A DYNAMIC STATE-SPACE MODEL FOR SIMULATING SCANNING PROBE MICROSCOPY

Omolabake A. Adenle & William J. Fitzgerald
email: oaa26, wjf1000@cam.ac.uk
Department of Engineering
University of Cambridge
Trumpington Street
Cambridge CB2 1PZ, U.K.

ABSTRACT
In this paper, we present an algorithm for simulating Scanning Probe Microscope (SPM) images motivated by sequential learning algorithms for dynamic state-space models. The resulting algorithm comprises a series of sequentially dependent optimisation problems allowing for specification of the various parameters associated with SPM - the density of the points along raster scan, probe geometry, and probe inhomogeneity. In conclusion, we present an application to simulating Scanning Electrochemical Microscopy.

KEY WORDS
Modeling, Simulation, Signal Processing, Image Processing

1 Introduction
In Scanning Probe Microscopy (SPM), a probe explores the surface of a sample by sequentially advancing the probe along a raster scan path. The corresponding topographical image is the probe’s trajectory as it follows the surface contour. As the probe is scanned relative to the substrate surface (or vice versa), it interacts with the specimen in a manner unique to each SPM. In Atomic Force Microscopy, the sample-probe interaction is a displacing force in the region between the sample and probe, and in Scanning Tunneling Microscopy (STM) the tunneling current collected over the region of sample-probe separation. The sample-probe interaction may be utilised to manipulate the probe position, depending on the operating mode, generating images that capture different properties of the sample surface. For instance, it is often required that the sample-probe interaction remain fixed at a given value, as in constant current and constant force mode for the STM and AFM, respectively. Consequently, at each point, the probe position is adjusted so as to restore the equilibrium sample-probe interaction, compensating for local changes in surface properties. Information collected in this manner along a path with a sufficiently dense coverage of the specimen, provides an accurate discrete topographical representation of the substrate surface.

While SPMs are very successful at extracting relevant topographical information, the interpretation of the image may be difficult as it exhibits distortions introduced by the probe geometry, which effectively limits the microscope’s resolving power. The extent of this effect depends on the size of the probe relative to that of the local features on the substrate. When the probe dimensions are significantly larger than those of the features, the observed image can be regarded as the output of a Linear Time-Invariant (LTI) system [8], implying convolution of the probe geometry with the local feature geometry. In this case, the so called tip effects can be removed by deconvolving the SPM image, or conversely simulate the SPM image by convolving the substrate topography with an appropriately chosen Point Spread Function (PSF). When the probe dimensions are comparable to the local size of the substrate features, the LTI model, however, is not a suitable model for SPM for several reasons:

- Distortions are non-linear, rendering the LTI model inappropriate. Deterministic methods previously developed for simulating SPM relied on non-linear transformations to describe the simulation and restoration of SPM images([11, 8, 13, 9, 5]).

- Probes are also often known to distort or change shape during scanning (as with AFM tip breakage), or the environment in which the specimen rests during imaging may alter, producing artefacts in the image. At best the process can be modelled as locally LTI and the entire image simulated/restored using a spatially variant PSF.

Simulating SPM is useful in that it creates a framework for analyzing the microscope’s behaviour under various configurations, and also for removing the tip effects. Previous methods have explored the latter more than the former. The former is important, however, as it provides a priori information on the effect the microscope’s configuration has on imaging specific topographies. For this reason, it is useful to develop a mathematical framework for modeling the generative process, which would provide a means of analyzing tip distortions and gaining insight regarding configurations that minimize these distortions and towards inverting algorithms for restoring SPM images.
Section 2 provides a brief review of the main methods for simulating SPM, their deficiencies and the need for a more robust approach. A review of dynamic linear models and sequential learning in these models can be found in Section 3. As our simulation approach is dependent on a sample-probe occlusion model requiring multidimensional integrals for which quadrature type methods are inefficient, a review of Monte Carlo methods for evaluating these integrals is presented in Section 4.3.1. Combining these concepts, a dynamic state-space model for SPM is presented in Section 4. Results on application of the algorithm and future extensions are presented in Sections 5 and 6, respectively.

