PROJECT SUMMARY

Overview:

This project will investigate and develop new methodologies, algorithms, and software to improve the individualized assessment of high-dimensional nonlinear dynamic systems by enabling the integration of complex domain knowledge--yielded by computer simulation of domain physical models--into the process of data-driven inference. Its overarching theme is flexibility and robustness. The flexibility is needed for a plug-and-play inclusion of domain physical models of any form without incurring prohibitive computation, reconciling between the conflicting needs for accuracy vs. efficiency. The robustness is needed when there is a lack of measurements to reliably resolve the error in model-generated prior knowledge. The driving application of this proposal is individualized modeling of in-vivo physiological systems using noninvasive biomedical data. Identified as one of the grand challenges for future healthcare, research in individualized physiological modeling is increasingly enabled by advances in computer simulation and sensor/imaging technology. The path to this end, however, is hindered by a separation between the current state of physiological simulations that are generally decoupled from individual data, and data analysis that struggles for realistic physiological contexts. This project aims to bridge this gap.

Intellectual Merit :

The proposed research builds and expands on the foundations of dynamic Bayesian inference, addressing two fundamental aspects for improving the quality of data-driven inference. First, a flexible inclusion of domain physical models will place the inference in a context that is relevant to the underlying physics of the domain problem. Second, for the inference to be robust to the error in the physical model, the proposed research goes outside the traditional regime of dynamic inference, and connects it with low-dimensional structures of the system (eg, sparsity in a basis) to further overcome the lack of measurements. Specifically, the integrated research and education plan address three challenges: 1) To enable a plug-and-play inclusion of domain physical models catering to different efficiency vs. accuracy needs: the key challenge is to combine the flexibility of Monte-Carlo methods and the efficiency of non-sampling analytical methods when extracting probabilistic priors from dynamic models. 2) To further overcome the lack of measurements by exploiting low-dimensional structures in high-dimensional systems: the key challenge is to strike a balance when integrating sparsity models with the dynamic context provided by domain physical models. 3) To enable a robust adaptation of the time-varying error in domain physical models: the key challenge is--given increased dimensionality of unknowns and limited measurements--to extend the concept of low-dimensional structural constraints to extract the significant information about model errors. Interdisciplinary collaborations are included to support both the application context and foundational infrastructure of this project. Scientific contributions of this project are two-fold. It will map out a new technical paradigm for individualized physiological modeling, bridging the ubiquitous gap between the current state of computer simulation and data analysis. It will also contribute to the foundations of statistical inference, bridge the current separation between the use of dynamic and sparsity models, and extend to a wide range of applications such as tumor modeling, climate modeling, systems biology, and finance.

Broader Impacts :

This project will deliver publicly-available multicore/GPU software that encapsulates the most effective algorithms developed for data-driven modeling and inference. This toolkit will contribute to the national effort toward noninvasive medicine and healthcare, and benefit a wide range of computational scientists. An integrated educational and outreach program is designed to foster interdisciplinary research training and to increase participation of underrepresented groups in STEM disciplines. Primary focuses include: 1) development and evaluation of "learning-by-doing" concept in graduate and undergraduate education; 2) research training for students at graduate, undergraduate, and high-school levels, with a focus on engaging women and underrepresented students at an early stage; and 3) broader outreach activities to area K-12 students and Paramedic communities. The participation of women, underrepresented, K-12, and Paramedic groups are reinforced through continued partnerships with programs offered in [redacted], local school district, and community college.

CAREER: Integrating Physical Models into Data-Driven Inference

1 Introduction

This project will investigate and develop new methodologies, algorithms, and software to improve the individualized assessment of high-dimensional nonlinear dynamic systems by enabling the integration of complex domain knowledge—yielded by computer simulation of domain physical models—into the process of data-driven inference. Its overarching theme is *flexibility* and *robustness*. The flexibility is needed for a *plug-and-play* inclusion of domain physical models of any form without incurring prohibitive computation, reconciling between the conflicting needs for efficiency *vs*. accuracy. The robustness is needed when there is a lack of measurements to reliably resolve the error in model-generated prior knowledge. The driving application is *individualized modeling* of *in-vivo* physiological systems—using noninvasive biomedical data—for improved disease prevention, diagnosis, and treatment. Outcomes of this project will contribute theoretically, algorithmically, and computationally to the foundations of statistical inference, and extend to a wide range of applications, *e.g.*, tumor modeling, climate modeling, systems biology, and finance.

1.1 Motivation

Motivations from Noninvasive Healthcare: The 2012 *IEEE Life Sciences Grand Challenges Conference* identified *individualized* quantitative modeling as one of the grand challenges for future medicine and healthcare: by offering a personalized, quantitative description of organic physiological activities, it can improve noninvasive health monitoring and prediction of adverse events, and thus better inform and engage patients [1].

To this end, on one hand, computer modeling and simulation are playing an increasingly important role in supporting quantitative understanding of multi-scale, multi-physics systems of organisms. A particular example is the *cardiac physiome model*—fruit of an international collaboration—that describes the coupling of metabolic, electrophysiological, and biomechanical processes of cardiovascular systems across from cell to organ levels [2]. However, making little usage of real time patient measurements, these *in-silico* assessments are traditionally decoupled from individual patients and their pathophysiological conditions [3].

On the other hand, owing to recent advances in unobtrusive body-sensor networks and high-resolution imaging [3], physiological data of individualized health conditions are growing both in quantity and quality. Nevertheless, there remains a major gap between what can be noninvasively measured from humans and what is needed to specify the underlying physiological processes [1]. Traditional methods for biomedical signal and image processing that are only data driven, therefore, face inherent challenges from the lack of realistic physiological contexts. For instance, in the attempt to assess cardiac bioelectrical activity using noninvasive body-surface recordings, it is known that—without meaningful physiological knowledge—a theoretically unique inference can only be made on the heart surface but not within the heart muscle [4].

To summarize, while modern computer simulation and sensor/imaging technology are offering enabling techniques for individualized physiological modeling, the gap between is yet to be bridged: this emerging research opportunity is ubiquitous in many fields, and it constitutes the focus of this proposal.

General Scientific Barriers: To bridge the above gap, the proposed research will build and expand on the foundations of dynamic Bayesian inference [5]. Though an emerging concept in individualized physiological modeling, dynamic Bayesian inference is popular in many traditional signal-processing applications—such as tracking [6] and speech recognition [7]—for its ability to include the underlying dynamics of a signal (in the form of mathematical models) to improve the quality of signal processing [5]. Its modern applications also extended to a wider variety of domains, such as *data assimilation* in weather forecasting [8] and oceanic modeling [9], or *data-driven simulation* in wildfire spread prediction [10]. However, for increasingly-emerging high-dimensional nonlinear applications, two fundamental issues remain to be resolved.

First, to include domain physical models, existing methods often trade flexibility for efficiency, or vice versa: Monte Carlo (MC) based methods allows a flexible inclusion of physical models of any form, yet their application to high-dimensional systems suffer from the *curse of dimensionality* and prohibitive computation [11, 12]; non-sampling analytical methods that are efficient in linear systems, on the other extreme, can be difficult or intractable when domain models have nontrivial or nonlinear forms.

Second, after prior knowledge about domain dynamics is included, it is important for the inference to be robust to the error in this knowledge. In high-dimensional systems, however, limited measurements puts the inference at the risk of being dominated by the model error. While model-error estimation has been researched for decades [13], its progress in high-dimensional nonlinear applications remains initial due to the increasing difficulty to overcome the lack of measurements as the number of unknowns increases [14].

1.2 Project Overview

This CAREER project aims to build a synergistic research and education program that advances knowledge about data-driven modeling and inference, contributes to the national effort toward noninvasive medicine and healthcare, and enhances learning and participation of underrepresented students—across from graduate, undergraduate, to K-12 levels—in STEM education and research.

Intellectual Merit: This project proceeds along two major thrusts—flexibility and robustness—to address the two critical issues of dynamic Bayesian inference in high-dimensional nonlinear applications. This encloses the fundamental contributions and novelties of the proposed research. First, a flexible inclusion of domain physical models is critical for improving the quality of data-driven inference by placing it within a context that is relevant to the underlying physics of the domain problem. Second, to be robust to the errors in domain physical models given limited measurements, the proposed research moves outside the traditional regime of dynamic inference, and integrates it with the exploitation of low-dimensional structures of the system (*e.g.* sparsity in a basis) to further over come the lack of measurements. Scientific contributions of this research are thus two-fold. It will map out a new perspective and technical paradigm for individualized physiological modeling, bridging the ubiquitous gap between the current state of computer simulation and data analysis. It will also contribute to the foundations of statistical inference, bridge the current separation between the use of dynamic and sparisity constraints, and benefit a broader range of scientific applications. The integrated research and education plan addresses the following challenges:

1. *To enable a* plug-and-play *inclusion of domain physical models for high-dimensional nonlinear applications with different efficiency vs. accuracy needs*: The key challenge is to combine the flexibility of MC simulation and the efficiency of non-sampling analytical methods.
2. *To further overcome the lack of measurements by exploiting the low-dimensional structure in high-dimensional systems*: The key challenge is to strike a balance when integrating the low-dimensional structure with the dynamic context provided by domain physical models.
3. *To enable a robust adaptation of the time-varying error in domain physical model*: The key challenge is—given the increased dimensionality of unknowns and limited measurements—to extend the concept of low-dimensional structural constraints to extract the significant information about model errors.

