Ruprecht Karl University of Heidelberg

Department of Mathematics and Computer Science

Bachelor-Thesis

Operator Learning Using Random Features for Phase-Field Tumor Models

Name: Jan Jakob

Matriculation number: 4065027

Supervisor: Prof. Vincent Heuveline

Submission date: 23.08.2025

Abstract

In mathematical oncology, tumor evolution can be described by phase-field equations that model the spatio-temporal dynamics of cell populations together with various biophysical mechanisms. Recent surveys outline multiple families of such models, ranging from the prototypical Cahn-Hilliard equation to multiphase formulations incorporating nonlocal cell adhesion, stochastic fluctuations, and mixed-dimensional couplings to vascular networks. All these partial differential equation systems are highly nonlinear, high-dimensional, and computationally demanding to solve with classical numerical schemes, motivating the development of efficient surrogate models. Supervised operator learning has emerged as a promising data-driven framework for approximating solution operators between infinite-dimensional function spaces. Among theses approaches, the random feature method for operator learning provides a well-balanced approach, combining scalability with rigorous theoretical guarantees: It frames training as a convex quadratic optimization problem, admits convergence and complexity bounds, and can be interpreted as a low-rank approximation of operator-valued kernel ridge regression with close connections to Gaussian process regression. Within this framework, we construct random-features for the Cahn-Hilliard equation as the prototypical model for tumor growth, demonstrating accurate prediction of its solution operator and proving convergence of the method in this setting. Building on these results, we try to extend the framework to the four-species phase-field tumor growth model. The random feature method has seen limited application to partial differential equations, with only a few isolated examples in the literature, such as the Burgers equation or Darcy flow. Our work provides new insights into the method's capabilities and extends its applicability to more complex partial differential equations. Our numerical experiments illustrate that random-feature operator learning provides a scalable, transferable, and discretization-invariant surrogate for the Cahn-Hilliard equation, offering a promising computational tool for mathematical oncology.

Zusammenfassung

In der mathematischen Onkologie lassen sich Tumorwachstumsvorgänge mithilfe von Phasenfeldgleichungen beschreiben, die die raum-zeitliche Dynamik von Zellpopulationen unter Einbezug verschiedener biophysikalischer Mechanismen modellieren. In der Literatur findet sich mittlerweile eine Vielzahl solcher Modelle - angefangen bei der prototypischen Cahn-Hilliard Gleichung bis hin zu Mehrphasenmodellen, die Effekte wie nichtlokale Zelladhäsion, stochastische Schwankungen oder die Kopplung an vaskuläre Netzwerke berücksichtigen. All diese Systeme partieller Differentialgleichungen haben gemein, dass sie hochdimensional und hochgradig nichtlinear sind, weshalb sie mit klassischen numerischen Verfahren nur schwer effizient lösbar sind. Dies motiviert die Entwicklung leistungsfähiger Surrogatmodelle. Supervised Operator-Learning hat sich in diesem Kontext als vielversprechender datengetriebener Ansatz etabliert. Es ermöglicht, Lösungsoperatoren zwischen unendlichdimensionalen Funktionsräumen effizient zu approximieren. Unter den verschiedenen Methoden zeigt sich die Random-Feature Methode als besonders ausgewogen: Sie verbindet effizientes Training mit theoretischen Fehler-Abschätzungen, indem sie den Trainingsvorgang als konvexes quadratisches Optimierungsproblem formuliert, Konvergenz- und Komplexitätsabschätzungen erlaubt und sich als niedrigrangige Approximation operatorwertiger Kernel-Ridge-Regression interpretieren lässt. Auf Grundlage dieser Methode konstruieren wir in dieser Arbeit Random Feature Funktionen für die Cahn-Hilliard-Gleichung als prototypisches Modell für Tumorwachstum. Wir zeigen, dass sich der Anfangsbedingung-zu-Lösungsoperator für die Gleichung damit präzise vorhersagen lässt und zeigen numerisch die Konvergenz der Methode in diesem Setting. Aufbauend auf diesen Ergebnissen versuchen wir, den Ansatz auf das sogenannte four-species model für Tumorwachstum zu erweitern. Bisher wurde die Random-Feature-Methode nur auf wenige, ausgewählte partielle Differentialgleichungen angewendet, etwa auf die eindimensionale Burgers-Gleichung oder auf Darcy Flow. Unsere Arbeit liefert neue Einblicke in die Leistungsfähigkeit der Methode und erweitert ihr Anwendungsspektrum auf komplexere Differentialgleichungen. Unsere numerischen Experimente demonstrieren, dass Random-Feature-Operator-Lernen einen skalierbaren, übertragbaren und diskretisierungsinvarianten Surrogatansatz für die Cahn-Hilliard-Gleichung bietet – und damit als ein vielversprechendes Werkzeug für die rechnerische Modellierung in der mathematischen Onkologie betrachtet werden kann.

.

Contents

1 Introduction

1.1 Overview

Cancer is one of the leading causes of death worldwide, and its incidence is projected to rise significantly in the coming decades. In 2020, an estimated 19.3 million new cancer cases and 9.96 million cancer-related deaths were reported globally. By 2040, these numbers are expected to increase to 30.2 million new cases and 16.3 million deaths per year [Sun+21]. Despite decades of research, each tumor remains unique - driven by a complex interplay of biological, genetic, and environmental factors - and there is still no universally effective treatment. As a result, mathematical models of tumor growth have become increasingly important in oncology. They aim to describe tumor progression in a mechanistic and predictive way, offering potential tools for prognosis, therapy planning, and personalized treatment strategies [Fri23]. Before numerical strategies can be developed, it is essential to ensure that the underlying model is mathematically well-posed; a comprehensive survey of the well posedness for different tumor growth models can be found in [Fri23].

Among the modeling approaches, phase-field models have emerged as particularly powerful tools for simulating tumor dynamics. These models represent tumors as multiphase systems consisting of multiple interacting species such as tumor cells, healthy tissue, nutrients, and extracellular matrix. Their mathematical formulation often involves higher-order partial differential equations (PDEs), such as the Cahn–Hilliard equation, which can naturally incorporate nonlinearities, spatial-temporal couplings, and nonlocal effects like long-range cell-cell interactions and memory phenomena [Fri23]. These features make phase-field models biologically realistic and versatile. However, their simulation is computationally demanding. This is particularly problematic when the model needs to be evaluated repeatedly - for instance, in uncertainty quantification, optimal treatment planning, or real-time decision support.

To overcome these challenges, data-driven surrogate models have gained increasing attention. One promising direction is operator learning, a paradigm designed to approximate solution operators of PDEs [KLS24]. Unlike classical regression methods, operator learning seeks to find approximations for mappings between infinite-dimensional input and output spaces, such as those arising from the evolution of tumor fields over time. Within the field of operator learning, the random feature method (RFM) offers a particularly appealing trade-off between simplicity, scalability, and theoretical rigor [NS24]. Instead of learning a deep neural network or using a computationally expensive kernel method, the RFM approximates the operator as a linear combination of fixed, randomly sampled nonlinear features. Only the linear output weights are trained, which makes the optimization process convex and computationally efficient [KLS24]. Moreover, RFM-based surrogates are discretization-invariant, allowing them to generalize across mesh resolutions once trained. This enables, for instance, studies of parameter variations or uncertainty quantification to be performed without the need to solve the full PDE system each time.

However, the performance of the RFM in learning operators such as the initial-condition-to-solution mapping at a certain timestep is heavily influenced by the design of the concrete random features for the problem considered. The selection of feature functions and the underlying probability distributions must be carefully adapted to the structure of the specific PDE and the characteristics of the solution at the target timestep. In [NS24], the RFM has been successfully applied to benchmark problems such as the one-dimensional viscous Burgers' equation and the two dimensional Darcy flow. However, its application to complex biological systems - particularly phase-field tumor models - has not yet been explored to the best of our knowledge.

1.2 Contributions.

Our primary contributions in this work are now listed.

(C1) We apply the RFM to the stationary, linearized Cahn-Hilliard equation on a circular disk, providing an explicit characterization of its kernel and the corresponding reproducing kernel Hilbert space.

(C2) We successfully apply the RFM to learn the initial-condition-to-solution operator for the Cahn-Hilliard equation by specifying the selected random features for an early timestep, where spinodal decomposition just has started. Furthermore, we perform numerical experiments that demonstrate two mesh-independent approximation properties that are built into the proposed methodology: invariance of relative error to mesh resolution and evaluation ability on any mesh resolution.

(C3) We develop a highly parallelized implementation of the RFM that can be run on multiple graphics processing units (GPUs). Our code is capable of computing an approximation of the solution operator of a PDE within minutes, even for a great number of random features and training pairs.

2 Theoretical Background

This section provides an overview of both the RFM and mathematical models used to describe tumor growth. In section 2.1, we first introduce the mathematical preliminaries of kernel methods, and then outline the RFM for operator learning based on [NS24]. Section 2.2 mainly follows [Fri23] by giving an overview of the most fundamental models for tumor growth, in particular the four-species model and the Cahn-Hilliard equation as the prototypical model for tumor growth, which will be the model we apply the RFM to later in sections 3 and 4.

2.1 Random Feature Method for Operator Learning

Many modern machine learning methods rely on the concept of a kernel to implicitly map data into high-dimensional feature spaces. The method, also known as the *kernel trick*, is widely used in applications such as classification, regression, and dimensionality reduction. Let $\mathbf{x}, \mathbf{x}' \in X \subset \mathbb{R}^d,\ d \in \mathbb{N}$ be samples and $\varphi : X \to \mathcal{H}$ be a feature map transforming samples to a high-dimensional (even infinite-dimensional) Hilbert space $\mathcal{H}$, where the mapped data can be learned by a linear model. In practice, the explicit expression of the feature map φ is not necessarily known and instead of computing the map $\varphi(x)$ explicitly, it is often sufficient to evaluate the inner product between two images $\varphi(x)$ and $\varphi(x')$ directly.

Kernel Functions and Reproducing Kernel Hilbert Spaces A *kernel* is a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ for which there exists a Hilbert space $\mathcal{H}$ and a mapping $\varphi : \mathcal{X} \to \mathcal{H}$ such that

$$k(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}}. \tag{1}$$

$\mathcal{X}$ does not necessarily be equipped with a vector structure. A central example of a kernel is the *Gaussian kernel*, defined on $\mathcal{X} = \mathbb{R}^d$ by

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\gamma^2} \right), \tag{2}$$

where $\gamma > 0$ controls the width of the kernel. This kernel corresponds to an inner product in an infinite-dimensional feature space $\mathcal{H}$, where the implicit map φ is described by infinitely many basis functions.

To systematically study functions defined via kernels, the concept of a *reproducing kernel Hilbert space* (RKHS) is introduced. An RKHS is a Hilbert space $\mathcal{H}_k$ of functions $f : \mathcal{X} \to \mathbb{R}$ with an associated kernel k satisfying the following two properties:

- For every $x \in \mathcal{X}$, the function $k(\cdot, x)$ lies in $\mathcal{H}_k$.
- For all $f \in \mathcal{H}_k$ and $x \in \mathcal{X}$, the so-called reproducing property holds:

$$f(x) = \langle f, k(\cdot, x)\rangle_{\mathcal{H}_k}.$$

This property allows the evaluation of a function at a point to be expressed as an inner product in the Hilbert space. It forms the foundation of many learning methods that minimize a functional over an RKHS. A fundamental property of all kernels that generate an RKHS is *positive definiteness*: A kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called positive definite if it is symmetric (i.e., $k(x, x') = k(x', x)$) and for arbitrary $x_1, \ldots, x_n \in \mathcal{X}$ and $a_1, \ldots, a_n \in \mathbb{R}$ it holds that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0. \tag{3}$$

Every positive definite kernel function uniquely defines an RKHS in which the kernel acts as the reproducing kernel. Conversely, every reproducing kernel is necessarily positive definite. This one-to-one correspondence is also known as the Moore–Aronszajn theorem [WS21].

Due to the ease of computing the inner product, the kernel method is effective for nonlinear learning problems with a wide range of successful applications. However, the kernel method does not scale well to extremely large datasets. For example, given n training samples, kernel regression requires $\mathcal{O}(n^3)$ training time and $\mathcal{O}(n^2)$ storage for the kernel matrix, which is often computationally infeasible when n is large. Using random features is one of the most popular techniques to overcome the computational challenges of the kernel method. The theoretical foundation of random (Fourier) features builds on Bochner's theorem, a central result from harmonic analysis.

From Bochner's Theorem to Scalar Regression *Bochner's theorem* [Lia24] states, that any continuous, shift invariant kernel $k(x, y) = k(x - y)$ on $\mathbb{R}^d,\ d \in \mathbb{N}$ is positive definite if and only if k is the Fourier transform of a probability measure ρ:

$$k(\mathbf{x}, \mathbf{x}') = \int_{\mathbb{R}^d} \exp\left(i\langle \boldsymbol{\omega}, \mathbf{x} - \mathbf{x}'\rangle\right) d\rho(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \exp\left(i\langle \boldsymbol{\omega}, \mathbf{x}\rangle\right) \overline{\exp\left(i\langle \boldsymbol{\omega}, \mathbf{x}'\rangle\right)}\, d\rho(\boldsymbol{\omega}). \tag{4}$$

Following Rahimi and Recht [RR07], we can approximate k using a Monte Carlo sampling: We draw N independent and identically distributed samples $\{\boldsymbol{\omega}_k\}_{k\in[N]}$ from $\rho(\boldsymbol{\omega})$, and define a corresponding *random Fourier feature map* $\varphi : \mathbb{R}^d \to \mathbb{C}^N$ as

$$\varphi(\mathbf{x}; \boldsymbol{\omega}) = \frac{1}{\sqrt{N}}\left[\exp(i\langle \boldsymbol{\omega}_1, \mathbf{x}\rangle), \ldots, \exp(i\langle \boldsymbol{\omega}_N, \mathbf{x}\rangle)\right]^T \in \mathbb{C}^N. \tag{5}$$

By using 5, we are able to define a finite-dimensional kernel function $\hat{k}(\mathbf{x}, \mathbf{x}') : X \times X \to \mathbb{R}$ as

$$\hat{k}(\mathbf{x}, \mathbf{x}') := \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}')\rangle_{\mathbb{C}^N}. \tag{6}$$

We can also use *random cosine features* to approximate any shift-invariant kernel [Lia24]: Let $\boldsymbol{\omega} \sim \rho(\boldsymbol{\omega})$ and $b \sim \text{Uniform}[-\pi, \pi]$. Using N i.i.d. samples $\{(\boldsymbol{\omega}_k, b_k)\}_{k=1}^N$, we define the feature map:

$$\phi(\mathbf{x}; \boldsymbol{\omega}) = \frac{1}{\sqrt{N}}\left[\cos(\langle \boldsymbol{\omega}_1, \mathbf{x}\rangle + b_1), \ldots, \cos(\langle \boldsymbol{\omega}_N, \mathbf{x}\rangle + b_N)\right]^T \in \mathbb{R}^N, \tag{7}$$

and approximate the kernel k this time by

$$\hat{k}(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathbb{R}^N}. \tag{8}$$

This reduces kernel computation to a low-dimensional inner product, enabling scalable learning with $\mathcal{O}(dN^2)$ time and $\mathcal{O}(dN)$ storage.