2 A Brief Review of SPM Simulation

Previous work on simulating SPM images focused on finding non-linear transforms, without recourse to the nature of the probe response, which model the distortive effect of the probe interaction under no-noise conditions. These methods are founded on the basic SPM scanning protocol requiring that a probe of known geometry maintain minimal contact with the sample surface. This has yielded two dominant frameworks for SPM simulation, which we now discuss briefly.

2.1 Simulation by Dilation

SPM image simulation and restoration as outlined by Villarubia [11] are described using concepts from mathematical morphology, where objects are described using sets and are manipulated using concepts from set theory. Mathematical morphology is the basis of many image-processing applications, including image restoration, noise reduction and edge detection. When the probe scans the substrate surface such that minimal contact is maintained, the observed image can be shown to be the result of the dilation of the true topography by the probe geometry [11]. Dilation of a reference set (the image), \( A \), by the structuring element, a set \( B \), is given by,

\[
A \oplus B = a + b : a \in A, b \in B
\]

\[
= \cup_{b \in B} (A + b)
\]  

(1) 

(2)

In applications on Euclidean space, this transform is equivalent to a local upward or downward scaling of the object, where the local size of the image is determined by the structuring element.

2.2 Simulation by Legendre Transform

Another approach to modeling the probe’s distortive effects arises upon considering that when the point at which the probe makes contact with the sample surface differs from the probe tip, as with wells in the topography whose radius of curvature is smaller than that of the probe, the resulting information recorded by the probe provides an inaccurate representation of that point on the sample topography. Knowledge of the apparent contact point can be used to determine tangent lines to the surface and probe tip at the contact point. Their respective intercepts correspond to the Legendre transform of the image and probe, and are related to the Legendre transform (intercept of the tangent line) of the simulated image, by the relation [8, 5, 9],

\[
\mathcal{L}[s(x)] = \mathcal{L}[s(x')] + \mathcal{L}[t(\Delta x)],
\]  

(3)

where \( s(x') \) is the substrate surface which is assumed locally convex, \( i(x) \) its image, \( t(x) \) the probe geometry, \( \Delta x \) the distance between the probe tip, \( x \), and the apparent point of contact, \( x' \), and \( \mathcal{L}[\cdot] \), the Legendre transform.

2.3 Motivation for a More Robust Approach

The previously mentioned methods are fairly generic and are restricted to simulating and restoring images when the probe maintains minimal contact with the substrate surface. This is sufficient when formulating a framework for image restoration under this specific scenario, but is limited in its application to image simulation. A more robust simulation method offers the possibility of

- Incorporating more information about the probe, aside from its geometry,
- Simulating and restoring images using information specific to a particular type of SPM, e.g. the operating mode
- Incorporating a noise model for both image simulation and restoration.
- Determining the “optimal” configuration of a given SPM (probe geometry and feedback mechanism) for imaging unique substrate topographies
- Perform a theoretical assessment of an SPMs resolving power

In this paper, we show that by treating the SPM as a stochastic system with a dynamic state-space model, it is possible to derive a recursive algorithm for simulating SPM images, which admits incorporation of SPM specific knowledge including the probe response (tunneling current, ion conductance).

3 Dynamic State-Space Models

Consider a set of observations, \( y_t \), which are assumed dependent on an unobserved Markov process, \( x_t \), such that

\[
x_t = f(x_{t-1}) + \eta_t
\]  

(4)

\[
y_t = g(x_t) + \nu_t
\]  

(5)
where \( f(\cdot) \) determines the transition of the unobserved state, \( \eta_t \) is the process noise, and \( \nu_t \) is observation noise. The function \( g(\cdot) \) relates the state process to the observations. Given a set of observations, \( y_{0:T} = \{y_0, y_2, \ldots, y_{T-1}, y_T\} \), the aim is to estimate the unobserved process, \( x_{0:T} \), and perhaps parameters associated with the process and observation noise.

