DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Reference herein to any social initiative (including but not limited to Diversity, Equity, and Inclusion (DEI); Community Benefits Plans (CBP); Justice 40; etc.) is made by the Author independent of any current requirement by the United States Government and does not constitute or imply endorsement, recommendation, or support by the United States Government or any agency thereof.

Final Technical Report

AEOLUS: Advances in Experimental Design, Optimal Control, and Learning for Uncertain Complex Systems

DOE Award Number DE-SC0021077

University of Texas at Austin
PI: Karen Willcox

Period of Performance: 09/01/2020 – 08/31/2023

Final Technical Report

AEOLUS: Advances in Experimental Design, Optimal Control, and Learning for Uncertain Complex Systems

The University of Texas at Austin
Senior Research Fellow: Max Gunzburger
PI: Karen Willcox
Grant number DE-SC0021077

Abstract

The AEOLUS Center is dedicated to developing a unified optimization-under-uncertainty framework for (1) learning predictive models from data and (2) optimizing experiments, processes, and designs governed by these models, all driven by complex, uncertain energy systems. AEOLUS addressed the critical need for principled, rigorous, scalable, and structure-exploiting capabilities for exploring parameter and decision spaces of complex forward simulation models—the so-called outer loop. This report summarizes the work done under DE-SC0021077 on (1) nonlocal models for solidification problems, (2) a multifidelity method for a nonlocal diffusion model, and (3) multifidelity Monte Carlo methods.

1 Nonlocal models for solidification problems

This work was published in [1, 2].

As has been noted, nonlocal models are better suited, compared to local PDE models, to model phenomena such as jump discontinuities. One setting in which jump discontinuities are desirable are in multi-phase problems for which the two phases appear only in their pure form, i.e., at each point, only one of the phases is present. Local models such as the Cahn-Hilliard equations can (approximately) model sharp transitions between pure phases only if a regularization parameter is very small and only if grids are seriously refined in the vicinity of the transitions. Here, we report on nonlocal versions of the Cahn-Hilliard phase-field model that we have developed, analyzed, and implemented and which can model sharp transitions independently of any parameter appearing in the local model and also without the need for severe grid refinement.

Local (PDE) Cahn-Hilliard models. The classical Cahn-Hilliard (C-H) model is derived as the gradient flow of the Ginzburg–Landau energy

$$E(u) = \frac{\varepsilon}{2}\int_\Omega |\nabla u|^2 + \int_\Omega W(u),$$

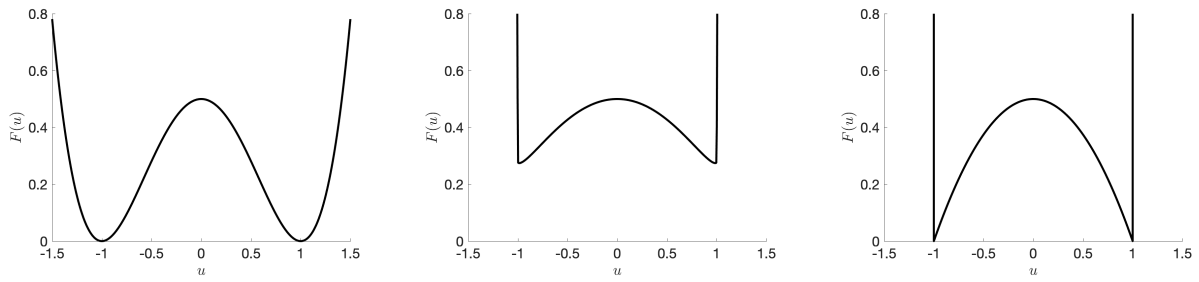
in the Hilbert space $H^{-1}(\Omega)$, where ε denotes and interface width parameter and $W(u)$ denotes a potential function. The resulting C-H equations are given by

$$\partial_t u + (-\Delta)(-\varepsilon\Delta u + W'(u)) = 0 \ \text{ in } \Omega \qquad \Longleftrightarrow \qquad \begin{cases} \partial_t u - \Delta w = 0 \\ w = -\varepsilon\Delta u + W'(u) \end{cases} \tag{1}$$

$$u(t=0) = u_0 \qquad \frac{\partial u}{\partial n} = \frac{\partial \Delta u}{\partial n} = 0 \tag{2}$$

Three choices for the potential $W(u)$ are given in the Figure 1.

The local C-H model may be generalized by replacing one or two of the Laplacian operators in (1) and (2) by general second-order elliptic PDE operators. Note that local C-H models result in sharp interfaces only in the limit $\varepsilon \to 0$.



regular double-well potential

$\frac{1}{2}(1-u^2)^2, \ \ u \in \mathbb{R}$

logarithmic potential ($c > 0$)

$\frac{1}{2}(1-u^2) + c(1+u)\ln(1+u)$
$+ (1-u)\ln(1-u) \ \ , u \in (-1,1)$

obstacle potential

$\begin{cases} \frac{1}{2}(1-u^2) & \text{if } |u| \le 1 \\ +\infty & \text{if } |u| > 1 \end{cases}$

Figure 1: *Three choices for the potential $W(u)$.*

Nonlocal (PDE) Cahn-Hilliard models. Let $\Omega \subset \mathbb{R}^n$ denote a bounded domain and, for a given $\delta > 0$, let the function the $\gamma_\delta(x,y)$ satisfy

$$\gamma_\delta(x,y) \neq 0 \quad \text{for all } y \in B_\delta(x) := \{y \in \mathbb{R}^n \colon |y-x| \le \delta, \ \ \delta > 0\}.$$

Thus, two points x and y interact with each other only if $|y-x| < \delta$; δ is referred to as the *horizon* or *interaction radius*. The *interaction domain* $\Omega_I \subset \mathbb{R}^n$ corresponding to Ω is then defined as

$$\Omega_I := \{y \in \mathbb{R}^n \setminus \Omega \ \ : \ \ \gamma_\delta(x,y) \neq 0, \quad x \in \Omega\}.$$

We also define the *nonlocal flux operator*

$$\mathcal{N}u(x) := \int_{\Omega \cup \Omega_I} (u(x) - u(y))\gamma_\delta(x,y)\,\mathrm{d}y = 0 \quad \forall x \in \Omega\}$$

which is the nonlocal analog of the local flux operator $\nabla u \cdot \mathbf{n}$.

We use the notation $\mathcal{B}$ to denote nonlocal operators that are analogs of local second-order elliptic PDE operators. We consider two types of nonlocal operators:

$$\textbf{type 1}\text{: Neumann type} \quad \begin{cases} \mathcal{B}u(x) = \varepsilon \displaystyle\int_{\Omega \cup \Omega_I} (u(x) - u(y))\gamma_\delta(x,y)\,\mathrm{d}y \quad \forall x \in \Omega \\ \text{along with } \mathcal{N}u = 0 \text{ for } x \in \Omega_I \end{cases} \tag{3}$$

$$\textbf{type 2}\text{: regional type} \quad \mathcal{B}u(x) = \varepsilon \int_{\Omega \cup \Omega_I} (u(x) - u(y))\gamma_\delta(x,y)\,\mathrm{d}y \quad \forall x \in \Omega. \tag{4}$$

For type 1 operators, nonlocal interactions occur in $\Omega \cup \Omega_I$ whereas for type 2, nonlocal interactions occur solely in Ω.

With $A = -\Delta$ and $\mathcal{B}$ either a nonlocal Neumann (with $\mathcal{N}u = 0$ on Ω_I) or nonlocal regional operator, *a nonlocal Cahn-Hilliard model with smooth or logarithmic potential* has the form

$$\partial_t u + Aw = 0 \ \forall x \in \Omega, \qquad w = \mathcal{B}u + W'(u) \ \forall x \in \Omega, \qquad \text{and} \qquad \frac{\partial w}{\partial n} = 0 \ \forall x \in \partial\Omega \tag{5}$$

For $c_f > 0$, the obstacle potential can be defined as

$$W(u) = W_0(u) + \mathbb{O}_{[-1,1]}(u) \ \text{ where } \ W_0(u) = \frac{c_F}{2}(1-u^2) \ \text{ and } \ \mathbb{O}_{[-1,1]}(u) := \begin{cases} 0 & \text{if } |u| \le 1 \\ +\infty & \text{otherwise.} \end{cases} \tag{6}$$

$W_0(u)$ is differentiable with, of course, $W_0' = -c_F u$. However, $\mathbb{O}_{[-1,1]}(u)$ is not differentiable so that subdifferentials are used. The subdifferential of $\mathbb{O}_{[-1,1]}(u)$ is given by

$$\partial \mathbb{O}_{[-1,1]}(u) = \begin{cases} (-\infty, 0] & \text{if } u = -1 \\ 0 & \text{for } u \in (-1,1) \\ [0, +\infty) & \text{if } u = 1 \end{cases}$$

so that the subdifferential of $W(u)$ is given by $\partial W(u) = -c_F u + \partial \mathbb{O}_{[-1,1]}(u)$. Then, a *nonlocal Cahn-Hilliard model with the obstacle potential* has the form

$$\partial_t u + Aw = 0 \qquad \text{and} \qquad w = \mathcal{B}u + W_0'(u) + \partial \mathbb{O}_{[-1,1]}(u). \tag{7}$$

This leads to a *variational inequality system* which can be written as a complementarity system with Lagrange multiplier $\lambda = \partial \mathbb{O}_{[-1,1]}(u)$.

To discretize the nonlocal C-H models, we use continuous piecewise linear finite element methods for spatial discretization and the backward-Euler method for temporal discretization.

Analysis and numerical analysis. We have analyzed the nonlocal Cahn-Hilliard models and their finite element approximation. In particular,

– we have proved that the energy decays in time for both the continuous problem and for the fully discrete problem

– we have proved that for certain specifically defined combinations of the parameters in the problem, the interface is sharp, that is, u can only take on the values -1 and 1

– we have proved well posedness

– we have derived error estimates for finite element approximations.