This project builds on the application context of individualized cardiac electromechanics [15–25] that the PI is experienced with. To ensure domain relevance, research evaluation will be supported by collaborators in cardiovascular imaging [redacted]), electrocardiographic mapping ([redacted]), and clinical electrophysiology [redacted]). For foundational and infrastructure support, complementary expertise will be sought from mathematical modeling [redacted]), hardware architecture [redacted]), and software infrastructure [redacted]). These collaborations will also foster the long-term success of the proposed research and the career development of the PI.

Broader Impact: This project will deliver publicly-available multicore/GPU software that encapsulates the most effective algorithms developed. These toolkits will enable an improved noninvasive, quantitative assessment of individualized cardiac conditions, with the potential to better assist physicians and inform patients with pre-procedural therapy planning, long-term health monitoring, and preemptive actions. They will also support numerous scientific applications involving data-driven modeling and inference.

The integrated educational and outreach program is designed to both foster interdisciplinary research training and to increase participation of underrepresented students in STEM disciplines. Primary focuses include: 1) development and evaluation of *"learning-by-doing"* concept in graduate and undergraduate education; 2) research training for students from graduate to high-school levels, with a focus to engage female and underrepresented students at an early stage; and 3) broader outreach activities to area K-12 students and Paramedic communities. The participation of women, underrepresented, K-12, and Paramedic groups will be reinforced through continued partnerships with McNair Scholars & Louis Stokes Alliance for Minority Participation programs, Career Exploration programs for [redacted] K-12 schools, and [redacted]

[redacted] Paramedic Education Program (see corresponding letters of support).

Qualifications: The PI's past research and educational activities have laid out the basis for this project. The PI's dissertation research initialized the concept of the proposed physics-constrained, data-driven inference framework [15], focusing primarily on its utility in individualized cardiac electrical activity [15–19] and integrated cardiac electromechanics [20–25]. Having accumulated successful experiences and having built a research group with external collaborators from various backgrounds, the PI has recently moved into the foundational work of generalizing the fundamental formalisms and techniques to suit many applications. At the same time, the PI has been actively integrating research into her curricula development and student mentoring activities; in particular, considering it her responsibility to be a role model for women and other underrepresented students in STEM disciplines, the PI has been actively involved with research programs offered for area high-school students and underrepresented undergraduates. The key focuses of the proposed educational activities are all built on the PI's past effort and involvement with the relevant programs.

2 Novelty with Respect to State of the Art: Background & Related

Individualized Physiological Modeling: Toward the need of quantitative modeling of individualized physiological systems, the ubiquitous gap between the current state of computer modeling and signal processing can be seen in the following example of cardiac electrical modeling and inference.

Computer modeling & simulation: Over the last 50 years, computational modeling of electrical excitation has become an important tool to generate and test hypotheses of 3D arrhythmia mechanisms that are difficult to address experimentally [26,27]. These multi-scale models simulate the electrophysiology of a cardiac cell, integrate them into current propagation through cardiac tissue, and eventually provide the whole-heart action potential excitation [28]. There are currently more than one hundred models available [28], ranging from simple macroscopic models with 2-10 variables [29, 30], to complex biophysical models with 30-100 variables [31, 32]. However, traditional simulation environments of these models have limited patient-specific features which, to date, is limited to cardiac anatomy on which the simulation is performed [26–28].

Inference from noninvasive data: Meanwhile, for the past three decades, inference of cardiac electrical excitation using noninvasive electrocardiographic (ECG) or magnetocardiographic (MCG) data has been intensively researched in another field. This so-called inverse problem of ECG/MCG [4] suffers from ill-posedness not only from the limited data, but also from the underlying biophysics: different 3D distributions of electrical sources inside the heart muscle can produce identical electromagnetic field on an external surface [33]. To avoid the second challenge, a common approach has been to formulate inverse solutions only on the heart surface [34–37]: the remaining ill-posedness is then tackled by various regularization techniques [34,38–40] that enforce greatly-simplified heuristic models on the spatial or temporal behaviors of the bioelectrical sources. To infer 3D electrical excitation within the myocardium, the importance to include *realistic physiologically constraints* has long been recognized [35]. The few existing attempts, however, generally follow an approach similar to surface-based inferences and employ regularization with simplified heuristic models such as the spatial and/or temporal smoothness of the solution [41, 42]. Physical-model-based approaches were recently proposed to optimize selected parameters of a realistic 3D cardiac conduction model to fit external measurements [43,44]. Though the optimization is straightforward in concept, the measurements are typically from direct invasive mapping on the heart surface [44]. It has proven difficult to use indirect, noninvasive data to reliably fit a deterministic model of complicated forms [35, 37].

To date, these two lines of work have not been well connected and such gap is not uncommon in other applications. Bridging this gap, this project aims *to bring together the fields of computer simulation and statistical inference to map out a new technical paradigm for the research in individualized physiological modeling, and beyond*.

Dynamic Bayesian Inference: While known in different fields by different terms, *online* Bayesian inference of discrete-time signals can be united into the recursive formalism summarized in Fig. 1(A). At each time point k: the *a priori* probabilistic distribution of the unknown $\mathbf{x}_k$ is first predicted based on its dependency on $\mathbf{x}_{k-1}$; its posterior density is then obtained after incorporating the dependency of measurement $\mathbf{y}_k$ on it. Various estimators, such as the *maximum a posteriori* (MAP) or *posterior mean* estimator, can then be applied to obtain posterior estimates of $\mathbf{x}_k$. One key advantage of this approach is that domain knowledge of the spatiotemporal dynamics, in the form of a physical model $\mathbf{x}_k = f(\mathbf{x}_{k-1})$, can be included in the prediction.

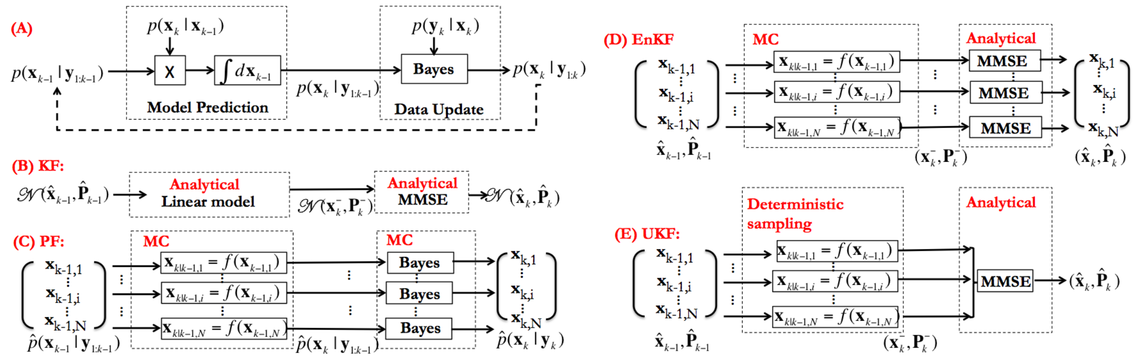

Figure 1: Schematics illustrating the overall structure of Bayesian filtering (A) and state-of-the-art variants (B-E).

Flexibility: The most popular techniques of dynamic Bayesian inference can be seen at the two ends of the spectrum of flexibility (Fig. 1(B)-(C)). In the simplest and most efficient version—the well-known Kalman filter (KF), both the prior and posterior density functions are analytically tractable, and the MAP estimator becomes equivalent to an efficient linear minimum-mean-square-error (LMMSE) estimator. However, KF and its variant (extended-KF) are only suitable for systems that are linear or suitable to be linearized.

On the other extreme, sequential Monte-Carlo—also known as *particle filtering* (PF)—can be used to fully implement the recursive Bayesian procedure, where density functions are numerically approximated by MC simulation using a set of random samples. In particular, prior distribution of $\mathbf{x}_k$ can be predicted by a straightforward, repetitive propagation of all the samples through $f(\mathbf{x}_{k-1})$ as a blackbox. However, MC-based methods suffer from *the curse of dimensionality*: the number of samples needed increases exponentially with the dimension of the system [11,12], leading to prohibitive computation on high-dimensional systems.