Depending on the chosen constraints, like Gaussian noise for instance, there are a variety of approaches to solving this problem. Of particular interest are recursive algorithms for estimating the states by filtering, such that \( \hat{x}_t = \mathbb{E}[x_t|x_{0:t-1}, Y_{0:t}] \), where \( \mathbb{E}(\cdot) \) denotes statistical expectation.

### 3.1 Recursive Learning in Dynamic Linear Models

A review of the statistical signal processing literature yields a wealth of algorithms for recursive estimation of dynamic state-space models, the most popular being the Kalman Filter\[10\]. Basic Kalman filtering requires that both \( g(\cdot) \) and \( f(\cdot) \) are linear, and both the process and observation noise are Gaussian. When the system is nonlinear, one may employ the Extended Kalman Filter (EKF), which provides a suboptimal linear filter to an approximate linearization of the system obtained by Taylor Series expansion.

Kalman-type filtering algorithms take a two-stage approach to sequentially estimating the hidden states. First, a prediction, \( x_{t|t-1} \), is made based on the previous time state estimate \( x_{t-1} \). Subsequently, the estimate is corrected to obtain the time \( t \) state estimate, \( \hat{x}_t = x_{t|t} \) by incorporating knowledge obtained from the current observation, \( y_t \). While the EKF is suboptimal for nonlinear systems, a more accurate approach to hidden state estimation for nonlinear systems involves first applying an unscented transformation \[7, 12\] to a set of appropriately chosen points determined from the previous time state estimate. Instead of projecting the predicted hidden state using a linearized version of the system, the Unscented Kalman Filter (UKF) considers a set of sigma-points, \((X_i)_{i=1}^{2L}\), which are deterministic perturbations of the previous time estimate, \( \hat{x}_{t-1} \in \mathbb{R}^L \).

\[
X_i = \hat{x}_{t-1} \pm (\lambda P_{x_{t-1}}^{1/2}), i = 1, \ldots, 2L
\]  

where \( \lambda \) is a chosen scale factor, and \( P_{x_{t-1}} \) is the estimated covariance of the state at the previous time point, and \( P_{x_{t-1}}^{1/2} \) is a suitable matrix square root, such as one obtained via Cholesky decomposition. Use of the sigma points allows the algorithm to better capture the underlying changes in the statistical properties of the state with respect to the system. The current state estimate is obtained from a weighted expectation of the sigma-points such that \( \hat{x}_t = \sum_{i=0}^{2L} W_i X_i \), where weights \( W_i \), are determined in some optimal sense \[7\]. We adapt the UKF framework for estimating the latent process (the probe position), in deriving a recursive in time (position) algorithm for simulating SPM images.

### 4 A Dynamic State-Space Model for SPM

Previous methods for simulating SPM relied on deterministic non-linear transforms to describe the distortive effect of the probe on the substrate surface (Section 2). We introduce a robust method for simulating SPM images motivated by sequential learning algorithms for dynamic state-spaces models, particularly the Unscented Kalman Filter, which admits robust analysis of SPM without recourse to a specific probe response or operating mode. We begin by formulating an observation model.

#### 4.1 The Observation Model

As with the points along the path of a raster scan, consider a discrete grid such that information about the sample topography is collected at points, \((x, y) \in \mathcal{G} = \mathcal{G}_x \otimes \mathcal{G}_y\), where \( \mathcal{G}_x = \{0, \Delta_x, \ldots, N_x \Delta_x\} \) and \( \mathcal{G}_y = \{0, \Delta_y, \ldots, N_y \Delta_y\} \) with \( \Delta_x \) and \( \Delta_y \) denoting the grid size. By noting that the sample-probe interaction, \( y_t \), is dependent on the height of the topography sensed by the probe at each \( t = (x, y) \), we can define an observation model such that

\[
y_t = g(p_t) + \eta_t,
\]