Numerical results. We provide some sample numerical results that illustrate the differences between solutions of the local C-H equations and the nonlocal C-H equations with obstacle potentials and also to illustrate the differences between solutions of the Type 1 and Type 2 nonlocal C-H equations. For the sake of brevity, we do not provide any details about the specifics of the calculation. Detailed descriptions can be found in [**?**]. We do point out that the kernel $\gamma_\delta(x, y)$ used for the computations is integrable which implies that the space in which nonlocal models are well posed include *functions with jump discontinuities* [11, 12].

We first consider the one-dimensional setting. Figure 2 corresponds to the Type 1 nonlocal "Neumann"-type C-H equations. We can clearly observe that the nonlocal solution has jump discontinuities (with respect to the employed computational grid) and, in for sufficiently large times, admits pure states (i.e., the solution values are either -1 or 1) whereas the local solution also takes on values in $(-1, 1)$ even for large times. These observations are in agreement with our theoretical results.

Figures 3 and 4 correspond to the Type 2 nonlocal "regional" C-H equations with the difference between the two figures being that the former corresponds to the obstacle potential (6) whereas the latter corresponds to a modification of that potential. In contrast to the example of Figure 2, in Figure 3 we no longer observe sharp interfaces in the nonlocal solution whereas in Figure 4 we do observe sharp interfaces in the nonlocal solution. Thus, Type 2 nonlocal "regional" C-H equations have the flexibility of being able to model both sharp and diffuse interface. These results are in agreement with our theoretical results.

Next, we consider the two-dimensional setting and the "Neumann"-type nonlocal C-H equations. In Figures 5 and 6, we plot the snapshots of the local and nonlocal solutions at different time steps. The difference between the two solutions is in their initial conditions. Figure 5 corresponds to a smooth initial condition $u_0(x)$ whereas for Figure 6 the initial condition is defined by sampling

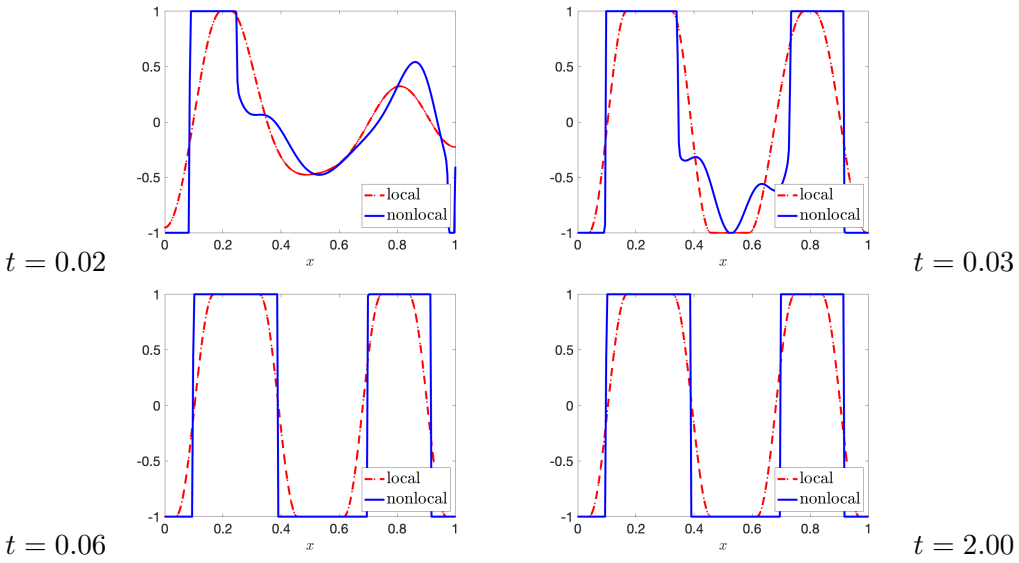


Figure 2: *Snapshots of the solutions of the nonlocal "Neumann"-type Cahn-Hilliard equations and solutions of the local PDE Cahn-Hilliard equation at different time instances.*

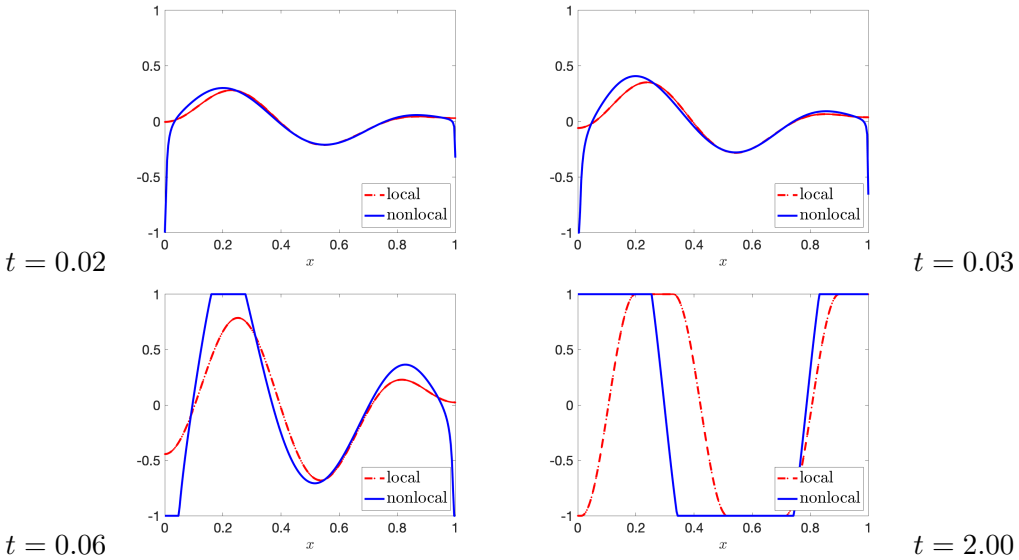


Figure 3: *Snapshots of the solutions of the nonlocal "regional"-type Cahn-Hilliard equations and solutions of the local PDE Cahn-Hilliard equation at different time instances.*

from a uniform random distribution on $[-1, 1]$ at each grid point. In Figure 5, we observe that local and nonlocal solutions look alike. However, whereas the nonlocal model can describe perfectly sharp interfaces up to the resolution of the discretization mesh, the interfaces for local solution are diffuse. In Figure 6, we observe that the local and nonlocal solution are quite different in a pointwise manner, but are qualitatively the same. We also again observe that the nonlocal model can describe perfectly sharp interfaces up to the resolution of the discretization mesh but the interfaces for local solution are again diffuse. These results are in agreement with our theoretical results.

Finally, in Figure 7 we consider the "regional"-type nonlocal C-H equations with the modified

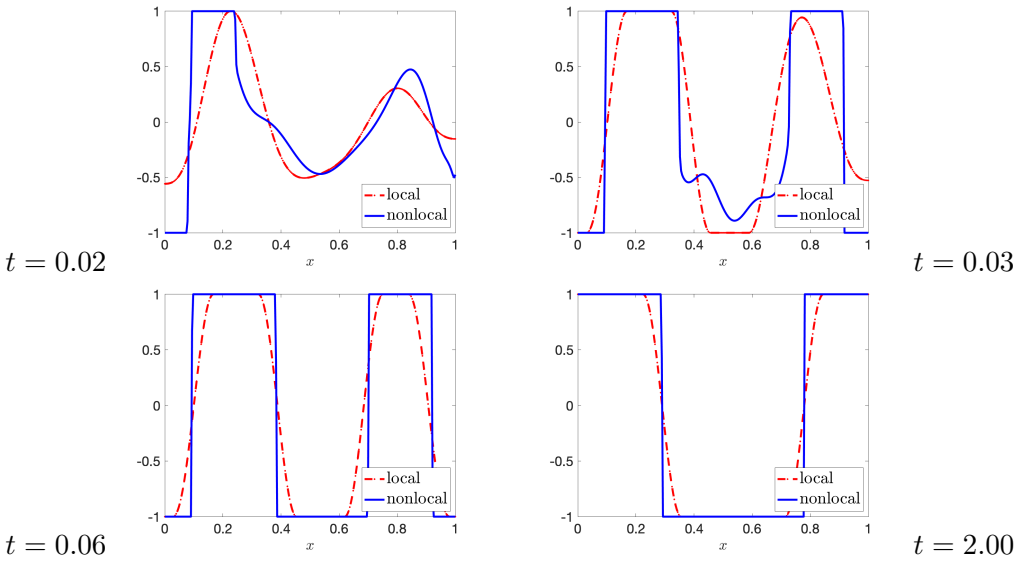


Figure 4: *Snapshots of the solutions of the nonlocal "regional"-type Cahn-Hilliard equations and solutions of the local PDE Cahn-Hilliard equation at different time instances.*

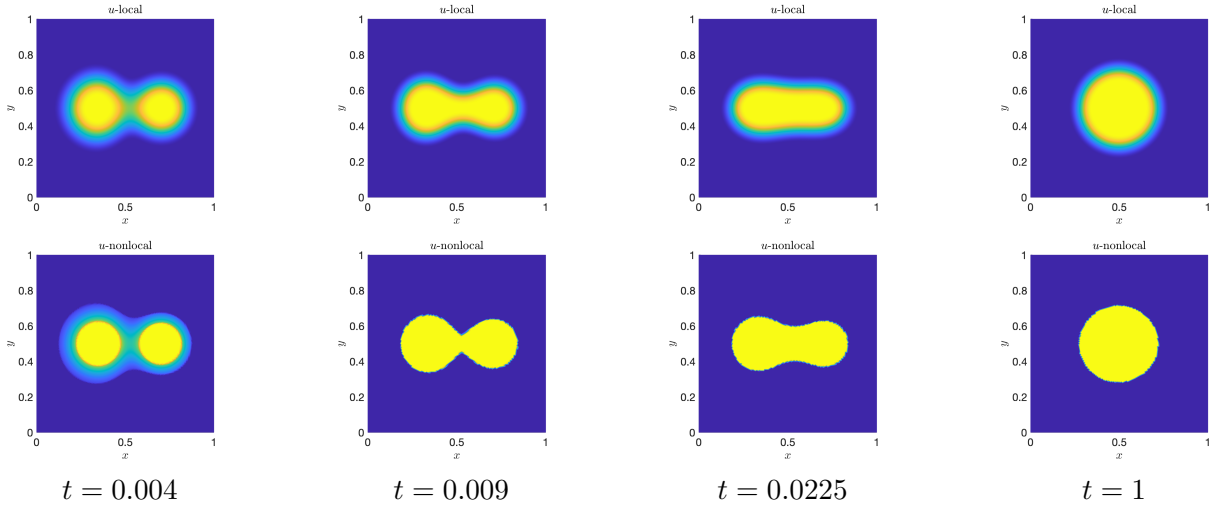


Figure 5: *For a smooth initial condition, snapshots of the solutions of the local C-H equations (top) and of the "Neumann"-type C-H equations (bottom) at different time instances.*

obstacle potential and with a random initial condition. The observations made corresponding to Figure 6 hold for this case and again the results are in agreement with our theoretical results.

