confined, and how the surrounding field is organized. Predicting these responses across configurations is important for design and process assessment, where repeatedly solving the underlying physical system can be costly. Data-driven surrogates offer a way to learn from available simulations and predict spatially resolved fields at new parameter settings.

Fourier neural operators (FNOs) provide an efficient framework for field prediction by combining spectral mixing with pointwise transformations (Li et al., 2021). Geometry-aware extensions incorporate geometric information through learned coordinate transformations or representations that connect physical and latent grids (Li et al., 2023a;b). We consider fields sampled on a common regular grid as internal geometry or operating conditions vary. The coordinates remain fixed, while the spatial response changes with the configuration. A predictor must learn how a compact parameter description produces these configuration-dependent spatial responses.

This setting motivates a specific design question: when should system parameters participate in constructing spatial features before those features undergo Fourier processing? Parameters can already enter an FNO through input channels or condition its intermediate transformations. A learned coordinate network offers an additional point of control. When it receives coordinates alone, it supplies one spatial feature map shared across configurations. Conditioning this network allows the features at each location to change with the physical configuration. Conditional neural representations provide established mechanisms for making coordinate-based functions depend on input information (Mehta et al., 2021; Lee et al., 2023). We investigate this choice as part of the representation presented to a Fourier model.

To study this question, we use a spatially conditioned FNO in which system parameters modulate a sinusoidal coordinate network. The resulting spatial features are processed by conditioned Fourier blocks to predict the physical fields. The coordinate branch constructs features from position and configuration, while the Fourier blocks exchange information across locations. All components are



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

trained jointly from field observations. This architecture gives parameters a direct role in shaping the spatial representation used for prediction.

We study this design through controlled heat transport and an industrial plasma application. Translating and expanding low-diffusivity material rings change heat transport around prescribed sources, providing interpretable tests of prediction across geometry configurations. Comparisons with a shared coordinate representation separate the benefit of learning spatial features from the additional benefit of conditioning their construction. Whole-field errors, localized errors, and responses between geometry configurations examine complementary aspects of the resulting predictions. The industrial setting is radio-frequency hollow cathode discharge in semiconductor manufacturing, where operating conditions determine coupled electric-potential and species-density fields within a fixed cathode geometry. Together, these settings connect a controlled architectural question to spatial prediction in a practical engineering system.

Our contributions can be summarized as follows:

- We examine when system parameters should participate in spatial feature construction. Controlled material-ring experiments distinguish the benefit of learning spatial features from the additional benefit of conditioning them: a shared representation is effective for expansion, while conditioning gives a larger gain for translation and improves localized structure and geometry responses.
- On fixed-geometry radio-frequency hollow cathode discharge, a shared-embedding implementation achieves the lowest mean error on each of four plasma fields and outperforms FNO, Transolver, and DeepONet in test loss in all five training seeds.

## 2 Technical Background

### 2.1 Geometry-dependent solution fields

We consider physical systems on a fixed rectangular coordinate domain $D \subset \mathbb{R}^2$. Internal material interfaces can move or expand while the sampling grid stays the same.

Let $q = (g, p) \in \mathcal{Q} \subset \mathbb{R}^{d_q}$ collect the system parameters, where $g$ describes the geometry and $p$ contains the remaining physical conditions. The target is a field with $d_u$ output channels, such as temperature, electric potential, or species density. Writing $\mathcal{U}$ for the space of these fields on $D$, the solution map is

$$\mathcal{S}^\dagger : \mathcal{Q} \longrightarrow \mathcal{U}, \qquad q = (g, p) \longmapsto u(\cdot; q). \tag{1}$$

Here $u(x; q) \in \mathbb{R}^{d_u}$ is the field value at coordinate $x$. Depending on the physical system, the target is a steady solution or a cycle-averaged field. Varying $g$ indexes different geometries; in fixed-geometry experiments, $g$ is held constant and only $p$ varies.

Training pairs provide $q$ and the corresponding field sampled on a common uniform grid $X_h = \{x_i\}_{i=1}^{N_h}$, giving an array $U_q \in \mathbb{R}^{N_h \times d_u}$. A predictor learns to produce this array for a new parameter vector. Known spatial inputs, such as a prescribed forcing field, are denoted by $I_q(x)$ when available. The prediction task therefore connects a compact description of the system to a spatially resolved response.

### 2.2 Coordinate-based spatial representations

A coordinate network maps each location to a feature vector $h(x) = H_\theta(\xi(x)) \in \mathbb{R}^{d_h}$. Here $\xi(x)$ is a prescribed coordinate encoding, such as normalized coordinates or sine–cosine features, and $\theta$ denotes the learned weights. Sinusoidal representation networks (SIRENs) use sine activations to construct this map (Sitzmann et al., 2020). Starting from $s_0(x) = \xi(x)$, their hidden layers take the form

$$s_{\ell+1}(x) = \sin(\omega_\ell \, [A_\ell s_\ell(x) + c_\ell]) \, . \tag{2}$$

The matrix $A_\ell$ and bias $c_\ell$ are learned, and $\omega_\ell$ is a prescribed frequency multiplier. The sine acts elementwise. Evaluating the network at all points in $X_h$ gives an $N_h \times d_h$ array: each row describes one location, and each column is a learned spatial feature. When the network receives coordinates alone, this feature map is shared across system configurations.

### 2.3 FOURIER NEURAL OPERATORS AND CONDITIONING

A Fourier neural operator (FNO) lifts its inputs to a feature field, mixes information across the grid, and projects the result to the output channels (Li et al., 2021). Write $a_q(x)$ for the input channels, which can include coordinates, broadcast parameters, and prescribed forcing fields. A pointwise lift gives $v_0(x;q) = P_\theta(a_q(x))$, with $w$ feature channels. A Fourier block then combines spectral mixing with a pointwise transformation:

$$v_{\ell+1}(x;q) = \sigma_\ell\big(W_\ell v_\ell(x;q) + b_\ell + \mathcal{F}_h^{-1}[R_\ell\, \mathcal{F}_h v_\ell(\cdot;q)]\,(x)\big)\,. \tag{3}$$

The transforms $\mathcal{F}_h$ and $\mathcal{F}_h^{-1}$ act on the grid's spatial axes. At each retained frequency $k$, the learned complex matrix $R_\ell(k) \in \mathbb{C}^{w\times w}$ mixes channels; unretained frequencies are set to zero in the spectral branch. The matrix $W_\ell$ and bias $b_\ell$ act at each grid point, and $\sigma_\ell$ is a pointwise activation, which may be the identity in the final block. A readout produces $\widehat{u}(x;q) = Q_\theta(v_L(x;q))$ after $L$ blocks.

Conditioning makes these computations depend on the system parameters. An encoder maps $q$ to a latent code $z = E_\theta(q)$, and small layer-specific networks, called modulation heads, turn $z$ into modulation coefficients. Feature-wise linear modulation (FiLM) (Perez et al., 2018) uses channel scales and shifts: $\widetilde{v}_\ell(x;q) = \gamma_\ell(z) \odot v_\ell(x;q) + \beta_\ell(z)$. The vectors $\gamma_\ell(z), \beta_\ell(z) \in \mathbb{R}^w$ are shared across grid locations, and $\odot$ denotes elementwise multiplication. The coefficients vary between configurations while the backbone weights remain shared. The coordinate network constructs spatial features at each location, while Fourier layers exchange information across locations. This distinction motivates conditioning not only feature processing, but also the construction of spatial features.

## 3 METHODOLOGY

Our model constructs spatial features from coordinates and system parameters, adds them to a conventional FNO input lift, and processes the combined field with conditioned Fourier blocks. This lets the spatial representation change with the system configuration before information is mixed across the grid. We describe the architecture used in our controlled PDE experiments.

The parameter encoder supplies the spatial and Fourier stages with condition codes $z_{\mathrm{sp}}(q)$ and $z_{\mathrm{op}}(q)$, respectively. The coordinate network produces $h_q(x) = H_\theta(\xi(x); z_{\mathrm{sp}}(q)) \in \mathbb{R}^{d_h}$. Using the input channels $a_q$ defined in Section 2.3, the forward pass is

$$\begin{aligned} v_0(x;q) &= P_\theta(a_q(x)) + B_\theta(h_q(x)), \\ v_L(\cdot;q) &= F_\theta(v_0(\cdot;q); z_{\mathrm{op}}(q))\,, \\ \widehat{u}(x;q) &= \mathcal{R}_q(Q_\theta(v_L(x;q)))\,. \end{aligned} \tag{4}$$

Here $P_\theta$ is the direct input lift, $B_\theta$ is a learned pointwise projection from $d_h$ spatial channels to $w$ Fourier channels, and $F_\theta$ contains $L$ Fourier blocks. The pointwise readout $Q_\theta$ produces all output channels; the prescribed conversion $\mathcal{R}_q$ returns them to physical units. Both terms in $v_0$ are evaluated on the same grid.