From Scalar Regression to Operator Learning Operator learning generalizes function regression to the setting where both input and output spaces are infinite-dimensional - typically functions defined on a continuous domain. In operator learning, the goal is to learn an approximation F for the map $F^\dagger : \mathcal{X} \to \mathcal{Y}$ between the Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, based on a finite training dataset $(a_i, y_i)_{i=1}^n \subset \mathcal{X} \times \mathcal{Y}$ of input-output pairs. Here, a_i represents the coefficient, boundary, or initial data of the PDE our surrogate model should approximate, and $y_i = F^\dagger(a_i)$ represents the corresponding PDE solution. The inputs $(a_i)_{i=1}^n$ are drawn from a data distribution ν on $\mathcal{X}$, which may either be a deliberate modeling choice at the chosen discretization or an unknown quantity.

To learn such operators in a nonparametric and nonintrusive fashion, Nelsen and Stuart [NS24] extended the random feature methodology to the function-valued setting. In [NS24], they introduced a random feature map $\phi : \mathcal{X} \times \Theta \to \mathcal{Y}$, where $\theta \in \Theta$ is drawn from a probability measure μ on Θ. This induces an operator-valued kernel

$$k_\mu(a, a') := \mathbb{E}^{\theta \sim \mu}[\phi(a; \theta) \otimes \phi(a'; \theta)], \tag{9}$$

that defines a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_{k_\mu} \subset L^2_\nu(\mathcal{X}; \mathcal{Y})$ of operator-valued functions in the following manner:

$$\mathcal{H}_{k_\mu} := \{\mathbb{E}^{\mu \sim \mu}[c(\theta)\phi(\cdot; \theta)] : c \in L^2_\mu(\Theta; \mathbb{R})\}. \tag{10}$$

Eq. 10 indicates that the space $\mathcal{H}_{k_\mu}$, defined by the feature pair (ϕ, μ) and coefficient functions $c \in L^2_\mu(\Theta; \mathbb{R})$, provides a very general nonparametric class of operators for approximation. However, directly working with this space is typically impractical, since implementing estimators in $\mathcal{H}_{k_\mu}$ leads to extremely high computational costs unless additional strong assumptions are made about the structure of (ϕ, μ). To overcome this limitation, Nelsen and Stuart [NS24] adopted a parametric approximation: instead of representing the operator with an infinite-dimensional feature space, they considered a finite number of randomly sampled features. According to result 2.6. in [NS24], under the assumption of $\phi \in L^2_{\nu \times \mu}(\mathcal{X} \times \Theta; \mathcal{Y})$ and linear independence of the random features $\{\phi(\cdot; \theta_j)\}_{j=1}^m$ in $L^2_\nu(\mathcal{X}; \mathcal{Y})$, the RKHS $\mathcal{H}_{k^{(m)}}$ is equal to the linear span of $\{\phi(\cdot; \theta_j)\}_{j=1}^m$. Here, $k^{(m)}$ is the empirical approximation to k_μ, that is,

$$k^{(m)}(a, a') = \mathbb{E}^{\theta \sim \mu^{(m)}}\left[\varphi(a; \theta) \otimes \varphi(a'; \theta)\right] = \frac{1}{m} \sum_{j=1}^m \varphi(a; \theta_j) \otimes \varphi(a'; \theta_j), \tag{11}$$

with $\mu^{(m)}$ denoting the approximation of the measure μ by the empirical measure

$$\mu^{(m)} := \frac{1}{m} \sum_{j=1}^m \delta_{\theta_j}, \text{ where } \theta_j \overset{\text{iid}}{\sim} \mu. \tag{12}$$

δ_{θ_j} in eq. 12 denotes the Dirac measure. By applying a Monte Carlo sampling approach to the elements in the RKHS 10 and replacing the probability measure μ by the empirical measure $\mu^{(m)}$ in eq. 12, we obtain

$$\frac{1}{m} \sum_{j=1}^m c(\theta_j)\, \varphi(\cdot; \theta_j) \;\approx\; \mathbb{E}^{\theta \sim \nu}[c(\theta)\, \varphi(\cdot; \theta)] \quad \text{for } c \in L^2_\mu(\Theta; \mathbb{R})\,. \tag{13}$$

Low-rank approximations using Monte Carlo sampling achieve the classical rate $\mathcal{O}(m^{-1/2})$ in expectation and belong to the finite-dimensional RKHS $\mathcal{H}_{k^{(m)}}$. However, in the present setting, the Monte Carlo approach is not directly implementable for learning a target map $F^\dagger \in \mathcal{H}_{k_\mu}$, since the operator $F^\dagger$, the kernel k_μ, and the associated RKHS $\mathcal{H}_{k_\mu}$ are all unknown. Furthermore, in practice, $F^\dagger$ may not even belong to $\mathcal{H}_{k_\mu}$, introducing a smoothness misspecification gap that the Monte Carlo method does not address. To overcome these limitations, the RFM employs a data-driven optimization strategy to construct an estimator of $F^\dagger$ within $\mathcal{H}_{k^{(m)}}$. We formalize the RFM as follows:

Random Feature Model (RFM) Let $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \nu)$ and $(\Theta, \mathcal{B}(\Theta), \mu)$ be probability spaces with $\mathcal{X}$ and Θ finite- or infinite-dimensional real Banach spaces. Let $\mathcal{Y}$ be a real separable Hilbert space, and let $\varphi \in L^2_{\nu\otimes\mu}(\mathcal{X} \times \Theta; \mathcal{Y})$. Then the Random Feature Model is the parametric map

$$F_m : \mathcal{X} \times \mathbb{R}^m \to \mathcal{Y}, \quad (a; \alpha) \mapsto F_m(a; \alpha) := \frac{1}{m} \sum_{j=1}^{m} \alpha_j \, \varphi(a; \theta_j), \quad \theta_j \overset{\text{iid}}{\sim} \mu, \tag{14}$$

where $\alpha = (\alpha_1, \ldots, \alpha_m)^T \in \mathbb{R}^m$ is the vector of trainable coefficients. The RFM aims to approximate $F^\dagger \in \mathcal{H}_{k_\mu}$ by $F_m(\cdot; \alpha) \in \mathcal{H}_{k^{(m)}}$, with $\alpha \in \mathbb{R}^m$ determined from the empirical input-output data. The method is itself random and may be viewed as a spectral method, since the randomized family $(\{\varphi(\cdot; \theta)\})$ in the linear expansion is defined ν-almost everywhere on $\mathcal{X}$. Determining α from data obviates the necessity of explicit kernel evaluation and circumvents the limitations of the Monte Carlo prediction, relying only on the pair (φ, μ) and sample input-oustput pairs from the target operator $F^\dagger$. F_m is linear in α, but highly nonlinear in a. Similar to operator learning architectures such as neural operators and Fourier neural operators [Li+21], the RFM is a nonlinear approximation, so its output $F_m(a; \alpha)$ lies on a nonlinear manifold in $\mathcal{Y}$, unlike methods such as PCA-Net [Bha+21] or DeepONet [LJK21], which are restricted to fixed linear subspaces.

The choice of the random feature pair (φ, μ) largely determines the approximation quality as will be further demonstrated in sections 3 and 4. While most kernel-based approaches first select a kernel and then derive a random feature map, the RFM perspective proposed in [NS24] is to select (φ, μ) first, which implicitly defines the kernel via eq. 11 and avoids explicitly computing it, reducing memory and computational costs. However, this approach leaves open questions about the universality of the resulting kernels and the structure of the corresponding RKHS $\mathcal{H}_{k_\mu}$.

Training Process One of the main advantages of the RFM resides in the structural simplicity of its training procedure. Under an L^2-type loss, minimization of the empirical risk constitutes a convex optimization problem, reducible to the solution of a finite-dimensional linear system. This is in sharp contrast to deep neural operators, whose training necessitates nonconvex stochastic gradient descent, rendering both theoretical analysis and numerical implementation substantially more challenging. Crucially, the RFM formulation does not require explicit knowledge of the operator-valued kernel k_μ, relying instead solely on the associated random feature pair (φ, μ).

Let us consider n input-output pairs $\{(a_i, y_i)\}_{i=1}^n \subset \mathcal{X} \times \mathcal{Y}$ with $a_i \sim \nu$ and $y_i = F^\dagger(a_i)$. Employing a penalized least-squares loss over the RFM hypothesis class is rigorously equivalent to performing kernel ridge regression restricted to the span of the random features [NS24]. The optimal coefficient vector $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$ is determined by the $m \times m$ normal equations

$$\sum_{j=1}^{m} \left(\frac{1}{m} \sum_{i=1}^{n} \langle \varphi(a_i; \theta_l), \varphi(a_i; \theta_j) \rangle_{\mathcal{Y}} + \lambda \, \delta_{lj} \right) \alpha_j = \sum_{i=1}^{n} \langle \varphi(a_i; \theta_l), y_i \rangle_{\mathcal{Y}}, \quad l = 1, \ldots, m, \tag{15}$$

where $\lambda \geq 0$ denotes an regularization parameter and δ_{lj} is the Kronecker delta. The solution of this system fully specifies the trained RFM $F_m(\cdot;\alpha)$ as a linear functional of the output data $\{y_i\}_{i=1}^n$. While this linearity implies statistical suboptimality for certain problem classes, adaptively selecting the feature pair (φ, μ) or the regularization parameter λ based on the data, for example via cross-validation, can restore near-optimal performance.

Error Bounds The RFM is one of the few operator learning methods where rigorous convergence and complexity guarantees could be proved. Under suitable assumptions on the boundedness and measurability of the data and the feature map φ, the RFM admits rigorous generalization guarantees: In the limit where the number of features m and the number of training samples n both tend to infinity at appropriate rates, while the regularization parameter λ simultaneously tends to zero at an appropriate rate, the trained model $F_{m_l}(\cdot;\hat{\alpha}^{(l)})$ converges almost surely to the true operator $F^\dagger$ in the $L^2_\nu(\mathcal{X};\mathcal{Y})$ norm (cf. theorem 2.11 in [NS24]). Also, up to constant factors, an appropriately tuned regularization parameter $\lambda \simeq n$ and a number of random features $m \simeq n$ are sufficient to guarantee a trained RFM generalization error of order

$$n^{-1/4} \simeq m^{-1/2}$$

with high probability (cf. theorem 2.12 in [NS24]). However, this quantitative result relies on the well-specified condition $F^\dagger \in \mathcal{H}_{k_\mu}$, which is generally difficult to verify in practice. On the other hand, this allows us to draw conclusions about whether $F^\dagger \in \mathcal{H}_{k_\mu}$ by analyzing the relative error w.r.t. m, as we will see in section 4 for the Cahn-Hilliard equation.

Practical Considerations As stated in [NS24], the success of the RFM depends critically on the choice of the random feature pair (φ, μ). In practice, φ and μ are often chosen to reflect the structure of the problem. For example, φ may be constructed using Gaussian processes, trigonometric bases, or even randomized neural networks. The concrete choices of (φ, μ) for the biological models described in the following subsections are one of the main contributions of this work.

2.2 Tumor Evolution Models

The growth and evolution of tumors are complex processes involving a multitude of interacting biological phenomena like nonlocal effects, nonlinearities, and stochasticity. To capture this complexity mathematically, we often turn to continuum mixture theory. This approach allows us to model the tumor and its surrounding environment as a mixture of different constituents or species (like cancer cells, healthy cells, and nutrients), where each species is represented by a field describing its volume fraction. The interactions and dynamics of these species are governed by fundamental physical laws, primarily balance laws and constitutive laws. For a comprehensive overview of tumor growth modeling, we refer the reader to [Fri23]. In what follows, we summarize the main models relevant to our work, in particular the Cahn–Hilliard equation and the four-species model.

2.2.1 Cahn-Hilliard Equation

The Cahn–Hilliard equation can be considered as the prototypical model for tumor growth [Fri23]. It is a parabolic partial differential equation of fourth order that describes phase separation in binary mixtures, such as the separation of two components in a fluid. For example, the dynamics of spinodal decomposition is commonly modeled using the Cahn–Hilliard equation. As a phase-field equation of the diffuse-interface type, it features the fundamental characteristic that the solution is either 0 or 1, or changes smoothly from one to the other. Therefore, it is well-suited as a base model for tumor growth if the 1-phase represents the presence of tumor cells, while the 0-phase indicates their absence.

The equation can be derived from the mass conservation law. If we consider two components with concentrations ϕ_1 and ϕ_2 such that $\phi_1 + \phi_2 = 1$, their concentrations obey the mass conservation law:

$$\partial_t \phi_i = -\mathrm{div} J_i,\ i \in \{1, 2\}, \tag{16}$$

where J_i denotes the mass flux of the i-th component. We assume that the flux is conserved, i.e. $J_1 + J_2 = 0$, and reduce the equations by defining the quantities $\phi = \phi_1 - \phi_2$ and $J = J_1 - J_2$ yielding

$$\partial_t \phi = -\mathrm{div} J. \tag{17}$$

J is given by the negative of the gradient of the chemical potential μ, i.e., $J = -\nabla\mu$. A more general form incorporates a mobility function m into the flux describing microscoping interactions and yielding to $J = -m(\phi)\nabla\mu$ (cf. [Gur95]). Following [CH58], The chemical potential μ can be calculated as the Gâteaux derivative of the Ginzburg-Landau energy functional

$$\mathcal{E}(\phi) = \int_\Omega \left(\Psi(\phi) + \frac{\epsilon^2}{2} |\nabla\phi|^2 \right) \mathrm{d}x, \tag{18}$$

with ϵ expressing the interfacial width and Ψ denoting a double-well potential, e.g., the Landau potential $\Psi(\phi) = \frac{1}{4}\phi^2(1-\phi)^2$. This leads to the Cahn-Hilliard equation with concentration-dependent mobility:

$$\partial_t \phi = \nabla \cdot (m(\phi)\nabla\mu) = \nabla \cdot (m(\phi)\nabla(\Psi'(\phi) - \epsilon^2 \Delta\phi)). \tag{19}$$

The mobility function is often of the form $m(\phi) = M\phi^2(1-\phi)^2$ for some constant $M > 0$. Ψ' can also be considered as the density of the free energy of the homogeneous system and ϵ^2 as the gradient energy coefficient, which describes the energy of the concentration gradient. The scenario of constant mobility has been exhaustively examined, and well-posedness can be demonstrated through the use of sufficient assumptions, as done in [Mir19]. For this reason, the present work will consider this scenario exclusively in sections 3 and 4.