In summary

– the numerical results reported are in agreement with out theoretical results

– stable computations without regularization are possible in the nonlocal case

– in general, "Neumann"-type nonlocal C-H equations result in sharp interfaces

– in general, "regional"-type nonlocal C-H equations can result in sharp or diffuse interfaces, depending on parameter choices

– in local case, the thickness of the interface depends on the parameter ε and sharp interfaces can be obtained only in the limit $\varepsilon \to 0$ so that to obtain nearly sharp interfaces ε must be small

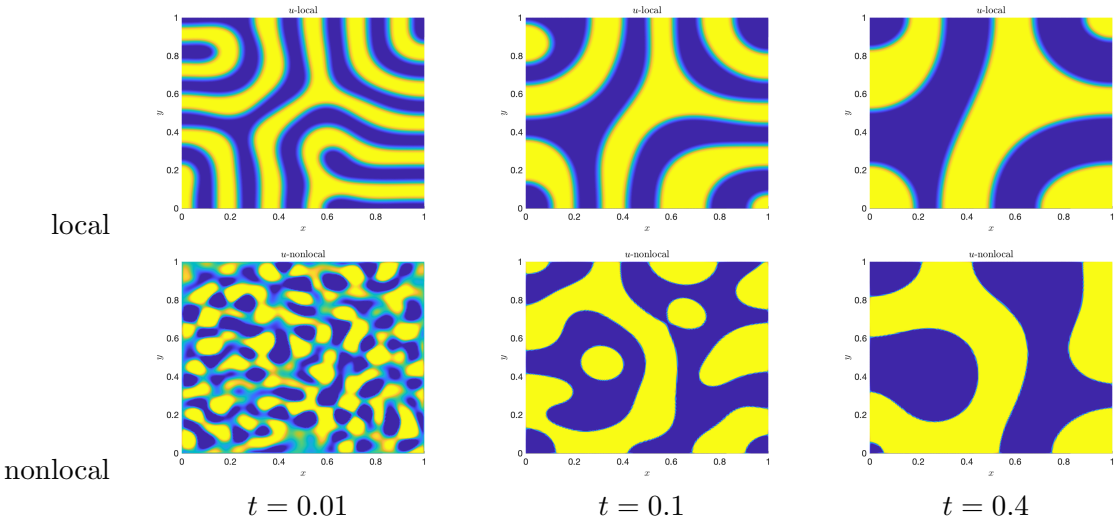


Figure 6: *For a random initial condition, snapshots of the solutions of the local C-H equations (top) and of the "Neumann"-type C-H equations (bottom) at different time instances.*

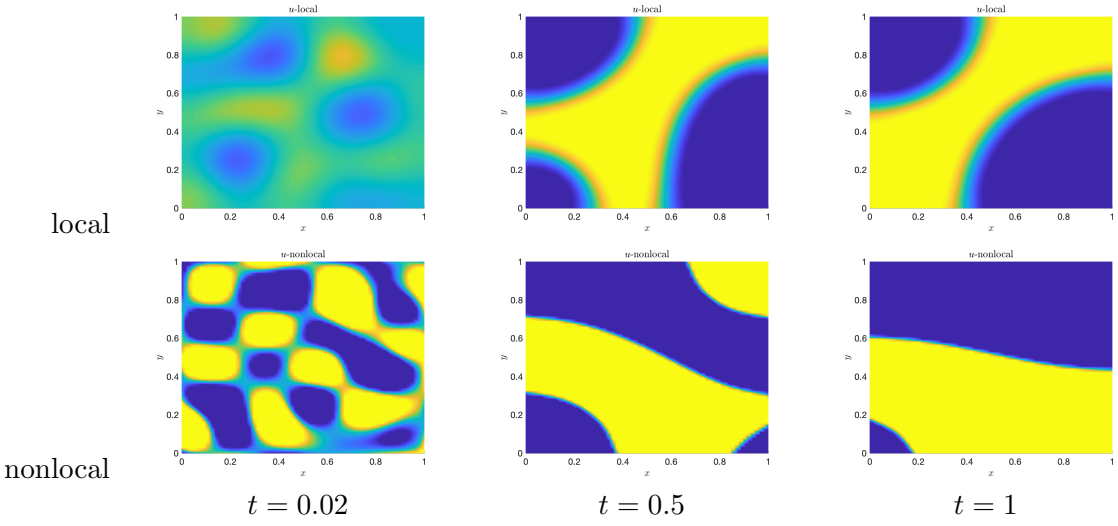


Figure 7: For a random initial condition, snapshots of the solutions of the local C-H equations (top) and of the "regional"-type C-H equations with a modified obstacle potential at different time instances.

and severe grid refinement near the interface is needed whereas, in the nonlocal case, parameter choices can be made so that

sharp interfaces can be obtained independently of ε

so that

no grid refinement is needed near the interface

and also for parameter choices that result in diffuse interfaces,

the thickness of those interfaces are also independent of ε.

Following up our previous efforts (see [10]) on nonlocal models for the Cahn-Hilliard and Allen-Cahn equations, during the second reporting period we defined a novel nonlocal model for the solidification of pure materials. For such materials, the Kobayashi model for the solidification

of pure materials [4] is described by a coupled system of partial differential equations equations involving two variables: a nonlinear Allen-Cahn model for the phase field variable u (with $u = 0$ corresponding to a liquid and $u = 1$ to a solid state) and a nonlinear partial differential equation for the temperature T. Specifically, the Kobayashi model is given by

$$\begin{aligned} \frac{\partial T}{\partial t} &= D\Delta T + K\frac{\partial u}{\partial T} && \text{(a)} \\ \tau\frac{\partial u}{\partial t} &= \nabla\cdot(\varepsilon(\theta)^2\nabla u) - \nabla\cdot\left(|\nabla u|^2\varepsilon(\theta)\frac{\partial\varepsilon(\theta)}{\partial(\nabla u)}\right) - F'_{reg}(u,T), && \text{(b)} \end{aligned} \tag{8}$$

where we have the double-well potential

$$F_{reg}(u,T) = \frac{1}{4}u^2(1-u)^2 + m(T)\left(\frac{1}{3}u^3 - \frac{1}{2}u^2\right) \tag{9}$$

and ε is a nonlinear function of the gradient which in the two-dimensional case $\Omega \subset \mathbb{R}^2$ is given by

$$\varepsilon(\theta) = \tilde{\varepsilon}\sigma(\theta(\nabla u)), \tag{10}$$

where D and K are constants, $\tilde{\theta}$ is a mean value of ε and $\sigma(\theta) = 1 + \delta\cos(j(\theta - \theta_0))$, and θ is an angle between ∇u and a certain direction. The coupling of the phase field and the temperature is effected via the $\partial u/\partial T$ term in (8a) and the function

$$m(T) = \left(\frac{\alpha}{\pi}\right)\tan^{-1}\left(\rho(T_e - T)\right), \quad 0 < \alpha < 1, \quad |m(T)| < 1/2 \tag{11}$$

that appears in (8b) through (10).

The novel nonlocal model for the solidification of pure materials we have developed can be considered to be analogue of the Kobayashi model in which the principal modification is the replacement of the *nonlinear* local differential operator in (8b) for the phase-field variable by the *linear* nonlocal operator

$$Bu := \int_{\Omega\cup\Omega_I}(u(\mathbf{x}) - u(\mathbf{y}))\gamma(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{y} = c_\gamma(\mathbf{x})u(\mathbf{x}) - (\gamma * u)(\mathbf{x}), \quad \mathbf{x}\in\Omega, \tag{12}$$

where $\gamma(\cdot,\cdot) \geq 0$ (which is referred to as the *kernel*), $(\gamma * u)(\cdot)$ denotes the convolution operation,

$$\Omega_I := \{\mathbf{y}\in\mathbb{R}^n \setminus \Omega \,:\, \gamma(\mathbf{x},\mathbf{y}) \neq 0, \;\; \mathbf{x}\in\Omega\}$$

(which is referred as the *interaction domain* corresponding to Ω), and $c_\gamma(\mathbf{x}) = \int_{\Omega\cup\Omega_I}\gamma(\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{y}$.

We make two additional modifications of the Kobayashi model.[1] First, we use of a Cahn-Hilliard type model instead of the Allen-Chan type model; we do this because Cahn-Hilliard type models allow for sharp interfaces whereas Allen-Cahn-type models do not. Second, we replace the smooth potential (9) by the non-smooth *obstacle potential*

$$F_{\text{obs}}(u,T) := \frac{c_F}{2}u(1-u) - c_m m(T)u + \mathbb{I}_{[0,1]}(u), \quad c_F,\; c_m > 0,$$

where $\mathbb{I}_{[0,1]}(u)$ denotes the indicator function; we do so because the obstacle potential only allows values of the phase field in the interval $[0,1]$ whereas the smooth potential allows for values less than zero and greater than one. See the figure below for plots of the two types of potentials.

