Optimal Control of Material Micro-Structures

Aayushman Sharma, Zirui Mao, Haiying Yang, Suman Chakravorty, Michael J Demkowicz, Dileep Kalathil

Abstract—In this paper, we consider the optimal control of material micro-structures. Such material micro-structures are modeled by the so-called phase field model. We study the underlying physical structure of the model and propose a data based approach for its optimal control, along with a comparison to the control using a state of the art Reinforcement Learning (RL) algorithm. Simulation results show the feasibility of optimally controlling such micro-structures to attain desired material properties and complex target micro-structures.

I. INTRODUCTION

In this paper, we consider the optimal control of complex and high DoF (Degree of Freedom) material micro-structure dynamics, by doing which we aim at developing a robust tool for exploring the fundamental limits of materials micro-structure control during processing. The micro-structure dynamics process is governed by the so-called Phase Field (PF) model [1], which is a powerful tool for modeling micro-structure evolution of materials. It represents the spatial distribution of physical quantities of interest by an order parameter field governed by one or a set of partial differential equations (PDEs). The phase field method can naturally handle material systems with complicated interface, thanks to which it has been successfully applied to the simulation of a wide range of materials micro-structure evolution processes, such as e.g., alloy solidification [2], alloy phase transformation [3], and failure mechanism of materials [4] etc.

The phase field model is complex, nonlinear, and high DoF, and does not admit analytical solutions. Thus, data based approaches are natural to consider for the control of these systems. In recent work [5], we proposed a novel decoupled data based control (D2C) algorithm for learning to control an unknown nonlinear dynamical system. Our approach introduced a rigorous decoupling of the open-loop (planning) problem from the closed-loop (feedback control) problem. This decoupling allows us to come up with a highly efficient approach to solve the problem in a completely data based fashion. Our approach proceeds in two steps: (i) first, we optimize the nominal open-loop trajectory of the system using a ‘blackbox simulation’ model, (ii) then we identify the linear system governing perturbations from the nominal trajectory using random input-output perturbation data, and design an LQR controller for this linearized system. We have shown that the performance of D2C algorithm is approximately optimal, in the sense that the decoupled design is near optimal to second order in a suitably defined noise parameter [5]. In this work, we apply the D2C algorithm to the optimal control of the evolution of material micro-structures governed by the phase field model. For comparison, we consider RL techniques [6]. The past several years have seen significant progress in deep neural networks based reinforcement learning approaches for controlling unknown dynamical systems, with applications in many areas like playing games [7], locomotion [8] and robotic hand manipulation [9]. A number of new algorithms that show promising performance are proposed [10][11][12] and various improvements and innovations have been continuously developed. However, despite excellent performance on a number of tasks, RL is highly data intensive. The training time for such algorithms is typically very large, and high variance and reproducibility issues mar the performance [13]. Thus, materials micro-structure control is intractable for current RL techniques. Recent work in RL has also focused on PDE control using techniques like the Deep Deterministic Policy Gradient (DDPG) with modifications such as action descriptors to limit the large action spaces in infinite dimensional systems [14], as such algorithms tend to degrade in performance with an increase in dimensionality of the action-space. However, such an approach, in general, is not feasible for the micro-structure problem.

The contributions of the paper are as follows: we show how to model the dynamics of a multi-phase micro-structure and unveil its underlying structure. We present the application of the D2C approach, and a state of the art RL technique (DDPG) to the control of such microstructure dynamics for the first time in the literature, to the best of our knowledge. Our results show that the local D2C approach outperforms the global RL approach such as DDPG when operating on higher dimensional phase field PDEs. We also show that the D2C approach is robust to noise in the practical regimes, and that global optimality can be recovered in higher noise levels through an open-loop replanning approach. Furthermore, we can exploit the peculiar physical properties of material micro-structure dynamics to optimize the control actuation architecture that leads to highly computationally efficient control that does not sacrifice much performance. We envisage that this work is a first step towards the construction of a systematic feedback control synthesis approach for the control of material micro-structures, potentially scalable to real applications.

The rest of the paper is organized as follows: In Section II, the micro-structure dynamics are expanded upon, describing the different PDEs tested. We propose the decoupled data based control (D2C) approach in Section III. In Section IV, the proposed approach is illustrated through custom-defined test problems of varying dimensions, along with testing robustness to noise, followed by improvements to the algorithms which exploit the physics of the system.
II. MICROSTRUCTURE DYNAMICS: CLASSES OF PDEs IMPLEMENTED

This section provides a brief overview of the non-linear dynamics model of a general multi-phase micro-structure. We assume an infinitely large, 2-dimensional structure satisfying *periodical boundary* conditions.