2.2.2 Multiple Constituent Model

In a more general framework, we can consider a medium with N interacting constituents. At a given place $x \in \Omega \subset \mathbb{R}^d,\ d \in \mathbb{N}$, we allow multiple mechanical and chemical species to coexist in accordance with the continuum mixture theory paradigm. We represent the volume fraction of each species by a field $\phi_\alpha,\ 1 \leq \alpha \leq N$. For the sake of convenience, we denote with $\phi_\mathbb{A} := (\phi_\alpha)_{\alpha\in\mathbb{A}}$, where $\mathbb{A} := \mathbb{CH} \uplus \mathbb{RD} \uplus \mathbb{OD}$ denotes the index set that disjointly separates between the phase-field index set $\mathbb{CH}$, the reaction-diffusion indices $\mathbb{RD}$, and the evolution indices $\mathbb{OD}$ that correspond to abstract ordinary differential equations (ODEs). Following the work of Lima et al. [LOA14; LAO15], the constituents, $\phi_\alpha, \alpha \in \mathbb{A}$, are subject to the extended mass balance law

$$\partial_t \phi_\alpha + \mathrm{div}(\phi_\alpha \mathbf{v}_\alpha) = -\,\mathrm{div}\, J_\alpha(\phi_\mathbb{A}) + S_\alpha(\phi_\mathbb{A}), \tag{20}$$

where v_α is the cell velocity, S_α is a mass source term dependent on the species, and J_α is the flux of the α-th constituent. If $\sum_{\alpha\in\mathbb{A}} S_\alpha(\phi_\mathbb{A}) = 0$ holds, we call the system closed. The flux is proportional to the negative gradient of the chemical potential, multiplied by a mobility function:

$$J_\alpha(\phi_\mathbb{A}) = -m_\alpha(\phi_\mathbb{A})\nabla\mu_\alpha, \tag{21}$$

with μ_α representing the chemical potential of the α-species, and m_α representing the mobility function, which may depend on all constituents. In [Fri23], Fritz proposed the system's energy

$$\mathcal{E}(\phi_\lambda) = \int_\Omega \left\{ \Psi(\phi_{\mathbb{CH}}) + \Phi(\phi_\mathbb{A}) + \sum_{\alpha\in\mathbb{CH}} \frac{\varepsilon_\alpha^2}{2} |\nabla\phi_\alpha|^2 + \sum_{\beta\in\mathbb{RD}} \frac{D_\beta}{2}\phi_\beta^2 \right\} \mathrm{d}x, \tag{22}$$

where ϵ_α ($\alpha \in \mathbb{CH}$) is a parameter that characterizes the thickness of the interface separating the different cell types. The function Φ is used to describe adhesion mechanisms such as chemotaxis and haptotaxis. Ψ represents a double-well potential as in the Cahn-Hilliard equation 19. Following the approach used for the Cahn–Hilliard equation, Fritz [Fri23] computed the Gâteaux derivative of the Ginzburg–Landau energy 22 with respect to the relevant constituents. This yields the corresponding chemical potentials:

$$\begin{aligned} \mu_\alpha &= \partial_{\phi_\alpha}\Psi(\phi_{\mathbb{CH}}) + \partial_{\phi_\alpha}\Phi(\phi_{\mathbb{A}}) - \varepsilon_\alpha^2\Delta\phi_\alpha, && \alpha \in \mathbb{CH}, \\ \mu_\beta &= D_\beta\phi_\beta + \partial_{\phi_\beta}\Phi(\phi_{\mathbb{A}}), && \beta \in \mathbb{RD}, \\ \mu_\gamma &= \partial_{\phi_\gamma}\Phi(\phi_{\mathbb{A}}), && \gamma \in \mathbb{OD}. \end{aligned} \tag{23}$$

Combining eq. 23 with the mass balance equations 20 and 21 yields the multiple constituent model:

$$\begin{cases} \partial_t\phi_\alpha + \operatorname{div}(\phi_\alpha v_\alpha) = \operatorname{div}\big(M_\alpha\phi_\alpha^2(1-\phi_\alpha)^2\nabla\mu_\alpha\big) + S_\alpha(\phi_{\mathbb{A}}) & \alpha \in \mathbb{CH}, \\ \qquad\qquad \mu_\alpha = \partial_{\phi_\alpha}\Psi(\phi_{\mathbb{CH}}) + \partial_{\phi_\alpha}\Phi(\phi_{\mathbb{A}}) - \varepsilon_\alpha^2\Delta\phi_\alpha & \alpha \in \mathbb{CH}, \\ \partial_t\phi_\beta + \operatorname{div}(\phi_\beta v_\beta) = \operatorname{div}\Big(M_\beta\nabla\big(D_\beta\phi_\beta + \partial_{\phi_\beta}\Phi(\phi_{\mathbb{A}})\big)\Big) + S_\beta(\phi_{\mathbb{A}}) & \beta \in \mathbb{RD}, \\ \qquad\qquad \partial_t\phi_\gamma = S_\gamma(\phi_{\mathbb{A}}) & \gamma \in \mathbb{OD}. \end{cases} \tag{24}$$

Here, M_α and M_β denote constants that arises from the mobilites, which are typically assumed to be:

$$\begin{aligned} m_\alpha(\phi_{\mathbb{A}}) &= M_\alpha\,\phi_\alpha^2(1-\phi_\alpha)^2, && \alpha \in \mathbb{CH}, \\ m_\beta(\phi_{\mathbb{A}}) &= M_\beta, && \beta \in \mathbb{RD}, \\ m_\gamma(\phi_{\mathbb{A}}) &= 0, && \gamma \in \mathbb{OD}. \end{aligned} \tag{25}$$

2.2.3 Four-Species Tumor Growth Model

A fundamental example of the multiple constituent model is the four-species model. Is was introduced by Hawkins-Daarud et al. in [HZO12], incorporating volume fractions of tumor cells, healthy cells, nutrient-rich extracellular water, and nutrient-poor extracellular water. The mathematical well-posedness of this model has been studied by Garcke and Lam [GL16; GL17a; GL17b], while Frigeri et al. [FGR15; FLR17] investigated the effects of degenerating mobility functions. Due to the fact that the model is based on a fourth-order PDE with concentration-dependent mobilities, the uniqueness of weak solutions is unresolved, even for the prototype model in 19 [EG96]. Further theoretical analysis includes optimal control aspects [CGR+17] and studies of long-time behavior [CGG11; Mir19].

The four-species model can be derived from the multiple-constituent model 23 by specifying the constituent sets as follows: $\mathbb{A} = \{T, \sigma\}$, $\mathbb{CH} = \{T\}$, $\mathbb{RD} = \{\sigma\}$ and $\mathbb{OD} = \emptyset$. The volume fraction of tumor cells, ϕ_T, is interpreted as an averaged cell density, providing a homogenized description of several thousands of cells. The field ϕ_σ represents the local nutrient concentration. Moreover, we introduce the adhesion function $\Phi(\phi_T, \phi_\sigma) = -\chi_c\phi_T\phi_\sigma$ in the energy functional (2.6), with $\chi_c > 0$ being a chemotaxis parameter. For both tumor cells and nutrients, we assume a volume-averaged velocity, which is a reasonable approximation given the dense packing of cells. Incorporating these assumptions into the multispecies framework leads to the so-called four-species model:

$$\begin{cases} \partial_t\phi_T + \operatorname{div}(\phi_T v) = \operatorname{div}\left(M_T\phi_T^2(1-\phi_T)^2\nabla\mu_T\right) + S_T(\phi_T, \phi_\sigma) \\ \qquad\qquad \mu_T = \Psi'(\phi_T) - \chi_c\phi_\sigma - \varepsilon_T^2\Delta\phi_T \\ \partial_t\phi_\sigma + \operatorname{div}(\phi_\sigma v) = \operatorname{div}\left(M_\sigma\nabla(D_\sigma\phi_\sigma - \chi_c\phi_T)\right) + S_\sigma(\phi_T, \phi_\sigma) \end{cases} \tag{26}$$

Indeed, system 26 can be derived from the four constituents volume fraction of tumor cells ϕ_T, healthy cells ϕ_C, nutrient-rich extracellular water ϕ_σ, and its nutrient-poor counterpart ϕ_{σ_0}. Source functions, expressed as sink and source terms, are of particular relevance. Tumor cells consume nutrients, making growth proportional to nutrient depletion, while apoptosis supplies additional nutrients through the recycling of dead cells. In the four-species model we often consider the source function

$$S_T(\phi_T, \phi_\sigma) = -S_\sigma(\phi_T, \phi_\sigma) = \lambda_T^{\text{pro}}\, \phi_\sigma \phi_T (1 - \phi_T) - \lambda_T^{\text{apo}}\, \phi_T, \tag{27}$$

where λ_T^{pro} denotes the proliferation rate and λ_T^{apo} the apoptosis rate.

To incorporate mechanical effects, velocity fields have been added under various assumptions. Initially modeled via Darcy's law [GL16; GLN+18], the velocity description has since been extended to the Brinkman law [EG19a; EG19b], unsteady Darcy–Forchheimer–Brinkman law [FLO+19], and the Navier–Stokes equations [LW17; He21]. The chosen velocity law significantly influences the morphology and growth direction of the tumor, with a comparative discussion and visualization of models presented in [GLS+16].

3 Application to PDE Solution Operators

In this section, we design the random feature maps $\phi : \mathcal{X} \times \Theta \to \mathcal{Y}$ and measures μ for the RFM approximation of three PDE parameter-to-solution maps: the stationary, linearized Cahn-Hilliard equation on a circular disk in subsection 3.1, the inital-condition-to-solution operator of the Cahn-Hilliard equation at an early timestep where spinodal decomposition just has started in subsection 3.2, and the initial-condition-to-solution operator of the Cahn-Hilliard equation for late timesteps where the two phases already have formed in subsection 3.3. In kernel methods, the specification of (φ, μ) is central to the accuracy of function reconstruction. While RFMs are purely data-driven and agnostic to the governing PDE, following [NS24], we emphasize that prior knowledge should inform the choice of (φ, μ). The systematic identification of effective random feature pairs, yielding data-adapted kernels, remains an open problem. In this work, we employ feature maps φ that are nonlinear in both arguments and specify a probability measure ν on the input domain $\mathcal{X}$ for each PDE setting. This choice is crucial: although we aim for out-of-distribution generalization, reliable performance can only be expected on inputs statistically aligned with those sampled from ν.

3.1 Stationary, Linearized Cahn-Hilliard Equation on a Disk

The linearized, stationary Cahn-Hilliard equation describes the behavior of small perturbations around an equilibrium state in a binary mixture. It is an important tool for analyzing phase stability in materials. It arises from the general Cahn-Hilliard equation 19 by assuming constant mobility $m(\phi) = m_0$ and considering small perturbations $\phi = \phi_0 + \eta\psi$ around a homogeneous equilibrium state ϕ_0, where $\eta \ll 1$. Linearizing the chemical potential and neglecting higher-order terms yields a fourth-order linear PDE in ψ. In the stationary case, where $\partial_t \psi = 0$, eq. 19 reduces to:

$$\Delta\left(-\epsilon^2 \Delta\psi + \Psi''(\phi_0)\psi\right) = 0. \tag{28}$$

By rescaling the spatial units such that $\epsilon^2 = 1$ and assuming $\Psi''(\phi_0) = 1$, the equation simplifies further to the canonical form of the linearized, stationary Cahn–Hilliard equation:

$$\Delta(\Delta + 1)\psi = 0 \tag{29}$$

We now apply the RFM to eq. 29 defined on a disk $\mathbb{D} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq R^2\}$ for some $R > 0$. Bessel functions are the natural spectral and eigenfunction system for the Cahn-Hilliard

problem in radially symmetric geometries [GLW01]: When separated in polar coordinates, eq. 29 decomposes into a Laplace and a Helmholtz part. The solutions of the Laplace part

$$\Delta\psi_L = 0 \tag{30}$$

can generally be expressed as

$$\psi_L(r,\theta) = \sum_{m=-\infty}^{\infty} \left(A_m r^{|m|} + B_m r^{-|m|}\right) e^{im\theta}, \tag{31}$$

with $A_m, B_m \in \mathbb{C}$. Regularity at the origin enforces $B_m = 0$. The Helmholtz part:

$$(\Delta + 1)\psi_H = 0, \tag{32}$$

yields as possible solutions

$$\psi_H(r,\theta) = \sum_{m=-\infty}^{\infty} \left(C_m J_m(r) + D_m Y_m(r)\right) e^{im\theta}, \tag{33}$$

where $C_m, D_m \in \mathbb{C}$, and J_m denotes the first and Y_m the second Bessel function. Again, requiring regularity of the solution enforces $D_m = 0$. By assuming Dirichlet zero boundary conditions on the boundary of the disk:

$$\psi(R,\theta) = 0, \tag{34}$$

only the Helmholtz part remains, whose admissible wavenumbers are discretized. Ψ can then be represented as a Fourier-Bessel series,

$$\psi(r,\theta) = \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} Z_{n,m} J_m\left(\rho_{n,m}\frac{r}{R}\right) e^{im\theta}, \tag{35}$$

where $\rho_{n,m}$ denotes the n-th zero of J_m and $Z_{n,m} \in \mathbb{C}$. The Fourier–Bessel series is a particular kind of generalized Fourier series based on Bessel functions, particularly used for cylindrical coordinate systems. In the following, let L denote the operator

$$L := \Delta(\Delta + 1) \tag{36}$$

on the disk $\mathbb{D} = \{(r,\theta)|r \leq R\}$ with Dirichlet boundary $\psi|_{r=R} = 0$. Using eq. 35, we can determine the Green's function, G_L as

$$G_L((r,\theta),(r',\theta')) = \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \lambda_{n,m}\psi_{n,m}(r,\theta)\overline{\psi_{n,m}(r',\theta')} \tag{37}$$

with the eigenfunctions

$$\psi_{n,m}(r,\theta) = \frac{1}{N_{n,m}} J_m\left(\rho_{n,m}\frac{r}{R}\right) e^{im\theta}, \tag{38}$$

and the eigenvalues

$$\lambda_{n.m} = \frac{1}{((\rho_{n.m}/R)^2)((\rho_{n,m}/R)^2 + 1)}. \tag{39}$$

Here, $N_{n,m}$ denotes a normalization factor. With this theoretical background, we are able to provide a two-dimensional instantiation of the RFM for the linearized, stationary Cahn-Hilliard equation: Take the input space as $\mathcal{X} := L^2(\mathbb{D})$, output space as $\mathcal{Y} := L^2(\mathbb{D})$, input space measure d. We define our random feature mapping $\phi : \mathcal{X} \times \Xi \to \mathcal{Y}$ by

$$\phi(a;\chi)(x) := \langle a, \psi_\chi\rangle\psi_\chi(x) \tag{40}$$

with

$$\psi_\chi(x) = \psi_\chi(r;\theta) := \sqrt{\frac{2}{\lambda_{n,m}}} J_m\left(\frac{\rho_{n,m}}{R} r\right) \cos(m\theta + b), \tag{41}$$

$\chi = (n, m, b) \in \mathbb{N} \times \mathbb{N} \times [0, 2\pi]$ and $\mu := \mu_{spec} \otimes \mathrm{Unif}[0, 2\pi)$. The measure μ_{spec} is defined via the spectral values $\lambda_{n,m}$ in eq. 39 by

$$\mu_{spec}(n, m) := \frac{\lambda_{n,m}}{\sum \lambda_{n',m'}}. \tag{42}$$

The induced kernel $k_\mu : \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{Y})$ can then be calculated as

$$\begin{aligned} k_\mu(a, a') &= \mathbb{E}^{\theta\sim\mu}\left[\phi(a;\chi)\phi(a';\chi)\right] \\ &= \sum_{n,m} \lambda_{n,m}^2 \langle a, \psi_{n,m}\rangle \langle a', \psi_{n,m}\rangle \psi_{n,m} \otimes \psi_{n,m} \\ &= \langle a, L^{-1}a'\rangle G_L, \end{aligned} \tag{43}$$

with the Greens function G_L from eq. 37. Using this fact, we may explicitly characterize the associated RKHS $\mathcal{H}_{k_\mu}$ as follows. First, we have

$$T_{k_\mu} f = \langle a, L^{-1}a'\rangle G_L[f] = \langle a, L^{-1}a'\rangle L^{-1} f \tag{44}$$

If we view T_{k_μ} as an operator from $L^2(\mathbb{D})$ into itself, we conclude with eq. 44 that for any elements f and g of H_{k_μ}, it holds that

$$\langle f, g\rangle_{\mathcal{H}_{k_\mu}} = \langle f, T_{k_\mu}^{-1} g\rangle_{L^2(\mathbb{D})} = \frac{1}{\langle a, L^{-1}a'\rangle} \langle f, Lg\rangle_{L^2(\mathbb{D})}. \tag{45}$$

By this argument, the reproducing kernel Hilbert space (RKHS) H_{k_μ} can be, up to a constant factor, identified with the energy space associated with the operator L, defined as

$$H_{k_\mu} = \{h \in L^2(\mathbb{D}) \mid \langle h, Lh\rangle < \infty\}. \tag{46}$$

In particular, the RKHS consists of those functions whose L-energy is finite. For the more complex problems we study numerically in the next two subsections, we do not have knowledge of effective, computable bases for general maps in infinite-dimensional spaces. The RFM approach leverages randomness to explore these maps, implicitly uncover their structure, and provide a means to represent them.