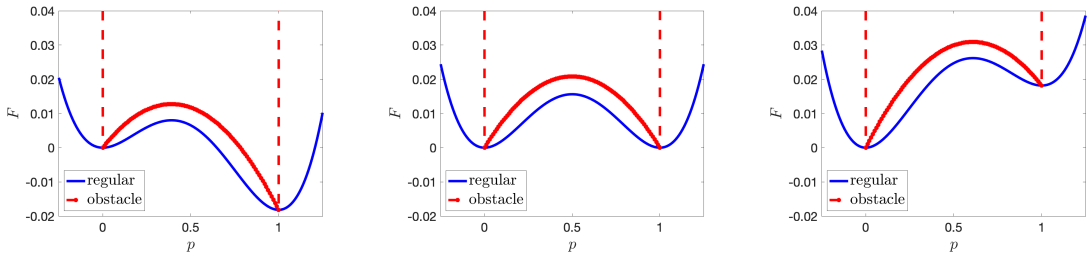
[1] These modifications can also be effected for the Kobayashi model itself.

The obstacle potential is non-smooth so we resort to the notion of subdifferentials to define the derivative of F_{obs}, i.e., we have

$$\partial F_{\text{obs}}(u,T) = -c_F u + \frac{c_F}{2} - c_m m(T) + \partial \mathbb{I}_{[0,1]}(u), \tag{13}$$

where $\partial \mathbb{I}_{[0,1]}(u)$ is a subdifferential of the indicator function defined as

$$\partial \mathbb{I}_{[0,1]}(u) = \begin{cases} (-\infty, 0] & \text{if } u = 0 \\ 0 & \text{for } u \in (0,1) \\ [0, +\infty) & \text{if } u = 1. \end{cases}$$



$m > 0$ $\qquad$ $m = 0$ $\qquad$ $m < 0$

Comparisons of the regular $F_{reg}(u,T)$ and obstacle $F_{\text{obs}}(u,T)$ double-well potentials.

The nonlocal solidification model then given by

$$\begin{cases} \dfrac{\partial T}{\partial t} = D\Delta T + K\dfrac{\partial u}{\partial t} \\ \tau \dfrac{\partial u}{\partial t} + (I - \beta\Delta)w = 0 \\ w = Bu + \partial F_{\text{obs}}(u,T), \end{cases} \tag{14}$$

where $\beta \geq 0$. Note that if $\beta = 0$ the last two equations above reduce to a nonlocal Allen-Cahn model with the double-well obstacle potential. The model (14) can be rewritten as [2]

$$\begin{cases} \dfrac{\partial T}{\partial t} = D\Delta T + K\dfrac{\partial u}{\partial t}, \\ \tau \dfrac{\partial u}{\partial t} + (I - \beta\Delta)w = 0 \\ w = -\xi u + \gamma * u - \dfrac{c_F}{2} + c_m m(T) - \lambda, \quad \lambda \in \partial \mathbb{I}_{[0,1]}(u), \end{cases} \tag{15}$$

where $*$ denotes the convolution operator and

$$\xi(x) = \int_{\widetilde{\Omega}} \gamma(\mathbf{x} - \mathbf{y})\, d\mathbf{y} - c_F \qquad \text{and} \qquad \gamma * u = \int_{\widetilde{\Omega}} u(\mathbf{y})\gamma(\mathbf{x} - \mathbf{y})\, d\mathbf{y}. \tag{16}$$

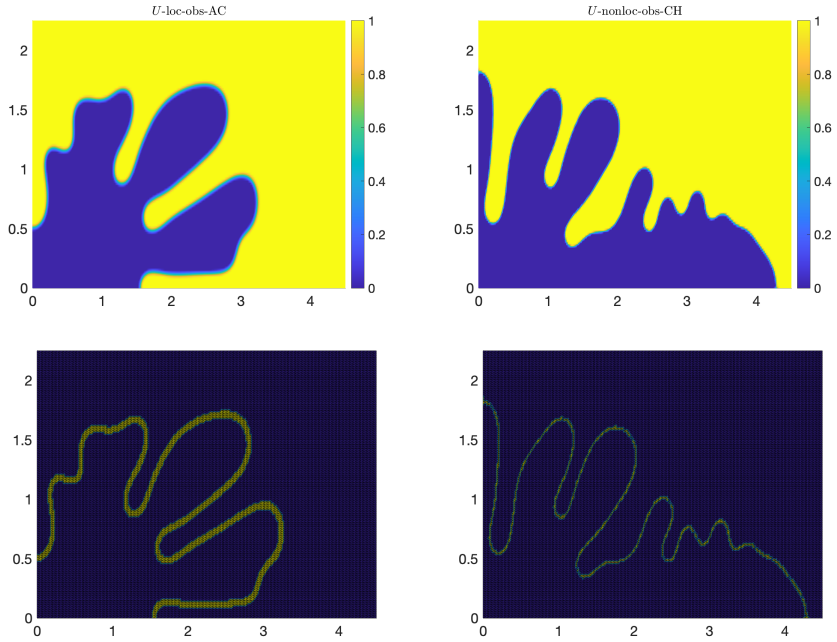
We supplement (14) or (15) with additional constraints placed on u on Ω_I and on w on the boundary of Ω. We set $c_F = 1/6 = c_m = 1/6$, so that as is the case for the regular potential, $F_{\text{obs}}(0,T) - F_{\text{obs}}(1,T) = m/6$.

To discretize the model (15), we employ an implicit-explicit time stepping scheme, where we discretize the nonlocal convolution term $\gamma * u$ together with the coupling term $m(T)$ explicitly.

This choice of the discretization allows us to decouple the problem and to obtain an efficient solution strategy. In particular, the phase field variable u can be obtained via a projection formula so that *no solution of the dense nonlocal system is necessary.* For spatial discretization we use piecewise-linear continuous finite elements.

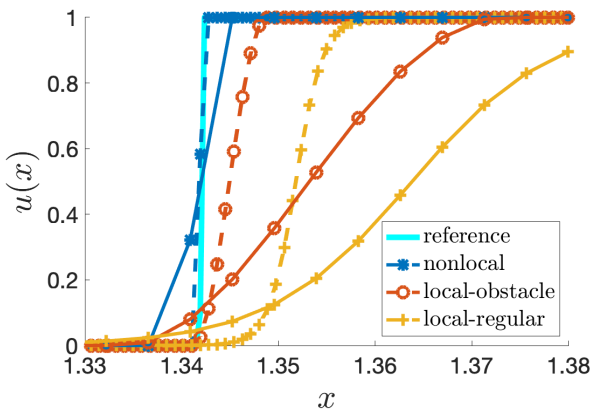
We have conducted several computational examples using different kernels and different model data. Here, we give results for two examples. For both, we use the kernel $\gamma(\mathbf{x}-\mathbf{y}) = C_\delta \max\left(0, 1 - \frac{|\mathbf{x}-\mathbf{y}|^2}{\delta^2}\right)$, where $C_\delta = 15/2\delta^3$ in one dimension and $C_\delta = 24/\pi\varepsilon^4$ in two dimensions. For the two-dimensional example, we use parameter values for pure nickel and the spatial-temporal domain is chosen as $\Omega = (4.5, 2.25) \times (4.5, 2.25)$ and $t = [0, 3.5]$. For the local model, we choose[2] $\epsilon = 10^{-2} \sim 5h$. For the nonlocal model we choose $\delta = 0.0753$ $(\sim 4h)$ and choose other model parameters to simulate nickel solidification. Th first figure below is illustrative of the 2D results we have obtained.

The 1D results given in the second figure below clearly illustrate the differences in how local and nonlocal models handle the interface between the liquid and solid states. Using the obstacle potential and the same grid spacing for both models, the interface thickness for the local model is 8 or 9 grid intervals wide whereas the nonlocal interface is 2 intervals wide. The disparity is even worse for the local model with the very common choice of the smooth potential for which the interface is 20 or so intervals wide. For the local model, reducing ϵ by a factor of 8 requires the refinement of the grid by a factor of 8 near the interface to produce an interface thickness akin to that of the nonlocal model.



Left: the local interface spreads over $\epsilon \sim 5$ or more grid cells. Right: the nonlocal interface spreads at most 1 or 2 grid cells.

[2]For the local model, the thickness of the interface is of order ϵ so that sharp interfaces can only occur in the limit of $\epsilon \to 0$. A a result, if one wants to obtain sharp approximate interfaces, one must refine the spatial grid in the vicinity of the moving interface.



Predicted 1D interfaces at a final time for different models (zoomed-in). For the local model, two grid spacings are used and both the regular and obstacle potentials are considered. For the nonlocal model, the obstacle potential is considered and the grid corresponds to the coarser grid used for the nonlocal model.

2 A multifidelity method for a nonlocal diffusion model

This work was published in [9].

Nonlocal models feature a finite length scale, referred to as the horizon, such that points separated by a distance smaller than the horizon interact with each other. Such models have proven, compared to differential equations models, to more faithfully agree with observations in settings such as fracture, subsurface flows, and image processing. However, due to the reduced sparsity of discretizations, they are also generally computationally more expensive compared to their local differential equation counterparts, so much so that in multi-query applications such as uncertainty quantification, optimization, and control the increased computational costs can render nonlocal modeling to be infeasible. Not surprisingly, there have been several proposed approaches such as reduced-order modeling (e.g., proper orthogonal decomposition) or better sampling strategies (e.g., alternatives to Monte Carlo sampling) aimed at mitigating this situation.

In local settings such as for the numerical solution of differential equations, *multifidelity methods* that judiciously combine a "truth" model of interest with less costly surrogate models have been shown, e.g., in multi-query settings, to effectively reduce costs without compromising accuracy; see, e.g., [5, 6]. We have developed, implemented, and tested a *a multifidelity Monte Carlo method* for nonlocal models; see [9]. We emphasize that due to the reduced sparsity of nonlocal models the need for approaches such the one we have developed is much greater than that of related PDE models and thus multifidelity methods have greater impact for nonlocal models compared to local models.