### 3.1 CONDITIONED SPATIAL FEATURES

Layer-specific heads use the spatial code to modulate the SIREN introduced in Section 2.2. Every channel of the spatial network receives the same code $z_{\mathrm{sp}}$. The resulting feature vector depends on both the location and the system configuration. The parameter inputs and preprocessing for each stage are specified with the experimental setup.

Starting from $s_0(x;q) = \xi(x)$, each SIREN layer applies condition-dependent amplitude and phase adjustments:

$$s_{\ell+1}(x;q) = \big(1 + \alpha_\ell^{\mathrm{sp}}(z_{\mathrm{sp}})\big) \odot \sin(\omega_\ell[A_\ell s_\ell(x;q) + c_\ell] + \beta_\ell^{\mathrm{sp}}(z_{\mathrm{sp}}))\,. \tag{5}$$

The heads produce one coefficient pair per hidden channel. The coefficients are shared across grid locations, while $A_\ell$ and $c_\ell$ are shared across configurations. Amplitude modulation changes the strength of a spatial feature, and phase modulation changes its sine response as a function of position.

The final sine layer supplies $h_q$ in Equation 4.

### 3.2 Fourier processing and prediction

The combined feature field is processed by the spectral and pointwise paths of Equation 3. The operator condition code supplies channelwise scale and shift through FiLM:

$$\mathcal{M}_\ell(v; z_{\mathrm{op}}) = \big(1 + \alpha_\ell^{\mathrm{op}}(z_{\mathrm{op}})\big) \odot v + \beta_\ell^{\mathrm{op}}(z_{\mathrm{op}}). \tag{6}$$

The coefficients are broadcast across grid locations. This transformation is applied to each block's input, before the spectral and pointwise paths. The learned Fourier kernels and pointwise weights are shared across configurations.

A static-spatial control replaces $h_q(x)$ with a coordinate-only $h(x)$ in the same additive path, keeping the direct parameter inputs and Fourier conditioning. Its learned spatial features are shared across configurations, whereas our model constructs them from the current system parameters.

The readout predicts all fields at every grid point, and all learned components are trained end to end with supervised field losses.

## 4 Physical Problems

We consider two physical settings with fields sampled on a fixed grid. They provide a direct test of whether conditioned spatial features improve field prediction at held-out parameter configurations. Material-ring heat transport offers controllable combinations of sources, material properties, and geometry. Hollow cathode discharge extends this evaluation to an industrial plasma with coupled transport, reactions, and electric fields.

### 4.1 Heat transport through a material ring

Low-diffusivity material slows the spread of heat. Moving a ring of this material across fixed heat sources changes where heat accumulates; expanding it changes the region it encloses (Figure 1). Predicting the resulting field requires resolving localized temperature variations near the ring as well as heat redistribution across the domain.

We model two coupled steady temperature fields on the dimensionless periodic square $D = [0, 2\pi)^2$:

$$-\nabla\cdot(\kappa_g(x)\nabla u_i(x)) + u_i(x) + \rho\big(u_i(x) - u_j(x)\big) = f_i(x), \qquad (i,j) \in \{(1,2),(2,1)\}. \tag{7}$$

Here $f_i$ is a prescribed heat source, the unit reaction term represents heat loss, and $\rho$ controls heat exchange between the two fields. The coupling transfers heat from the higher-temperature field to the lower-temperature field at each location. The diffusivity $\kappa_g$ is lower within the ring wall than in its surroundings. Heat sources act throughout the domain, including the wall, so their overlap with the material directly affects the temperature response.

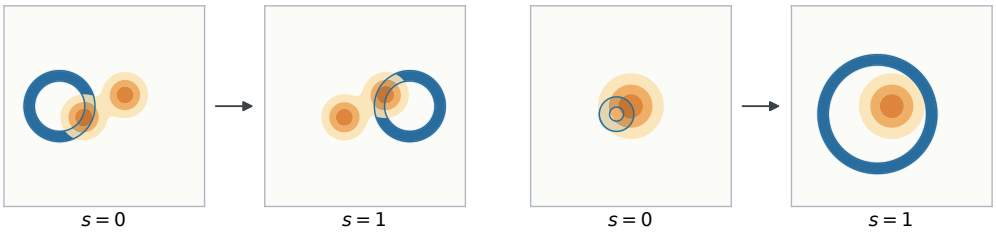


Figure 1: **Changing the material geometry around fixed heat sources.** Blue shows the low-diffusivity ring; orange contours show the localized part of $f_1$. Each pair displays the endpoints of a geometry sweep, indexed by the normalized geometry parameter $s \in [0, 1]$. The left pair translates the ring past two sources; the right pair expands it around a single source. These are coefficient-and-forcing schematics on the common periodic square.

Along each geometry sweep, the sources remain fixed. The predictor receives coordinates, geometry parameters, and the prescribed source fields, and returns both temperature fields over the whole square.

The material ring provides an idealized setting for heat transport in heterogeneous materials. A practical example is thermal protection of metallic components, where low-conductivity coatings reduce heat transfer into the underlying metal (Miller & Zhu, 2003). Such applications make the arrangement of materials an important part of thermal design. Our ring geometry expresses this dependence through two simple changes—position and size—while keeping the relationship between material placement, heat sources, and temperature patterns easy to interpret.

### 4.2 Radio Frequency Hollow Cathode Discharge

Radio frequency hollow cathode discharge (HCD) concerns plasma sustained within an electrode cavity. Simulations of a plasma etching reactor linked locally enhanced discharges at showerhead gas-hole outlets to gas-hole erosion over the showerhead's life (Denpoh, 2009). Bera et al. (2024b) characterized RF HCD with particle-in-cell/Monte Carlo collision simulations and built a neural-network reduced-order model from those simulations. Here we learn spatially resolved fields from fluid simulations across operating conditions in a fixed cathode geometry.

The dataset consists of simulations of an axisymmetric argon cavity (Figure 2). A 19.5 mm-diameter, 40 mm-deep bore is located at $z = 41.75$ mm. The cap, bore sidewall, and annular face at the mouth ($9.75 \le r \le 15$ mm) form the RF-powered electrode; the base is grounded. The drive is $V_{\mathrm{el}}(t) = V_{\mathrm{RF}} \sin(2\pi f t)$, with frequency $f = 13.56$ MHz and amplitude $V_{\mathrm{RF}}$. The self-bias $V_{\mathrm{dc}}$ develops without applied DC. The ion-induced secondary-electron emission coefficient (SEEC), $\gamma_{\mathrm{se}}$, specifies the number of electrons emitted per incident ion. The three operating parameters are pressure, RF voltage, and SEEC.

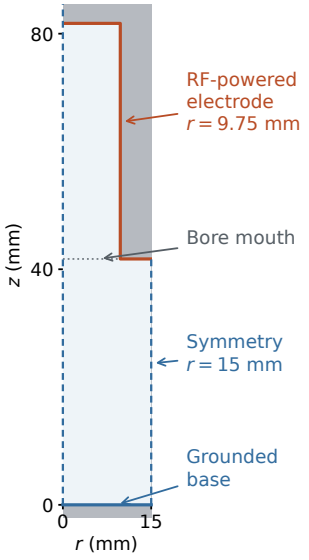


Figure 2: **Hollow cathode geometry.** Axisymmetric half-section. Orange marks the powered cap, sidewall ($r = 9.75$ mm), and annular face; blue marks ground and dashed symmetry boundaries. Grey denotes solid. Base location and thickness are schematic.