3.2 Cahn-Hilliard Equation - Early Timesteps

The Cahn–Hilliard equation is representative of the mass-conserving, diffusion-dominated PDE problem class; these time-dependent equations arise as gradient flows of free energy functionals, and the nonlinear, higher-order diffusion drives phase separation and interface dynamics. The initial value problem we consider in this work is

$$\begin{aligned} \partial_t \phi &= m\nabla^2(\Psi'(\phi) - \epsilon^2\Delta\phi) && \text{in } (0,\infty)\times[0,L]^2, \\ \phi(\cdot, x) &= \phi(\cdot,\, x + L\, e_1)\ , && \text{in } (0,\infty)\times\{0,L\}\times[0,L], \\ \phi(\cdot, x) &= \phi(\cdot,\, x + L\, e_2)\ , && \text{in } (0,\infty)\times[0,L]\times\{0,L\}, \\ \phi(0,\cdot) &= a && \text{in } [0,L]^2, \end{aligned} \tag{47}$$

with

$$\Psi(\phi) = W\phi^2(1-\phi)^2 \tag{48}$$

as the bulk free energy and L as the length of the quadratic domain $[0, L]^2 \subset \mathbb{R}^2$. The constant W can be regarded as the height of the thermodynamic barrier. The initial condition a serves as the input and is drawn according to a Gaussian measure defined by

$$a \sim v := N(0.5, 0.05\mathrm{Id}),$$

which represents white noise after subtracting the mean value of $\mu = 0.5$. For the Cahn–Hilliard equation, mass is conserved and the gradient contribution to the free energy is finite; therefore, we choose the space

$$H^1_{(0)}(\Omega) := \Big\{ u \in H^1(\Omega) \ : \ \int_\Omega u(x)\, dx = C \Big\},$$

with $C = \text{const.}$ for both the initial condition and the solution at a timestep T:

$$F^\dagger : H^1_{(0)}(\Omega) \to H^1_{(0)}(\Omega), \ a \mapsto F^\dagger(a) := \phi(T, \cdot) \tag{49}$$

We now describe our found random feature map for use in the RFM that worked best for approximating the solution operator 49 at a timestep, where spinodal decomposition just has begun. Our chosen random features can be categorized as *Fourier space random features* and have some parallels to the random features used for the Burgers' equation in [NS24]. Let $\mathcal{F}$ denote the Fourier transform over the two-dimensional spatial domain $\mathcal{D}$. We define our random feature mapping $\varphi : \mathcal{X} \times \Theta \to \mathcal{Y}$ by

$$\varphi(a; \theta) := \sigma\left(\mathcal{F}^{-1}(\chi \mathcal{F} a \mathcal{F} \theta)\right), \tag{50}$$

where $\sigma(.)$ denotes the ELU function defined below as a mapping on $\mathbb{R}$ and applied pointwise to the inverse Fourier transform. Note that the product $\mathcal{F}a\mathcal{F}\theta$ corresponds to applying the Fourier transform to the convolution of a with θ, i.e. $\varphi(a; \theta) := \sigma\left(\mathcal{F}^{-1}(\chi \mathcal{F}(a * \theta))\right)$. The map $\varphi(\cdot; \theta)$ essentially performs a filtered random convolution with the initial condition.

Considering the parameter space $\Theta \subset L^2(\Omega; \mathbb{R})$, the randomness enters through $\theta \sim \mu := N(0, \mathrm{Id})$ with Id denoting the identity covariance operator. $\theta \in \Theta$ is white noise on Ω: At each spatial location we draw a random variable out of a standard normal distribution independent of the draws on the other spatial location within the domain Ω. Choosing the right *wavenumber filter function* $\chi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is now crucial in order to achieve a good approximation accuracy with a reasonable number of random features. The trick for the Cahn-Hilliard equation is to incorporate only the physically relevant spectral properties into the random features, analogous did in [NS24] for the one-dimensional Burgers' equation. In our case, this is achieved by exploiting the dispersion relation of the Cahn-Hilliard equation, which can be deduced from eq. 47 by linearizing around a homogeneous, stationary state ϕ_0:

$$\phi(x, t) = \phi_0 + \delta\phi(x, t), \quad \text{with } \delta\phi \ll 1. \tag{51}$$

Expanding Ψ' to first order yields

$$\Psi'(\phi) \approx \Psi'(\phi_0) + \Psi''(\phi_0)\delta\phi. \tag{52}$$

Substituting eq. 51 and 52 into eq. 19 and neglecting constant terms provides the linearized equation

$$\partial_t \delta\phi = m\nabla^2 \left(\Psi''(\phi_0)\delta\phi - \epsilon^2 \Delta\delta\phi\right). \tag{53}$$

We now apply a Fourier mode ansatz

$$\delta\phi(x, t) = \hat{\phi}(k) e^{ik\cdot x + \lambda(k)t}, \tag{54}$$

which implies

$$\nabla^2 \delta\phi = -|k|^2 \delta\phi, \quad \Delta\delta\phi = \nabla^2 \delta\phi. \tag{55}$$

Inserting this into the linearized eq. 53 yields the dispersion relation

$$\lambda(k) = -m|k|^2 \left(\Psi''(\phi_0) + \epsilon^2|k|^2\right). \tag{56}$$

This relation determines the growth or decay rate of perturbations of wavenumber k. By unifying the constants, eq. 56 can be written as

$$\lambda(k) = |k|^2(\beta - \gamma|k|^2). \tag{57}$$

Maximum growth occurs at

$$k^* = \sqrt{\frac{\beta}{2\gamma}}, \tag{58}$$

i.e., at an early timestep the patterns predominantly have wavenumber k^*. In a random realization pattern, the spectral power density $S(k)$ is distributed around this value. To selectively enhance these intermediate-wavelength patterns while suppressing unwanted fluctuations, we design our filter function χ to act on the spectrum $\lambda(k)$ in the following manner: For small values of $|k|$, we have $\lambda(k) \approx \beta|k|^2 > 0$, and allow these modes to pass through the filter. In contrast, for large values of $|k|$, $\lambda(k) < 0$ and should therefore be clipped to zero, effectively blocking these modes. For very small $k \approx 0$, $\lambda(k) \to 0$, and these modes should strongly be suppressed. Hence, the filter should effectively act as a band-pass: it should suppresses low-frequency modes corresponding to large-scale patterns, suppresses high-frequency noise, and retains intermediate frequencies, such as those associated with interface patterns. We thus choose as wavenumber filter function χ:

$$\chi(k) := \sigma_\chi(2\pi|k|), \text{ where } \sigma_\chi(r) := \sqrt{\max\left(0, r^2(\beta - \gamma r^2)\right)}, \tag{59}$$

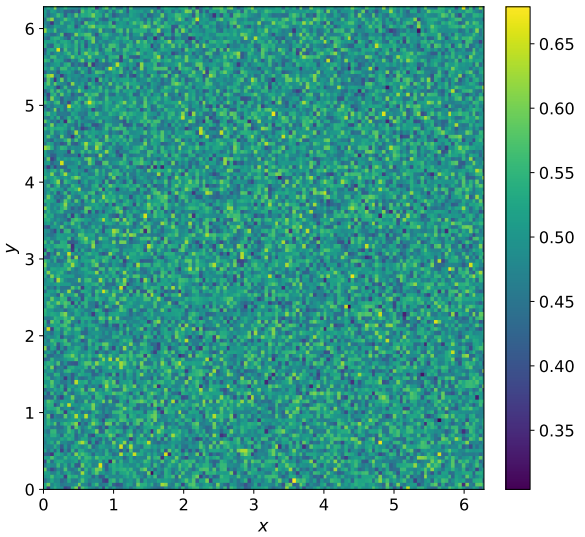
where $\beta > 0$ and $\gamma > 0$. Figure 1 illustrates a sample input and output from φ. The parameters β and γ are simply hand-tuned for performance, but not strictly optimized. We choose the activation function σ in eq. 50 to be the exponential linear unit:

$$x \mapsto \mathrm{ELU}(x) := \begin{cases} x, & \text{if } x \geq 0 \\ e^{-x} - 1, & \text{if } x < 0. \end{cases} \tag{60}$$

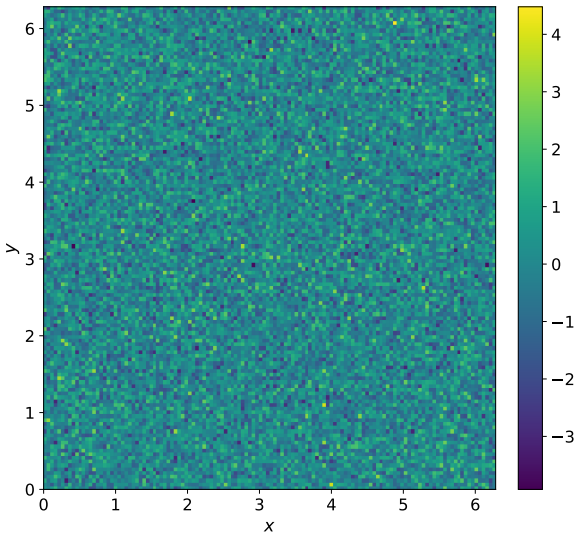
By conducting empirical tests, we also found ELU$(\cdot)$ to perform better in the RFM framework than several other commonly used activation functions in machine learning frameworks like ReLU$(\cdot)$, tanh$(\cdot)$, sigmoid$(\cdot)$, SeLU$(\cdot)$, and softplus$(\cdot)$.

3.3 Cahn-Hilliard Equation - Late Timesteps

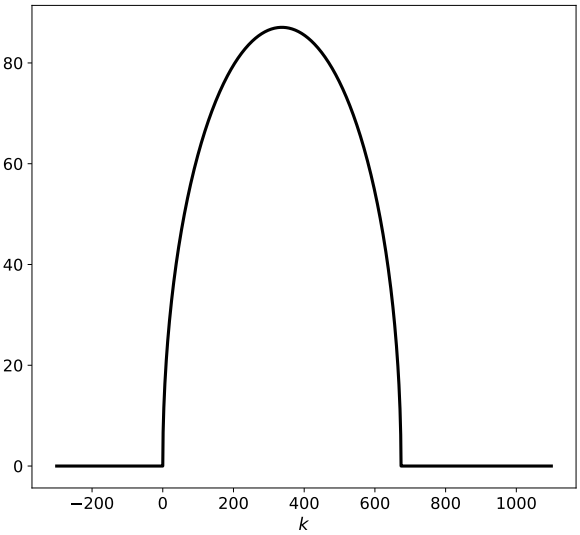
For late timesteps, the filter function defined in eq. 59 does no longer provide an adequate representation of the underlying physics, as the two phases have already formed. We need to find random features that model the multiscale behavior of the phase separation well. In this work, we tried out an approach to approximate the solution operator 49 at these later timesteps by using random features built up of cosine functions, which we call *cosine random features*. Their choice was inspired by the results Liao presented in [Lia24], in which Liao introduced a random feature–based framework for solving specific PDEs by directly approximating the solution at collocation points. In this framework, random features are employed to approximate the solution of a PDE directly, rather than learning a general operator. Our idea was by adapting the design principles of the random features in [Lia24] to operator learning, we can construct highly accurate features for approximating the solution operator 49 of the Cahn–Hilliard equation at late timesteps. Unfortunately, this approach has proven to be ineffective. Nevertheless, we provide in the following a short overview of that approach by first giving an overview about



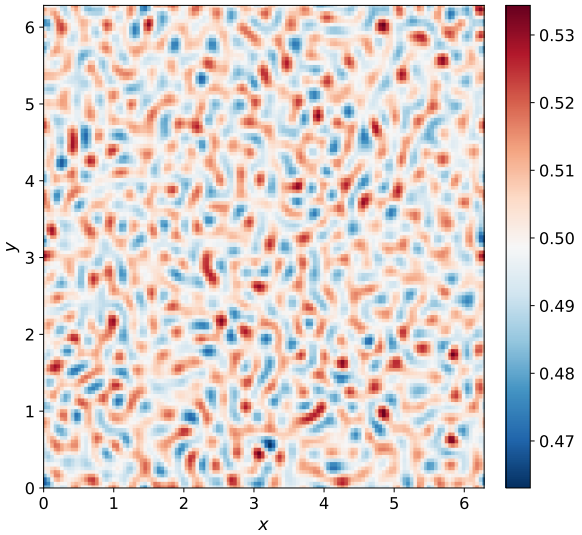
(a) Representative Input $a \sim N(0.5, 0.05\mathrm{Id})$



(b) Representative $\theta \sim N(0, \mathrm{Id})$



(c) Filter function



(d) Random Feature

Figure 1: Random feature map construction for the Cahn-Hilliard equation at an early timestep (T = 5.0): (a) displays a representative input draw $a \sim \nu = N(0.5, 0.05\mathrm{Id})$; (b) shows a representative parameter draw $\theta \sim \mu = N(0, \mathrm{Id})$, while (d) displays the corresponding random feature $\varphi(a; \theta)$ belonging to that a and θ. (c) show the filter function $k \to \chi(k)$. Here, $\beta = 45.0$, and $\gamma = 0.06678$.

the main results in [Lia24] relevant for our choice of (φ, μ):

Consider a PDE of the form

$$P[u](x) = 0, \quad x \in \Omega \subset \mathbb{R}^n, \qquad B[u](x) = 0, \quad x \in \partial\Omega, \tag{61}$$

where P is the interior differential operator and B the boundary operator. The idea presented in [Lia24] is to represent the solution u^* of 61 as a randomized feature expansion

$$u^\sharp(x) = \sum_{k=1}^{N} c_k^\sharp \, \varphi(x; \omega_k), \tag{62}$$

where $\{\omega_k\}_{k=1}^N$ are i.i.d. samples drawn from the spectral density $\rho(\omega)$ guaranteed by Bochner's theorem, and $\varphi(x; \omega_k)$ are random features such as

$$\varphi(x; \omega_k) = \cos(\langle \omega_k, x \rangle + b_k), \qquad b_k \sim \mathrm{Unif}[-\pi, \pi]. \tag{63}$$

The coefficients $c_k^\sharp$ are then determined by enforcing the PDE and boundary conditions at a finite set of collocation points. For interior collocation points $\{x_j\}_{j=1}^{M_\Omega}$ and boundary points $\{x_j\}_{j=M_\Omega+1}^{M}$, the training problem takes the form of a convex regression problem:

$$\min_{c \in \mathbb{R}^N} \ \|c\|_2^2 + \lambda_1 \sum_{j=1}^{M_\Omega} \big(P[u^\sharp](x_j)\big)^2 + \lambda_2 \sum_{j=M_\Omega+1}^{M} \big(B[u^\sharp](x_j)\big)^2, \tag{64}$$

with the hyperparameters $\lambda_1, \lambda_2 > 0$. This approach leverages the fact that random features approximate the RKHS associated with k, ensuring convergence to the true PDE solution as the number of collocation points and features increases.