A tradeoff between the KF and PF has led to two modern techniques: unscented KF (UKF) [45] and ensemble KF (En-KF), the latter mostly used in atmospheric and oceanic *data assimilation* [8,46]. Both techniques share a similar concept that—instead of complete density functions—only the first two moments are of utmost interest to real-world applications (*i.e.*, mean and covariance provide an estimate and its uncertainty measure). Therefore, both techniques assume a *combined sampling-analytical structure* as illustrated in Fig. 1(D)-(E): sample-based methods are used for the prediction of the *a priori* moments of $\mathbf{x}_k$; a simple LMMSE estimator is adopted for the data-driven update. The UKF also further improves the efficiency by deterministic instead of random sampling to reduce the number of samples needed in the prediction. However, neither UKF nor EnKF has seen a wide application beyond their initial application areas [8,46].

Robustness: When strong knowledge about domain dynamics is included, it becomes important for the inference to be able to adapt to the error in this knowledge, especially when the measurements are limited. The common approach is to augment the system state with an auxiliary variable that represents unknown error or parameter of the dynamic model [13]. In high-dimensional applications, however, this path leads to an even higher-dimensional random space to sample, track and estimate [14,47–49]. Research along this line is thus still initial [14], hindered by not only the computation cost but more importantly, the increasing difficulty to overcome the lack of measurements as the dimensionality of unknowns increases substantially. To move forward, additional structural constraints on the unknowns need to be sought.

Outside the regime of dynamic inference, the low-dimensional structure of a signal, *i.e.*, its sparsity in a certain basis, has become a powerful regularizer to overcome the lack of measurements [50,51]. Recent success of compressive sensing [50] is an example that significantly increased the popularity of sparsity models. However, these studies mostly focus on static signals. To date, there has been little effort in exploiting low-dimensional structural constraints for improving the robustness of dynamic inference [52–54].

3 Research Plan

3.1 Vision: Theoretical Framework

Current state of dynamic inference research shows evidence that the *combined sampling-analytical structure* is suitable for most scientific applications that have high-dimensional nonlinear dynamics, with a relatively-simpler mapping process to the data space (*e.g.* a direct, partial measurement or governed by certain physics principles). The sample-based prediction is able to include prior knowledge of domain dynamics in any forms; the analytical approach for data-driven update avoids prohibitive computation without compromising practical application needs. However, on sample-based prediction, thoeretical foundations are yet to be generalized in order to have a widespread influence on a large range of scientific applications with different accuracy *vs*. efficiency needs. On analytical-based inference, there is a need to connect with the low-dimensional structure of the system so as to further overcome the lack of measurements and to improve the robustness of the inference.

Overcoming these challenges, the overall vision of the proposed research is a theoretical framework for the *combined sampling-analytical inference structure* that support both flexible and robust inclusion of domain physical models on a variety of systems that are high-dimensional, nonlinear, spatiotemporal-varying. It will be achieved through research tasks outlined in Fig. 2.

The first aim (A1) focus on the theoretical foundations of sample-based predication methods for a *plug-and-play* inclusion of domain physical models. Built on the success of deterministic sampling methods used in the UKF, we propose: 1) algorithmic reduction for the needs of higher efficiency; 2) generalization to different probabilistic distributions and higher accuracy.

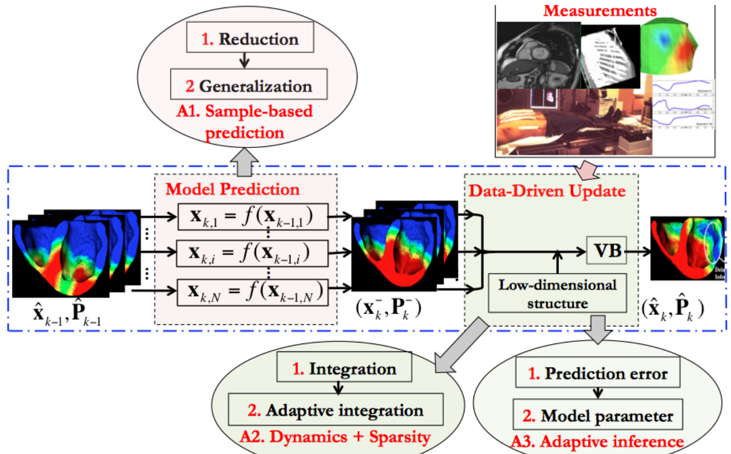

Figure 2: Theoretical framework (blue dashed frame) and research roadmap (oval frame). VB standards for Variational Bayes.

The second aim (A2) focus on integrating low-dimensional structural constraints into dynamic inference to further overcome the lack of measurements. We propose: 1) methods for truly integrating the sparsity model with statistics predicted by domain dynamic models; and 2) methods for adaptively integrating them for improved robustness. We also propose the use of variational Bayesian methods as a means to relax the restrictions of LMMSE-based approaches, while accelerating sample-based approach used in PF.

The final aim (A3) focuses on an enhanced robustness to estimate the time-varying model-error from measurements. We propose to extend the concept of low-dimensional structural constraints to: 1) extract the significant portion of model-predicted error; and 2) estimate the physical parameters of the system.

3.2 Sample-Based Prediction: *Plug-and-Play* Inclusion of High-Dimensional Nonlinear Models

Overview: The flexibility to accommodate a large variety of domain physical models essentially comes from two aspects: 1) the ability to include the model as a *blackbox*; and 2) not to incur prohibitive computation as the complexity or the dimensionality of the model increases (*e.g.,* at the order of 10^3 or up). This leaves us with the option of sample-based prediction methods, with sample locations fixed deterministically *a prior* with the aid of analytical methods to avoid the curse of dimensionality in random sampling.

Hence, investigations here will extend on the concept of *unscented transform* (UT) that was proposed in the UKF [45]. In brief, to track a random variable undergoing a nonlinear model: $\mathbf{u} = f(\mathbf{x})$, the UT keeps the nonlinear function intact, and approximates the statistical moments of $\mathbf{x}$ with the fewest possible number of samples under a prescribed order of accuracy. For example, if the selected samples $\{\mathcal{X}_i, \omega_i\}_{i=0}^N$ correctly captures the mean and covariance of $\mathbf{x}$, it can be shown that the mean and covariance captured by the corresponding samples $\{\mathcal{U}_i, \omega_i\}_{i=0}^N$ are correct at least to the second-order of accuracy. This was proven achievable using a minimal $n+1$ samples for an n-dimensional $\mathbf{x}$ [55]. By extending it to a $2n+1$ symmetrical set, third-order accuracy can be achieved. Most UT applications are seen with these two sample sets.

However, though seen in different applications [48, 56], the influence of deterministic sampling methods on nonlinear inference is not as widespread as expected; neither are we aware of foundational works that generalize this concept to fit a wider variety of applications. Building on the concept of UT, it is our intention to generalize on the theoretical foundations and algorithms of deterministic sampling methods, and to provide a *plug-and-play* software module for the use of a wide variety of scientific applications.

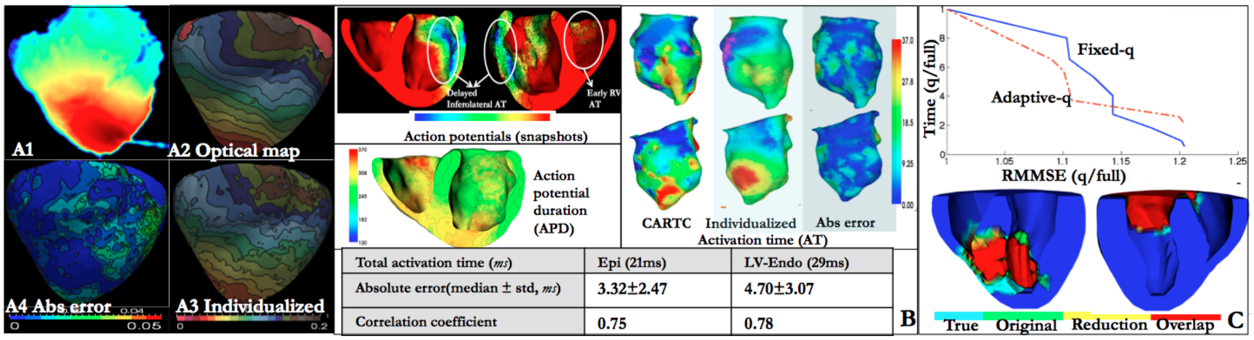

Figure 3: A: Preliminary data of activation isochrones (A3-A4) obtained on a healthy porcine model, verified by optical mapping data (A1-A2). B: Preliminary data on a post-infarction porcine model, verified by *in-vivo* catheter mapping (CARTO). C: Preliminary data (computer simulation study) on computational cost *vs.* quality of approximation.

Preliminary Work: We have successfully applied UKF to both noninvasive inference of individualized cardiac electrical excitation [15–18, 57, 58] and cardiac electromechanics [20, 23, 25]. For the former, stochastic simulation of a macroscopic two-variable cardiac excitation model [30] based on the $2n+1$ UT sample set was used ($n \approx 2-3\times 10^3$); owing to the electrophysiological context provided by this model, our technique has become one of the first —verified both *in silico* and *in vivo*—to be able to infer electrical activity beneath the heart surface using body-surface voltage data. For the latter, a multi-physics cardiac physiome model that includes electrical, electromechnical coupling, and biomechanical components was used to guide the inference; the $n+1$ UT scheme was used due to the increased cost of a more sophisticated model. Fig. 3A shows examples of our pilot study on a healthy porcine model during endocardial pacing [17]. Fig. 3B illustrates different physiological features of individualized cardiac excitation on a porcine model 5-week after infarction [18], where the delay and shortening of activation was consistent with the region of infarction seen in histological data. Current GPU implementation can finish one cardiac cycle within two hours [59].