where \( y_t \) is the probe response as determined by the probe wavefunction, where the possibly non-linear function \( g(\cdot) \) relates the probe position to the observed response \( p_t \) the probe position determined with respect to a reference plane (the “floor” upon which the substrate is placed), and \( \eta_t \) the observation noise, which captures inconsistencies in the recording device measuring the probe response. In addition, if we consider the probe position to evolve slowly such that \(|p_t - p_{t-1}| < \epsilon\) for a small number \( \epsilon > 0 \), the probe position can be treated as hidden states such that

\[
p_t = p_{t-1} + \nu_t,
\]

where \( \nu_t \) encodes the process noise. It is clear that (4-5) defines a dynamic state-space model for SPM image simulation. In Section 4.2, we will show that under certain, general assumptions the probe response, \( g(\cdot) \) is a multi-dimensional integral, which may have no closed-form solution but can be approximated by Monte Carlo integration. One note of importance is that the generative model explicitly accounts for noise, unlike previous methods. Also the observation model is robust in its application to any form of SPM, and does not assume a specific sample-probe separating distance.

Adopting state-space formulation for SPM allows us to treat simulating an image, i.e. the probe trajectory \( p_{0:T} \), dependent on the probe response as an inference problem. A recursive algorithm can be defined where the model parameters, operating mode, probe dimensions, etc., and the observed sample-probe interactions, \( y_{0:T} \), guide inference.

Implementation of an observation model necessitates description of changes the probe wavefunction experiences upon contact with the sample topography. In the following
section, we provide a simple occlusion model, describing the effect of the topography on the probe wavefunction.

4.2 A Sample-Probe Occlusion Model

We begin by assuming the sample topography and probe response interact in a manner that introduces no deformation to the probe geometry and consequently its wavefunction. We also assume that the interaction is such that the response is diminished by that part of the wavefunction that interacts with the substrate topography. The observed probe response, $y_t$, can then be said to be related to the latent probe position, $p_t$, such that

$$\hat{g}(p_t) = P_{tot} - \int_{V(p_t)} P(x, y, z) \rho(x, y, z) dx dy dz, \quad (9)$$

where $P(\cdot)$ is the probe wavefunction, $P_{tot}$ the probe response\footnote{We differentiate the wavefunction i.e., the equation describing the spatial variation in the probe properties, from its response, the combined result upon interaction with a body.} when no occlusions are present, $\rho(\cdot)$ the substrate wavefunction, and $V(p_t)$ probe-position dependent volume over which the integral is performed (see Figure 1). When either wavefunction is unavailable, it may be replaced with the indicator function, $I(b)_{V(p_t)}$, for which $I(b)_{V(p_t)} = 1$ if $b \in V(p_t)$, and 0, otherwise.

By incorporating a functional representation for the spatial variation of the substrate properties, the simulation can be used to observe images of inhomogeneous substrates, as in many Scanning Electrochemical Microscopy (SECM) [3] applications, for example. Where inappropriate, this sample-probe occlusion model may be replaced with a suitable alternative amenable to analytic evaluation.

4.3 Evaluating the Probe Response

While quadrature-type methods offer a viable approach to computing the integral (9) in two dimensions, the required computation time becomes prohibitive in three dimensions. Consequently, we compute the required integrals using Monte Carlo integration. Monte Carlo methods also offer modularity across various forms of SPM and permit evaluation of the probe response when the probe wavefunction does not admit analytic integration. A brief review of Monte Carlo methods follows.

4.3.1 Monte Carlo Simulation

Monte Carlo simulation provides a means for analysing stochastic systems, which are intractable to analytic evaluation. Observations from a stochastic system are governed by often non-standard probability distributions, which could, for example, describe the posterior density over the system configuration. Random variables simulated from this distribution can be used to perform the required analyses.