A. The material system

The evolution of material micro-structures can be represented by two types of partial differential equations, i.e., the Allen-Cahn equation [15] representing the evolution of a non-conserved quantity, and the Cahn-Hilliard equation [16] representing the evolution of a conserved quantity. The Allen-Cahn equation has a general form of

$$\frac{\partial \phi}{\partial t} = -M(\frac{\partial F}{\partial \phi} - \gamma \nabla^2 \phi)$$  \hspace{1cm} (1)

while the Cahn-Hilliard equation has the form

$$\frac{\partial \phi}{\partial t} = \nabla \cdot M(\frac{\partial F}{\partial \phi} - \gamma \nabla^2 \phi)$$  \hspace{1cm} (2)

where $\phi = \phi(x, t)$ is called the ‘order parameter’, which is a spatially varying quantity. In Controls parlance, $\phi$ is the state of the system, and is infinite dimensional, i.e., a spatio-temporally varying function. It reflects the component proportion of each phase of material system. For a two-phase system studied in this work, $\phi = -1$ represents one pure phase and $\phi = 1$ represent the other, while $\phi \in (-1, 1)$ stands for a combination state of both phases on the boundary interface between two pure phases; $M$ is a parameter related to the mobility of material, which is assumed *constant* in this study; $F$ is an energy function with a non-linear dependence on $\phi$; $\gamma$ is a gradient coefficient controlling the diffusion level or thickness of the boundary interface.

In essence, the Phase Field Method is one of gradient flow methods, which means the evolution process follows the path of steepest descent in an energy landscape starting from an initial state until arriving at a a local minimum. So, the behavior of micro-structures in phase field modeling highly depends on the selection of energy density function $F$. For instance, the double-well potential function owing to two minima at $\phi = \pm 1$ and separated by a maximum at $\phi = 0$ (as plotted in Fig. 1) will cause the field $\phi$ to separate into regions where $\phi \approx \pm 1$, divided by the boundary interface, while the single-well potential function owing to a single minimum at $\phi = 0$ (see Fig. 1) predicts a gradual smoothing of any initial non-uniform $\phi$ field, yielding a uniform field with $\phi \approx 0$ in the long-time limit.

Accordingly, the evolution of material micro-structure can be governed by selecting proper form of energy density function $F$. Controlling the evolution process of micro-structures represented by Allen-Cahn equation and Cahn-Hilliard diffusion equation are completely distinct. The former one is governed through generating or deleting order parameter straightforwardly, while the latter is done through guiding the transport and redistribution of the conserved order parameter across the whole domain.

![Fig. 1: Illustration of double-well and single-well potential function F.](image)

In this study, we adopt the following general form of energy density function $F$:

$$F(\phi; T, h) = \phi^4 + T \phi^2 + h \phi$$  \hspace{1cm} (3)

Here, the parameters $T$ and $h$ are not governed by the A-C and C-H equations. Instead, they may be externally set to any value and may be spatially as well as temporally varying.

III. DECOUPLED DATA BASED CONTROL (D2C) ALGORITHM

In this section, we propose a data-based approach to solve the material phase-separation control problem. In particular, we apply the so-called decoupled data-based control (D2C) algorithm that we recently proposed [5], and extend it for high dimensional applications. The D2C algorithm is a highly data-efficient Reinforcement Learning (RL) method that has shown to be much superior to state of the art RL algorithms such as the Deep Deterministic Policy Gradient (DDPG) in terms of data efficiency and training stability while retaining similar or better performance. In the following, we give a very brief introduction to the D2C algorithm (please see [5], [17], [18] for the relevant details).

Let us reconsider the dynamics of the system given in Eq. (1) and (2) and rewrite it in a discrete time, noise perturbed state form as: $\Phi_{t+1} = f(\Phi_t) + g(\Phi_t)(U_t + \epsilon w_t)$, where $\Phi_t$ is the state, $U_t$ is the control, $w_t$ is a white noise perturbation to the system, and $\epsilon$ is a small parameter that modulates the noise in the system. Suppose now that we have a finite horizon objective function: $J(\Phi_0) = E[\sum_{t=0}^{T-1} c(\Phi_t, U_t) + \Psi(\Phi_T)]$, where $c(\Phi, u)$ denotes a running incremental cost, $\Psi(\Phi)$ denotes a terminal cost function and $E[.]$ is an expectation operator taken over the sample paths of the system. The objective of the control design is then to design a feedback policy $U_t(\Phi_t)$ that minimizes the cost function above and is given by $J^*(\Phi_0) = \min_{U_t(\cdot)} E[\sum_{t=0}^{T-1} c(\Phi_t, U_t) + \Psi(\Phi_T)]$.