Proposed Work: To generalize on the deterministic sampling methods, the proposed investigations will be united under the following optimization problem to find the location and weight of N samples $\{\mathcal{X}_i, \omega_i\}_{i=0}^N$:

$$\min_{\{\mathcal{X}_i,\omega_i\}_{i=0}^N} c[\{\mathcal{X}_i,\omega_i\}_{i=0}^N, p(\mathbf{x}), \bar{\mathcal{U}}, \mathbf{P}_{uu}] \quad \text{subject to} \quad g[\{\mathcal{X}_i,\omega_i\}_{i=0}^N, p(\mathbf{x}), \bar{\mathcal{U}}, \mathbf{P}_{uu}] = 0 \tag{1}$$

where $g[\cdot]$ include criteria the sample set must satisfy, *e.g.*, the statistics of $p(\mathbf{x})$ it needs to match; $c[\cdot]$ include additional constraints, such as higher-order errors, to be minimized. In comparison to the general formulation [55, 60], we include additional constraints on the sample mean and covariance of $\{\mathcal{U}_i, \omega_i\}_{i=0}^N$.

Reduction for higher efficiency: Standard UT samples lay at mean $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}} \pm d_i\sqrt{\mathbf{P_{xx}}}$, *i.e.*, a distance d_i along the axes determined by any square-root of the covariance matrix $\mathbf{P_{xx}}$ [45, 60]. Because our knowledge of $\mathbf{P_{xx}}$ is essentially non-exact during inference, we propose that the most efficient way to reduce computation at a minimal lost of statistical accuracy is to neglect the smallest variances, *i.e.*, to sample along the principal axes of $\mathbf{P_{xx}}$ in a decreasing order until the *dominant* uncertainty is accounted for, and ignore the remaining $n-q$ axes. Initially, singular value decomposition can be used to extract principal axes. We are currently developing two alternatives to decide the value of q: to fix the number of q (fixed-q), or to change the number of q for a prescribed percentage of variance (adaptive-q). Preliminary data in Fig. 3C illustrates their potential to provide a principled tradeoff between computation cost (y-axis) and quality of inference (x-axis) [19], where more than 60% of computation was reduced within 10% accuracy decrease in individualized action potentials. Fig. 3C also shows infarction regions detected by clustering the estimated bioelectrical features [19], where similar results were obtained with and without the proposed computation reduction.

To proceed, we will conduct the following studies. First, alternative methods will be investigated for locating the principal axes that are used to move the samples. For example, we can apply independent component analysis on the standard UT sample sets, and to identify a set of q statistically *most independent* samples to go through model simulations. Second, the proposed reduction leads to an under-estimation of $\mathbf{P_{xx}}$ that will accumulate and impact subsequent iterations. To compensate for this error, we will place the study under the optimization problem (1). Fixing the principal axes $\{\mathbf{p}_i\}_{i=1}^q$ along which to place the samples, we will find the optimal values for the distance d_i each sample should be placed away from the mean (*i.e.*,

$\mathcal{X}_i = \bar{\mathbf{x}} \pm d_i \mathbf{p}_i$), or the weight ω_i associated with each sample, so that the residual $\mathbf{P}^r = \sum_{i=1}^{q} \omega_i (d_i \mathbf{p}_i)^2 - \mathbf{P_{xx}}$ has the minimal l-2 norm. We will consider the common UT strategy where all samples (except the mean) share the same weight ω and distance d, as well as alternative strategy to iteratively optimize the value of ω_i or d_i starting with the *most important* sample located along the first principal axis of $\mathbf{P_{xx}}$. In a way, instead of empirical *covariance inflation* that multiplies the covariance matrix with a parameter slightly larger than 1 [61], we are seeking a more principled way to optimally distribute the inflation parameter to the samples. We will also investigate the optimization problem with respect to different values of q, and establish the theoretical bounds of accuracy lost in the approximated mean and covariance after the nonlinear model.

Generalization for different distributions & higher accuracy: An important nature of UT is to match the statistical moments without regard to the type of probabilistic distributions. In Bayesian inference, nevertheless, the distribution type of a prior plays an important role. Therefore, we will explore to establish different sampling schemes that are optimal for different distribution types, especially a few distributions that are commonly used in inference such as Gaussian, Laplacian, exponential, truncated Gaussian, and Gamma distributions. These distributions are important for implicitly imposing constraints on the random variable, such as sparsity or the empirical range of values, which will be utilized in sections 3.3 and 3.4.

Sample sets that match given mean and covariance up to a prescribed order of accuracy are not unique. We propose that additional constraints on distribution types can be imposed by matching both higher-order moments (*e.g.* skewness, kurtosis, etc) and entropy of the samples. In other words, the more property of a distribution the sample set satisfies, the less likely that it will fit other types of distributions. To include entropy and higher-order statistics of $\mathbf{x}$ into the optimization (1), we will extend the work of [62] to compose the sample set with several subsets: each sub-set contributes to part of $\mathbf{P_{xx}}$ and, similar to a standard UT set, shares the same weight and located at $\pm d\sqrt{\mathbf{P_{xx}}}$. Different subsets have different weights ω and distance d, which are to be found by the optimization (1). The match in higher-order moments and entropy can be enforced either through minimization or equality constraints: both alternatives will be investigated. For optimal sample sets derived for the same distribution types, we can further investigate the scaling of their weights so as to minimize higher-order differences in the approximated mean $\bar{\mathcal{U}}$ and covariance $\mathbf{P}_{uu}$ [60].

Because of the parallel nature of sample-based prediction, parallel algorithms will be developed for GPU and multicore machines. The end product of these investigations is a set of deterministic sampling algorithms and software that allow domain scientists—as they plug in their physical models for data-driven inference—to choose the types of distributions to impose, as well as the efficiency *vs.* accuracy tradeoff.

Exploratory Objective & Integration with Curricula: As computer models take on a new role in guiding data-driven inference, there arises a complex and multifaceted question: *what levels of complexity are desired in these models?* To make *in-sillico* simulation as realistic as possible, high model complexity is required. For data-driven inference, however, model factors that cannot be measured by limited data will lead to uncontrolled assumptions and increased uncertainties. Therefore, our working hypothesis is that we should use models that are as simple as possible, and yet accurate enough to reflect the underlying dynamics that can be measured. The PI plans to integrate this exploratory research theme into her courses on computational modeling and simulation (both graduate and undergraduate). In course projects and in-class practices, students will apply sensitivity analysis [63] to study how the uncertainty of a model output is related to the uncertainty of each model factor. Students will also use objective statistical criteria to determine which factors are most responsible for model uncertainty and deserves further analysis in data-driven inference. Students will experiment with domain-specific dynamic models at a wide range of complexity. Using the cardiac excitation model as an example, students will compare two-variable macroscopic models used in our preliminary study, *reduced biophysical models* that provide a closer match of cardiac cell properties without ion-channel details (*e.g.*, [redacted]), and complex detailed ionic models [32]. The PI plans to include different models and revise analysis methods for each offering of her courses throughout this 5-year project, hoping to accumulate a relatively comprehensive understanding about the issue of model selection for the purpose of data-driven inference.

3.3 Robustness: Exploiting Low-Dimensional Structure in High-Dimensional Dynamic Inference

Overview: In a typical approach of dynamic inference, the dynamic prior as generated in section 3.2 would be directly combined with the measurements for approximating the posterior distributions of $\mathbf{x}_k$. However,

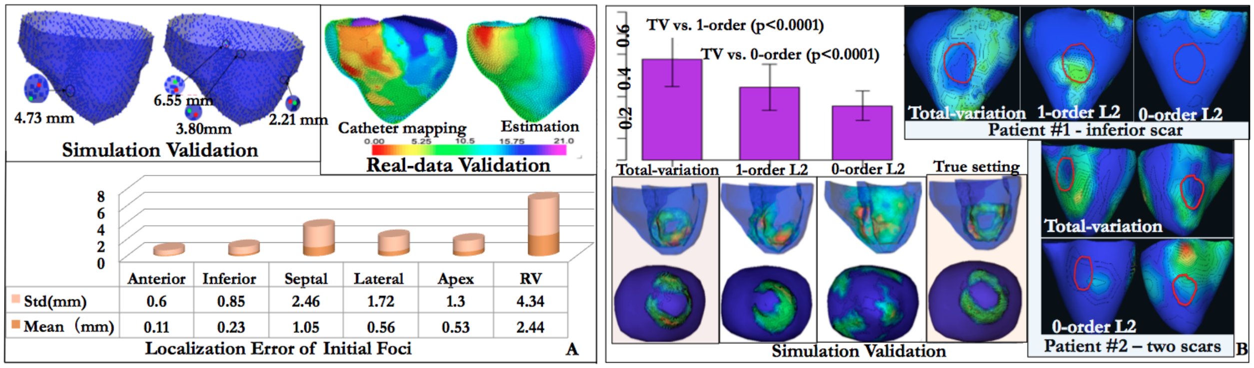

Figure 4: A: Preliminary data of using sparsity model to localize initial foci for electrical excitation. Top left & bottom: localization error from 215 cases of simulation study. Top right: activation isochrone simulated from the estimated foci verified by invasive mapping. B: Preliminary data of using gradient sparsity to promote steep gradients at the border between infarct region and normal heart muscle, in comparison to existing methods of 1-order and 0-order quadratic regularization [41, 42]. Left: simulation study on 137 cases of infarcts located at different regions of the heart. Right: real-data validation on patients with prior infarction; red contour shows infarct delineated from MRI.