However, operator learning does not aim to approximate a single PDE solution but instead aims to learn a mapping from problem data (e.g., an initial condition a) to the corresponding solution u. Consequently, the random features cannot depend only on the spatial variable x but must also encode the dependence on the input function a. Extending this to the RFM for operator learning raises the question of how to represent a in the random features φ so that $\varphi(\cdot; \theta)$ becomes a highly nonlinear function of a. As shown in section 4, this is a challenging problem, and we did not resolve it fully in this work.

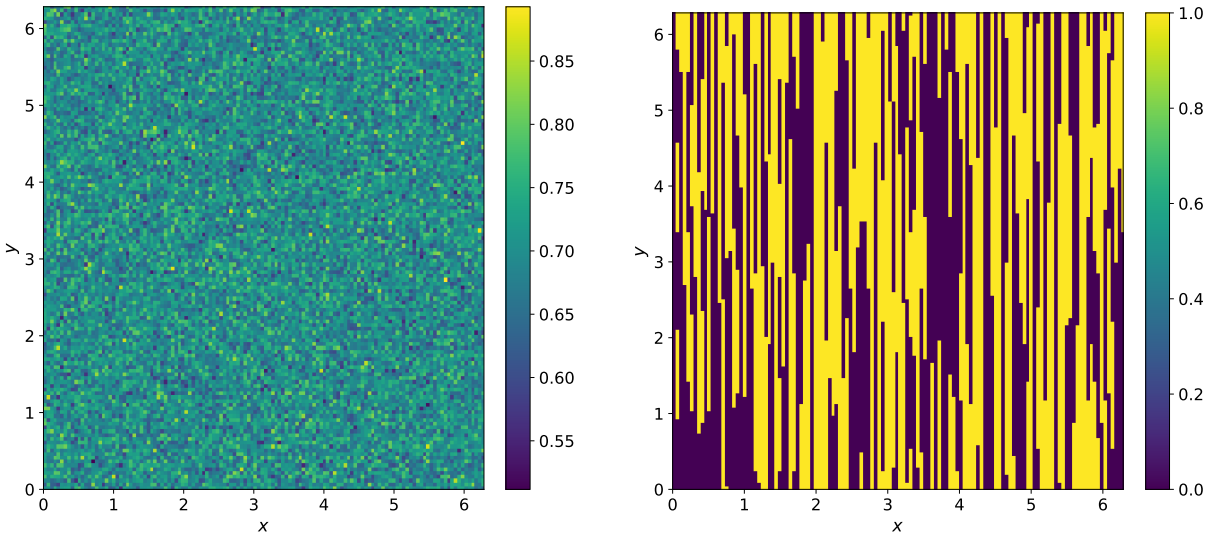
One possible approach is to aggregate a and incorporate it into the frequency or phase information. Our method instead couples a pointwise with the cosine functions. Specifically, in section 4.2, we consider the *cosine random feature mapping* $\varphi : \mathcal{X} \times \Theta \to \mathcal{Y}$ defined as

$$\varphi(a; \theta)(x) := \sigma\Big(a(x) \cdot \cos(\langle x, w \rangle + b)\Big), \tag{65}$$

where σ is a nonlinear activation and $a \sim \nu = N(0.7; 0.05\mathrm{Id})$. 65 makes φ a nonlinear function of a, enabling a random feature model that could, in principle, approximate the initial-condition-to-solution operator 49. For σ, we used a sharpened sigmoid function

$$\sigma : \mathbb{R} \to [0, 1], \ x \mapsto \sigma(x) = \frac{1}{1 + e^{-\beta x}} \tag{66}$$

with $\beta > 0$, that is applied pointwise to take into account the binary structures. The parameter space Θ consists of $\theta = (w, b)$, where $w \in \mathbb{R}^2$ are two random frequencies drawn from $N(0, 1/l^2\mathrm{Id})$ and $b \in [0, 2\pi)$ is a random phase. An example $(a, \varphi(a; \theta))$ pair is shown in figure 2. However, as discussed in section 4.2, applying this random features to the solution operator 49 for late timesteps shows to be only suited for fitting a single input-output pair and does not generalize well to operator learning.



(a) Representative Input $a \sim N(0.7, 0.05\mathrm{Id})$ (b) Random Feature

Figure 2: Random feature map construction for the Cahn-Hilliard equation at a late timestep (T = 50.0): (a) displays a representative input draw $a \sim \nu = N(0.7, 0.05\mathrm{Id})$, while (b) displays the corresponding random feature $\varphi(a;\theta)$. Here, $\beta = 100.0$ in the activation function σ.

4 Numerical Experiments

We evaluate the performance of the RFM for approximating the operator $F^\dagger : \mathcal{X} \to \mathcal{Y}$ in eq. 49. Implementing the approach on a computer requires discretization of the input-output function spaces $\mathcal{X}$ and $\mathcal{Y}$. In the following numerical experiments, all infinite-dimensional objects - including training data, evaluations of random feature maps, and random fields - are discretized on an equispaced mesh with K^2 points. This equispaced choice enables the fast Fourier transform (FFT) with $O(K^2 \log K)$ computational complexity in two spatial dimensions.

However, the use of equispaced points does not limit the method, since the RFM is formulated in function space and can be implemented numerically with any spatial discretization. This discretization leads to a high - but finite - dimensional approximation of operators mapping $\mathbb{R}^{K^2}$ to $\mathbb{R}^{K^2}$ via similarly discretized RFMs. Importantly, K can vary, and we study the properties of the discretized RFM as a function of K. Because the RFM is conceptually defined in function space without reference to discretization, its numerical realization preserves approximation quality in the limit $K \to \infty$. Consequently, the same trained model can be applied across the entire hierarchy of spaces $\mathbb{R}^{K^2}$ for $K \in \mathbb{N}$, provided K is sufficiently large. In this section, we omit explicit dependence on K in the notation for the discretized RFM and target operator. We demonstrate these properties numerically.

Our code is highly optimized, leveraging JIT compilation via `jax` and parallelization across four A100 GPUs using `joblib`. The RFM map is efficiently evaluated with the FFT and requires no additional tools for discretization. The input functions and the selected random feature map 50 require i.i.d. samples of Gaussian random fields, which is achieved using `numpy.random.randn`. Given the discretization and the input-output pairs (ϕ, μ) from section 3, the final algorithmic step is to train the RFM via eq. 15 and evaluate its performance. For a fixed number of random features m, a single realization of the RFM is trained and tested. In experiments with varying m, the parameters $\{\theta_j\}_{j=1}^m$ are resampled i.i.d. from μ.

To quantify the approximation error, we use the expected relative test error

$$e_{n',m} := \frac{1}{n'} \sum_{j=1}^{n'} \frac{\|F^\dagger(a'_j) - F_m(a'_j;\hat{\alpha})\|_{L^2}}{\|F^\dagger(a'_j)\|_{L^2}} \approx \mathbb{E}^{a'\sim\nu}\left[\frac{\|F^\dagger(a') - F_m(a';\hat{\alpha})\|_{L^2}}{\|F^\dagger(a')\|_{L^2}}\right], \tag{67}$$

where $\{a'_j\}_{j=1}^{n'}$ are i.i.d. samples from ν and n' denotes the number of test pairs. All $L^2(\Omega;\mathbb{R})$ norms are approximated via a composite trapezoid quadrature:

$$\int_0^{2\pi}\int_0^{2\pi} u(x,y)\,dy\,dx \approx \frac{1}{(K-1)^2}\sum_{i=0}^{K-1}\sum_{j=0}^{K-1} w_i w_j u(x_i,y_j), \tag{68}$$

where the trapezoidal weights w_i and w_j are defined as

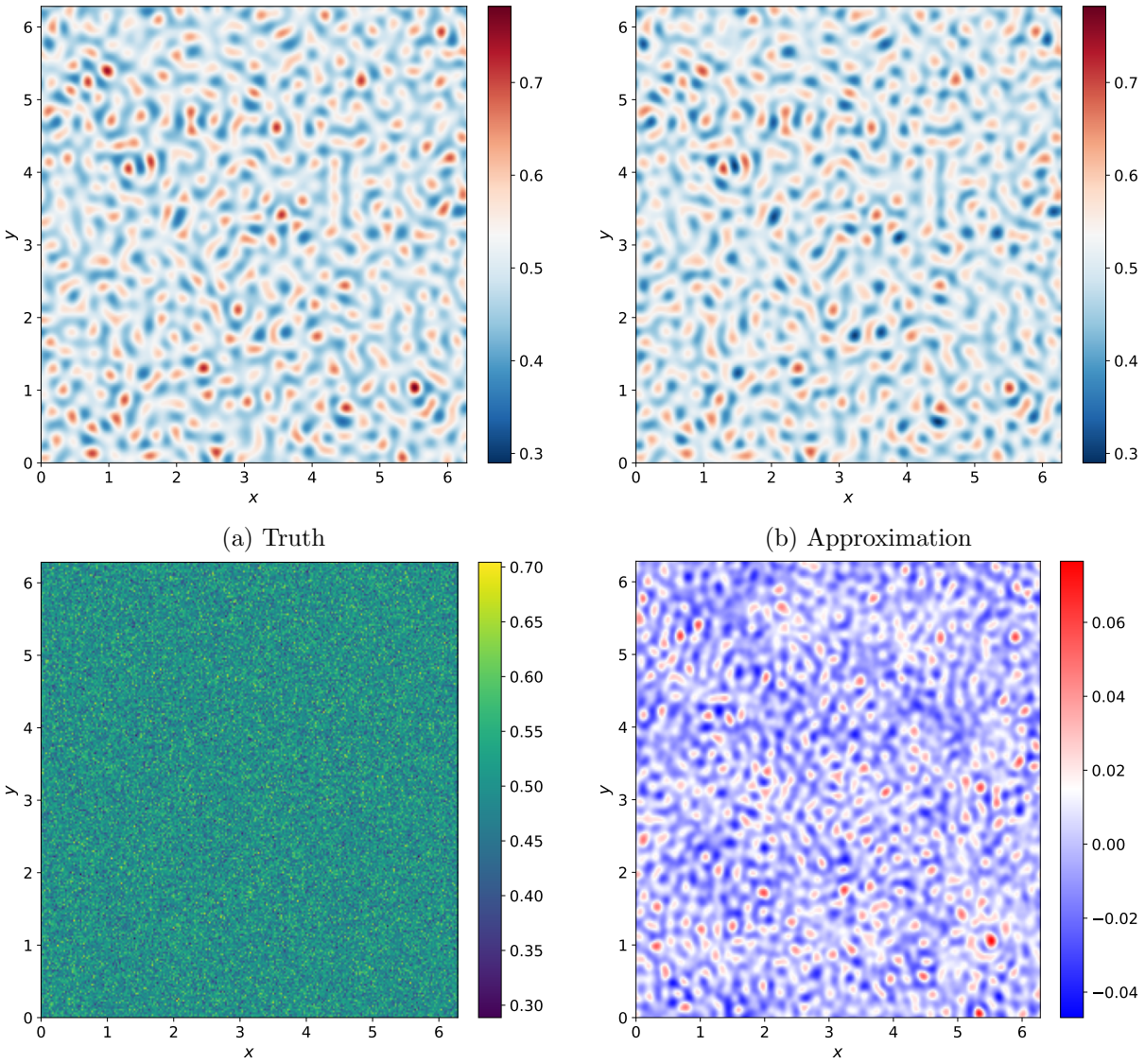
$$w_i = \begin{cases} \frac{1}{2}, & \text{for } i = 0 \text{ or } i = K-1, \\ 1, & \text{for } 1 \le i \le K-2, \end{cases} \qquad w_j = \begin{cases} \frac{1}{2}, & \text{for } j = 0 \text{ or } j = K-1, \\ 1, & \text{for } 1 \le j \le K-2. \end{cases}$$

Since $\mathcal{Y} \subset L^2$ for the PDE solution operator 49, all required inner products during training are performed in L^2, which reduces the relative test error $e_{n',m}$.

4.1 Cahn-Hilliard Equation - Early Timesteps

In this section, we consider the Cahn-Hilliard equation 19 with constant mobility term on the physical domain $\Omega = [0, L]^2$, where $L = 2\pi$. We generate a high-resolution dataset of input-output pairs by solving the Cahn-Hilliard equation on a periodic, equispaced mesh of size $K = 1024$ with random initial conditions drawn from $\nu = \mathcal{N}(0.5, 0.05\mathrm{Id})$, i.e ν corresponds to Gaussian white noise with variance $\sigma^2 = 0.05$ after subtracting a mean of $\mu = 0.5$. Spatial discretization is performed using an FFT-based pseudospectral method, while temporal integration employs a backward Euler method with a timestep size of $\mathrm{d}t = 0.1$ [SBT23]. All datasets on smaller mesh sizes $K < 1024$ used in training and testing are obtained by subsampling this high-resolution dataset. Consequently, numerical realizations of $F^\dagger$ are considered up to the map $\mathbb{R}^{1024^2} \to \mathbb{R}^{1024^2}$. We fix $n = 512$ training pairs and $n' = 2000$ testing pairs unless otherwise noted, and set the mobility constant to $M = 3.0$, while assuming a gradient energy coefficient of $\epsilon^2 = 2.5$ and $W = 2.0$ as height of the thermodynamic barrier in eq. 48. The parameters M, W and ϵ are chosen in such a way, that for an early timestep $T = 5.0$ the solution represents the stadium where spinodal decomposition just has started and the two phases begin to form. The solution maintains a mean of $\mu = 0.5$ at all times and converges to a to a binary distribution when the two phases have completely formed at later timesteps. For the random feature map in eq. 50, we fix the hyperparameters as $\beta = 45.0$ and $\gamma = 0.06678$. The map is evaluated efficiently using the FFT and requires no additional tools for discretization. Hyperparameters were hand-tuned rather than optimized. We train the RFM with $\lambda = 10^{-8}$ as regularization parameter by solving the normal equations in 15. However, as shown in figure 6a, training with the RFM with regularization parameters has negligible impact on the relative test error.When a regularization parameter λ is set, we obtain the coefficient vector α by solving $(A + \lambda I)\,\alpha = b$ using a direct linear solver (`jnp.linalg.solve`), while in the case of absent regularization we use the pseudoinverse to obtain the minimum-norm least squares solution, implemented with the truncated singular value decomposition in `scipy.linalg.pinv2`.