For example, a set of $N$ samples $(x_1, \ldots, x_N)$, $x_i \sim p(x_i)$ can be used to estimate various moments of the distribution such as the mean, $\mu \approx \frac{1}{N} \sum_{i=1}^N x_i$, or the normalization constant, $Z = \int_{-\infty}^{\infty} p(x) dx \approx \sum_{i=1}^N p(x_i)$, of $p(\cdot)$ for a sufficiently large sample size, $N$, with exact estimates being obtained almost surely as $\lim N \rightarrow \infty$. Non-standard distributions can be simulated with any one of a number of algorithms such as Rejection Sampling, Importance Sampling, and Markov Chain Monte Carlo. We consider Importance Sampling for evaluating integrals associated with the sample-probe occlusion.

4.3.2 Importance Sampling

With Importance Sampling candidate samples are simulated from an appropriately chosen importance distribution, $\pi(x)$, which closely approximates the distribution of interest, $p(x)$. Of particular importance is that the proposal distribution exhibit positive density $\pi(x) > 0$ in the portions of the support of the true distribution where $p(x) > 0$, and that the tails of $\pi(x)$ decay at a rate slower than that of $p(x)$ [6], so as to ensure full exploration of the state space. Each sample is assigned an importance weight, $\hat{w}(x) = \frac{p(x)}{\pi(x)}$, which encodes how well the variate mimics one obtained from the true distribution, $p(x)$. An empirical approximation to $p(x)$ is given by $[\hat{w}(x^{(1)}), x^{(1)}]$, with $x^{(i)} \sim \pi(x)$,

$$\hat{\pi}(x) = \sum_{i=1}^N \hat{w}(x^{(i)}), \quad (10)$$
where \( \tilde{w}(x^{(i)}) \) are the normalized importance weights, 
\[
\tilde{w}(x^{(i)}) = \frac{w(x^{(i)})}{\sum_{i=1}^{N} w(x^{(i)})}.
\]
Integrals of a bounded and \( \pi \)-integrable function, such that \( f(x) < +\infty \), of the form \( \int f(x)\pi(x)dx \) can be approximated by 
\[
E_{\pi(x)}(f(x)) \approx \sum_{i=1}^{N} \tilde{w}(x^{(i)})f(x^{(i)}).
\]
Success of this method is highly dependent on the structure of the proposal distribution and how closely it mimics the true distribution [6].

Importance sampling can be used to obtain approximations to the multidimensional integrals associated with determining the sample-probe occlusion. If samples distributed according to the probe wavefunction can be obtained, (9) can be approximated as
\[
y_t = P_{tot} - \int_{V(p_t)} P(x,y,z)\rho(x,y,z)dxdydz
\approx \hat{P}_{tot} - \frac{1}{N} \sum_{i=1}^{N} w(t^{(i)}, z^{(i)}) \rho(t^{(i)}, z^{(i)}),
\] 
(11)

where \( (t^{(i)}, z^{(i)}) \sim \mathcal{U}(A_t, B_t) \), and importance weights 
\[
w(t^{(i)}, z^{(i)}) = P(x^{(i)}, y^{(i)}, z^{(i)}).
\]
A three dimensional uniform distribution with upper and lower bounds \( (A_t, B_t) \in (\mathbb{R}^3 \times \mathbb{R}^3) \) about the probe tip, serves as the proposal distribution (see Figure 1).

Figure 1 shows a 2D example of an inhomogeneous hemispherical probe experiencing significant contact with a single-feature substrate. Samples are color coded according to their importance weights, which correspond to the probe’s wavefunction, which in this case is the charge density around the surface of an SECM electrode.

### 4.4 Sequential SPM Simulation Algorithm

The observation model (8-7) is non-linear due to the sample-probe occlusion model. Depending on the form of SPM under consideration, the noise distribution may also be non-Gaussian. Consequently, inference should be performed using any of the methods available for non-linear and non-Gaussian models. Practical consideration of the upwards or downwards adjustments to the probe position bear physical resemblance to the UKF’s sigma points, suggesting the UKF as a candidate method.