The data-generating model is a two-dimensional axisymmetric application of the fluid model described by Rauf et al. (2009); Agarwal et al. (2012), using argon plasma with an electrostatic description. Electron transport is modeled using a drift–diffusion approximation, whereas ion dynamics are governed by fluid momentum equations. In the gas region $\Omega_{\mathrm{gas}}$, with radial–axial coordinates $x = (r, z)$, its balances are

$$\begin{aligned} &\partial_t n_\nu + \nabla\cdot \mathbf{\Gamma}_\nu = S_\nu, \\ &-\nabla\cdot(\varepsilon_0 \nabla\phi) = \rho_c = e(n_{\mathrm{Ar}^+} - n_e), \qquad \mathbf{E} = -\nabla\phi, \\ &\partial_t w_e + \nabla\cdot \mathbf{Q}_e = \mathbf{J}_e\cdot\mathbf{E} - \mathcal{L}_e. \end{aligned} \tag{8}$$

Here $\nu \in \{e, \mathrm{Ar}^+, \mathrm{Ar}^*\}$ denotes electrons, argon ions, and metastable argon; $n_\nu$, $\mathbf{\Gamma}_\nu$, and $S_\nu$ are density, particle flux, and net production. The remaining quantities are vacuum permittivity $\varepsilon_0$, elementary charge $e$, electron energy density $w_e$, energy flux $\mathbf{Q}_e$, current $\mathbf{J}_e = -e\mathbf{\Gamma}_e$, and collisional energy loss $\mathcal{L}_e$. Derivatives are axisymmetric. Metastable kinetics include stepwise ionization, pooling, and quenching (Lymberopoulos & Economou, 1993). Boundary conditions impose $\phi = V_{\mathrm{el}}(t)$ on the powered surfaces, $\phi = 0$ on the base, ion-induced emission with coefficient $\gamma_{\mathrm{se}}$, and axis of symmetry at $r = 0$ and symmetry $r = 15$ mm.

The targets average one RF period $T = 1/f$ after periodic steady state is reached at $t_0$. The four physical fields are

$$u(x;q) = \frac{1}{T}\int_{t_0}^{t_0+T} \left[\phi(x,t;q),\, n_e(x,t;q),\, n_{\mathrm{Ar}^+}(x,t;q),\, n_{\mathrm{Ar}^*}(x,t;q)\right]^\top dt. \tag{9}$$

Densities span several orders of magnitude in the gas. Predicting the four fields therefore requires resolving both localized charged-particle structure and broader variations in potential and metastable density.

# 5 EXPERIMENTS

We evaluate whether conditioning the construction of spatial features improves prediction across parameter configurations. We compare whole-field accuracy, examine errors near localized physical responses, and use a static spatial representation to identify the contribution of parameter dependence.

## 5.1 EXPERIMENTAL SETUP

**Heat transport.** Each ring family contains 128 training, 64 validation, and 128 test configurations with disjoint geometry parameters. Both temperature fields are predicted on the full $256 \times 256$ grid. The translating ring passes two fixed sources, while the expanding ring encloses a single source.

All models receive periodic coordinates, prescribed source fields, and broadcast system parameters. FNO uses a direct input lift and Fourier processing; FiLM FNO also modulates the Fourier blocks. Static Spatial adds a coordinate-only SIREN branch, while our model conditions that branch on the current parameters. The Fourier-stage condition code is the same in Static Spatial and our model. The models have approximately 117,000 trainable real scalar parameters (Table 1) and use the same training budget and validation-based checkpoint selection.

We report scaled MSE using training-set channel scales and physical relative $L^2$, each averaged equally over test cases. Results are means and sample standard deviations over training seeds 17, 29, and 43. Appendix A.1 specifies the source and material parameters, architectures, training schedule, and metrics.

**Hollow cathode discharge.** The fixed-cathode argon dataset varies pressure, RF voltage, and SEEC over five levels each, giving 125 operating conditions. RF voltage ranges from 50 to 150 V and SEEC from 0 to 0.1. Each case contains four cycle-averaged fields on the same $31 \times 171$ radial–axial grid with 0.5 mm spacing. We use a fixed split of 101 training, 12 validation, and 12 test cases, with five training seeds (0–4) for every architecture.

Our HCD predictor uses a shared embedding of all three operating parameters to condition both its SIREN and Fourier layers. We compare with FNO, Transolver (Wu et al., 2024), and DeepONet (Lu et al., 2021). The implementation and common training procedure are given in Appendix A.2. All models predict and are evaluated on the entire grid. Densities are represented in standardized log space and potential in standardized linear space. Table 3 reports aggregate test loss and fieldwise relative $L^2$, with means and sample standard deviations over the five seeds.

## 5.2 PREDICTION ACROSS GEOMETRY CONFIGURATIONS

Table 1: **Held-out heat-transport accuracy.** Mean $\pm$ sample standard deviation over three training seeds, on fixed geometry-disjoint splits. Lower is better. Bold identifies the lowest mean in each column.

| | Real | Translation | | Expansion | |
|---|---|---|---|---|---|
| Model | parameters | MSE $\times 10^{-4}$ | $L^2 \times 10^{-3}$ | MSE $\times 10^{-6}$ | $L^2 \times 10^{-4}$ |
| FNO | 117,062 | $3.07 \pm 0.89$ | $3.23 \pm 0.46$ | $18.39 \pm 6.65$ | $6.374 \pm 1.083$ |
| FiLM FNO | 116,162 | $4.13 \pm 1.76$ | $3.68 \pm 0.81$ | $12.85 \pm 2.14$ | $5.426 \pm 0.455$ |
| Static Spatial | 116,946 | $3.99 \pm 0.59$ | $3.66 \pm 0.26$ | $10.10 \pm 1.21$ | $4.861 \pm 0.280$ |
| Ours | 117,058 | $\mathbf{1.85 \pm 1.07}$ | $\mathbf{2.42 \pm 0.69}$ | $\mathbf{9.38 \pm 1.83}$ | $\mathbf{4.647 \pm 0.460}$ |

Our model improves both whole-field metrics over FNO and FiLM FNO in both geometry families (Table 1). The translating-ring result also improves on Static Spatial in every training seed. Thus, this setting benefits from making the coordinate representation depend on the configuration, beyond supplying geometry parameters to the direct input lift or modulating Fourier features.

The expanding-ring family also favors learned spatial features over the two FNO baselines. Our model has the lowest mean error, improving on Static Spatial in two of three seeds. The additional gain from conditioning is smaller here than in the translating, two-source family. This contrast

persists when the heat sources and material contrast are matched (Appendix C.2), linking the value of conditioning to the geometry path.

### 5.3 Resolving localized field structure

Whole-field scores summarize accuracy but do not show which structures are recovered. Figure 3 examines the central configuration of the translating-ring sweep. The material wall restricts diffusion near the sources, concentrating the temperature response. Enlarging the same material window for every model reveals errors around the ring that are difficult to distinguish in full-field predictions.

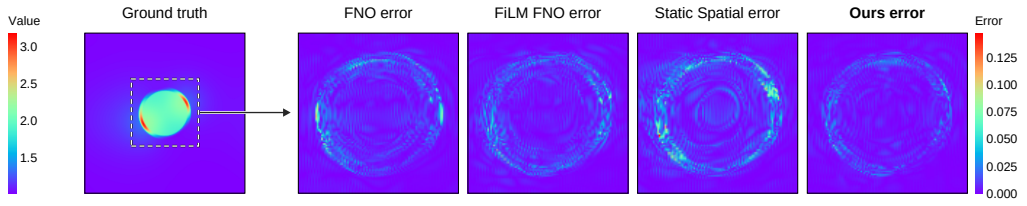


Figure 3: **Localized errors around the material ring.** Field $u_1$ at $s = 0.5$. The dashed box in the full reference field marks the material-ring window enlarged in the error panels. The four models share a linear absolute-error scale; reference values use a separate scale. Panels use seed 29, the middle seed by our model's full-test MSE.

A held-out configuration provides a closer view of the concentrated thermal response (Figure 4). The crop is centered on the reference maximum. Our model reduces the spatial extent of error around the hotspot and has the lowest mean absolute error over this window. This view connects the aggregate improvement to reconstruction of the hotspot's surrounding structure.