Our experiments investigate the RFM approximation of the initial-condition-to-solution operator $F^\dagger$ in 49. To provide a visual reference for this high-dimensional problem, figure 3 presents a representative input-output pair alongside the corresponding test prediction produced by a trained RFM. For $n = 512$ training pairs, $m = 2048$ random features and a spatial discretization



(a) Truth

(b) Approximation

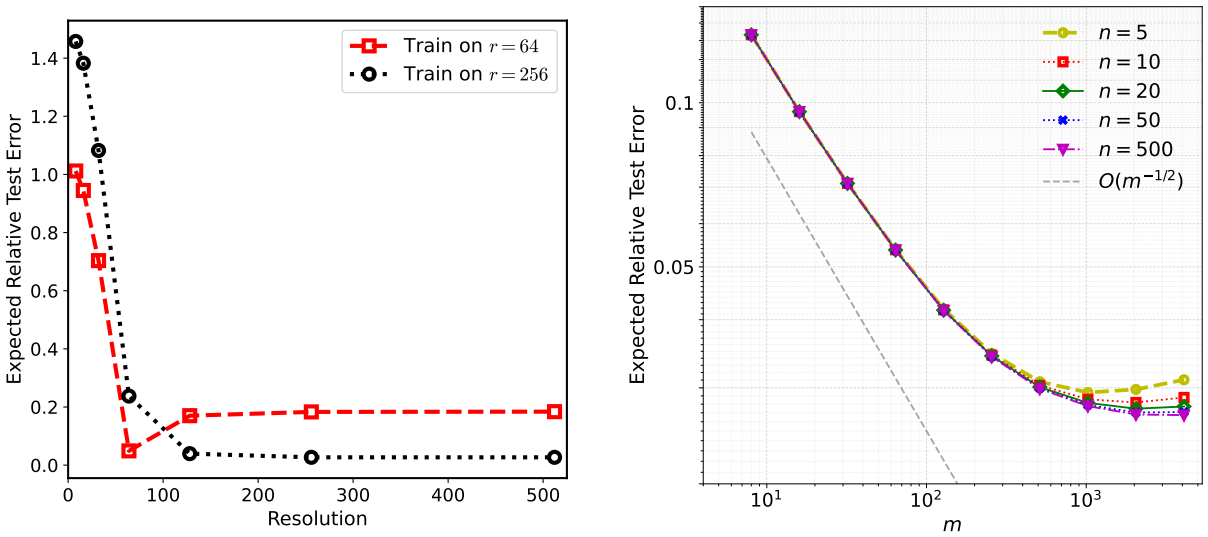
(c) Input

(d) Pointwise error

Figure 3: Representative input-output test sample for the solution of the Cahn-Hilliard equation at an early timestep: (c) shows a sample test input, (a) the resulting output (truth) from a pseudospectral solver, (b) the the tranined RFM prediction, while (d) displays the pointwise error. The relative L^2 error for this single prediction is 0.02593. Here, $n = 512$, $m = 2048$, and $K = 256$.

of $K = 256$ for each direction, the relative L^2 error has the order of magnitude $\mathcal{O}(10^{-2})$.

We first studied the ability of the RFM to transfer its learned coefficients $\hat{\alpha}$, obtained from training on a mesh of size K, to different mesh resolutions K' for evaluation at a timestep of $T = 5.0$ (figure 4a): The lowest test error occurs when $K = K' = 64$, i.e., when the training and testing resolutions are identical. This behavior was also observed for the Burgers' equation and Darcy flow in [NS24] and was reported in other coeval research [Li+20]. At very low resolutions, such as $K = 16$ or $K = 32$, the test error is mainly dominated by the discretization error, which can become quite large and even exceeds 1.0. For instance, resolving conceptually infinite-dimensional objects, such as our Fourier space-based feature map 50 or the L^2-norms in eq. 67, with only 16 grid points results in poor accuracy. Outside this regime, however, the errors remain essentially constant across resolutions, irrespective of the training resolution K. This suggests that the RFM determines optimal coefficients regardless of the resolution, enabling it to generalize effectively to any mesh size. In fact, the trained model can be deployed on different discretizations of the domain Ω (e.g., various choices of finite elements, graph-based, or particle methods), not just on different mesh sizes. Practically, this means that high-resolution training sets for the Cahn-Hilliard equation can be subsampled to smaller mesh sizes K (provided they are still sufficiently fine to avoid large discretization errors) to enable faster training, yielding a trained model with nearly identical accuracy at all higher resolutions ("superresolution") [NS24].

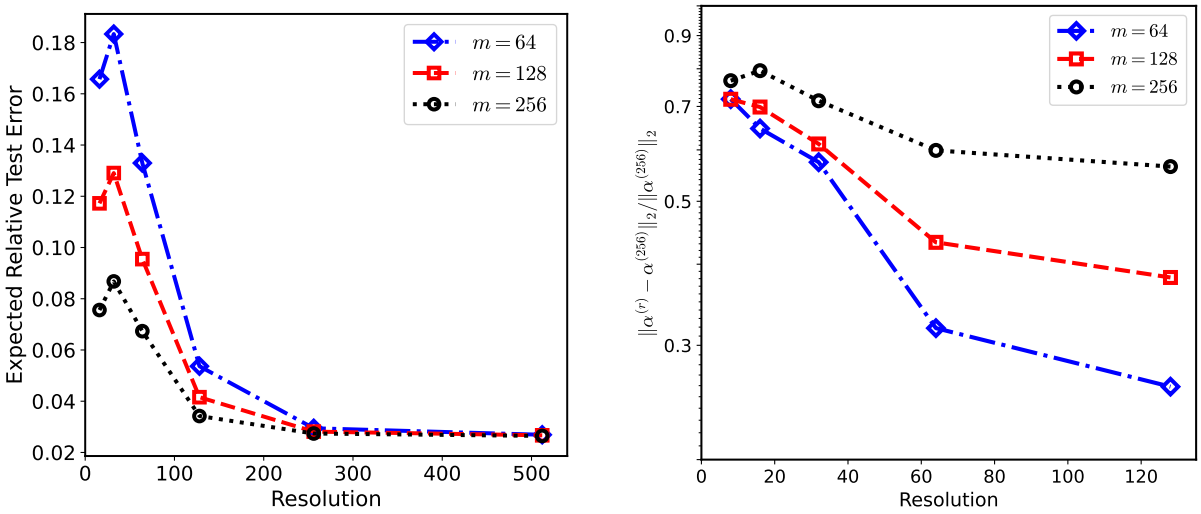


(a) Test error vs. resolution

(b) Test error vs. m and n

Figure 4: Expected relative test error of a trained RFM for the Cahn-Hilliard equation with $n' = 2000$ test pairs: (a) displays the invariance of test error w.r.t. training and testing on different mesh sizes without significant loss of accuracy. (b) shows the decay of the test error for resolution $K = 128$ fixed as a function of m and n; the smallest error achieved is 0.02673 for $n = 500$ and $m = 4096$.

As illustrated in 4b, the RFM reaches a smallest expected relative test error of 0.02673 for $n = 500$ and $m = 512$, while being evaluated on $n' = 2000$ test pairs. This result is particularly encouraging, as the observed error is comparable in magnitude to that reported for the considerably simpler Burgers' equation and Darcy flow solution operator in [NS24]. The error we report is also of the same order of magnitude as that reported in [Esh+25] for the Cahn-Hilliard solution operator. In that work, the authors used a Multi-Head Neural Operator, which is able to capture both coarse and fine features of the solution simultaneously through its different heads. However, the solution in [Esh+25] was considered at a later timestep with different



(a) Test error vs. resolution

(b) Minimizer error vs. resolution

Figure 5: Results of a trained RFM for the Cahn-Hilliard solution operator $F^\dagger = u(5.0; \cdot)$: (a) demonstrates the resolution-invariant test error for various m, while (b) displays the relative error of the learned coefficient α^r at resolution r w.r.t. the coefficient learned on the resolution $r = 256$. Here, $n = 256$ training and $n' = 2000$ testing pairs were used.

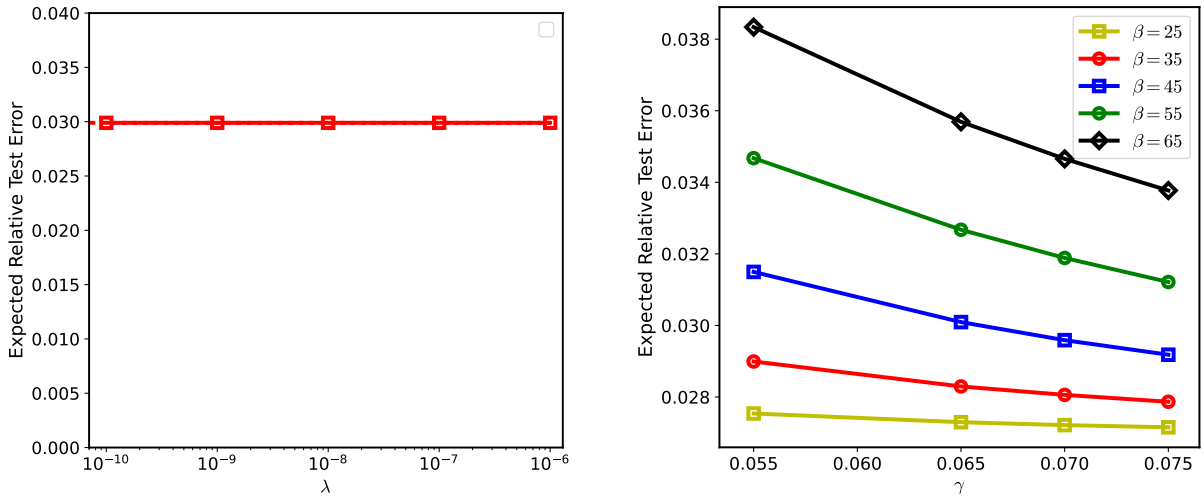
parameter choices - comprehensive parameter studies for the Cahn-Hilliard equation, as well as a comparison of the RFM against Fourier Neural Operators [Li+21] or DeepONets [LJK21], are left for future work. We expect that RFMs, compared to other neural operator approaches based on deep learning, which require training large neural networks via nonconvex optimization with stochastic gradient descent, can achieve similar performance for the Cahn-Hilliard solution operator. In contrast, RFMs use orders of magnitude fewer trainable parameters, which can be efficiently optimized through convex optimization.

Figure 5b shows that for sufficiently large n, the error first empirically follows the $O(m^{-1/2})$ parameter complexity bound suggested by theorem 2.5 in [NS24]. However, this theorem does not strictly apply here, since it requires $F^\dagger$ to lie in the RKHS of (φ, μ) - a condition that we did not verify. For $m > 256$, the error appears to start saturating and does not follow the $O(m^{-1/2})$ rate anymore. This behavior is likely due to fixing the regularization parameter λ to a constant value instead of scaling it with m as suggested by theorem 2.5 in [NS24]. It is also possible that the Cahn-Hilliard solution map does not belong to the RKHS $\mathcal{H}_{k_\mu}$, introducing an additional misspecification error. In contrast, if we train the RFM on a single input-output pair, the relative error for this pair converges to zero, as illustrated in Figure 7, at rates that increase with m.

Figure 5a illustrates that the expected relative test error remains essentially unchanged when varying the mesh resolution used for training and evaluation. This robustness arises from the formulation of the RFM directly in function spaces; In contrast, other machine learning-based surrogate models defined in finite-dimensional spaces typically show increasing test errors as the mesh is refined (see [Bha+21, section 4] for a numerical discussion of this effect). Figure 5a presents the expected relative test error as a function of mesh resolution for three different values of m. At very coarse resolutions, the error exhibits slight fluctuations, but it stabilizes to a constant value for large K. This value differs only slightly on the number of random features that are taken into account. In figure 5b, we plot the relative error of the learned parameter $\alpha(r)$ at resolution r w.r.t. the parameter learned at the highest resolution trained, which was $r = 256$. Figure 5b suggests that the learned coefficients $\alpha(K)$ converges toward a limiting

value $\alpha(\infty)$ as $K \to \infty$, further highlighting the RFM's construction as a mapping between infinite-dimensional spaces.

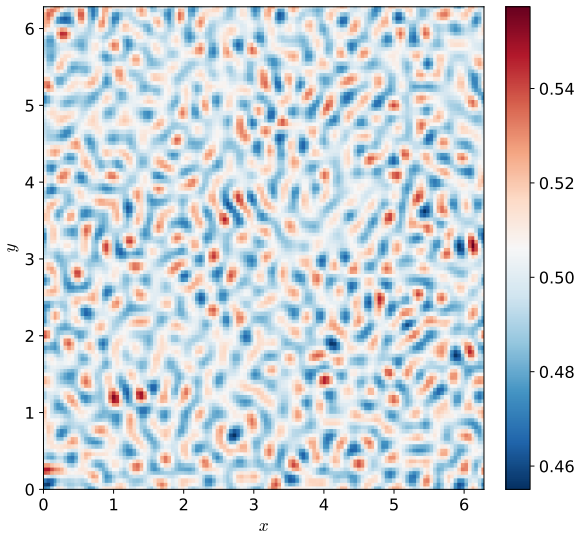
Finally, we investigated the effect of both different regularization parameters λ and varying hyperparameters β and γ on the relative test error. As illustrated in 6a, it was found that regularization has a negligible impact on the relative test error in our setting. Figure 6b shows that the relative error decreases as the hyperparameter β in the filter function 59 is chosen smaller. The influence of γ in the considered range is only minor. The smallest relative error is obtained for a choice of $\gamma = 0.075$ and $\beta = 25.0$. Overall, however, the relative error depends only weakly on the exact choice of γ and β within a fairly large range: for example, the relative error for the pair $(\gamma = 0.055, \beta = 65)$ is 0.03833, while for the pair $(\gamma = 0.075, \beta = 25)$ it is 0.02715. This suggests that the structure of the filter function is the most important factor, and as mentioned at the beginning, precise parameter optimization can be omitted.



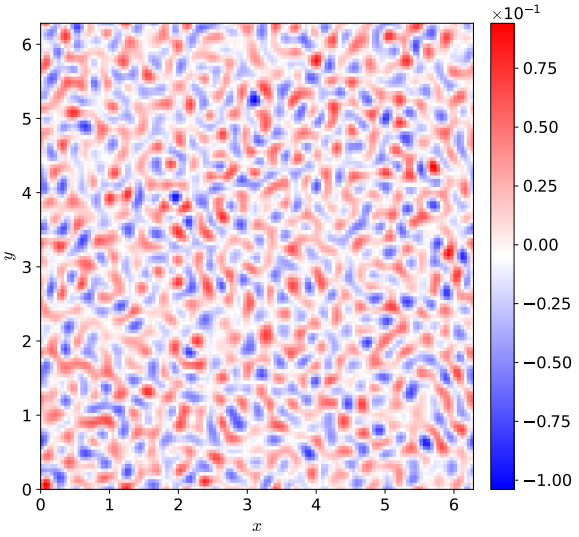
(a) Test error vs. λ

(b) Test error vs. γ and β

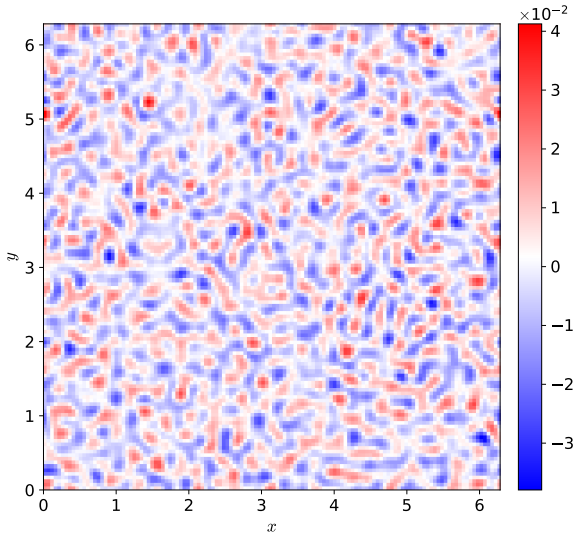
Figure 6: Dependence of the relative test error on the hyperparameters of the trained RFM for the Cahn-Hilliard solution operator $F^\dagger$ with $m = 512$, $n = 256$ training and $n' = 2000$ test pairs: (a) demonstrates the invariance of the test error for various regularization parameters λ, while (b) shows the decay of the test error as a function of the hyperparameters γ and β; the smallest error achieved is 0.02715 for $\gamma = 0.075$ and $\beta = 25.0$.