The D2C algorithm then proposes a 3 step procedure to approximate the solution to the above problem.

First, a noiseless open-loop optimization problem is solved to find an optimal control sequence, $U_t^*$, that generates the nominal state trajectory $\Phi_t^*$:

$$J^*(\Phi_0) = \min_{U_t} \sum_{t=0}^{T-1} c(\Phi_t, U_t) + \Psi(\Phi_T)).$$  \hspace{1cm} (4)

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
subject to the zero noise nominal dynamics: \( \Phi_{t+1} = f(\Phi_t) + g(\Phi_t)U_t \).

Second, the perturbed time varying linear system around the nominal given by \( \delta\Phi_{t+1} = A_r\delta\Phi_t + B_r\delta U_t \) is identified in terms of the time varying system matrices \( A_r, B_r \).

Third, an LQR controller for the above time varying linear system is designed whose time varying gain is given by \( K_t \). Finally, the control applied to the system is given by \( U_t(\Phi_t) = U_t^* + K_t\delta\Phi_t \).

The following subsections provide details for each of the above-mentioned steps of the D2C algorithm.

A. Open Loop Trajectory Optimization

We utilize an ILQR based method to solve the open-loop optimization problem. ILQR typically requires the availability of analytical system Jacobian, and thus, cannot be directly applied when such analytical gradient information is unavailable (much like Nonlinear Programming software whose efficiency depends on the availability of analytical gradients and Hessians). In order to make it an (analytical) model-free algorithm, it is sufficient to obtain estimates of the system Jacobians from simulations, and a sample-efficient randomized way of doing so is described in the following subsection. Since ILQR is a well-established framework, we skip the details for the paucity of space.

1) Estimation of Jacobians: Linear Least Squares by Central Difference (LLS-CD): Using Taylor’s expansions of the dynamics ‘h’, where \( h(\Phi, U) = f(\Phi) + g(\Phi)U \) is the non-linear model of Section 2), about the nominal trajectory \( (\bar{\Phi}_t, \bar{U}_t) \) on both the positive and the negative sides, we obtain the following central difference equation difference:

\[
h(\Phi_t + \delta\Phi_t, \bar{U}_t + \delta U_t) - h(\Phi_t - \delta\Phi_t, \bar{U}_t - \delta U_t)
= 2[h_{\Phi_t}, h_{U_t}]^T \left[ \begin{array}{c} \delta\Phi_t \\ \delta U_t \end{array} \right]
+ O(\|\delta\Phi_t\|_3^3 + \|\delta U_t\|_3^3).
\]

Multiplying by \( [\delta\Phi_t^T \delta U_t^T]^T \) on both sides to the above equation and apply standard Least Square method:

\[
[h_{\Phi_t}, h_{U_t}] = H \delta Y_t^T (\delta Y_t \delta Y_t^T)^{-1},
\]

where \( H =
\[
\begin{pmatrix}
h(\Phi_t + \delta\Phi_t^1, \bar{U}_t + \delta U_t^1) - h(\Phi_t - \delta\Phi_t^1, \bar{U}_t - \delta U_t^1)
h(\Phi_t + \delta\Phi_t^2, \bar{U}_t + \delta U_t^2) - h(\Phi_t - \delta\Phi_t^2, \bar{U}_t - \delta U_t^2)
\vdots
h(\Phi_t + \delta\Phi_t^n, \bar{U}_t + \delta U_t^n) - h(\Phi_t - \delta\Phi_t^n, \bar{U}_t - \delta U_t^n)
\end{pmatrix}
\]

and \( 'n' \) be the number of samples for each of the random variables, \( \delta\Phi_t \) and \( \delta U_t \). Denote the random samples as \( \delta X_t = [\delta\Phi_t^1 \delta\Phi_t^2 \ldots \delta\Phi_t^n], \delta U_t = [\delta U_t^1 \delta U_t^2 \ldots \delta U_t^n], \) and \( \delta Y_t = [\delta Y_t^1 \delta Y_t^2 \ldots \delta Y_t^n] \).