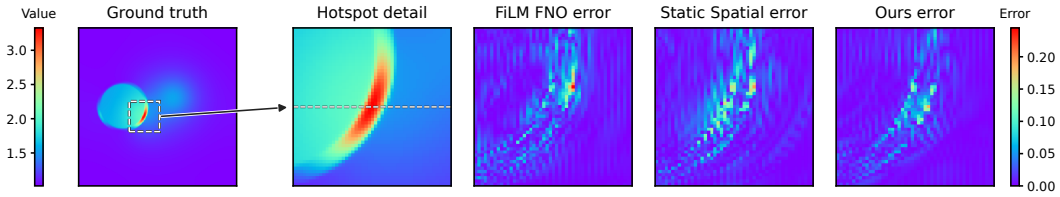


Figure 4: **A held-out hotspot in detail.** Field $u_1$ for the first held-out translating-ring case, training seed 17 ($s \approx 0.00252$). The boxed $49 \times 49$-node window around the reference maximum is enlarged in the second panel. The three error panels use that same window and a shared linear absolute-error scale; the reference and its detail share a separate value scale. The dashed line in the detail locates the profiles in Appendix B.1.

We quantify error in two regions across all held-out configurations: the material transition and the highest $5\%$ of reference values in each field. The first tests interface structure; the second tests concentrated thermal response. Region definitions and aggregation follow Appendix A.1.

Table 2: **Localized accuracy and geometry response for translation.** Means over three seeds. Transition and hotspot scaled MSE are measured on the test set. Response relative $L^2$ compares adjacent differences along nine fixed geometry states. Lower is better.

| Model | Transition MSE $\times 10^{-3}$ | Hotspot MSE $\times 10^{-3}$ | Response $L^2$ $\times 10^{-2}$ |
|---|---|---|---|
| FNO | 1.041 | 2.325 | 5.706 |
| FiLM FNO | 1.314 | 2.687 | 6.446 |
| Static Spatial | 1.331 | 2.808 | 6.371 |
| Ours | **0.618** | **1.344** | **4.275** |

Our model has lower error in both regions (Table 2), connecting the whole-field gain to the material transition and concentrated thermal response. For the illustrated held-out case, Appendices B.2 and B.3 also find smaller errors at a non-peak boundary and in the smooth exterior. The local improvement therefore includes structures beyond the dominant hotspot.

### 5.4 Responses to geometry changes

Accurate fields should also reproduce how the solution changes between configurations. We compare adjacent predicted field differences with reference differences along nine evenly spaced geometry states. Response relative $L^2$ measures the discrepancy relative to the reference change (Appendix A.1).

Our model gives the lowest response error (Table 2), connecting spatial conditioning to accurate changes between fields as well as individual predictions. Appendix C.1 reports the corresponding localized and response metrics for single-source expansion.

### 5.5 Accuracy of the coupled plasma fields

Table 3: **Five-seed prediction of cycle-averaged plasma fields.** Mean $\pm$ sample standard deviation on the same 12 test cases. Test loss and normalized-space relative $L_2$ are in units of $10^{-3}$; lower is better. Densities use standardized $\log_{10}$ values and potential uses standardized linear values. Ours uses shared conditioning. Parameter counts sum trainable tensor elements, counting complex entries once.

| | Params | Test loss | Normalized relative $L_2$ | | | |
|---|---|---|---|---|---|---|
| Model | (M entries) | | Potential | Electrons | $Ar^+$ | $Ar^*$ |
| DeepONet | 6.671 | $27.24\pm12.00$ | $12.27\pm3.40$ | $20.44\pm6.80$ | $27.88\pm14.00$ | $36.25\pm19.00$ |
| Transolver | 6.669 | $5.859\pm0.180$ | $6.693\pm0.720$ | $4.931\pm0.330$ | $4.075\pm0.580$ | $5.064\pm0.540$ |
| FNO | 6.421 | $3.282\pm0.230$ | $3.914\pm1.100$ | $2.901\pm0.130$ | $2.449\pm0.370$ | $3.041\pm0.260$ |
| Ours | 6.672 | $\mathbf{2.739}\pm0.085$ | $\mathbf{3.246}\pm0.370$ | $\mathbf{2.313}\pm0.190$ | $\mathbf{2.092}\pm0.160$ | $\mathbf{2.491}\pm0.100$ |

Our shared-embedding model achieves the lowest mean error in each of the four output fields (Table 3). The aggregate test-loss ordering is identical in all five seeds: ours, FNO, Transolver, then DeepONet. The gain spans electric potential and both charged and neutral species densities, rather than being confined to a single output. Appendix D.2 connects these test results to the five-seed validation trajectories under the common training budget.

Figure 5 connects these aggregate errors to the cycle-averaged plasma structure. The prediction recovers the narrow charged-particle layer beneath the cavity, the broader metastable-density profile, and the potential variation across the chamber. These distinct spatial patterns reflect the coupling between species transport, reactions, and charge described above. Appendix D.1 compares reference fields across operating conditions, separating changes in amplitude from changes in their spatial distribution within the fixed device.

## 6 Related Work

**Conditioning neural operators.** Fourier neural operators learn nonlocal interactions through spectral convolutions (Li et al., 2021). System parameters can enter these models through input channels, feature modulation, or weight generation. FiLM supplies condition-dependent channelwise scales and shifts (Perez et al., 2018); HyperFNO uses hypernetworks to generate weights in the physical and Fourier domains (Alesiani et al., 2022). The hpFNO architecture introduces shifts that depend jointly on spatial coordinates and system parameters within Fourier layers (Jing et al., 2026). Our focus is how the location of conditioning affects field prediction: we test whether parameters should also construct the spatial features supplied to Fourier processing.

**Conditional spatial representations.** Coordinate networks represent signals as functions of position. SIREN uses periodic activations to represent detailed spatial variation (Sitzmann et al., 2020), and modulated periodic networks let a latent code change these representations across signals (Mehta

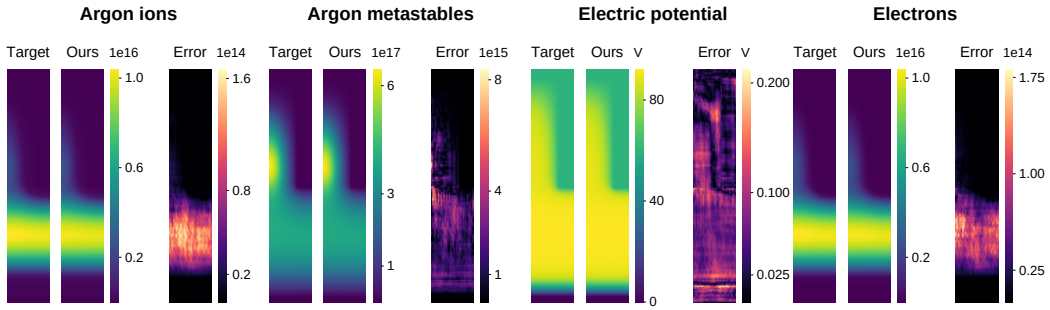


Figure 5: **Cycle-averaged plasma fields at a held-out operating condition.** Targets, our predictions, and absolute errors for electric potential and the three species densities. Each target is averaged over one RF period after periodic steady state is reached. This illustration uses the separate seed-1000 run. Axial position runs vertically and radial position horizontally, preserving physical aspect ratio. Each target–prediction pair shares a linear value scale; each error map has its own linear scale. Densities are in $\mathrm{m}^{-3}$ and potential is in volts.

et al., 2021). This idea also appears in scientific learning. Neural Implicit Flow separates spatial representation from temporal or parametric dependence (Pan et al., 2023). HyperDeepONet generates the weights of a coordinate-query network from an input function (Lee et al., 2023), while CORAL learns operators between latent codes of modulated neural fields (Serrano et al., 2023). SirenFNO uses a SIREN for a different role: parameterizing the Fourier kernel as a function of frequency (Shi et al., 2026). In our architecture, the conditioned coordinate network produces intermediate features in physical space. Fourier blocks then exchange information across locations to predict the solution field. The static-spatial control isolates the effect of letting the current system parameters change these features, with operator conditioning retained in both models.

**Geometry-aware operator learning.** Geometry can enter an operator through the coordinate map, integration domain, or representation used for nonlocal interactions. Geo-FNO learns a deformation from physical geometry to a regular latent grid (Li et al., 2023a). DAFNO incorporates a smoothed domain indicator within its integral layers (Liu et al., 2023). GINO combines signed-distance geometry information and graph operators with Fourier processing on a regular latent grid (Li et al., 2023b). Transolver instead groups spatial points into learned physics-aware slices and applies attention between their tokens (Wu et al., 2024). Our setting keeps the sampling grid fixed while internal geometry or operating conditions change the physical field. We use system parameters to construct spatial features on the original coordinates, then process the features with Fourier layers. The ring experiments test this choice under geometry variation, and HCD evaluates it across operating conditions in a fixed device.