Accordingly, we use two-stage an algorithm motivated by the UKF (Section 3). First, a set of \( k \) sigma points, \( \mathcal{P}_{t,i} = \hat{\rho}_{t-1} \pm \lambda \sigma_t \) are generated from the estimated probe position at the previous point in the raster scan. The unscented transform, i.e., the sample-probe occlusion model, is applied to determine the predicted observations, \( X^{k}_{t,i} = g(P_{t,i}) \). Unlike the UKF, however, we determine \( X^{*}_{t,i} = \arg \min_{\lambda_i} \left( \mathcal{Q}_{t,i} - y_t \right)^2 \), which negates the effect of the Kalman update as the error term goes to zero [1]. Therefore, the current estimate for the probe position is chosen as \( p_{t|t} = \mathcal{P}_{t,i} \).

Essentially, instead of taking the expectation of the probe position, we propagate points that best predict the observations. Also, we do not use a fixed number of proposals at each \( t \), but sequentially simulate and test the response of each proposal until a minimum error criterion is attained. A practical analogue of this procedure is the practice of tip approach, where the probe tip is moved close to the surface topography until the sample-probe response registers an equilibrium position. The resulting variation is similar to an adaptive optimization and has two benefits - (1) it ensures smoothness of the simulated topography by preserving local distances with respect to the observed sample-probe position and (2) minimizes computational expense by using the fewest possible sigma points at each point in the simulation.

More details on the theory and implementation of the simulation algorithm can be found in [1], and a deterministic formulation of the algorithm with application to Scanning Ion Conductance Microscopy in [2].


5 Application: Scanning Electrochemical Microscopy

We present a novel application of the described method for simulating SPM images to Scanning Electrochemical Microscopy (SECM). Similar to the STM, the SECM simulates images by controlling the current flowing between the probe tip and the substrate surface. The SECM, however, uses an electrolysis current generated by oxidation-reduction processes taking place between an insulated ultramicroelectrode and the sample surface. The solution to Poisson’s equation for an inlaid disc microelectrode [4], provides a model for the probe’s wavefunction - the spatial variation of charge density near the surface of the ultramicroelectrode.

\[ P(x, y, z) = \frac{2c_0}{\pi} \times \arccos\left(\frac{2R}{(z^2 + (R + r)^2)^(1/2) + (z^2 + (R - r)1^{1/2})}\right), \]

where \( c_0 \) is the initial concentration at the microelectrode surface, \( R \) the electrode radius, \( z \) the perpendicular distance from the electrode surface, and \( r \) the magnitude displacement from the electrode center.

An SECM image of a 500nm \times 500nm surface with two Gaussian features, each with a full-width at half maximum of 50nm and height of 50nm, separated by a distance of 100nm was simulated in Constant Current mode. The probe was regulated at 99.9% occlusion, so that it meets the constant current criterion. As the probe descends from the peak it must drop drastically in order to drop on the right side of the simulated image. As the probe proceeds noise was Gaussian with variance 0.01, and we used no process noise. Results are shown in Figure 2. Note the sharp noise was Gaussian with variance 0.01, and we used no process noise. Results are shown in Figure 2. Note the sharp drop on the right side of the simulated image. As the probe descends from the peak it must drop drastically in order to meet the constant current criterion.

6 Conclusion & Future Work

We presented a dynamic state-space model for simulating SPM and described a recursive learning algorithm for simulating images using an adaption of the Unscented Kalman Filter. Utility of the algorithm was demonstrated in an application to Scanning Electrochemical Microscopy (SECM). The method described for simulating SPM images can also be adapted to removing tip-effects from SPM images. Consider a set of observations \( y_{0:T} \) denoting the sample probe positions along a raster scan path, \( \mathcal{G} \). By encoding our \textit{a priori} knowledge of the SPM in the function \( f(\cdot) \), and assuming a Markov transition model, \( s_t = s_{t-1} + \nu_t \), describing the hidden process, the true image \( s_{0:T} \), one can define a similar state-space model and restoration algorithm.

Another application involves analysing various simulated images, either for different feature sizes or under different microscope configurations, in order to analyze an SPMs resolving power. This work has already been done for an SECM with a circular ultramicroelectrode, and we are pursuing applications to other SPMs.

References