given limited measurements in a high-dimensional system, it becomes difficult for the inference to do so without being overly-dominated by the dynamic knowledge. To solve this problem, we propose that the low-dimensional structure of the system (*i.e.*, its sparsity in a basis) needs to be exploited. While dynamic and sparsity constraints are two most successful approaches for overcoming the lack of measurements in inference, their integration has not received much attention until recently. Initial work often directly enforce sparse solutions before [52] or after the standard KF [53]. More recent work began to include sparsity-enforcing norms into the fundamental posterior optimization problem [54], for example, as follows:

$$\hat{\mathbf{x}}_k = \arg\min_{\mathbf{x}_k}[||\mathbf{y}_k - g(\mathbf{x}_k)||_2^2 + \lambda_1||\mathbf{x}_k - f(\mathbf{x}_{k-1})||_1 + \lambda_2||\mathbf{x}_k||_1 \tag{2}$$

which approximates an MAP estimation with Gaussian likelihood (1$^{\text{st}}$ term), and Laplacian distributions to enforce sparsity on the dynamic signal (3$^{\text{rd}}$ term) or occasionally the prediction error (2$^{\text{nd}}$ term), or both. Still at an early stage of development, research along this line faces several challenges. First, the utilization of sparisty and dynamic models needs to be balanced; as shown in (2), when sparsity is enforced, it is difficult to utilize the *a priori* statistics of $\mathbf{x}_k$ predicted by the dynamic model. Second, objective function (2) implies that it is tractable to directly include the dynamic model $f(\mathbf{x})$ in the optimization, with little consideration that—in most applications—stochastic simulations are needed to *implicitly* include $f(\mathbf{x})$.

The proposed investigations will address the challenges identified above. We are targeting at general natural or engineered systems, where the observation matrix does not necessarily satisfy the restricted isometry property (RIP) in compressive sensing [50] (which also further justifies the need to combine complementary dynamic constraints). We are also interested in a broad definition of low-dimensional structures that are inherent in most high-dimensional systems [65]. In some systems, they could be of physical meanings and known from domain knowledge or nonlinear dynamics theory [66]; in others, they may be the target being learned through various methods of dimensionality reduction [65].

Preliminary Work: We have obtained promising data on exploiting the sparsity of intramural action potential pattern when inferring it from noninvasive data [67, 68]. Initially, we directly enforced the spatial sparsity of electrical excitation at the initial stage of a cardiac cycle, using an *l*-0 penalty realized through an iteratively re-weighted algorithm [67, 69]. As illustrated in Fig. 4A, this method showed excellent performance in pinpointing the initial foci deep within the heart muscle. Later, we continued to uncover the sparsity hidden in the localized steep gradient of intramural action potential; a total-variation penalty was used to enforce this sparsity, again realized through an iterative re-weighted algorithm [68, 70]. This initial work has showed substantial improvement in localizing the border of infarcted or ischemic areas (Fig. 4B).

Proposed Work: Continuing on our preliminary study, we will integrate sparsity models into the dynamic context. It is assumed that the basis $\mathbf{\Phi}$ is given so that $\mathbf{\Phi}\mathbf{x}_k = \mathbf{z}_k$, where $\mathbf{z}_k$ consists of mostly zeros, and $\mathbf{\Phi}$ gathers input from domain knowledge on the low-dimensional structure of high-dimensional systems [65].

Integrating dynamic and low-dimensional structures: The first question to explore is how to properly combine the sparsity constraint with statistics of $\mathbf{x}_k$ predicted from the dynamic model. Two alternatives will be explored, both starting from the fundamental structure of the posterior density function.

The first approach is to directly modify the prior $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ to include sparsity models. Three alternative models will be investigated. First, we propose a mixture model using a Gaussian from the sample-based prediction, and a Laplacian enforcing sparsity of $\mathbf{\Phi x}_k$: $\mathcal{N}(\mathbf{x}_k|\mathbf{x}_k^-, \mathbf{P_{xx}}) \cdot \frac{\gamma}{2} e^{-\gamma||\mathbf{\Phi x}_k||_1}$ ($||\bullet||_1 = \sum_i |\bullet_i|$). To illustrate the difference between the proposed model and (2), we can interpret it in an MAP estimation:

$$\hat{\mathbf{x}}_k = \arg\min_{\mathbf{x}_k}[||\mathbf{y}_k - g(\mathbf{x}_k)||^2_{2,\mathbf{R}_k} + ||\mathbf{x}_k - \mathbf{x}_k^-||^2_{2,\mathbf{P_{xx}}} + \gamma||\mathbf{\Phi x}_k||_1] \tag{3}$$

where $p(\mathbf{y}_k|\mathbf{x}_k)$ is assumed Gaussian for the sake of comparison, with error covariance $\mathbf{R}_k$. It corresponds to (2) when the sparsity is only enforced on the dynamic signal, with an important difference that the objective function now also includes mean $\mathbf{x}_k^-$ and covariance $\mathbf{P_{xx}}$ of $\mathbf{x}_k$ predicted by stochastic simulation of domain dynamic models. The second model to be explored is a direct Laplacian distribution of $\mathbf{\Phi x}_k$ (hence $\mathbf{x}_k$) to both enforce sparsity and include statistics predicted by the dynamic model: $\mathcal{L}(\mathbf{x}_k|\mathbf{x}_k^-, \mathbf{b}), 2\mathbf{b}\mathbf{b}^T = \mathbf{P_{xx}}$. Similarly, the MAP estimation with this new prior will correspond to (2) when sparsity is only enforced on the error of predicted dynamics, but statistics predicted by dynamic models are included through the term: $||\mathbf{x}_k - \mathbf{x}_k^-||_{1,\frac{1}{2}\sqrt{\mathbf{P_{xx}}}}$. Naturally, the last option is to combine the above two models, so that $\mathbf{x}_k$ is a mixture of one Laplacian incorporating dynamic statistics and one standard Laplacian. The MAP will correspond to (2) when sparsity is enforced both on the dynamic signal and its error.

The second approach is to introduce a second layer of two independent random variables $\mathbf{s}_k$ and $\mathbf{t}_k$ for the low-dimensional structure and dynamic structure, respectively. A decomposition model can connect them to $\mathbf{x}_k$. Initially, we will consider a simple model $\mathbf{x}_k = \mathbf{s}_k \cdot \mathbf{t}_k + \delta$, where δ is a pre-defined zero-mean Gaussian error representing the residual of this model. The joint prior of all the unknowns become $p(\mathbf{x}_k, \mathbf{s}_k, \mathbf{t}_k|\mathbf{y}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{s}_k, \mathbf{t}_k)p(\mathbf{t}_k|\mathbf{y}_{1:k-1})p(\mathbf{s}_k)$, where the first term on the right hand side comes from the decomposition model, the second predicted by stochastic simulation of domain dynamic models, and the third enforcing sparsity. Other forms of decomposition models will be explored further into the study.

Adaptive integration: The relative importance of the sparsity prior with respect to the dynamic model is controlled by the hyperparameter γ of the Laplacian distribution. Initially, γ will be pre-defined and its optimal value empirically determined from large studies. In a second stage, we will investigate methods to automatically adjust this weight during the inference. Primarily, we will adopt the concept of Laplacian scale mixture (LSM) introduced in [71] to consider a hierarchical model: in short, γ will be modeled as the second layer of unknown random variables, where each element γ_i in the vector (modulating variance) are independent and identically distributed Gamma distributions (conjugate prior for the Laplacian). During the inference, γ will be inferred along with $\mathbf{x}_k$, which is expected to provide higher robustness by avoiding pre-defining the value of γ, at the cost of additional estimations at each data-driven update.