We are free to choose the distribution of \( \delta\Phi_t \) and \( \delta U_t \). We assume both are i.i.d. Gaussian distributed random variables with zero mean and a standard deviation of \( \sigma \). This ensures that \( \delta Y_t \delta Y_t^T \) is invertible.

Let us consider the terms in the matrix \( \delta Y_t \delta Y_t^T = \delta Y_t \delta X_t \delta X_t^T \delta Y_t \).

Similarly, \( \delta U_t \delta U_t^T = \delta U_t \delta X_t \delta X_t^T \delta U_t \).

From the definition of sample variance, for a large enough \( n \), we can write the above matrix as

\[
\delta Y_t \delta Y_t^T = \sum_{i=1}^{n} \delta\Phi_t i \delta\Phi_t i^T \approx \frac{\sigma^2(n-1)\delta Y_t \delta Y_t^T}{\sigma^2(n-1)\delta X_t \delta X_t^T}.
\]

B. Linear Time Varying System Identification

Closed-loop control design in step 2 of D2C requires the knowledge of the linearized system parameters \( A_t \) and \( B_t \) for \( 0 \leq t \leq T - 1 \). Here we use the standard least square method to estimate these parameters from input-output experiment data.

First start from the perturbed linear system about the nominal trajectory and estimate the system parameters \( A_t \) and \( B_t \) from: \( \delta\Phi_{t+1} = A_t \delta\Phi_t + B_t \delta U_t \), where \( \delta\Phi_t^{(n)} \) is the state perturbation vector and \( \delta U_t^{(n)} \) is the control perturbation vector we feed to the system at step \( t \), \( \delta Y_t^{(n)} \) simulation. All the perturbations are zero-mean, i.i.d. Gaussian noise with covariance matrix \( \sigma \). This ensures a \( \delta Y_t \delta Y_t^T \) small value selected by the user. \( \delta\Phi_t^{(n)} \) denotes the deviation of the output state vector from the nominal state after propagating for one step.

Run \( N \) simulations for each step and collect the input-output data: \( Y = [A_t | B_t]X \) and write out the components: \( Y = [\delta Y_t^{(1)} \delta Y_t^{(2)} \ldots \delta Y_t^{(N)}], X = [\delta\Phi_t^{(1)} \delta\Phi_t^{(2)} \ldots \delta\Phi_t^{(N)}]. \)

Finally, using the standard least square method, the linearized system parameters are estimated as \( [\hat{A}_t | \hat{B}_t] = YX^T(XX^T)^{-1} \).

C. Closed Loop Control Design

Given the estimated perturbed linear system, we design a finite horizon, discrete time LQR [19] along the trajectory for each time step to minimize the cost function, and use the standard Riccati equations to solve for the feedback gains.

D. D2C Algorithm: Summary

The Decoupled Data Based Control (D2C) Algorithm is summarized in Algorithm 1.

IV. SIMULATION RESULTS

In this section, we compare the training and performance of the data-based D2C approach, with the RL-based Deep Deterministic Policy Gradient (DDPG), on a material with control inputs as the temperature and external field inputs on subsets of grid points.

A. Structure and Task

We simulated the phase-separation dynamics in Python, through calling an explicit, second-order solver subroutine in FORTRAN. The system and its tasks are defined as follows:
Algorithm 1: D2C Algorithm
1) Solve the deterministic open-loop optimization problem for optimal open-loop control sequence and state trajectory \((\hat{U}_t)_{t=0}^{T-1}, \{\hat{\Phi}_t\}_{t=0}^{T-1}\) using ILQR (Section III-A).
2) Linearize and identify the LTV system parameters \((A_t, B_t)\) via least square (Section III-B).
3) Solve the Riccati equations for each step along the nominal trajectory for feedback gain \(\{K_t\}_{t=0}^{T-1}\).
4) Apply the closed-loop control policy,

\[
\begin{align*}
U_t &= \hat{U}_t^* + K_t \delta \Phi_t, \\
\delta \Phi_{t+1} &= \Phi_{t+1} - \hat{\Phi}_{t+1}^* \\
t &= t + 1.
\end{align*}
\]

end while

C. Training and Testing

D2C implementation is done in three stages corresponding to those mentioned in the previous section, and a black box phase field model is simulated in Python.