## 7 Conclusion and Limitations

Building on established sinusoidal representations and Fourier operators, this work examines when parameter-dependent spatial features improve physical-field prediction. Controlled ring experiments distinguish the benefit of conditioning from that of learning a shared representation, while fixed-geometry HCD extends the evaluation to an industrial plasma problem. Our conclusions are empirical and limited to the studied settings and FNO-based architectures. Theoretical criteria for when spatial conditioning is beneficial, and evaluation on HCD with changing cathode geometry, remain future directions.

## AI USE STATEMENT

Generative AI tools assisted with experiment management, code development, and manuscript preparation. Uses included organizing and scheduling computational jobs, monitoring runs, prototyping and debugging code, preparing analysis and visualization scripts, and editing text for grammar, clarity, and consistency. AI assistance also supported cross-checking implementation descriptions against experimental configurations and reported results. The authors directed the research and retained responsibility for experimental design, methodological choices, interpretation of results, and the final presentation. The authors are responsible for reviewing and validating AI-assisted code and text and take full responsibility for the correctness, accuracy, and integrity of the work.

## REPRODUCIBILITY STATEMENT

We describe the problem formulations in Section 4 and report the experimental setups in Section 5. Appendix A details the architectures, data splits and preprocessing, training schedules, random seeds, and evaluation metrics. The material-ring equations, source and material parameters, geometry sweeps, and numerical reference procedure are specified in Section 4.1 and Appendix A.1 to support independent reproduction. The HCD dataset is proprietary and cannot be publicly released. To guide the generation of comparable simulation data, Section 4.2 provides the device geometry, governing physical balances, transport assumptions, boundary conditions, and cycle-averaged target definition, with references to the underlying plasma models. The code for this work, excluding the generation code specific to the proprietary HCD dataset, will be publicly released upon acceptance.

## REFERENCES

Ankur Agarwal, Leonid Dorf, Shahid Rauf, and Ken Collins. On the scaling of rf and dc self-bias voltages with pressure in electronegative capacitively coupled plasmas. *Journal of Vacuum Science & Technology A*, 30(2):021303, 2012. doi: 10.1116/1.3676182.

Francesco Alesiani, Makoto Takamoto, and Mathias Niepert. HyperFNO: Improving the generalization behavior of Fourier neural operators. In *NeurIPS 2022 Workshop on Machine Learning and the Physical Sciences*, 2022. URL https://ml4physicalsciences.github.io/2022/files/NeurIPS_ML4PS_2022_89.pdf.

Kallol Bera, Han Luo, Xingyi Shi, Ihor Korolov, Abhishek Verma, Shahid Rauf, Jan Guttmann, and Julian Schulze. Radio-frequency hollow cathode discharge: Modeling and experimental diagnostics. AVS 70 International Symposium and Exhibition, contribution PS-ThM-1, 2024a. URL https://www.avssymposium.org/AVS2024/Sessions/Schedule/82430.

Kallol Bera, Abhishek Verma, Sathya Ganta, and Shahid Rauf. A neural-network-based model of radio-frequency hollow cathode discharge characterized using particle-in-cell/monte carlo collision simulation. *Journal of Physics D: Applied Physics*, 57(21):215201, 2024b. doi: 10.1088/1361-6463/ad2835.

Kazuki Denpoh. Locally enhanced discharges at gas hole outlets of a showerhead in a plasma etching reactor. *Journal of Physics D: Applied Physics*, 42(3):032003, 2009. doi: 10.1088/0022-3727/42/3/032003.

Cheng Jing, Uvini Balasuriya Mudiyanselage, Abhishek Verma, Kallol Bera, Shahid Rauf, and Kookjin Lee. Extending Fourier neural operators for modeling parameterized and coupled PDEs. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2607.23466.

Jae Yong Lee, Sung Woong Cho, and Hyung Ju Hwang. HyperDeepONet: Learning operator with complex target function space using the limited resources via hypernetwork. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OAw6V3ZAhSd.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for PDEs on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023a. URL https://jmlr.org/papers/v24/23-0064.html.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems*, volume 36, 2023b. URL https://arxiv.org/abs/2309.00583.

Ning Liu, Siavash Jafarzadeh, and Yue Yu. Domain agnostic Fourier neural operators. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2305.00478.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Dimitris P. Lymberopoulos and Demetre J. Economou. Fluid simulations of glow discharges: Effect of metastable atoms in argon. *Journal of Applied Physics*, 73(8):3668–3679, 1993. doi: 10.1063/1.352926.

Ishit Mehta, Michaël Gharbi, Connelly Barnes, Eli Shechtman, Ravi Ramamoorthi, and Manmohan Chandraker. Modulated periodic activations for generalizable local functional representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. URL https://arxiv.org/abs/2104.03960.

Robert A. Miller and Dongming Zhu. Lower-conductivity thermal-barrier coatings. NASA Tech Briefs, June 2003. URL https://ntrs.nasa.gov/citations/20110024150.

Shaowu Pan, Steven L. Brunton, and J. Nathan Kutz. Neural implicit flow: A mesh-agnostic dimensionality reduction paradigm of spatio-temporal data. *Journal of Machine Learning Research*, 24(41):1–60, 2023. URL https://jmlr.org/papers/v24/22-0365.html.

J. D. P. Passchier and W. J. Goedheer. A two-dimensional fluid model for an argon rf discharge. *Journal of Applied Physics*, 74(6):3744–3751, 1993. doi: 10.1063/1.354487.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11671.

Shahid Rauf, Jason Kenney, and Ken Collins. Three-dimensional model of magnetized capacitively coupled plasmas. *Journal of Applied Physics*, 105:103301, 2009. doi: 10.1063/1.3126718.

Louis Serrano, Lise Le Boudec, Armand Kassaï Koupaï, Thomas X Wang, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Operator learning with neural fields: Tackling PDEs on general geometries. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2306.07266.

Pengqing Shi, Jie Yin, Stephen Tierney, and Junbin Gao. SirenFNO: Efficient and full frequency learning of Fourier neural operators. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence*, pp. 4804–4812, 2026. doi: 10.24963/ijcai.2026/535.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473, 2020.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53681–53705, 2024.

# A EXPERIMENTAL DETAILS

## A.1 HEAT TRANSPORT

**Reference solutions and sampling.** Reference solutions are computed in float64 on a $512 \times 512$ grid and sampled on the common $256 \times 256$ prediction grid. The training set includes the two sweep endpoints and stratified interior samples; validation and test configurations are sampled independently within the same parameter interval. The targets contain both temperature fields over the entire square.

The coupling is $\rho = 0.5$, and diffusivity is one away from the ring wall. The wall has width $0.32$ at half contrast and a $C^4$ transition of width $0.20$ at each edge. For translation, the center is $(\pi - 1.4 + 2.8s, \pi)$, the radius is $0.95$, and wall diffusivity is $10^{-3}$. Two Gaussian sources with standard deviation $0.35$ are centered at $(\pi - 0.65, \pi - 0.35)$ and $(\pi + 0.65, \pi + 0.35)$. For expansion, the center is fixed at $(\pi - 0.45, \pi - 0.25)$, the radius is $0.38 + 1.32s$, and wall diffusivity is $0.1$; a single Gaussian source of standard deviation $0.50$ is centered at $(\pi, \pi)$. In both families, $s \in [0, 1]$ indexes independently solved steady configurations. The second field's source pattern is shifted left by $\pi/2$. Each source field has background one and total excess power $7 \cdot 2\pi \cdot 0.35^2$, fixed throughout its geometry sweep.

**Architecture and initialization.** All four models use three Fourier blocks of width 16 with $6 \times 6$ retained modes. The spatial networks have three sine layers of width 16 and frequency multiplier one. The pointwise readout has hidden width 128, increased to 284 for FNO to match the overall model size. Table 1 reports the number of trainable real scalar parameters.

In our model, source and geometry entries supply $z_{\mathrm{sp}}$; propagation and coupling entries supply $z_{\mathrm{op}}$. Within each ring family, only geometry varies, so the spatial code changes across configurations while the Fourier-stage code is shared. Static Spatial retains the same direct-input and Fourier-conditioning paths. Modulation heads start at neutral values, and the added spatial projection starts at zero.