Variational Bayes with sparse priors: Instead of focusing only on an MAP estimation, our primary framework will utilize variational Bayesian (VB) methods for a higher flexibility in approximating posterior density functions given the above priors. VB was developed and mostly applied in *static* inference of complex statistical models that have both latent variable and unknown parameters [72]. Typically, it provides an analytical approximation $q(\Theta)$ of intractable posterior distribution $p(\Theta|\mathbf{y}_k)$ by minimizing their *Kullback-Leibler* divergence (KLD), where $\Theta = \{\theta_i\}_{i=1}^l$ includes the set of unknown variables. It thus fits the purpose of the proposed framework to both accelerate sample-based approach (*e.g.*, PF), and to relax the restrictions of an LMMSE-based approach (*e.g.*, UKF for EnKF). In dealing with sparse prior such as the Laplacian, for efficiency, we will extend the majorization method proposed in [73]. Essentially, the KLD minimization is relaxed to a recursive procedure where we can obtain a sequence of distributions $\hat{q}(\Theta)$ that leads to a monotonically decreasing upper bound to the original KLD, converging to an approximation of the true posterior distribution [73]. It is interesting to note that, similar to [73,74], at each iteration the Laplacian prior can be seen as being approximated by a Gaussian prior weighted by the square-root of its value estimated from last iteration. This approximation method becomes similar to the iterative re-weighted method recently proposed to solve l-1 norm optimization [70], and their relationship will be investigated. Alternative approaches to solve the optimization with sparsity models [71] will also be exploited. The relation between the VB approach and other types of Bayesian optimization methods, (*e.g.*, MAP estimations) will be studied.

Software module developed here will allow users to input their knowledge on the low-dimensional structure of the system (through the basis $\mathbf{\Phi}$). Overall, the proposed methods and software obtain an important advantage that: 1) the data-driven update does not need any explicit information on the dynamic model, and 2) the prior statistics predicted by stochastic simulation of domain dynamic models can be naturally passed into the data-driven update as the flow outlined in Fig. 2. As mentioned in section 3.1, we are focusing on applications with strong nonlinear dynamics and assuming the observation process to be relatively simple, *i.e.*, $p(\mathbf{y}_k|\mathbf{x}_k)$ can be approximated with an analytical form.

3.4 Robustness: Adapting the Dynamic Model Error

Overview: To attain a higher level of robustness, we need to use the measurements to extract information about the unknown error in the domain physical model. Along this lone, the common approach is to introduce additional unknown variables, often in forms of prediction error or the parameter of the dynamic model, and assume them to be Gaussian and time-invariant [14, 48, 49]. The inference is then ended up with a system with substantially augmented unknowns [14]. Essentially, this approach imposes little constraint on the model-error except its initial value (and a smooth distribution around it). In high-dimensional applications where the increased unknowns substantially increase the difficulty to overcome the lack of measurements, more information and constraints on the model-error need to be sought.

To meet these challenges, the theme of the proposed investigations is to leverage and extend the concept in section 3.3, and to exploit the low-dimensional structure in the quantities that account for the dynamic model error. First, for estimating the error in model-predicted dynamics, we argue that it is beneficial to focus only on the significant portion of the error and thus to exploit its sparsity during the inference. While an l-1 sparse error is not uncommon in compressive sensing [50], to our knowledge there has been little to no effort in exploiting sparsity models in adaptive dynamic inference. Second, we will investigate the low-dimensional structure in the physical parameters of the dynamic model, and include it to improve parameter estimation. This will contribute toward the desirable, yet still elusive, goal of parameter estimation: once the parameter is individualized, the dynamic model can be used for predictive simulation without needs of additional individual measurements.

Preliminary Work: We have previously investigated traditional methods for model error estimation during dynamic inference. In [75], we developed an algorithm that estimates model bias during the noninvasive inference of intramural cardiac excitation. In [76], the standard UKF was applied to the augmented unknown vector including both the intramural action potential and a physical parameter that represents tissue *excitability*. Because of the increased degree of freedom and the limited body-surface data, however, the accuracy of the estimated parameters relies significantly on their initial values. To reduce the number of unknowns, we also developed another method for parameter estimation that is currently popular [17, 44]: the heart muscle was pre-divided into a small number of zones, each with a uniform local parameter; derivative-free constrained optimization was then performed with respect to these parameters [21]. While less dependent on initial values, the need for pre-divided zones with uniform parameters is empirical and is likely to violate actual physiological conditions. These previous experiences have motivated the proposed methods to exploit the low-dimensional structural constraints in overcoming the lack of measurements.

Proposed Work

Estimation of sparse model error: Given the dynamic model: $\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{e}_k$, in previous sections and most applications, $\mathbf{e}_k$ is assumed Gaussian with pre-defined covariance $\mathbf{Q}_k$, *i.e.*, the predicted $\mathbf{P}_{\mathbf{xx}}= \mathbf{P}^-_{\mathbf{xx}} + \mathbf{Q}_k$, where $\mathbf{P}^-_{\mathbf{xx}}$ is sample covariance from the noise-free model. Here, we assume that the values in $\mathbf{e}_k$ is generally small but occasionally large, *i.e.*, $\mathbf{e}_k$ is a sparse vector. Two major questions will be answered.

The first question regards the sparsity model for $\mathbf{e}_k$, *i.e.*, the distribution of $p(\mathbf{e}_k)$. We will consider two alternative methods. First, similar to earlier investigations, we can directly model the sparsity of $\mathbf{e}_k$ using a Laplacian distribution. We will consider both a fixed hyperparameter and the LSM with unknown variance. Second, we extend on the sparse Bayesian model proposed in [77] where the sparsity is enforced implicitly through a hierarchical model of Gaussian $p(\mathbf{e}_k|\alpha)$, with *unknown* hyperparameters α (inverse of the variance) in forms of Gamma distributions. Similar in concept to the LSM, α is a vector so that there is an individual α_i associated independently with each $e_{k,i}$. It was shown in [77] that, although $p(\mathbf{e}_k|\alpha)$ appears to be a smooth Gaussian, its sparsity structure is disguised under its hierarchical structure: if we integrate out α, $p(\mathbf{e}_k)$ will assume a product of Student-t distributions that encourage sparsity. Such ability to enforce sparsity hidden behind a regular Gaussian is appealing because it will simply the KLD-minimization.

The second question to be exploited is how to incorporate $\mathbf{e}_k$ into the posterior density functions of interest. A reasonable assumption is that model-prediction $\mathbf{x}_p = f(\mathbf{x}_{k-1})$ and its error $\mathbf{e}_k$ are independent. This allows us to derive the prior joint distribution $p(\mathbf{x}_k, \mathbf{e}_k|\mathbf{y}_{1:k-1})$ as $p(\mathbf{x}_k|\mathbf{e}_k, \mathbf{y}_{1:k-1})p(\mathbf{e}_k)$, where $p(\mathbf{x}_k|\mathbf{e}_k, \mathbf{y}_{1:k-1})$ has mean $\mathbf{x}_k^- + \mathbf{e}_k$ and covariance $\mathbf{P}_{\mathbf{xx}}^-$ predicted by a noise-free dynamic model. Interestingly, when assuming a Laplacian model for $\mathbf{e}_k$ and a Gaussian model for $\mathbf{x}_p$, the proposed approach is similar to the Laplacian-Gaussian mixture recently proposed for speech enhancement [78]. We can also model $\mathbf{x}_p$ with one of the sparsity models proposed in section 3.3, and exploit the low-dimensional structure in both the signal and error space. Solving the KLD-minimization will give us the posterior estimate of both $\mathbf{x}_k$ and $\mathbf{e}_k$. We will also study whether to use the posterior estimation of $\mathbf{e}_k$ to initialize its prior at time $k+1$.

Estimation of physical model parameter: Assuming the parameter θ unknown gives us a modified form of the dynamic model: $\mathbf{x}_k = f(\mathbf{x}_{k-1}, \theta) + \mathbf{e}_k$. Similarly, two questions remain to be investigated.

First, in comparison to prediction error $\mathbf{e}_k$, it is not as straightforward to derive the joint prior $p(\mathbf{x}_k, \theta)$ because the two variables are highly correlated and, given a complex nonlinear model, $p(\mathbf{x}_k|\theta)$ often does not have an analytical form. In order to be able to enforce low-dimensional structural constraints on $p(\theta)$, we propose to build on a suboptimal approach similar to the structure of dual-estimation. Two estimators, one for $\mathbf{x}_k$ and the other θ, will run simultaneously and, in between each interaction, will rely on each other's current estimate as known values. The estimator for $\mathbf{x}_k$ will be as developed in the previous two sections: we will compare whether or not to include also the estimation of the prediction error (as developed above); it is possible that, given limited data, the estimator will decrease in accuracy as more unknowns are being inferred. For the parameter, because the intricate relationship between θ and $\mathbf{y}_k$ prohibits an analytical form of $p(\mathbf{y}_k|\theta)$, a sample-based prediction plus LMMSE will be used.