A nonlocal diffusion model. Given a domain[3] $\Omega = (0,\ L) \subset \mathbb{R}$ and a length scale δ (referred to as the *interaction radius* or *horizon*), we define the *interaction domain* $\Omega_{\mathcal{I}} = \{y \in \mathbb{R} \setminus \Omega\ :\ |y-x| \leq \delta \text{ for some } x \in \Omega\} = [-\delta, 0] \cup [L,\ L+\delta]$, i.e., $\Omega_{\mathcal{I}}$ consists of all points in $\mathbb{R} \setminus \Omega$ that interact with points in Ω. For given functions $b(x)$, $g(x)$, and $\gamma(x, x')$ defined on Ω, $\Omega_{\mathcal{I}}$, and $(\Omega \cup \Omega_{\mathcal{I}}) \times (\Omega \cup \Omega_{\mathcal{I}})$,

[3] To simplify the exposition, we confine ourselves to a one-dimensional setting (as a proof of concept); generalization to higher dimensions is a straightforward exercise.

respectively, the one-dimensional, steady-state nonlocal diffusion model we consider is given by

$$\begin{cases} -2\displaystyle\int_{x-\delta}^{x+\delta} \big(u_\delta(x') - u_\delta(x)\big)\,\gamma(x,x')\,dx' = b(x) \quad \forall x \in \Omega = (0,\,L) & \text{(a)} \\ \qquad\qquad u_\delta(x) = g(x) \quad \forall x \in \Omega_{\mathcal{I}} = [-\delta,\,0] \cup [L,\,L+\delta]\,. & \text{(b)} \end{cases} \tag{17}$$

Note that (17) is a nonlocal analog of the PDE Dirichlet boundary-value problem $-\nabla\cdot(a(x)\nabla u) = b(x)$ in Ω and $u = g(x)$ on the boundary of Ω. Indeed, as the horizon $\delta \to 0$, the nonlocal model, when properly formulated, reduces to such a PDE problem; see, e.g., [11, 12, 13].

The function $\gamma(x,x')$ in (17) is referred to as the *kernel*; the choice made for $\gamma(x,x')$ determines the properties of solutions including their smoothness properties. As a result of the flexibility available in the choice of $\gamma\,(x,x')$, nonlocal models can model a wide variety of observed behaviors. For example, for bounded kernels, the nonlocal problem admits solutions having jump discontinuities; of course, such solutions are not obtainable from second-order elliptic PDEs.

Nonlocal models such as (17) have been subject to discretization via the same approaches as those used for discretizing PDEs. Which of these approaches one selects is not germane to our goals. For the sake of concreteness, we use a finite difference method. The reduced sparsity of the discretizations is now easily explained. For a fixed horizon δ, as the grid size h is reduced, the number of nodes that interact with a given node, i.e., the stencil associated with a give node, increases and the same occurs if h is fixed and δ increases. This is in contrast to the local setting for which the stencil remains the same as h is reduced.

A brief review of multifidelity Monte Carlo methods. We provide a brief account of the multifidelity Monte Carlo (MFMC) method of [5]. For the sake of concreteness, we consider the uncertainty quantification setting.

The goal is to determine approximate statistical information about a desired "truth" output of interest (OoI) $f^{(1)}(\mathbf{z}) : \mathcal{Z} \to \mathcal{F}$, where $\mathbf{z}$ is a random input selected from the input domain $\mathcal{Z} \subset \mathbb{R}^d$ and $\mathcal{F} \subset \mathbb{R}$ denotes the corresponding output domain. Again, for the sake of concreteness, we consider the (statistical) quantity of interest (QoI) to be the expected value $\mathbb{E}[f^{(1)}(\mathbf{z})]$; other QoIs such as higher moments of the OoI can be treated in a similar manner. We approximate the QoI via Monte Carlo (MC) sampling, i.e., for M_{MC} independent and identically distributed random samples $\mathbf{z}_m \in \mathcal{Z}$, $m = 1, \ldots, M_{MC}$, drawn from a given probability density function, we have the *MC estimator*

$$f^{MC}_{M_{MC}} = \frac{1}{M_{MC}} \sum_{m=1}^{M_{MC}} f^{(1)}(\mathbf{z}_m) \approx \mathbb{E}\left[f^{(1)}(\mathbf{z})\right]. \tag{18}$$

If $M^{MC} \gg 1$ and $f^{(1)}(\cdot)$ is expensive to evaluate, (18) may be prohibitive in cost.

Suppose we have in hand a set of lower-cost surrogate OoIs $f^{(2)}, \ldots, f^{(K)} : \mathcal{Z} \subset \mathcal{F}$ with corresponding costs w_k, $k = 2, \ldots, K$. Then, for a given set of weights $\alpha_2, \ldots, \alpha_K$ and a given set of sample sizes $0 < M_1 \le M_2 \le \cdots \le M_k$, the unbiased *MFMC estimator* is defined as

$$f^{MFMC}_{M_{MF}} = f^{(1)}_{M_1} + \sum_{k=2}^{K} \alpha_k \left(f^{(k)}_{M_k} - f^{(k)}_{M_{k-1}}\right) \approx \mathbb{E}\left[f^{(1)}(\mathbf{z})\right], \tag{19}$$

where $f^{(k)}_{M_k}$ and $f^{(k)}_{M_{k-1}}$ denote the MC estimators for the OoI $f^{(k)}(\cdot)$, respectively, using M_k and M_{k-1} input samples drawn from the input domain $\mathcal{Z}$, i.e.,

$$f^{(k)}_{M_k} = \frac{1}{M_k} \sum_{m=1}^{M_k} f^{(k)}(\mathbf{z}_m) \quad \text{and} \quad f^{(k)}_{M_{k-1}} = \frac{1}{M_{k-1}} \sum_{m=1}^{M_{k-1}} f^{(k)}(\mathbf{z}_m). \tag{20}$$

Note that the M_{k-1} samples taken in the second sum can be reused in the first sum so that the total number of evaluations of $f^{(k)}(\cdot)$ for each term in (19) is M_k. Thus, the cost incurred to determine the MFMC estimator (19) is given by $\sum_{k=1}^{K} w_k M_k$.

For $k = 1, \ldots, K$, let $\sigma_k^2 = \mathrm{Var}\big(f^{(k)}(\mathbf{z})\big)$, $\rho_{1,k} = \mathrm{Cov}\big(f^{(k)}(\mathbf{z}), f^{(1)}(\mathbf{z})\big)/(\sigma_k \sigma_1)$ = Pearson correlation coefficient. Then, the optimal values for the weights α_k and the number of samples M_k are determined, for a given budget p, by *minimizing the variance of* $f_{M_{MF}}^{MFMC}$ *subject to the constraints* $\sum_{k=1}^{K} w_k M_k = p$ *and* $0 < M_1 \leq M_2 \leq \cdots \leq M_k$. If one assumes that

$$|\rho_{1,1}| > \cdots > |\rho_{1,K}| \qquad \text{and} \qquad \mu_k = \frac{w_{k-1}}{w_k} > \nu_k = \frac{\rho_{1,k-1}^2 - \rho_{1,k}^2}{\rho_{1,k}^2 - \rho_{1,k+1}^2} \quad \text{for} \quad k = 2, \ldots, K \tag{21}$$

with $\rho_{1,K+1} = 0$ hold, this optimization problem has the unique analytic solution (see [5])

$$\alpha_k = \frac{\rho_{1,k}\sigma_1}{\sigma_k} \quad \text{and} \quad M_k = M_1 r_k \quad \text{for} \quad k = 2, \ldots, K \qquad \text{with} \quad M_1 = \frac{p}{\sum_{k=1}^{K} w_k r_k}, \tag{22}$$

where $r_k = \Big(w_1(\rho_{1,k}^2 - \rho_{1,k+1}^2)/w_k(1 - \rho_{1,2}^2)\Big)^{1/2}$ for $k = 1, \ldots, K$. We note that the constraint $M_1 > 0$ ensures that the accuracy of the MFMC estimator corresponds to that of the truth model $f^{(1)}(\cdot)$, i.e., the accuracy is not damaged by the use of surrogate models.

The first requirement in (21) is easily satisfied by re-ordering the OoIs $f^{(k)}(\cdot)$. The second requirement is violated whenever the decrease in the accuracy of the low-fidelity model k (in the order of decreasing correlation coefficient) is more significant that the reduction of its cost of evaluation. The models that violate the second requirement in (21) are then eliminated from the set of models used in the MFMC estimator. It is worth mentioning that the second requirement is defined by comparing each model to the preceding one in the order of decreasing correlation coefficient. One then continues the winnowing process of checking the second requirement and eliminating the models that violate this requirement until all the remaining models satisfy the second requirement.

Note that (22) does not, in general, yield integer values for the number of samples M_k. In this study, we have chosen to round up to the nearest integer. Also, the values of σ_k and $\rho_{1,k}$ are generally not known a priori, so that in practice they are estimated by computing a very few realizations of the models $f^{(k)}(\cdot)$ for $k = 1, \ldots, K$, where "very few" is relative to the very large number of samples one would have to use if one were to use a straightforward MC estimator for $\mathbb{E}[f^{(1)}(\cdot)]$. Such realizations can even be reused to estimate the costs w_k of evaluating $f^{(k)}(\cdot)$ should those costs not be known.[4]

A numerical illustration. Consider the discretization of the nonlocal diffusion model (17) with $L = 1$, a random constant source term b for $x \in (0, 1)$, the constraints $u(x) = g$ with g a random constant for $x \in [-\delta, 0]$, $u(x) = 0$ for $x \in [1, 1+\delta]$, and the constant kernel function $\gamma(x, x') = 1/2\delta^3$ for $|x - x'| < \delta$. The random inputs are independently, identically, and uniformly distributed realizations of $\mathbf{z} = (b, g)$ within the input domain $\mathcal{Z} = [-1.1\ -0.9] \times [0.9\ 1.1] \subset \mathbb{R}^2$. Note that the dependence of the kernel on δ is a scaling factor that ensures that as $\delta \to 0$, i.e., as the extent of nonlocal interactions vanish, the nonlocal problem reduces to its local differential equation counterpart $-d^2u/dx^2 = b$.[5]

[4] We note that *multilevel Monte Carlo methods* [14, **?**] are a special case of multifidelity Monte Carlo methods; the former usually involve surrogates that depend on coarser grid sizes and for which a hierarchy of costs, correlations, etc. are known a priori. Multifidelity Monte Carlo methods do not assume that such a hierarchy is available and also, in general, can involve surrogates of many different types.

[5] This is not a requirement for the proposed multifidelity framework. We choose the specific model we consider only for validation purposes.