**Training and whole-field metrics.** For heat transport, each model is trained for 32,000 AdamW updates with batch size 16, weight decay $10^{-5}$, and gradient clipping at norm one. The learning rate warms up to $10^{-3}$ over 800 updates and then follows a cosine schedule to $10^{-5}$. We evaluate validation error every 400 updates and retain its minimum. Training uses float32 with complex64 Fourier weights; metric accumulation uses float64. All models use the same data split and output normalization. The readout predicts scaled field fluctuations, and $\mathcal{R}_q$ restores the spatial mean implied by the forcing and coupling balance. Channel scales are the root-mean-square training-field fluctuations about their spatial means.

We report scaled MSE, averaged over grid points, channels, and test cases after dividing each channel's error by its training scale. Physical relative $L^2$ is computed for each case using the joint norm of the two temperature fields and then averaged across cases. Both families use training seeds 17, 29, and 43, with means and sample standard deviations over seeds.

**Localized and response metrics.** The material-transition region consists of points where the wall's smooth occupancy lies between $0.1$ and $0.9$. The hotspot region contains the highest $5\%$ of reference values in each field. Both use the same channel scaling as the whole-field MSE and average each case equally. These two regions ask complementary questions: whether the model resolves the material transition and whether it reconstructs the concentrated thermal response.

At nine evenly spaced states $s = 0, 1/8, \ldots, 1$, we compare each adjacent predicted field difference with the corresponding reference difference. Response relative $L^2$ is the norm of their discrepancy divided by the norm of the reference difference, averaged over the eight intervals.

## A.2 HOLLOW CATHODE DISCHARGE

**Grid, split, and preprocessing.** The radial–axial grid covers $r \in [0, 15]$ mm and $z \in [0, 85]$ mm. Split seed 0 fixes the 101/12/12 training/validation/test cases used by all architectures and training seeds.

Inputs are standardized. Density targets are clipped below $10^{-30}$ and transformed by $\log_{10}$; potential remains linear. All four output channels are then standardized separately. Every model uses nor-

malization statistics fitted over all 125 cases before splitting. Losses and metrics include the entire grid, including low-density material regions. Field visualizations invert the preprocessing to recover physical quantities.

**Architecture.** Our HCD predictor uses a shared 128-dimensional embedding of all three operating parameters to condition a three-layer, width-128 SIREN and two width-34 Fourier stacks retaining $(85, 8)$ axial–radial modes. The shared embedding supplies both spatial and Fourier conditioning. The SIREN receives normalized coordinates and Gaussian Fourier features, with a frequency multiplier of 30. Its modulation heads apply condition-dependent scales and phase shifts inside the sine argument. The first two-block Fourier stack predicts four fields through separate pointwise heads. A conditioned $4 \times 4$ mixing matrix and bias couple these fields, and the second two-block stack provides a residual correction. Each modulation site has an independent head receiving the complete shared code. The fixed cathode layout is learned from the supervised fields. Each HCD Fourier block applies feature modulation after summing its spectral and pointwise paths and before the GELU activation.

FNO uses four width-34 Fourier blocks with $(85, 8)$ modes; Transolver (Wu et al., 2024) uses four width-240 layers, eight attention heads, and 32 slices. Both receive the complete parameter embedding at the input and use layerwise feature modulation. DeepONet (Lu et al., 2021) combines a parameter branch with a coordinate trunk; both have width 640 and depth four, with 1,440 basis coefficients per output field.

**Training and evaluation.** All four architectures use AdamW with learning rate $10^{-4}$, zero weight decay, batch size 16, and a 3,000-epoch budget. For each case, the field loss averages the four standardized-space relative $L_2$ errors and adds 0.1 times the joint relative Frobenius error; case losses are then averaged equally. Each denominator includes $10^{-8}$. Training adds 0.25 times the mean squared adjacent-grid difference of the predicted metastable field along the radial axis. Validation selects the lowest field loss without this penalty.

For the fieldwise relative $L_2$ in Table 3, squared errors and targets are pooled over test cases and grid points before taking the norm ratio. We report means and sample standard deviations over the five training seeds. Figure 5 uses the separate seed-1000 run.

## B Localized prediction structure

The whole-field and regional scores in Tables 1 and 2 establish an aggregate gain. The following views ask what that gain looks like within a field. We move from the hotspot's steep flanks to a non-peak material boundary and then to the smooth exterior, separating peak reconstruction, transition shape, and small background residuals. All views use the first held-out translation case from seed 17, at $s = 0.00252$, as in Figure 4.

### B.1 Errors around the temperature maximum

Figure 6 examines the horizontal line through the reference maximum in Figure 4. The profiles reproduce the same sharp temperature peak, so their differences are difficult to distinguish on the field's value scale. Signed residuals separate the smaller surrounding deviations from the larger excursions near the steep peak flanks. Our model has the lowest median absolute error along this line: $0.0063$, compared with $0.0112$ for FiLM FNO and $0.0147$ for Static Spatial. Its largest pointwise excursion is nevertheless higher. The profile thus distinguishes reduced typical error from peak error; the two-dimensional window in Figure 4 shows how these residuals are distributed around the hotspot.

### B.2 Boundary structure away from the maximum

The hotspot combines large values with steep local variation. To examine transition shape away from that peak, we turn to the upper rim. Its temperature changes across the material wall, but the window excludes the global maximum in both fields. Figure 7 shows a smooth, curved transition in the target. The baseline error maps contain alternating high- and low-error bands along and around this transition. Our model reduces their amplitude in both fields. The improvement therefore extends to the shape of the transition away from the dominant hotspot.

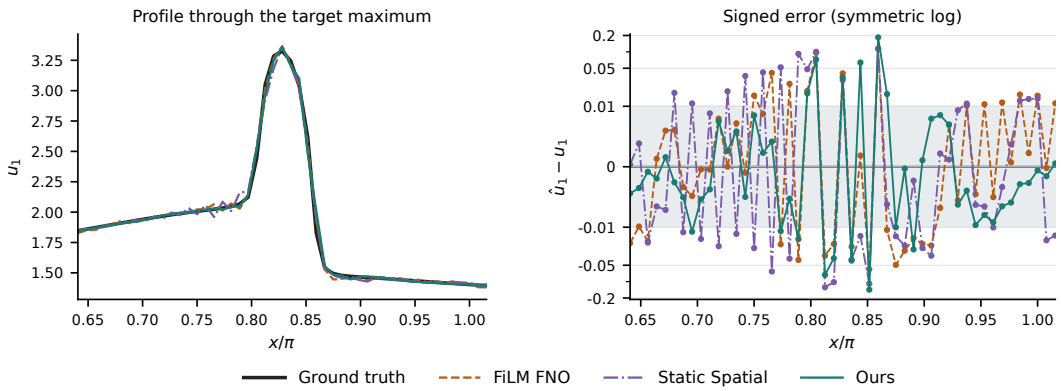


Figure 6: **Field profiles and signed prediction errors.** The horizontal line passes through the reference maximum at $y/\pi = 0.875$ in the held-out configuration of Figure 4. Left: reference and predicted values of $u_1$. Right: signed errors at the original grid nodes, connected without smoothing. The shared symmetric-logarithmic vertical axis is linear within the shaded interval $[-0.01, 0.01]$ and logarithmic outside it, making small residuals visible while retaining every error peak. The horizontal black line marks zero error.

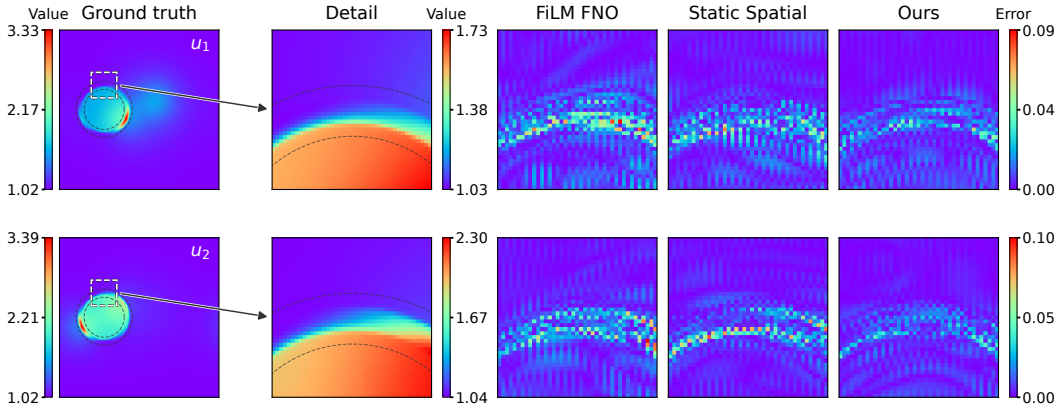


Figure 7: **Prediction errors at a non-peak material boundary.** Top: $u_1$; bottom: $u_2$. The dashed box in the full target marks a $41 \times 41$-node window centered on the upper rim; the arrow leads to its target detail. The three right panels show absolute prediction errors. Dashed curves mark the ring's inner and outer half-contrast edges. Full targets use their whole-field ranges, while details use local value ranges. Each row's three error maps share one linear, zero-based scale covering every displayed error. Pixels retain the original grid values.