The key to the proposed research, therefore, is to include low-dimensional structure of θ into $p(\theta)$ so the constraint can be imposed when samples are taken from $p(\theta)$. Our primary interests lie in two constraints. One is the gradient sparsity of θ, motivated by our preliminary study that the physical parameters of a system are usually piecewise smooth with localized gradients in between. For example, in physiological systems, the gradient of these parameters often divide different tissue types, or more importantly, tissues of different health status. Another important constraint for θ is its empirical range of values. For example, if a parameter takes only positive values, an exponential distribution can be assigned. If a local parameter is known to be tightly distributed around a value, a shaper rather than smooth distribution can be assigned. If a parameter will lose physical meaning beyond a range of values, truncated Gaussian or uniform distributions can be used. Imposition of these prior distributions are then ensured by a proper generation of samples from these distributions, which ties back to the first part of the proposed research (section 3.2).

By the time we reach this stage, curricula activities proposed in section 3.2 are also expected to shed some light on which parameters in cardiac electromechanical models have higher priority to be estimated.

Exploratory Objective & Integration with Curricula: Another major approach to improve the robustness of dynamic inference is known as the interacting multiple model (IMM) paradigm where a concurrent set of filters is run simultaneously to combine and prune prior knowledge from multiple dynamic models [79,80]. This has the potential to produce a higher-level robustness to the structural error in the dynamic models. Recent work also saw the integration of both PF and UKF into an IMM framework [81–83]. Because its implied computation cost, however, the study of IMM are generally restricted within tracking applications. The PI plans to include an exploratory research theme that investigates the future integration of the proposed framework into an IMM paradigm, which will be designed into course projects for her graduate course at a later stage of this project. The main goal is to identify the feasibility and critical issues of this research direction, and thus pave the road for the next-stage research beyond this 5-year CAREER project.

3.5 Data-Driven Inference & Modeling Toolkit

We will develop a toolkit for data-driven modeling/inference applications, encapsulating the most successful algorithms developed in the proposed research (sections 3.2 - 3.4). The software will be developed for both GPU and multicore machines, exploiting the data-parallel nature of the stochastic simulation (section 3.2) and large-size matrix operations. It is acknowledged that the proposed algorithms also have components of recursive computation; assisted by collaborator [redacted] we will explore alternative computing environments. The software will be released in two forms: 1) software for foundational data-driven inference algorithms with a *plug-and-play* interface for domain scientists to directly plug in the code of their domain physical models and their knowledge on the low-dimensional structure of the system; and 2) software specifically for individualized cardiac electrophysiology & electromechanics, for professional research and healthcare communities. The software will be released in our project website under an open-source software license, along with the software manual and tutorial. This toolkit can be used by a wide range of computational scientists, and will be used in the proposed educational activities when proper.

3.6 Evaluation

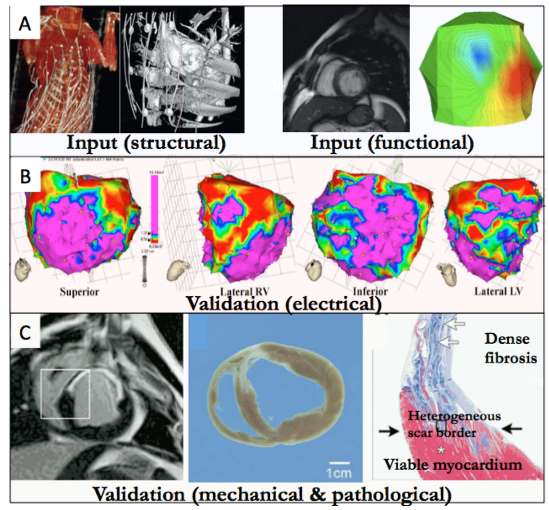

Figure 5: Examples of experimental data for evaluating the proposed research.

The proposed research and software will be evaluated on individualized modeling of cardiac electromechnical activities using noninvasive body-surface recordings and medical images. At the early stage of algorithm development and software instantiation, we will use computer-simulated environment for a controlled, systematic, and large-size evaluation. It will extend on 23 animal and human models currently available in the PI's laboratory [16], each with whole-heart electromechanical activities simulated during various pathological conditions such as myocardial infarction and cardiac dyssynchrony. In a second stage, evaluation will be performed on real experimental data contributed from collaborators [18, 58], with a focus on hearts with prior infarction and at risk for sudden cardiac death. These datasets include approximately: 10 porcine hearts with chronic infarction from [redacted], 30 patients with chronic infarction from [redacted], and 10 patients who underwent ablation of ventricular tachycardia due to prior infarction from [redacted]. The input data include multi-lead body-surface ECG, chest tomographic scans (CT or MRI), and cardiac MRI sequence. Validation data include late gadolinium-enhanced MRI that delineates the true shape of infarction, and/or contact catheter mapping that records heart-surface voltage data. In animal data, postmortem histology is available for infarction morphology. In patient data, clinically decided ablation sites are available for verifying ablation sites predicted by the method. Fig. 5 gives examples of the input and validation data involved. All validation studies involve only retrospective review and analysis of existing patient data, which will be provided to the PI and her team in a de-identified manner. Thus approval of IRB Exempt Category 4 will be sought.

Evaluation will focus on the following metrics. First, the accuracy of the individualized *intramural* electrical excitation and myocardial deformation. Second, the accuracy in infarction detection using the estimated electromechanical activity, and using parameter estimation. Third, the efficiency in terms of computation time. Finally, the robustness to different cardiac conditions (meaning different errors in the dynamic model); in simulation studies, we will also intentionally create errors in the prior dynamic models and test the robustness of the proposed methods to such errors.

4 Integrated Education Plan & Broader Impact Activities

As an integral part of the PI's career development, the overall goal of the proposed educational and outreach activities is to provide multi-disciplinary education and training (with STEM foundations) to students from graduate to high-school levels, with a special focus on involving women and underrepresented students at an early stage in order to increase and retain their future participation in STEM disciplines.

4.1 *"Learning-By-Doing"* in Graduate & Undergraduate Classrooms: Development & Evaluation

In her past a few years as one of the first-generation junior faculty in [redacted] PhD Program [redacted]), the PI has poured energy into integrating her research with curriculum development. She revised and taught one undergraduate course for the Department of Computer Science (CS); she introduced and taught two new courses in the PhD Program and one new graduate research seminar in the CS Department. These courses have received positive feedbacks in past student course evaluations, and attracted graduate and undergraduate students not only from CIS but across the institute, such as Imaging Science, Electrical Engineering, Applied Mathematics, and Physics.

In particular, the PI has experimented and obtained initial success with integrating concepts of *"learning-by-doing"* [84–86] into one of the foundational courses for first-year PhD students ([redacted]). Observing the difficulty for students to learn abstract theories and algorithms without concrete contexts, the PI introduced a new class structure to divide the course content into several themes (*e.g.*, linear regression, optimization, *etc*), each taught and tied with a student team-project over a period of 3-5 weeks. During this period, a set of lectures mixed with in-class practices are used for students to acquire both theoretical foundations and application capabilities toward their completion of the project. The PI found that the process of *doing* stimulated the learning process for students to better grasp otherwise abstract and elusive concepts. Student course evaluation and feedbacks from the past years have reported a substantial and steady increase in learning outcomes and enthusiasms. Motivated by this initial success, one of the primary educational goals of the PI is to extend and evaluate the concept of *"learning-by-doing"* in both graduate and undergraduate [redacted] education. This CAREER project includes this effort in the following aspects.

Graduate Course Development: The first course is a graduate-level course on *computational modeling and simulation of complex systems*, which the PI introduced into the [redacted] PhD curriculum since 2009. This course is closely tied to the research expertise of the PI, focusing on fundamental mathematical and computational methods for mechanistic *vs*. data-driven modeling of complex systems; real-world applications in various domains are used as case studies. It also features a term-long course project for students to apply learned methods on a domain problem. During its offering last year, the PI obtained initial success in integrating the *"learning-by-doing"* component into the instruction of several mathematical concepts that were considered most challenging in the past, such as *variational principle* in system modeling. To continue to increase the portion of practice- and enquiry-based learning, the PI will design the proposed research and cardiovascular applications into domain examples and term projects. Small in-class practices will also be crafted out of the proposed research as described in sections 3.2 and 3.4. Appropriate results will be incorporated into research development. According to past enrollment, 5-10 students are expected for this annual course.

Undergraduate Course Development: The second course will be a new undergraduate-level course on *modeling, simulation, and scientific visualization* for the CS curriculum, developed as an introductory version of the graduate course described above. From her past experience, the PI believed that the *"learning-by-doing"* concept is of particular importance in undergraduate classroom for stimulating student interest and increasing the likelihood for them to pursuit graduate research. The PI will draw heavily on her previous success with the *"learning-by-doing"* structure of *Discovery*, while adapting to the fundamental differences between graduate and undergraduate courses in terms of depth and breadth. For example, hands-on practice will be increased to more actively engage students, while the difficulty and size of the practice will be reduced with closer guidance. Students who take this class will be well positioned to pursue their Honors Capstone projects with the PI, or continue to take the graduate version of this course for advanced training.