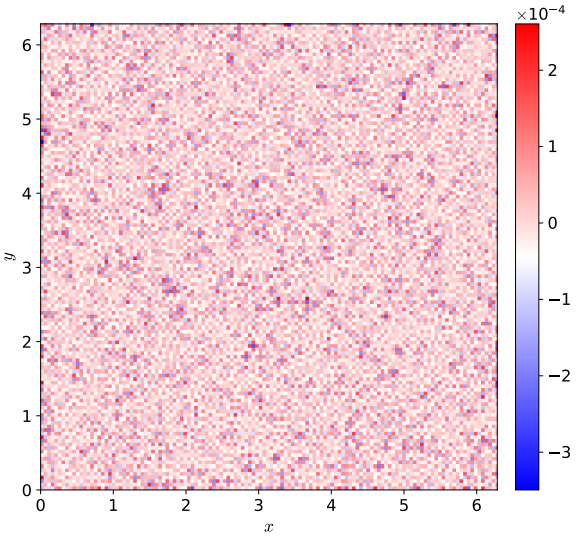
(a) Truth



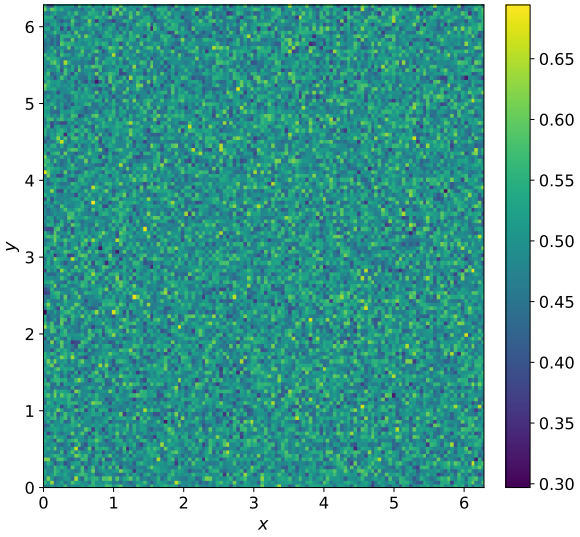
(b) Pointwise error for $m = 64$



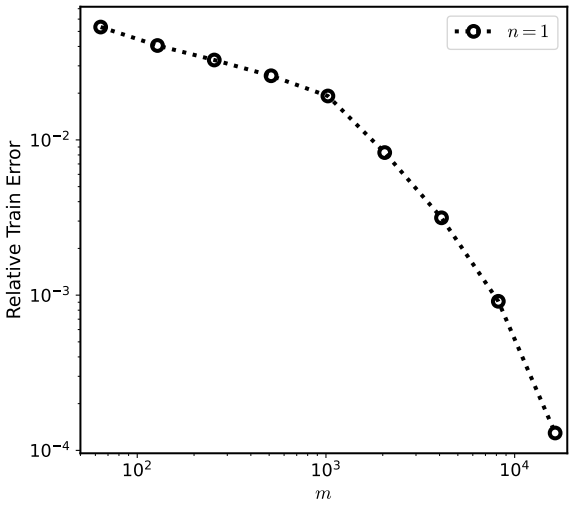
(c) Pointwise error for $m = 1024$



(d) Pointwise error for $m = 16384$



(e) Input



(f) Train error vs. m

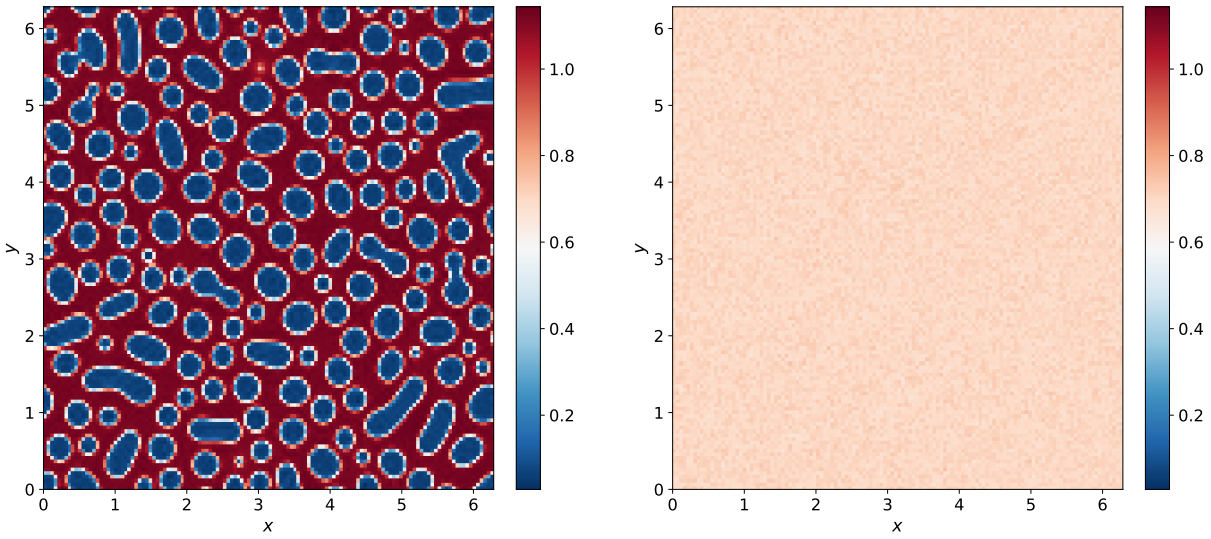
Figure 7: Results of a trained RFM for the Cahn-Hilliard solution operator $F^\dagger$ for one single input-output pair (a_1, y_1) at an early timestep: (a) displays y_1, (e) displays a_1. (b) - (d) show the pointwise error for different m, while (f) displays the relative error as a function of m. Here, $K = 128$ was chosen as spatial discretization.

4.2 Cahn-Hilliard Equation - Late Timesteps

In this section, we again consider the solution operator $F^\dagger$ 49 for the Cahn-Hilliard equation, but now for a late timestep where the two phases are already formed. The computational domain considered is again $\Omega = [0, 2\pi]^2$, and we generate a high-resolution dataset of input-output pairs using the same pseudospectral solver as in section 4.1. We assume a mobility constant of $M = 1.0$, $W = 2.0$ as height of the thermodynamic barrier and $\epsilon^2 = 0.5$ as gradient energy coefficient. For the adapted sigmoid function, we set the hyperparameter to $\beta = 100$.

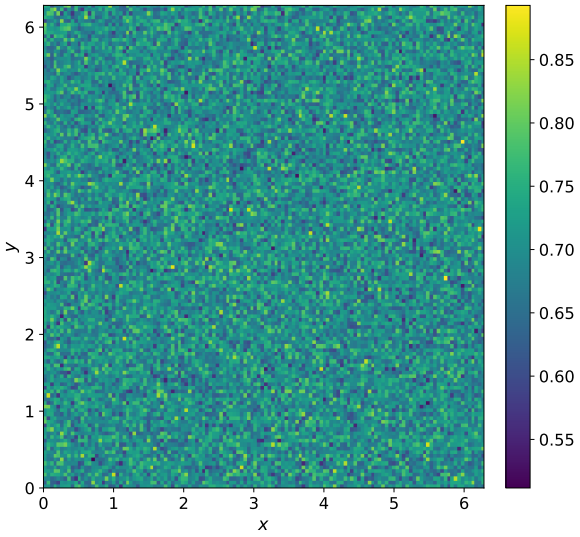
As shown in figure 8, our chosen random features are not suited to approximate the solution operator 49 at all. Even for $m = 8192$ random features, the operator cannot be approximated reliably - the randomness of a is overly dominant in the result. A potential remedy could be smoothing a, e.g. with a Matérn kernel. This will be subject to future work. An other alternative is to apply Fourier random features again also for late timesteps, but to adapt the filter function to the binary distribution. However, our experiments in this direction have not been successful so far.

However, if one considers only a single input–output pair (a_i, y_1) and approximates the solution using the random features in 65, as illustrated in figure 9, the result converges to y_1, but this time significantly more slowly compared to early timesteps and only for $\mathcal{O}(10^4)$ random features m. This convergence is consistent with the findings of [Lia24]. For operator learning, however, the random feature mapping 65 appears unsuitable, unless both m and n are chosen in the order of $\mathcal{O}(10^4) - \mathcal{O}(10^5)$, which however exceeds the VRAM capacity of our GPUs used in our experiments.

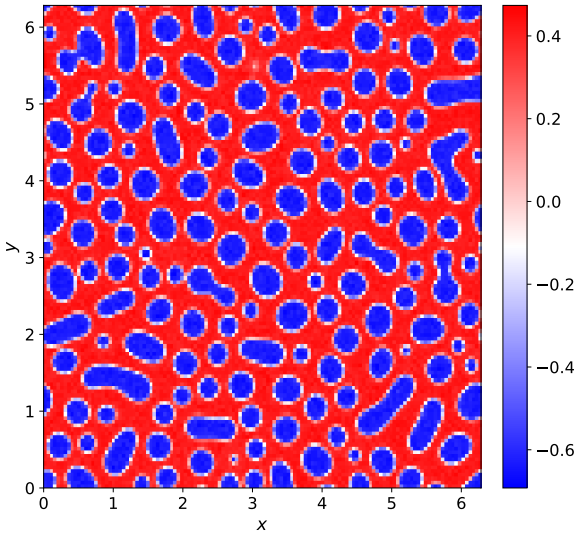


(a) Truth

(b) Approximation

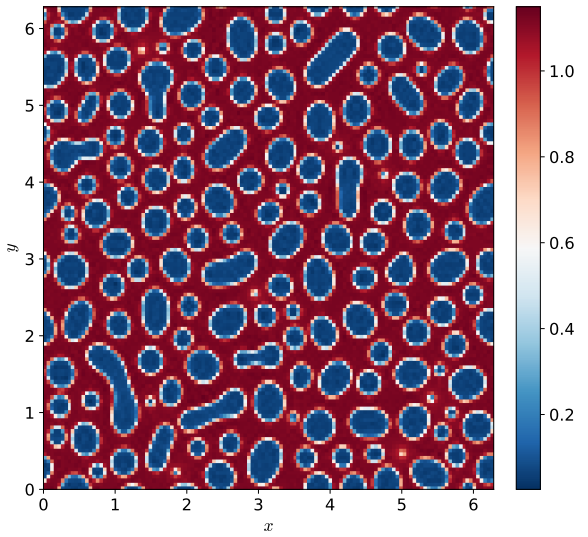


(c) Input

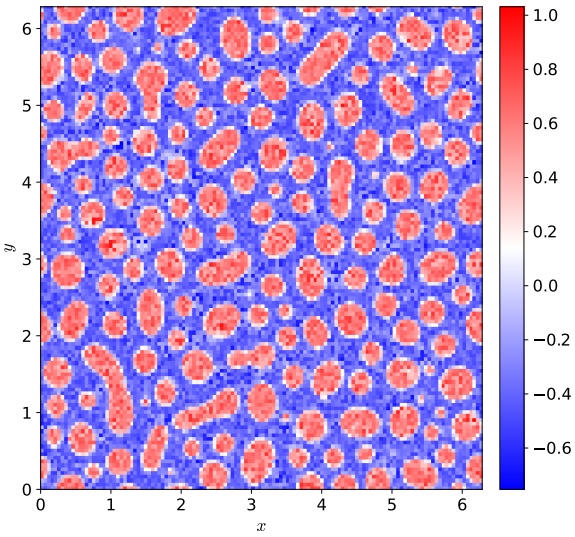


(d) Pointwise error

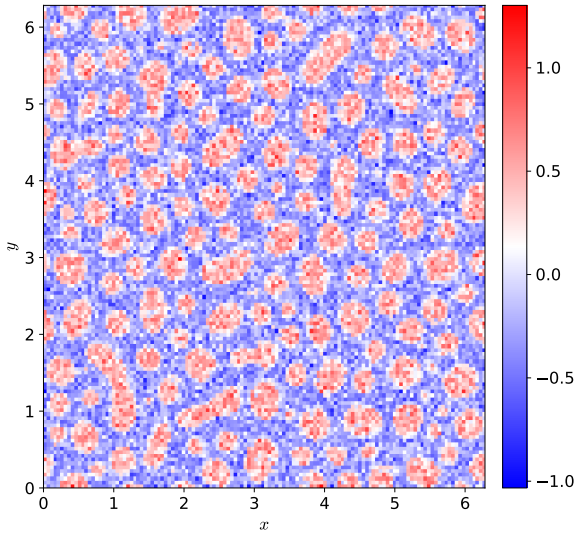
Figure 8: Representative input-output test sample for the solution of the Cahn-Hilliard equation at a late timestep: (c) shows a sample test input, (a) the resulting output (truth) from a pseudospectral solver, (b) the the tranined RFM prediction, while (d) displays the pointwise error. The relative L^2 error for this single prediction is 0.549. Here, $n = 1024$, $m = 8192$, and $K = 128$.