Table 4 quantifies error magnitude and spatial variation over equal-sized, geometry-selected windows. Alongside mean absolute error (MAE), we report $R_x$, the root-mean-square of differences between signed errors at horizontally adjacent grid nodes. This quantity is expressed in field units without division by grid spacing. At the upper rim, our model reduces both measures in both fields: the smaller errors also vary less from node to node.

### B.3 Smooth continuation outside the ring

Moving outside the material transition provides a complementary test. The target is low in amplitude and changes gradually, so the relevant structure is a smooth continuation of the field rather than a sharp peak or interface. Figure 8 places the entire window beyond the ring's outer material transition. The local value scale reveals the gentle target gradient, while the error maps expose residual bands that are difficult to see on the whole-field scale. Our model has lower MAE and lower adjacent-node residual variation in both fields. Absolute errors here are much smaller than at the rim, as indicated

by the separate error scales. This comparison shows that the local gain includes the weak background surrounding the material object.

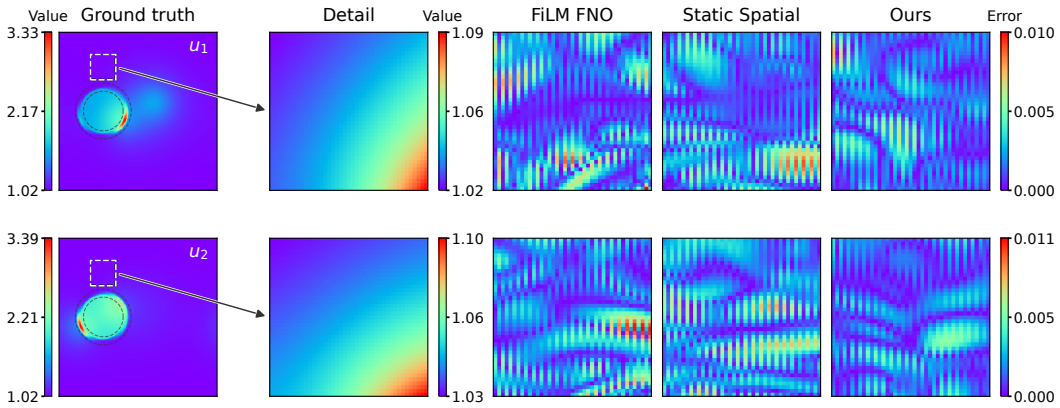


Figure 8: **Prediction errors in a smooth, low-value exterior region.** The two rows show $u_1$ and $u_2$ for the same held-out case as Figure 7. The $41 \times 41$-node window lies above the outer material transition. The layout and color conventions are unchanged; error limits are set from this window. The locally scaled target detail resolves its gentle variation, while the common error scale within each row permits direct comparison of all three models.

Table 4: **Error magnitude and variation across local regions.** MAE and adjacent-node residual variation $R_x$, both multiplied by $10^3$, for the held-out case shown above. Lower is better. All windows contain $41 \times 41$ original nodes. Rim windows are centered on the upper, left, or lower wall; the interior window is centered in the cavity. These centers depend only on geometry. The left-rim window for $u_2$ also contains its temperature maximum. Bold marks the lowest value within each three-model comparison.

| | | MAE $\times 10^3$ | | | $R_x \times 10^3$ | | |
|---|---|---|---|---|---|---|---|
| Region | Field | FiLM FNO | Static | Ours | FiLM FNO | Static | Ours |
| Upper rim | $u_1$ | 8.58 | 7.18 | **5.70** | 21.06 | 15.67 | **11.04** |
| | $u_2$ | 8.15 | 6.76 | **5.35** | 19.86 | 17.95 | **10.34** |
| Left rim | $u_1$ | 11.19 | 12.92 | **8.51** | 22.21 | 27.57 | **19.80** |
| | $u_2$ | 12.63 | 12.10 | **8.71** | 32.97 | 34.61 | **23.99** |
| Lower rim | $u_1$ | 13.25 | 8.94 | **8.07** | 24.98 | 17.86 | **15.62** |
| | $u_2$ | 10.63 | 7.81 | **6.92** | 21.02 | **14.83** | 16.66 |
| Exterior | $u_1$ | 1.92 | 1.98 | **1.60** | 2.82 | 2.69 | **2.41** |
| | $u_2$ | 2.23 | 2.43 | **1.61** | 3.48 | 2.85 | **1.87** |
| Cavity | $u_1$ | 10.54 | 7.47 | **6.35** | 18.69 | 14.05 | **10.78** |
| | $u_2$ | 7.51 | **4.66** | 5.42 | 13.29 | **5.45** | 9.28 |

Across these ten region–field pairs, our model has the lowest MAE in nine. The cavity of $u_2$ favors Static Spatial; at its lower rim, our model has smaller MAE but greater residual variation than Static Spatial. The clearest shared pattern is at the upper boundary and exterior, where both error magnitude and residual variation decrease in both fields. Together with the hotspot view, this locates the gain in both the material-associated structure and its smooth continuation. The comparison with Static Spatial ties these local differences to conditioning the coordinate representation. Whether the same benefit persists along a different geometry path is examined next in Appendix C.

# C SPATIAL CONDITIONING ACROSS RING MOTIONS

The translation views locate the spatial improvement. We now ask how it depends on the geometry path: first through local and response errors for the original expansion, then through an expansion with sources and material contrast matched to translation.

## C.1 LOCALIZED ACCURACY IN THE SINGLE-SOURCE EXPANSION

Table 5 extends the single-source expansion results in Table 1 to material transitions, hotspots, and changes between radii, using the definitions in Appendix A.1.

Table 5: **Localized accuracy and geometry response for single-source expansion.** Mean $\pm$ sample standard deviation over seeds 17, 29, and 43. Transition and hotspot scaled MSE use the held-out test set; response relative $L^2$ uses adjacent differences along nine fixed geometry states. Lower is better.

| Model | Transition MSE $\times 10^{-5}$ | Hotspot MSE $\times 10^{-5}$ | Response $L^2$ $\times 10^{-2}$ |
|---|---|---|---|
| FNO | $10.78 \pm 2.77$ | $8.42 \pm 2.61$ | $3.348 \pm 0.521$ |
| FiLM FNO | $7.93 \pm 1.27$ | $6.68 \pm 0.49$ | $2.724 \pm 0.190$ |
| Static Spatial | $6.76 \pm 0.93$ | $5.95 \pm 1.01$ | $2.493 \pm 0.152$ |
| Ours | $\mathbf{5.87} \pm 1.03$ | $\mathbf{5.29} \pm 0.82$ | $\mathbf{2.421} \pm 0.272$ |

Our model has the lowest mean error on all three measures. Relative to Static Spatial, the transition and hotspot MSE improvements are larger than the whole-field MSE improvement. Thus, a modest domain-averaged gain includes more pronounced refinement of the localized thermal structure. The response errors of the two spatial models are closer: a shared representation already captures much of the field change along this radial sweep.

## C.2 GEOMETRY PATHS UNDER MATCHED PHYSICAL CONDITIONS

The two main ring families differ in their heat sources and material contrast as well as their geometry paths. To compare the paths under matched physical conditions, we also expand a centered ring using the translation task's two sources and wall diffusivity $10^{-3}$. The radius is $0.38 + 1.32s$; source locations and power, wall width and transition, sampled $s$ values, model sizes, and training and selection budgets match the translating-ring comparison. The split contains 128/64/128 training/validation/test configurations. Only the geometry path changes from a translating ring of fixed radius to centered expansion.