Training: The open-loop training plots in Fig. 3 show the cost curve during training. After the cost curves converge, we get the optimal control sequence that could drive the systems to accomplish their tasks. The training parameters and outcomes are summarized in (Tables I, II, III). We note that the training time is dominated by the Python simulation code, and can be made significantly faster with an optimized and parallelized code. Thus, we envisage this work as an initial proof of the concept for controlling material microstructures.

The open-loop optimal trajectory learned by the D2C algorithm after convergence (for goal state-III, Allen-Cahn PDE) is shown in Figure 4.

### TABLE I: Comparison of the training outcomes of D2C with DDPG for the Allen-Cahn Model.

| Goal State  | Training time (in sec.) |
|-------------|-------------------------|
| Goal-I(10X10) | 13.83                  |
| Goal-II(20X20) | 55.206                 |
| Goal-III(50X50) | 3289.354              |

The open-loop training is run on a laptop with a 2-core CPU@2.9GHz and 12GB RAM. No multi-threading at this point.

* implies algorithm cannot converge to the goal state.
** implies system running out of memory at run-time.

### TABLE II: Comparison of the training outcomes of D2C with DDPG for the Cahn-Hilliard Model.

| Goal State  | Training time (in sec.) |
|-------------|-------------------------|
| Goal-I(10X10) | 141.573                |
| Goal-II(20X20) | 3083.223              |

The open-loop training is run on a laptop with a 2-core CPU@2.9GHz and 12GB RAM. No multi-threading at this point.

* implies algorithm cannot converge to the goal state.

### TABLE III: Parameter size comparison between D2C and DDPG

| System | No. of steps | No. of actuators | No. of parameters optimized in D2C | No. of parameters optimized in DDPG |
|--------|--------------|------------------|------------------------------------|-------------------------------------|
| Goal-I | 10            | 200              | 2000                               | 461901                              |
| Goal-II| 10            | 800              | 8000                               | 1122501                             |
| Goal-III| 10              | 5000             | 50000                              | 3746701                             |

Testing Criterion: For the closed-loop design, we proceed with the system identification and feedback gain design step of the D2C algorithm mentioned in the previous section to get the closed-loop control policy. For testing, we compare the performance between the open-loop D2C control policy and the closed-loop D2C control policy under different noise levels. The open-loop control policy is to apply the optimal control sequence solved in the open-loop training step without any feedback. Hence, the perturbation drives the model off the nominal trajectory and increases the episodic cost as the noise level increases. Zero-mean Gaussian i.i.d. noise is added to every control channel at each step. The standard
deviation of the noise is proportional to the maximum control signal in the designed optimal control sequence for D2C.

The feedback policy obtained appears to be near-optimal and highly robust for low noise levels (Fig 5), which is as expected from the \(O(e^4)\)-optimality of the algorithm used [18]. But when operating in high noise regimes, the robustness of the feedback policy quickly degenerates. Thus, to reduce the variance of the policy in these domains, we advocate the use of a rapid open loop solver along with replanning at every step, i.e. a Model-Predictive Control approach, to solve the RL problem, since MPC should be able to recover the global optimality for these problems [20].

The recursive MPC algorithm implemented, which uses a fast and reliable local-planner, is summarised in Algorithm 2, and Figure 6 compares the robustness to noise between the stochastic policy obtained from the recursive MPC with the decoupled closed-loop feedback policy at varying noise regimes.

**Algorithm 2: Recursive MPC Algorithm**

1) Given: Initial state \(\Phi_0\), time horizon \(T\), cost \(c(\Phi, U) = l(\Phi) + \frac{1}{2} r U^2\), and terminal cost \(c_T(\Phi)\).
2) Set \(H = T\), \(\Phi_1 = \Phi_0\).

while \(H > 0\) do

1) Solve the open-loop(deterministic) optimal control problem for initial state \(\Phi_1\) and horizon \(H\). Let the optimal sequence be \(U^* = \{U_0, U_1, \ldots, U_{H-1}\}\).
2) Apply the first control \(U_0\) to the stochastic system, and observe the next state \(\Phi_n\).
3) Set \(H = H - 1\), \(\Phi_t = \Phi_n\). 

end while
We seek to implement a smarter actuator-selection scheme that can quickly generate the closed-loop feedback policy. We observe from experiments that the DDPG algorithm is not able to converge to the goal state for the material system of state-dimension over 5x5 discretization (25 states, 50 controls), while the D2C algorithm can quickly generate the closed-loop feedback solution even for the more complex systems.