The *truth* quantity of interest (QoI) is the expected value of the output of interest (OoI) $f^{(1,1)}(\mathbf{z}) = \left(\int_\Omega u(x,\delta_1,h_1;\mathbf{z})^2 dx\right)^{1/2}$ for $\delta_1 = 0.25$ and $h_1 = 1/N_1 = 2^{-10}$. An approximation of the truth value of the QoI $\mathbb{E}[f^{(1,1)}(\mathbf{z})]$ is defined as the Monte Carlo (MC) estimator (19) using $M_{MC} = 5 \times 10^8$ samples of $(b, g) \in \mathcal{Z}$.

We define eight lower-cost surrogate estimators by using two smaller horizon values $\delta_2 = 0.25/2$ and $\delta_3 = 0.25/2^2$ and/or two larger grid length values $h_2 = 2^{-9}$ and $h_3 = 2^{-8}$. The models are indexed using a pair of indices (i, j) with the first index representing the δ-model, and the second index the h-model, e.g., $f^{(i,j)}(\mathbf{z}) = \left(\int_\Omega u(x,\delta_i,h_j;\mathbf{z})^2 dx\right)^{1/2}$. Including the truth estimator, we specifically solve the discretized nonlocal model (17) for samples $\mathbf{z}_m = (b_m, g_m) \in \mathcal{Z}$ to obtain $f^{(i,j)}(\mathbf{z}_m)$ for $i, j = 1, \ldots, 3$. Letting $k = j + 3(i-1)$ for $i, j = 1, \ldots, 3$, the costs w_k and correlation coefficients $\rho_{1,k}$ for the surrogates are approximated by the averages over 50 samples of the random inputs; cost measurements are defined as wall-clock times. Using 500 or even 1000 samples to estimate costs and correlation coefficients results in similar numbers as given below for 50 samples.

We consider four specific cases.

1. The MC estimator (18) with the truth values $\delta = 0.25$ and $h = 2^{-10}$ and with the number of samples M_{MC} limited by a given budget p.

2. $h = 2^{-10}$ is fixed at the truth value and $\delta_1 = 0.25$, $\delta_2 = 0.25/2$, and $\delta_3 = 0.25/2^2$ so that the first of these gives the truth model $f^{(1,1)}$ and the other two define cheaper and lower-fidelity surrogates $f^{(2,1)}$ and $f^{(3,1)}$, respectively. These three models satisfy both requirements in (21) in their current order ($f^{(1,1)}$, $f^{(2,1)}$, $f^{(3,1)}$). We refer to this case as the "three-δ case."

3. $\delta = 0.25$ is fixed at the truth value and $h_1 = 2^{-10}$, $h_2 = 2^{-9}$, and $h_3 = 2^{-8}$ so that the first of these gives the truth model $f^{(1,1)}$ and the other two defining cheaper and lower-fidelity surrogates $f^{(1,2)}$ and $f^{(1,3)}$, respectively. Similar to Case 2, these three models satisfy both requirements in (21) in their current order ($f^{(1,1)}$, $f^{(1,2)}$, $f^{(1,3)}$). We refer to this case as the "three-h case."

4. $\delta_1 = 0.25$, $\delta_2 = 0.25/2$, and $\delta_3 = 0.25/2^2$ and $h_1 = 2^{-10}$, $h_2 = 2^{-9}$, and $h_3 = 2^{-8}$ so that in addition to the truth model, there are eight lower-cost surrogate models. In this case, the second requirement in (21) is not satisfied. As a result, only the four surrogates $f^{(1,2)}$, $f^{(1,3)}$, $f^{(2,3)}$, and $f^{(3,3)}$ and, of course, the truth model $f^{(1,1)}$, survive the winnowing process mentioned above. We refer to this case as the "five-δ, h case."

Discussion of results. For each surrogate and for the truth model, the number of samples taken for each model is determined from (22). Figure 8 provides the distribution of number of samples for the four cases. Because the pair of "three-h" surrogates $f^{(1,2)}$ and $f^{(1,3)}$ are better correlated to the truth model $f^{(1,1)}$ than are the pair of "three-δ" surrogates $f^{(2,1)}$ and $f^{(3,1)}$, fewer number of samples of the truth model are needed for the former compared to that for the latter. Of course, the "five-δ-h" case requires an even smaller number of samples of the truth model, but the improvement over the "three-h" case is small compared to the improvement between the "three-δ" and "three-h" cases. The relative effectiveness of the MC estimator and the three MFMC estimators as a function of the budget available and as measured by the mean-squared error (MSE) is illustrated Figure 9. The MSE is estimated by comparison to the Monte Carlo estimator with 5×10^8 samples. Viewing the plots in Figure 9 vertically, we observe that for the same chosen budget value p, all three MFMC estimators, when compared to the use of only the MC estimator, result in better accuracy for the same costs, with the "three-δ" estimators yielding about one order of magnitude and the "three-h" and "five-δ, h" estimators yielding two or three orders of magnitude reductions in the error for the same cost. Viewing the plots horizontally, we observe that for a desired MSE, all three MFMC estimator result in lower costs, with again the "three-h" and "five-δ, h" estimators yielding the largest gains. We also observe from Figure 9 that the use of surrogate models defined for larger values of h outperforms the use of surrogate models defined for smaller values of δ. Most

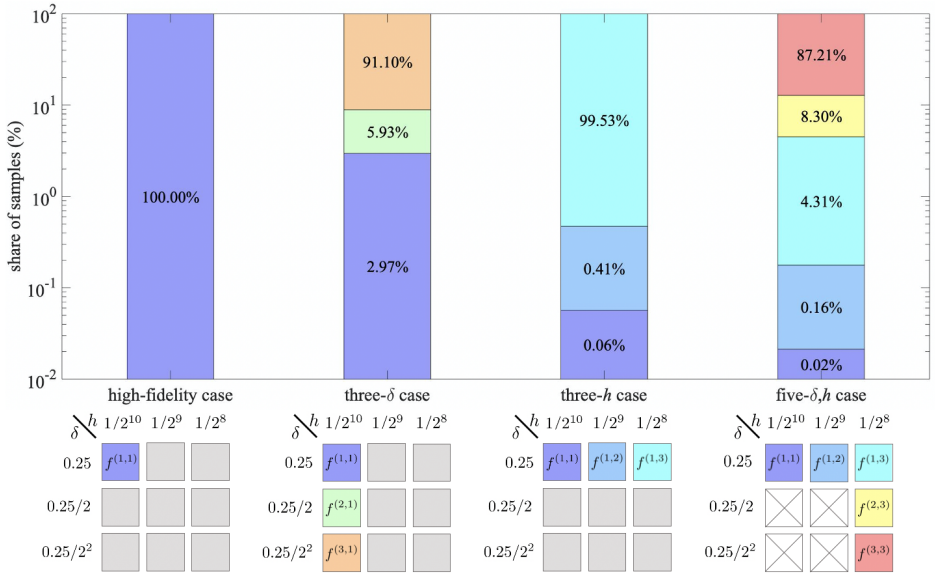


Figure 8: *The figure shows the distribution of the total number of samples across different models used in each MFMC case. Note the logarithmic scale on the vertical axis. The gray boxes indicate surrogates that were not included for each particular case. The boxes with crosses indicate surrogates that were eliminated because they violated the second requirement of* (21).

of the gain that the "five-δ, h" estimator achieves over the single-model case is due to the inclusion of different h surrogates relative to the gains affected by instead using different δ surrogates.

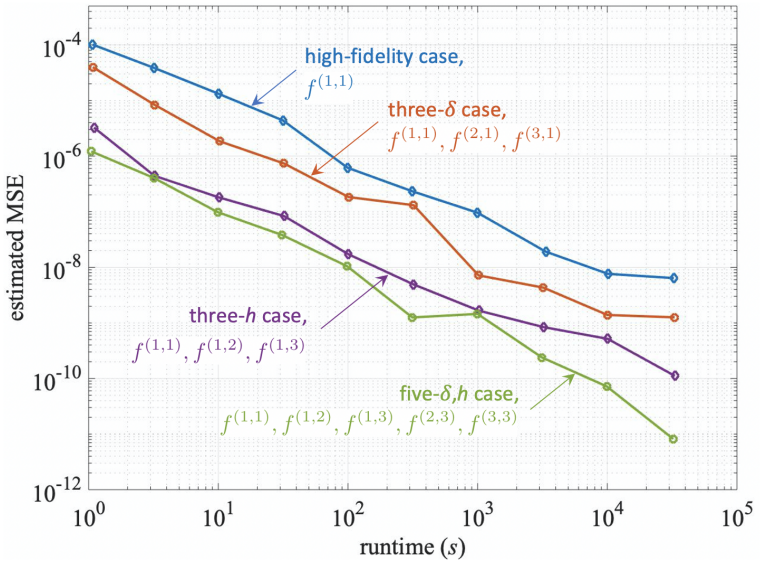


Figure 9: *Estimated mean-squared errors vs. the budget p in wall-clock seconds for the three MFMC estimators and the MC estimator.*

3 Multifidelity Monte Carlo methods for "realistic" computational budgets[6]

This work was published in [3].

Multifidelity Monte Carlo (MFMC) methods have shown great promise for the efficient and flexible estimation of statistical quantities. As experimental data can take a variety of forms, the primary advantage of MFMC estimation is its ability to accommodate a diverse ensemble of information sources which may be unrelated apart from predicting the same quantity of interest. However, despite their success, MFMC methods can "break down" when the model simulation costs are large relative to the available computational budget.

MC and MFMC estimators. Let $\{f^1(\boldsymbol{z}), f^2(\boldsymbol{z}), \ldots, f^K(\boldsymbol{z})\}$ denote scalar outputs of interest of a set of K computational models that depend on the random vector of parameters $\boldsymbol{z} \in \Gamma$ with decreasing computational complexity, where here Γ denotes a parameter domain. Recall that Monte Carlo (MC) sampling yields the K unbiased estimators

$$Q_k^{MC} = \frac{1}{M_k} \sum_{m=1}^{M_k} f^k(\boldsymbol{z}_m) \approx \mathbb{E}\left[f^k\right] \quad \text{for } k = 1, \ldots, K,$$

where $\boldsymbol{z}_1, \ldots, \mathbf{z}_{M_k}$ denote M_k i.i.d. samples in Γ and $\mathbb{E}[\,\cdot\,]$ denotes the expected value.