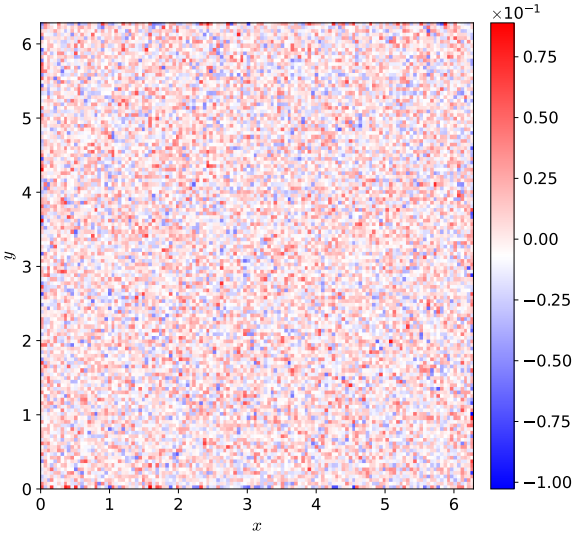
(a) Truth



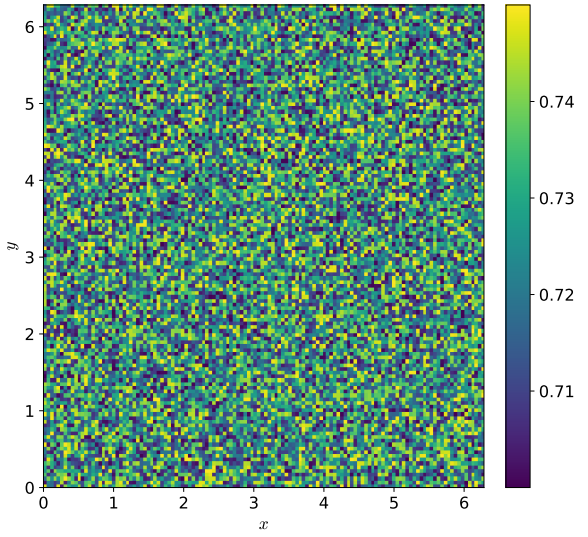
(b) Pointwise error for $m = 1024$



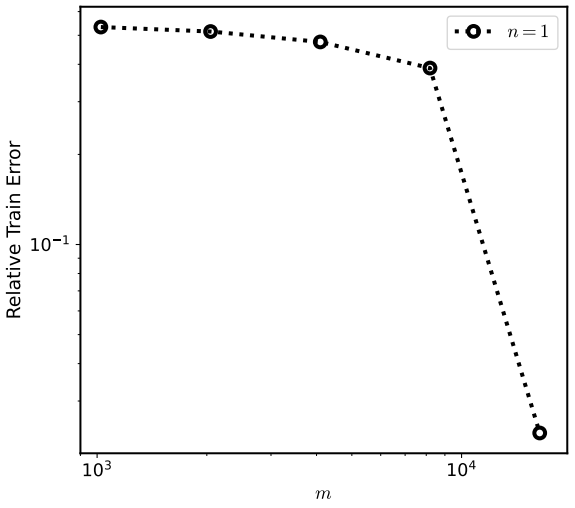
(c) Pointwise error for $m = 4096$



(d) Pointwise error for $m = 16384$



(e) Input



(f) Train error vs. m

Figure 9: Results of a trained RFM for the Cahn-Hilliard solution operator $F^\dagger$ for one single input-output pair (a_1, y_1) at a late timestep T: (a) displays y_1, (e) displays a_1. (b) - (d) show the pointwise error for different m, while (f) displays the relative error as a function of m. Here, $K = 128$ was chosen as spatial discretization.

5 Conclusion

In this work, we investigated the RFM as a scalable operator learning technique for approximating tumor growth models. We focused on the Cahn–Hilliard equation as the prototypical phase-field model for tumor growth and successfully demonstrated that the RFM is capable of accurately learning the initial-condition-to-solution operator, particularly at an early timestep where the two phases are just beginning to form.

Our first contribution (C1) involved applying the RFM to the stationary, linearized Cahn-Hilliard equation on a disk, where we provided a rigorous spectral characterization of the corresponding RKHS. In doing so, we established a solid theoretical foundation for applying random features in operator learning for fourth-order PDEs. In our second contribution (C2), we applied the RFM to the Cahn–Hilliard equation at early timesteps, where spinodal decomposition just begins. We found out, that by constructing a Fourier-space filter deduced from the dispersion relation of the Cahn–Hilliard equation, we selectively retained the physically relevant wavenumbers responsible for interface formation while suppressing both low-frequency background modes and high-frequency noise. Together with the ELU activation function, this filtering strategy allowed the model to accurately approximate the initial-condition-to-solution operator of the equation at early timesteps. The numerical experiments demonstrated that the random feature method achieves high accuracy in approximating the initial-condition-to-solution operator of the Cahn–Hilliard equation already with a moderate number of features, yielding relative errors on the order of $\mathcal{O}(10^{-2})$. The trained model also generalized seamlessly across different spatial resolutions, confirming the mesh-invariance property built into the methodology. Furthermore, the approximation error consistently decreased as the number of random features increased, in line with the theoretical convergence guarantees of the RFM. Our third contribution (C3) was the implementation of a highly parallelized and GPU-accelerated pipeline for training the RFM on large PDE datasets. The code is capable of approximating solution operators in a matter of minutes, demonstrating the method's practical viability.

Although the random feature method has proven effective for the Cahn–Hilliard equation at early timesteps, its application to more complex biological models, such as the four-species tumor growth model, remains an open challenge and could not be addressed in this work. Our results suggest that applying the RFM to such coupled systems will only be meaningful if the solution operator of the underlying Cahn-Hilliard equation can be approximated accurately also at late timesteps, when the binary interface has nearly fully developed. Future work should therefore focus on analyzing the Cahn–Hilliard equation at late timesteps. One possibility would be to rely on Fourier random features again and construct a suitable filter function depending on the solution spectrum of the equation for the timestep under consideration. Future work could also try to approximate the Cahn-Hilliard equation with a phase-dependent mobility. If this succeeds, the next logical step would be to consider weakly coupled systems before tackling the four-species model in its entirety.

Taken together, this thesis provides new insights into the capabilities of the RFM in the context of operator learning for phase-field models and contributes to the ongoing effort to develop efficient, physics-aware surrogates for models in mathematical oncology.

Data and Code Availability Links to datasets and code used to produce the numerical results and figures in this work are available at:
`https://github.com/JanJakob1/rfm-tumor-models`

Acknowledgments I thank Prof. Vincent Heuveline for providing the necessary framework and guidance while working on this exciting topic. I also want to thank Valentin Schmid and Isabel Gernand for helpful discussions, as well as their direct and spontaneous support regarding both feedback and computational resources - without their help, this work would not have been possible. Finally, I am very grateful to Lydia Cyriax for her invaluable help with administrative matters and for making sure everything ran smoothly, allowing me to focus on the research itself.

[Bha+21] K. Bhattacharya et al. “Model Reduction and Neural Networks for Parametric PDEs”. In: *SMAI J. Comput. Math.* 7 (2021), pp. 121–157.

[CH58] J. W. Cahn and J. E. Hilliard. “Free energy of a nonuniform system: I. Interfacial free energy”. In: *The Journal of Chemical Physics* 28.2 (1958), pp. 258–267. DOI: 10.1002/9781118788295.ch4. URL: https://doi.org/10.1002/9781118788295.ch4.

[CGG11] C. Cavaterra, C. G. Gal, and M. Grasselli. “Cahn–Hilliard equations with memory and dynamic boundary conditions”. In: *Asymptotic Analysis* 71.3 (2011), pp. 123–162. DOI: 10.3233/ASY-2010-1019.

[CGR+17] P. Colli, G. Gilardi, E. Rocca, et al. “Optimal distributed control of a diffuse interface model of tumor growth”. In: *Nonlinearity* 30.6 (2017), p. 2518. DOI: 10.1088/1361-6544/aa6e5f.

[EG19a] M. Ebenbeck and H. Garcke. “Analysis of a Cahn–Hilliard–Brinkman model for tumour growth with chemotaxis”. In: *Journal of Differential Equations* 266.9 (2019), pp. 5998–6036. DOI: 10.1016/j.jde.2018.10.045.

[EG19b] M. Ebenbeck and H. Garcke. “On a Cahn–Hilliard–Brinkman Model for Tumor Growth and Its Singular Limits”. In: *SIAM Journal on Mathematical Analysis* 51.3 (2019), pp. 1868–1912. DOI: 10.1137/18M1228104.

[EG96] C. M. Elliott and H. Garcke. “On the Cahn–Hilliard equation with degenerate mobility”. In: *SIAM Journal on Mathematical Analysis* 27.2 (1996), pp. 404–423. DOI: 10.1137/S0036141094267662.

[Esh+25] Mohammad Sadegh Eshaghi et al. *Multi-Head Neural Operator for Modelling Interfacial Dynamics*. 2025. arXiv: 2507.17763 [physics.comp-ph]. URL: https://arxiv.org/abs/2507.17763.

[FGR15] S. Frigeri, M. Grasselli, and E. Rocca. “On a diffuse interface model of tumour growth”. In: *European Journal of Applied Mathematics* 26.2 (2015), pp. 215–243. DOI: 10.1017/S0956792514000436.

[FLR17] S. Frigeri, K. F. Lam, and E. Rocca. “On a diffuse interface model for tumour growth with non-local interactions and degenerate mobilities”. In: *Solvability, Regularity, and Optimal Control of Boundary Value Problems for PDEs*. Ed. by P. Colli, A. Favini, E. Rocca, et al. Springer, 2017, pp. 217–254. DOI: 10.1007/978-3-319-64489-9_9.

[Fri23] M. Fritz. “Tumor Evolution Models of Phase-Field Type with Nonlocal Effects and Angiogenesis”. In: *Bulletin of Mathematical Biology* 85.6 (2023), p. 44. DOI: 10.1007/s11538-023-01151-6. URL: https://doi.org/10.1007/s11538-023-01151-6.

[FLO+19] M. Fritz, E. Lima, J. T. Oden, et al. “On the unsteady Darcy–Forchheimer–Brinkman equation in local and nonlocal tumor growth models”. In: *Mathematical Models and Methods in Applied Sciences* 29.9 (2019), pp. 1691–1731. DOI: 10.1142/s0218202519500325.

[GL16] H. Garcke and K. F. Lam. “Global weak solutions and asymptotic limits of a Cahn–Hilliard–Darcy system modelling tumour growth”. In: *AIMS Mathematics* 1.3 (2016), pp. 318–360. DOI: 10.3934/Math.2016.3.318.

[GL17a] H. Garcke and K. F. Lam. “Analysis of a Cahn–Hilliard system with non-zero Dirichlet conditions modeling tumor growth with chemotaxis”. In: *Discrete Continuous Dynamical Systems Series A* 37.8 (2017), pp. 4277–4308. DOI: 10.3934/dcds.2017183.

[GL17b] H. Garcke and K. F. Lam. “Well-posedness of a Cahn–Hilliard system modelling tumour growth with chemotaxis and active transport”. In: *European Journal of Applied Mathematics* 28.2 (2017), pp. 284–316. DOI: 10.1017/S0956792516000292.

[GLN+18] H. Garcke, K. F. Lam, R. Nürnberg, et al. “A multiphase Cahn–Hilliard–Darcy model for tumour growth with necrosis”. In: *Mathematical Models and Methods in Applied Sciences* 28.3 (2018), pp. 525–577. DOI: 10.1142/S0218202518500148.

[GLS+16] H. Garcke, K. F. Lam, E. Sitka, et al. “A Cahn–Hilliard–Darcy model for tumour growth with chemotaxis and active transport”. In: *Mathematical Models and Methods in Applied Sciences* 26.6 (2016), pp. 1095–1148. DOI: 10.1142/S0218202516500263.

[GLW01] Shuguang Guan, C.-H Lai, and G.W Wei. “Fourier–Bessel analysis of patterns in a circular domain”. In: *Physica D: Nonlinear Phenomena* 151.2 (2001), pp. 83–98. ISSN: 0167-2789. DOI: https://doi.org/10.1016/S0167-2789(01)00223-8. URL: https://www.sciencedirect.com/science/article/pii/S0167278901002238.

[Gur95] M. E. Gurtin. “Generalized Ginzburg–Landau and Cahn–Hilliard equations based on a microforce balance”. In: *Physica D: Nonlinear Phenomena* 92.3-4 (1995), pp. 178–192. DOI: 10.1016/0167-2789(95)00173-5.

[HZO12] A. Hawkins-Daarud, K. G. van der Zee, and J. T. Oden. “Numerical simulation of a thermodynamically consistent four-species tumor growth model”. In: *International Journal for Numerical Methods in Biomedical Engineering* 28.1 (2012), pp. 3–24. DOI: 10.1002/cnm.1467.

[He21] J. He. “Global weak solutions to a Navier–Stokes–Cahn–Hilliard system with chemotaxis and singular potential”. In: *Nonlinearity* 34.4 (2021), pp. 2155–2184. DOI: 10.1088/1361-6544/abc596. URL: https://doi.org/10.1088/1361-6544/abc596.

[KLS24] Nikola B. Kovachki, Samuel Lanthaler, and Andrew M. Stuart. *Operator Learning: Algorithms and Analysis*. 2024. arXiv: 2402.15715 [cs.LG]. URL: https://arxiv.org/abs/2402.15715.

[LW17] K. F. Lam and H. Wu. “Thermodynamically consistent Navier–Stokes–Cahn–Hilliard models with mass transfer and chemotaxis”. In: *European Journal of Applied Mathematics* 29.4 (2017), pp. 595–644. DOI: 10.1017/S0956792517000298.

[Li+20] Zongyi Li et al. “Neural Operator: Graph Kernel Network for Partial Differential Equations”. In: *arXiv preprint arXiv:2003.03485* (2020).

[Li+21] Zongyi Li et al. “Fourier Neural Operator for Parametric Partial Differential Equations”. In: *International Conference on Learning Representations (ICLR)*. 2021.

[Lia24] Chunyang Liao. *Solving Partial Differential Equations with Random Feature Models*. 2024. arXiv: 2501.00288 [math.NA]. URL: https://arxiv.org/abs/2501.00288.

[LAO15] E. Lima, R. C. Almeida, and J. T. Oden. “Analysis and numerical solution of stochastic phase-field models of tumor growth”. In: *Numerical Methods for Partial Differential Equations* 31.2 (2015), pp. 552–574. DOI: 10.1002/num.21934.

[LOA14] E. Lima, J. T. Oden, and R. C. Almeida. “A hybrid ten-species phase-field model of tumor growth”. In: *Mathematical Models and Methods in Applied Sciences* 24.13 (2014), pp. 2569–2599. DOI: 10.1142/S0218202514500304.

[LJK21] Lu Lu, Pengzhan Jin, and George Em Karniadakis. “Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators”. In: *Nature Machine Intelligence* 3.3 (2021), pp. 218–229.

[Mir19] Alain Miranville. *The Cahn–Hilliard Equation: Recent Advances and Applications*. Society for Industrial and Applied Mathematics, 2019. DOI: 10.1137/1.9781611975925.

[NS24] Nicholas H. Nelsen and Andrew M. Stuart. "Operator Learning Using Random Features: A Tool for Scientific Computing". In: *SIAM Rev.* 66.3 (2024), pp. 535–571. URL: https://doi.org/10.1137/24m1648703.

[RR07] Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines". In: *Advances in Neural Information Processing Systems*. Ed. by J. Platt et al. Vol. 20. Curran Associates, Inc., 2007. URL: https://proceedings.neurips.cc/paper_files/paper/2007/file/013a006f03dbc5392effeb8f18fda755-Paper.pdf.

[SBT23] Elvis do A. Soares, Amaro G. Barreto, and Frederico W Tavares. "Exponential Integrators for Phase-Field Equations using Pseudo-spectral Methods: A Python Implementation". In: (May 2023), pp. 1–12. arXiv: 2305.08998. URL: http://arxiv.org/abs/2305.08998.

[Sun+21] H. Sung et al. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". In: *CA: a cancer journal for clinicians* 71.3 (2021), pp. 209–249. DOI: 10.3322/caac.21660. URL: https://doi.org/10.3322/caac.21660.

[WS21] Henk J. van Waarde and Rodolphe Sepulchre. *Training Lipschitz continuous operators using reproducing kernels*. 2021. arXiv: 2112.03344 [math.OC]. URL: https://arxiv.org/abs/2112.03344.