Evaluation Methods: An additional objective of the PI's educational effort is to obtain a relatively complete assessment of the *"learning-by-doing"* teaching. In addition to traditional metrics including student course evaluations, quality of student research projects, course enrollment trends, and retention rates, the PI will leverage her current role in the [redacted]), and develop two additional methods for tracking student progresses. First, *pre- and post-semester surveys* [87] will be designed as a *summative* assessment of student learning before and after they take the classes. Second, methods of *systematic progression of assignments* [88] will be used for *formative* assessment to closely monitor student progress and to inform modifications of teaching activities. Two primary learning objectives will be assessed: 1) the depth of student understanding of the knowledge taught, monitored by a group of weekly quizzes; and 2) the ability of students to transfer the knowledge into problem solving, assessed using a specific subset of in-class practices built in the PI's lectures. To ensure the validity of both assessment methods, their design and implementation will follow relevant guidances [88,89] as well as consult with the [redacted]

4.2 From Graduate to High-School Research: Early Engagement of Underrepresented Groups

Student mentoring is another important aspect of the PI's educational activities. She has advised five undergraduate students on senior project and graduated one master student as the co-advisor. She is currently supervising two female PhD students and one master student, co-supervising two senior PhD students, and serving on advising or thesis committees for several graduate students. Her students have publications in reputable venues such as the Medical Image Computing and Computer-Assisted Intervention (MICCAI) [67, 68], International Symposium on Biomedical Imaging (ISBI) [90, 91], and Computing in Cardiology (CinC) [59, 92, 93], or in preparation for premier journals such as the IEEE Transactions on Medical Imaging [94] and IEEE Transactions on Biomedical Engineering [95]. Starting summer 2012, the PI partnered with the [redacted] (PCE) program and hosted their female high-school students for summer research training. In 2013, the PI also started to mentor McNair scholars. Mentoring activities in this CAREER project will build and expand on these past experiences of the PI.

Graduate and Undergraduate Research: This project will support the dissertation research of one PhD student; each year, it will also support the thesis project of at least one master student and the Honors Capstone project of one undergraduate student. In the past, the PI has been successful with attracting and advising female students; these effort will continue. The impact of student research will be measured by the quality and quantity of student publications, completion of the degree, and placement after graduation.

Multi-Level Summer Team Research: In order to engage women and underrepresented students at an early stage, in summer 2013 the PI experimented with a multi-level team-research project in partnership with the McNair Scholars and PCE programs. This team includes: 1) CIS PhD students, 2) a senior undergraduate student from Biomedical Sciences, 3) a junior undergraduate student from the McNair Scholars, and 4) a female high-school student from the PCE. The benefits of such multi-level team projects are multifold. First, the junior undergraduate and high-school students are both trained in basic aspects of research skills, and exposed to the overall picture of the research and its exploratory nature. Second, the senior undergraduate student will gain both research training and leadership experience from this project. Third, the PhD students will gain mentoring experience that can better prepare them for future career placement in academics. Overall, through weekly group meetings and daily interactions, the PI found that (by the time of the writing)—for students at the early stage of their higher-education—the multi-level team-research environment appears to promote their confidence and interests in graduate education and scientific research.

During the course of this five-year project, the PI plans to continue this effort to design and mentor a multi-level team research each summer, involving the PI's PhD students, senior undergraduate or master student who is doing thesis research under the PI's supervision, and junior undergraduate and female high-school student from the McNair Scholars & LSAMP and PCE programs. In individual cases where team-research theme or style does not fit the student, individual research project will be planned. The undergraduate students will be encouraged to submit their research work to the annual [redacted] Undergraduate Research Symposium, as well as other venues for undergraduate research publications. The high-school students will be encourage to participate in the *Siemens Competition in Math, Science & Technology* and the *Intel Science Talent search*, both of which are national programs promoting high-quality high-school research.

4.3 Other Dissemination and Outreach Activities

Outreach to Middle- and High-school Students: As a female faculty and the advisor of two female PhD students, the PI considers it her responsibility to be a role model for promoting women in STEM disciplines. In addition to hosting a female high-school student each year in her lab, the PI will extend her partnership with the PCE to build a broader outreach program to area K-12 students, with a special focus on girls:

- Semiannual seminars: The PI and her students will organize semiannual events to interact with local middle & high school students. Each year, locations of the events will alternate between local schools and [redacted] campus, where the PI and students will demonstrate connections between sciences, technologies, and real-life problems. Health-related examples, interactive visualizations, Q&A sessions, and games will be heavily used to simulate student interests. Two sessions will be offered each time for grades 6-8 (20 students) and 9-12 (20 students), respectively (80 students at grades 6-12 each year).
- Career Shadow: The PI will participate in this program currently offered by PCE, to host students for a day or half day (excused absence) for them to observe and experience the excitement as well

as challenges in a future career of education, research, or engineering. The PI expects to host two students each semester during this CAREER project (4 students at grades 11-12 each year).

- Community Service by Career Cluster: This is another program in PCE where students volunteer for services in the type of careers they are interested in. The PI plans to engage interested students into her dissemination and outreach activities, including those outlined in this proposal, as well as the organization of regional meetings within professional community (*e.g.*, annual meetings of the Upstate New York Cardiac Electrophysiologist Society (UNYCES), the 31st meeting of which was organized by the PI in [redacted]). These activities will help the students understand the connection and contribution of scientific research to the community (6-10 students at grades 8-12 each year).

Outreach to General Public & Paramedic Community: The PI will partner with the Paramedic Education (PE) program at [redacted] to disseminate outcomes of the proposed research to local paramedic groups. Annual seminars and lab visits will be hosted at [redacted] by the PI for her group and other [redacted] faculty to interact with faculty and students in the [redacted] PE program as well as area Paramedic groups, to demonstrate the potential of the proposed research in future healthcare, and to gather professional feedback on their needs in emergency medical services. These outreach activities will increase the societal impact of the proposed research, while providing real-world contexts for its applications. Other dissemination venues to the general public will include a public website dedicated to showcase this project to general audience, and exhibitions at the [redacted] event, an annual festival freely open to the public and well celebrated by the local community with over 30,000 attendees each year.

Academic Dissemination: Research findings will be disseminated through publications and, built on the PI's past experiences, special sessions, tutorials, and workshops organized in international venues, *e.g.*, ISBI and MICCAI meetings. A dedicated public project website will be developed and regularly updated to document the progress of the project. The website will also host the dissemination of educational and outreach materials, as well as software products as described in section 3.5.

Training and professional development will be emphasized for students involved in the project. Travel funds are requested for students to attend and present at national and international conferences. Female students of the PI have been and will continue to be encouraged to attend annual events such as Grace Hopper Celebration of Women in Computing and other events organized by [redacted] Women in Computing.

5. Timeline

Fig. 6 outlines the timeline and important milestones for the proposed research and education plans.

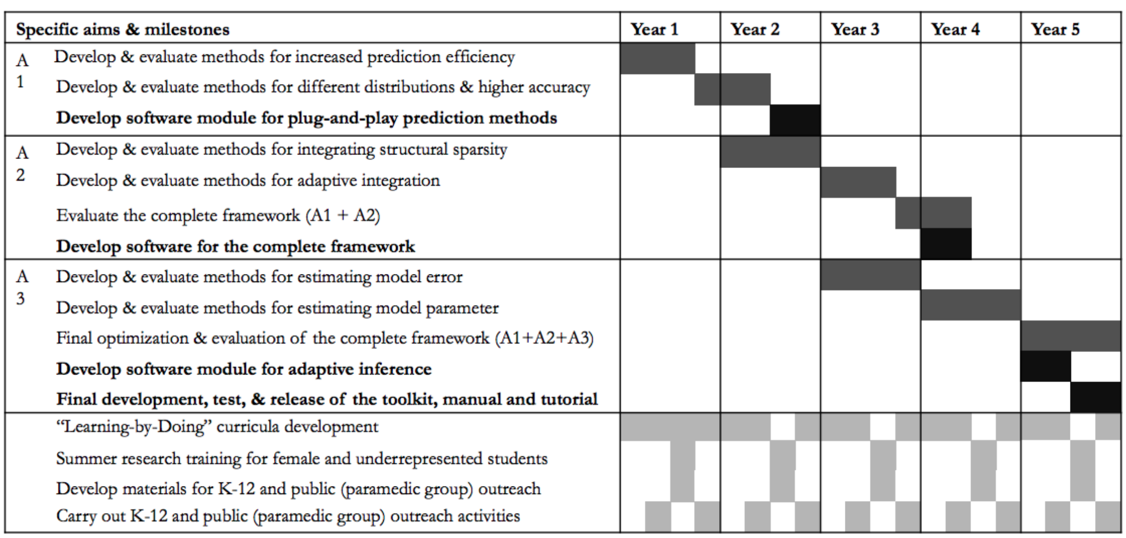

Figure 6: Timeline and important milestones for the proposed research and education activities.

6. Results of Prior NSF Support

The PI has received no prior NSF support.