The goal of MFMC is to estimate the expectation quantity of interest $Q = \mathbb{E}\left[f^1\right]$ using a linear combination of standard MC estimators. These can be used to define the MFMC estimator

$$Q^{MFMC} = Q_1^{MC} + \sum_{k=2}^{K} \alpha_k \left(Q_k^{MC} - Q_{k-1}^{MC}\right) \approx \mathbb{E}\left[f^1\right] \tag{23}$$

which is unbiased by linearity and where $\{\alpha_k\}_{k=2}^K$ denotes a set of scalar weights and $\{M_k\}_{k=1}^K$ denotes a set of increasing integers defining the number of samples.

Letting C_k detnote the cost of evaluating the k^{th} model output f^k, the costs of computing the respective MC and MFMC estimators are given as

$$C_k^{MC} = M_k C_k \qquad \text{and} \qquad C^{MFMC} = \sum_{k=1}^{K} C_k M_k = \mathbf{C} \cdot \mathbf{M},$$

where $\mathbf{C}$ and $\mathbf{M}$ denote K-vectors having components $\{C_k\}_{k=1}^K$ and $\{M_k\}_{k=1}^K$, respectively.

For a fixed computational budget $b = \mathbf{C} \cdot \mathbf{M}$, MFMC aims to construct an optimal sampling strategy $\mathbf{M} = \{M_k\}_{1=2}^K$ along with an optimal set of weights $\boldsymbol{\alpha} = \{\alpha_k\}_{k=2}^K$ so that the mean squared error (MSE) of the estimator Q^{MFMC} is lower than that of the Monte Carlo estimator Q_1^{MC}.

Denoting the variance of the k^{th} model output f^k and the correlation between models f^k and $f^{k'}$ by

$$\sigma_k^2 = \mathrm{Var}\left[f^k(\boldsymbol{z})\right] \qquad \text{and} \qquad \rho_{k,k'} = \frac{\mathrm{Cov}\left[f^k(\boldsymbol{z}), f^{k'}(\boldsymbol{z})\right]}{\sigma_k \sigma_{k'}},$$

[6]Anthony Gruber (postdoc) at Florida State University and Professors Lili Ju and Zhu Wang at the University of South Carolina were collaborators on this project; they were not supported by AEOLUS related grants. FSU and USC personnel were supported by U.S. Department of Energy grants DE-SC0020418 and DE-SC0020270, respectively.

respectively, we have that the MSE of the MC estimator Q_k^{MC} of the quantity of interest Q is simply

$$e(Q_k^{MC}) = \mathbb{E}\left[\left(\mathbb{E}\left[f^k(\boldsymbol{z})\right] - Q_k^{MC}\right)^2\right] = \mathbb{E}\left[f^k(\boldsymbol{z})\right]^2 - \mathbb{E}\left[f^k(\boldsymbol{z})^2\right] = \frac{\sigma_k^2}{M_k},$$

whereas the MSE of Q^{MFMC} is given by (see [5]) as

$$e(Q^{MFMC}) = \mathrm{Var}\left[Q^{MFMC}\right] = \frac{\sigma_1^2}{M_1} + \sum_{k=2}^{K}\left(\frac{1}{M_{k-1}} - \frac{1}{M_k}\right)\left(\alpha_k^2\sigma_k^2 - 2\alpha_k\sigma_k\sigma_1\rho_{1,k}\right). \tag{24}$$

Then, the optimal weights $\{\alpha_k^*\}_{k=2}^K$ and sampling numbers $\{M_k^*\}_{k=1}^k$ of the MFMC sampling strategy are defined through as the solution to the mixed integer minimization problem (MIP)

$$\underset{\mathbf{M}\in\mathbb{Z}^k,\boldsymbol{\alpha}\in\mathbb{R}^{k-1}}{\operatorname{argmin}} \quad e(Q^{MFMC}) \qquad \text{such that} \quad M_1 \geq 1, \quad M_{k-1} \leq M_k, \quad 2 \leq k \leq K, \quad \mathbf{C}\cdot\mathbf{M} \leq b. \tag{25}$$

Analytic solutions to the problem (25) are not easy to obtain (if they even exist). However, the continuous relaxation obtained by letting $M_k \in \mathbb{R}^+$ for all $k = 1, 2, \ldots, K$ has an analytic solution (see [5, Theorem 3.4]). Unfortunately, the optimal sampling numbers $\mathbf{M}^*$ computed according to this relaxation are not guaranteed to be integers. Thus, in practice, MFMC is implemented by rounding downwards, i.e., we make the substitution $\mathbf{M}^* \leftarrow \lfloor \mathbf{M}^* \rfloor$ for the optimal sampling number (see [5, Algorithm 2]).

For large budgets b, this is of little consequence – each model output is evaluated at least once and often some of the models are evaluated many times, so replacing M_k^* by largest smaller integer provides a simple and reasonable solution which is guaranteed to be feasible (if potentially suboptimal) for the original problem (25). However, this approach could be unsuitable in cases for which the cost of obtaining model outputs is large but computational budgets do not allow for a large number of samples of those outputs. Rounding in this case not only can violate the constraints of the problem but also the biases the estimator Q^{MFMC}, destroying the performance of the method. To remove the bias, one could instead round up, i.e., use $\lceil M_k^* \rceil$. However, in the case of expensive outputs, one finds that $M_1^* \ll 1$ and perhaps likewise for other sample numbers so that rounding up to the nearest integer results in violation of the available computational budget.

A modified MFMC estimator. Whereas simple rounding is not sufficient to repair MFMC estimation, a slightly modified MFMC algorithm always yields a solution which is budget constraint-preserving, feasible for (25), and optimal up to the rounding downwards of M_k^* of some M_k^*. First, note that it can be shown by repeating the arguments in [5] that the Lagrangian

$$L^k = e(Q^{MFMC}) + \lambda\left(\sum_{k'=k}^{K} C_{k'}M_{k'} - \left(B - \sum_{k'=1}^{k-1} C_{k'}\right)\right) + \sum_{k'=k+1}^{K}\mu_{k'}\left(M_{k'} - M_{k'-1}\right) - \xi M_k \tag{26}$$

has a unique global minimizer corresponding to a pair $(\mathbf{M}^*, \boldsymbol{\alpha}^*) \in \mathbb{R}^{K-k+1} \times \mathbb{R}^{K-k}$ for any $1 \leq k \leq K$.

Lemma. *Let $\{f^k\}_{k=1}^K$ denote the computational model outputs of interest with correlation coefficients $\{\rho_{1,k}\}_{k=2}^K$ and computational costs $\{C_k\}_{k=1}^K$ respectively satisfying*

$$|\rho_{1,k-1}| > |\rho_{1,k}| \qquad \textit{and} \qquad \frac{C_{k-1}}{C_k} > \frac{\rho_{1,k-1}^2 - \rho_{1,k}^2}{\rho_{1,k}^2 - \rho_{1,k+1}^2}, \qquad k = 2, \ldots, K. \tag{27}$$

Let

$$r_k^* = \frac{M_{k'}^*}{M_k^*} = \sqrt{\frac{C_k}{C_{k'}}\left(\frac{\rho_{1,k'}^2 - \rho_{1,k'+1}^2}{\rho_{1,k}^2 - \rho_{1,k+1}^2}\right)}, \qquad k = k', \ldots, K.$$

Then, for each $k = 1, \ldots, K$*, the unique global minimum* $(\mathbf{M}^*, \boldsymbol{\alpha}^*) \in \mathbb{R}^{K-k+1} \times \mathbb{R}^{K-k}$ *of the Lagrangian* L^k

$$M_k^* = \frac{B - \sum_{k'=1}^{k-1} C_{k'}}{\sum_{k'=k}^{K} C_{k'} r_{k'}}, \quad M_{k'}^* = M_k^* r_{k'}^*, \quad \textit{and} \quad \alpha_{k'}^* = \frac{\rho_{1,k'}\sigma_1}{\sigma_{k'}}, \qquad k = k'+1, \ldots, K. \qquad \square$$

The solution pair $(\mathbf{M}^*, \boldsymbol{\alpha}^*) \in \mathbb{R}^K \times \mathbb{R}^{K-1}$ may or may not contain components $M_k^* < 1$. If

> if no component of $\mathbf{M}^*$ is less than one, then no modification is needed and simple rounding downwards produces the model evaluation numbers $\mathbf{M}^* \leftarrow \lfloor \mathbf{M}^* \rfloor$ which are nearly optimal and *preserves the computational budget.*

Otherwise,

> if the first component $M_1^* < 1$, then that component is redefined to be $M_1^* = 1$ and the function L^2 in (26) is minimized for the remaining $K-1$ components of $\mathbf{M}^*$.

This gives a new solution vector $(M_2^*, \ldots, M_k^*)$ which may or may not contain entries less than 1. If it does not, the model evaluation vector becomes $\mathbf{M}^* = (1, M_2^*, \ldots, M_k^*)$, where here M_k, $k = 2, \ldots, K$ are have been rounded downwards. Otherwise,

> this process is repeated and the functions L^k are minimized until there are no more indices k satisfying $M_k^* < 1$.

The result is

> a pair $(\mathbf{M}^*, \boldsymbol{\alpha}^*) \in \mathbb{R}^K \times \mathbb{R}^{K-1}$ which represents a small perturbation from continuous optimality at each step and is guaranteed to be nearly optimal for the original problem (25).

In fact, there is the following guarantee.