**V. CONCLUSIONS**

In this article, we have provided an overview of the modeling of multi-phase micro-structures in accordance with the Allen-Cahn and Cahn-Hilliard equations. We have also compared learning/data-based approaches to the control of such structures, outlining their relative merits and demerits. We seek to implement a smarter actuator-selection scheme to reduce the action space, which may improve viability of DDPG for higher order systems, while also significantly improving performance of D2C. We also plan to extend the methodology for higher-dimensional materials, leading up to an implementation on a realistic full-scale model.

**VI. ACKNOWLEDGEMENTS**

This work was supported by the NSF CDS&E program under grant number 1802867.

**REFERENCES**

[1] K. E. Nikolias Provatas, *Phase Field Methods in Materials Science and Engineering*. John Wiley & Sons, Ltd, 2010.

[2] W. J. Boettingter, J. A. Warren, C. Beckermann, and A. Karma, “Phase-field simulation of solidification,” *Annual Review of Materials Research*, vol. 32, no. 1, pp. 163–194, 2002.

[3] L. Nguyen, R. Shi, Y. Wang, and M. De Graef, “Quantification of rafting of γ’ precipitates in Ni-based superalloys,” *Acta Materialia*, vol. 103, pp. 322–333, 2016.

[4] S. Khaderi, P. Murali, and R. Ahluwalia, “Failure and toughness of bio-inspired composites: Insights from phase field modelling,” *Computational Materials Science*, vol. 95, pp. 1 – 7, 2014.

[5] R. Wang, K. S. Parunandi, D. Yu, D. Kalathil, and S. Chakravorty, “Decoupled data-based approach for learning to control nonlinear dynamical systems,” *IEEE Transactions on Automatic Control*, vol. 67, no. 7, pp. 3582–3589, 2022.

[6] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[7] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot et al., “Mastering the game of go with deep neural networks and tree search,” *nature*, vol. 529, no. 7587, p. 484, 2016.

[8] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.

[9] S. Levine, C. Finn, T. Darrell, and P. Abbeel, “End-to-end training of deep visuomotor policies,” *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1334–1373, 2016.

[10] W. Yuhuai, M. Elman, L. Shun, G. Roger, and B. Jimmy, “Scalable trust-region method for deep reinforcement learning using kroenecker-factored approximation,” *arXiv:1708.05144*, 2017.

[11] S. John, L. Sergey, M. Philipp, J. Michael I., and A. Pieter, “Trust region policy optimization,” *arXiv:1502.05477*, 2017.

[12] S. John, W. Filip, D. Prafulla, R. Alec, and K. Oleg, “Proximal policy optimization algorithms,” *arXiv:1707.06347*, 2017.

[13] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meier, “Deep reinforcement learning that matters,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[14] Y. Pan, A.-m. Farahmand, M. White, S. Nabi, P. Grover, and D. Nikovski, “Reinforcement learning with function-valued action spaces for partial differential equation control,” in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 3986–3995. [Online]. Available: https://proceedings.mlr.press/v80/pan18a.html

[15] S. M. Allen and J. W. Cahn, “A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening,” *Acta Metallurgica*, vol. 27, no. 6, pp. 1085 – 1095, 1979.

[16] J. W. Cahn and J. E. Hilliard, “Free energy of a nonuniform system. i. interfacial free energy,” *J. Chem. Phys.*, vol. 28, no. 2, p. 258–267, 1958.

[17] D. Yu, M. Rafieisakhaei, and S. Chakravorty, “Stochastic Feedback Control of Systems with Unknown Nonlinear Dynamics,” in *56th IEEE Conference on Decision and Control (CDC)*, 2017.

[18] R. Wang, K. S. Parunandi, A. Sharma, R. Goyal, and S. Chakravorty, “On the search for feedback in reinforcement learning,” in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 1560–1567. [Online]. Available: https://arxiv.org/abs/2102.09478.

[19] A. E. Bryson and Y.-C. Ho, *Applied Optimal Control: Optimization, Estimation, and Control*. Routledge, New York, 1975.

[20] M. N. G. Mohamed, S. Chakravorty, R. Goyal, and R. Wang, “On the optimal feedback law in stochastic optimal nonlinear control,” *arXiv preprint arXiv:2004.01041*, 2020.

---

**Fig. 5:** Performance comparison between D2C open-loop and closed-loop control policy.

**Fig. 6:** Comparison of Robustness to process noise between MPC and closed loop policy.