Table 6: **Two-source expansion with matched physical and training settings.** Mean $\pm$ sample standard deviation over seeds 17, 29, and 43. The first four metrics use held-out configurations; response relative $L^2$ uses the nine fixed geometry states. Lower is better.

| Metric | Static Spatial | Ours |
|---|---|---|
| Scaled MSE ($\times 10^{-4}$) | $1.094 \pm 0.155$ | $\mathbf{0.989} \pm 0.138$ |
| Physical relative $L^2$ ($\times 10^{-3}$) | $1.952 \pm 0.133$ | $\mathbf{1.839} \pm 0.135$ |
| Transition MSE ($\times 10^{-4}$) | $3.620 \pm 0.555$ | $\mathbf{2.942} \pm 0.100$ |
| Hotspot MSE ($\times 10^{-4}$) | $7.805 \pm 0.698$ | $\mathbf{6.950} \pm 0.662$ |
| Response $L^2$ ($\times 10^{-2}$) | $3.591 \pm 0.241$ | $\mathbf{3.347} \pm 0.306$ |

Our model improves scaled MSE and physical relative $L^2$ in all three seeds, with lower mean transition, hotspot, and response errors (Table 6). Comparing each task's improvement relative to its Static Spatial control, conditioning remains more beneficial for translation than for centered expansion. The difference therefore persists when the heat sources and material contrast are matched. Expansion moves the boundary radially around a fixed center; translation moves the ring relative to the heaters. These results support a larger role for conditional spatial features when the material-associated structure changes position, while showing a smaller positive gain along the expanding path.

# D HOLLOW CATHODE DISCHARGE

## D.1 REFERENCE FIELD STRUCTURE

The ring experiments vary material layout; HCD varies operating conditions within a fixed device. Figure 9 compares two reference solutions from the pressure, RF-voltage, and SEEC sweep. Common physical scales reveal changes in field magnitude, while peak normalization exposes changes in the relative distribution between the bore and chamber. The predictor must therefore learn operating-dependent spatial profiles as well as field amplitudes.

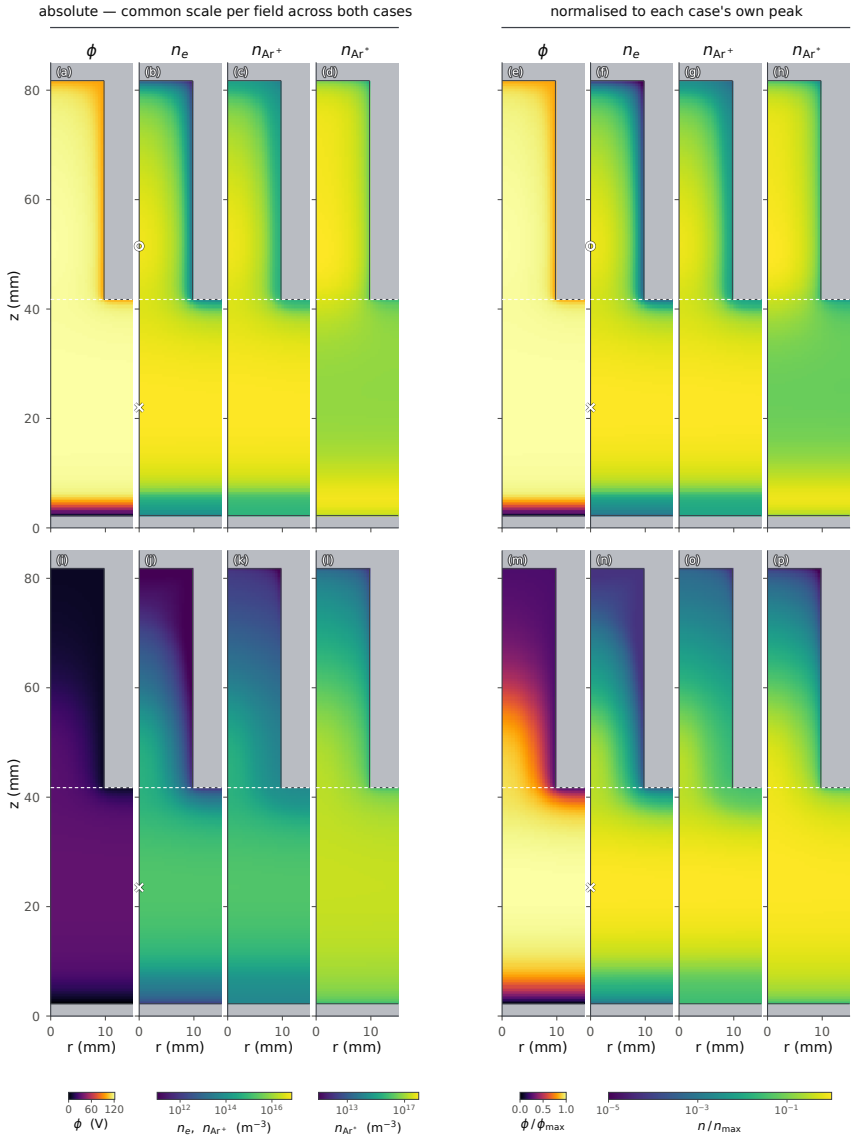


Figure 9: **Reference plasma fields in a fixed cathode geometry.** Two operating conditions from the pressure–voltage–SEEC dataset. The columns show cycle-averaged potential, electron density, argon-ion density, and metastable density. Left: common physical scales across the two cases. Right: each field divided by its own case-specific peak. Potential uses linear scales and densities logarithmic scales. Grey denotes electrode material; the dashed line marks the bore mouth. Both rows use the same geometry and mesh, and all panels preserve the physical aspect ratio.

### D.2 VALIDATION TRAJECTORIES

The reference maps show the field variation to be learned. Figure 10 then follows the four architectures in Table 3 during training. Our model and FNO follow similar learning trends, with our model maintaining a lower mean validation loss over much of training. Transolver exhibits larger epoch-to-epoch fluctuations, while DeepONet descends more gradually. The sustained separation between our model and FNO connects the held-out accuracy advantage to their validation behavior throughout much of optimization.

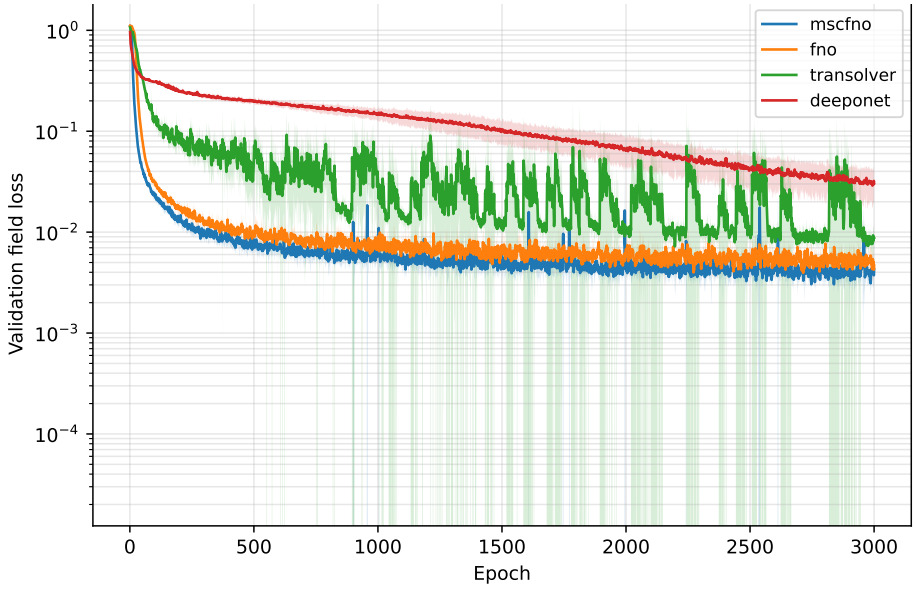


Figure 10: **Validation learning curves for fixed-geometry HCD.** Lines and shaded bands show the mean and standard deviation over training seeds 0–4 on the same data split. The vertical axis is logarithmic. The source legend's `mscfno` denotes our model. All runs use the common 3,000-epoch budget; test results in Table 3 use the minimum-validation checkpoint of each run.

The mean selected epochs are 2,698 for our model, 2,713 for FNO, 2,864 for Transolver, and 2,978 for DeepONet. The curves and selected epochs characterize performance under this shared training budget.