Theorem. *Suppose* $\mathbf{M}^*$ *is the solution to the relaxation of* (25) *with* $M_1^* < 1$*. Then,* $M_1^* \leftarrow 1$ *is optimal for* (25)*. Similarly,* $M_k^* \leftarrow 1$ *is optimal for* (25) *whenever* $M_1^* = \cdots = M_{k-1}^* = 1$ *are fixed and* $M_k^* < 1$ *is defined by the global minimizer of* L^k*.* $\square$

This theorem provides a guarantee that the modification is optimal for the original MIP (25), and leads to the immediate conclusion that the modified MFMC algorithm is at least as optimal for (25) as the original algorithm from [5]. Indeed, the above shows that $M_1^* \leftarrow 1$ is the integer value leading to the smallest Q^{MFMC} whenever $M_1^* < 1$ solves the relaxation of (25). Similarly, if $M_1^* = \cdots = M_{k-1}^* = 1$ and $M_k^* < 1$ minimizes L^k, then $M_k^* \leftarrow 1$ leads to the smallest Q^{MFMC} and satisfies the constraints of the MIP (25). Repeating this sequentially until there are no more $M_k^* < 1$ immediately yields the following corollary.

Corollary. *The modified MFMC algorithm is optimal up to the rounding* $\mathbf{M}^*$ *downwards inherent in the original procedure.* $\square$

This corollary is analogous to the similar guarantees in [5, 6] which establish optimality of the model evaluation vector $\mathbf{M}^*$ for the relaxation of (25).

A numerical illustration. A simple example is presented which compares the performance of the modified MFMC estimator to both the simple MC estimator and the original MFMC estimator of [5] on a benchmark experiment use in that paper. Consider the analytic model of a short column with rectangular cross-sectional area subject to bending and axial force. The high-fidelity model is

$$f^1(\boldsymbol{z}) = 1 - \frac{4z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5}{z_1 z_2 z_3}\right)^2,$$

where $\boldsymbol{z} \in \Gamma = [5,15] \times [15,25] \times \mathbb{R}^+ \times \mathbb{R}^2 \subset \mathbb{R}^5$. The random variable $\boldsymbol{z}$ is chosen such that the width and depth z_1, z_2 are distributed uniformly, the yield stress z_3 is distributed log-normally with mean 5 and standard deviation 0.5, the bending moment z_4 is distributed normally with mean 2000 and standard deviation 400, and the axial force z_5 is distributed normally with mean 500 and standard deviation 100. Surrogates are taken to be

$$f^2(\boldsymbol{z}) = 1 - \frac{3.8 z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5 \left(1 + \frac{z_4 - 2000}{4000}\right)}{z_1 z_2 z_3} \right)^2 \qquad f^3(\boldsymbol{z}) = 1 - \frac{z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5 (1 + z_4)}{z_2 z_3} \right)^2$$

$$f^4(\boldsymbol{z}) = 1 - \frac{z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5 (1 + z_4)}{z_1 z_2 z_3} \right)^2 \qquad f^5(\boldsymbol{z}) = 1 - \frac{z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5}{z_1 z_2 z_3} \right)^2,$$

where we note that f^3 and f^5 are reversed from their definitions in [5]. Letting $\boldsymbol{\rho}_1 = \{\rho_{1,k}\}_{k=1}^K$, the costs and model correlation coefficients are

$$\mathbf{C} = \begin{pmatrix} 100 & 50 & 20 & 10 & 5 \end{pmatrix}^\intercal \quad \boldsymbol{\rho}_1 = \begin{pmatrix} 1.0000000 & 0.9998791 & 0.6621724 & 0.8603847 & 0.9967698 \end{pmatrix}^\intercal,$$

where 1000 i.i.d. samples of $\boldsymbol{z}$ are used to compute the displayed (approximate) correlations. With this data, the model selection procedure [5, Algorithm 1] selects the subset $\{f^1, f^2, f^5\}$ to be used in constructing the MFMC estimator for $Q = \mathbb{E}\left[f^1\right]$.

Now, recall that the original MFMC procedure [5, Algorithm 2] can break down in the case that the budget b is "small". On the other hand, the optimal solution to (25) can be naïvely approximated by rounding M_k upwards when $M_k < 1$ and rounding M_k downwards otherwise, at the expense of potentially overshooting the budget constraint $\mathbf{C} \cdot \mathbf{M}^* = b$.

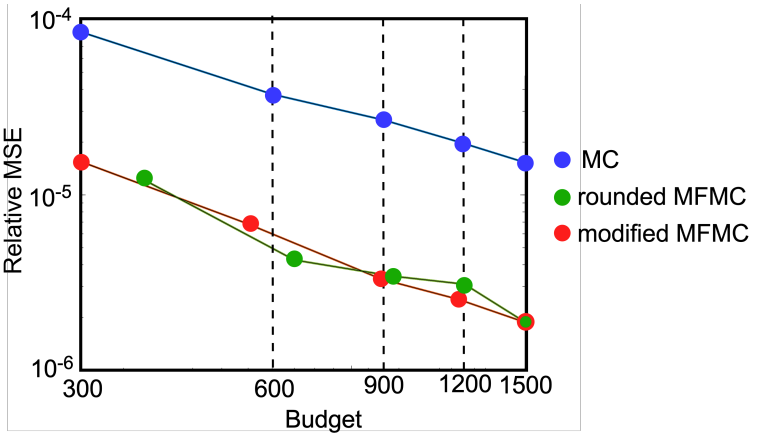
Consider using simple MC and both versions of MFMC to estimate $Q_{\text{ref}} = \frac{1}{10^7} \sum_{n=1}^{10^7} f^1(\boldsymbol{z}_n) \approx \mathbb{E}\left[f^1(Z)\right]$ with budgets $b = 300i$ for $i = 1, \ldots, 5$. To further mitigate the inherent dependence on stochasticity, the expectations and MSEs predicted for the three estimation methods are themselves averaged over 1000 independent runs of each method, so that

$$\mathbb{E}\left[Q_j^\sharp\right] \approx \frac{1}{1000} \sum_{j=1}^{1000} Q_j^\sharp, \qquad \mathbb{E}\left[\left(Q_j^\sharp - Q_{\text{ref}}\right)^2\right] \approx \frac{1}{1000} \sum_{j=1}^{1000} \left(Q_j^\sharp - Q_{\text{ref}}\right)^2,$$

where, $Q_j^\sharp \approx \mathbb{E}\left[f^1\right]$ for $Q_j^\sharp$ with $\sharp = \{\text{MC, rounded MCMF, modified MFMC}\}$ denotes the approximation produced at run j.

The table and figure below illustrate the results of the comparison between MC and both variants of MFMC. As expected, both MFMC procedures effect an MSE reduction of about an order of magnitude relative to MC and remain similar to each other in performance as the budget increases. However, the rounding procedure necessary for a direct application of [5, Algorithm 2] clearly exceeds the budget for the low budget cases $b = 300$ and 600. Also note that the high-fidelity model is only queried once and that almost all the sampling is done using the least costly model. Also, as expected, the prescribed budget is never exceeded when using the modified MFMC algorithm.

MC		Modified MFMC (Rounded MFMC)				
b	$e(Q^{MC}) \times 10^5$	Budget	# f^1	# f^2	# f^5	$e(Q^{MFMC}) \times 10^6$
300	8.183	300 (375)	1 (1)	1 (1)	30 (45)	14.91 (11.87)
600	3.612	550 (650)	1 (1)	1 (2)	80 (90)	6.579 (4.102)
900	2.572	890 (930)	1 (1)	3 (3)	128 (136)	3.198 (3.315)
1200	1.896	1180 (1205)	1 (1)	4 (4)	176 (181)	2.442 (2.984)
1500	1.475	1470 (1485)	1 (1)	5 (5)	224 (227)	1.856 (1.857)



References

[1] O. Burkovska and M. Gunzburger; *On a nonlocal Cahn–Hilliard model permitting sharp interfaces*, Mathematical Models and Methods in Applied Sciences **31** 2021 1749–1786.

[2] O. Burkovska, S. DeWitt, M. Gunzburger, and B. Radhakrishnan; *Nonlocal models for solidification* in preparation.

[3] A. Gruber, M. Gunzburger, L. Ju, and Z. Wang, *A multifidelity Monte Carlo method for realistic computational budgets*; Journal of Scientific Computing **94** 2023, 1145709.

[4] R. Kobayashi, *Modeling and numerical simulations of dendritic crystal growth*; Physica D: Nonlinear Phenomena **63** 1993, 410–423.

[5] B. Peherstorfer, K. Willcox and M. Gunzburger; *Optimal model management for multifidelity Monte Carlo estimation*; SIAM Journal on Scientific Computing **38** 2016, A3163–A3194.

[6] B. Peherstorfer, M. Gunzburger, and K. Willcox; *Convergence analysis of multifidelity Monte Carlo estimation*; Numerische Mathematik **139** 2018, 683–707.

[7] P. Khodabahshi, O. Burkovska, K. Willcox, and M. Gunzburger; Multifidelity methods for uncertainty quantification of a nonlocal model for phase changes in materials; *Computers & Structures* 297 (2024): 107328.

[8] A. Gruber, R. Lan, M. Gunzburger, L. Ju, and Z. Wang, Multifidelity Monte Carlo estimation for efficient uncertainty quantification in climate-related modeling; *EGUsphere* 2022 (2022): 1-27..

[9] P. Khodabakhshi, K. Willcox, and M. Gunzburger, A multifidelity method for a nonlocal diffusion model; *Applied Mathematics Letters* 121 (2021) 107361.

[10] O. Burkovska and M. Gunzburger, On a nonlocal Cahn-Hilliard model permitting sharp interfaces; *Mathematical Models and Methods in Applied Sciences* 31.09 (2021): 1749-1786.

[11] Q. Du, M. Gunzburger, R. B. Lehoucq, K. Zhou, Analysis and approximation of nonlocal diffusion problems with volume constraints, *SIAM Review* 54 (2012) 667–696.

[12] Q. Du, M. Gunzburger, R. B. Lehoucq, K. Zhou, A nonlocal vector calculus, nonlocal volume-constrained problems, and nonlocal balance laws, *Mathematical Models and Methods in Applied Sciences* 23 (2013) 493–540.

[13] M. D'Elia, Q. Du, C. Glusa, M. Gunzburger, X. Tian, Z. Zhou, Numerical methods for nonlocal and fractional models, *ACTA Numerica* 30, (2020)

[14] S. Heinrich, Multilevel Monte Carlo methods; *Large-Scale Scientific Computing* 2179 (2001) 58-67; Springer.