Final Report

SERDP Project MM-1591

Optimal Sensor Management for Next-Generation EMI Systems

Lawrence Carin, Nilanjan Dasgupta, and Hui Li
Signal Innovations Group, Inc.
Suite 200, 1009 Slater Road
Durham, North Carolina 27703

Phone: 919-660-5270
Fax: 919-660-5293

Contact Email: lcarin@siginnovations.com

June 13, 2008

Distribution Statement A: Approved for Public Release, Distribution is Unlimited

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

This document serves as the final report on the project titled "Optimal Sensor Management for Next-Generation EMI Systems" (SERDP Project MM-1591). This project is a collaboration between SIG, Dr. T. C. Bell of AETC, and Dr. Herb Nelson of NRL.

This research is directed toward developing the adaptive sensor-management architecture needed for next-generation electromagnetic induction (EMI) systems. Specifically, SERDP and ESTCP are currently funding multi-coil EMI systems that provide significant capability and diversity with respect to the shape of the incident magnetic field, as well as in how the induced magnetic fields are measured (*e.g.*, multi-field-component measurements). Moreover, systems can operate in the frequency and/or time domain, with prescribed data sampling rates.

The large number of sensor parameters (number of transmit/receive coils, as well as the time/frequency sample rate) often necessitate hardware design tradeoffs, with the goal of achieving practical sensing costs (*e.g.*, sensing time). By making these sensor-design tradeoffs in hardware, one necessarily loses functionality, limiting the utility of the system (*e.g.*, the system may have to be tailored in hardware to particular classes of UXO, and UXO depths). We are therefore developing here an adaptive EMI-sensing framework, for next-generation EMI systems; this framework adaptively tailors the *use* of sensor assets to the target under test. Sensor functionality is preserved by making fewer compromises in hardware, with practical sensing costs achieved through optimal and selective use of sensor assets. The algorithm also adaptively determines when to terminate sensing, defined when the data measured thus far are sufficient for classification within user-defined risk constraints.

This research is highly relevant for the full exploitation of current SERDP/ESTCP investments in next-generation EMI systems, and therefore there are many transition opportunities. We have had particularly close interactions with the Lawrence Livermore National Laboratory (LBNL) team, in the context of the Berkeley UXO discriminator (BUD) system, for which the algorithms developed here are particularly relevant.

The remainder of the document is organized as follows. We provide an overview of the problem and the description of sensor system in Section I, followed by descriptions of the two algorithmic approaches in Sections II and III. The performance of the proposed approaches is analyzed in Section IV, followed by conclusions in Section V. The algorithms are explicitly applied to data of the type measured by a state-of-the-art active electromagnetic prototype developed by Lawrence Berkeley National Laboratory (LBNL), thereby improving the efficiency with which such data may be collected in practical UXO-sensing missions.

I. Introduction

Electromagnetic induction (EMI) has been widely used for detection and characterization of buried conducting and/or ferrous objects. In order to accurately identify a buried UXO from other non-UXO metallic fragments (clutter), it is necessary to accurately estimate the parameters that characterize the buried objects. A search mechanism is required that estimates the parameters, such as the size, shape, orientation, shell thickness and metal content (ferrous or non-ferrous) of the buried object without explicit excavation. The search for UXO consists of two steps. In the first step, a buried object (UXO or UXO-like metal fragments) needs to be detected and its location needs to be identified. We have developed an active learning-based greedy search algorithm that involves sensing using EMI systems. The second phase involves the estimation of physical parameters, such as shape, size, thickness etc, in terms of induced magnetic moments and polarizabilities of the buried object.

A typical EMI-based sensor configuration consists of both transmitter and receiver coils, placed close to the ground and in the vicinity of the target. If the operating frequency corresponds to wavelengths that are typically much larger than target length, it allows one to develop simple models for the target response to an EMI sensor. Detection of secondary magnetic fields, produced by currents induced in a metallic object by time-varying magnetic fields from a source current coil, is a popular choice for detecting buried metallic objects such as unexploded ordnance (UXO). Detection of the secondary magnetic fields is complicated by the fact that they might be a few orders of magnitude weaker than the primary magnetic field. One way of reducing that problem is to use a time-domain system, which allows the transmitter to operate for a finite period of time and activate the receiver only after the effect of the primary inducing field has diminished. Another way is to design the location and orientation of the receiver coil such that it is null-coupled to the primary inducing field. The algorithms developed at SIG have been designed in the context of the state-of-the-art Berkeley UXO discrimminator (BUD) system developed at Lawrence Berkeley National Laboratory (LBNL).

(a) Berkeley UXO Discriminator

In order to fully characterize the inductive response of an isolated conducting object, it is generally desirable to measure its response to primary magnetic fields in three orthogonal directions. As mentioned above, the receiver coil needs to be null coupled to the primary magnetic field, in order to measure the weak secondary magnetic field from the buried objects. For a single or a pair of orthogonal transmitters, the receiver coil needs to be at right angles to the primary magnetic fields from both transmitters. It has been shown by Huang *et al.* [1] that when transmitter systems are constructed symmetrically with respect to a central point, and receiver pairs are similarly constructed, the differences between receiver pairs are insensitive to the primary magnetic fields, and thus null coupled in a difference mode, for as many transmitter loops as needed.

The magnetic field produced at point ***r*** due to a current element ***Idl*** located at point ***q*** is given by Biot-Savart's law as

$$dB(r) = \frac{\mu_0 Idl \times (r-q)}{4\pi |r-q|^3} \tag{1}$$

For a transmitter pair, placed symmetrically with respect to the origin (see Figure 1 below, where two current elements are placed diagonally opposite on the current loop), the magnetic field induced at location ***r*** is given by

$$dB(r) + dB'(r) = \frac{\mu_0 Idl \times (r-q)}{4\pi |r-q|^3} - \frac{\mu_0 Idl \times (r+q)}{4\pi |r+q|^3} \tag{2}$$

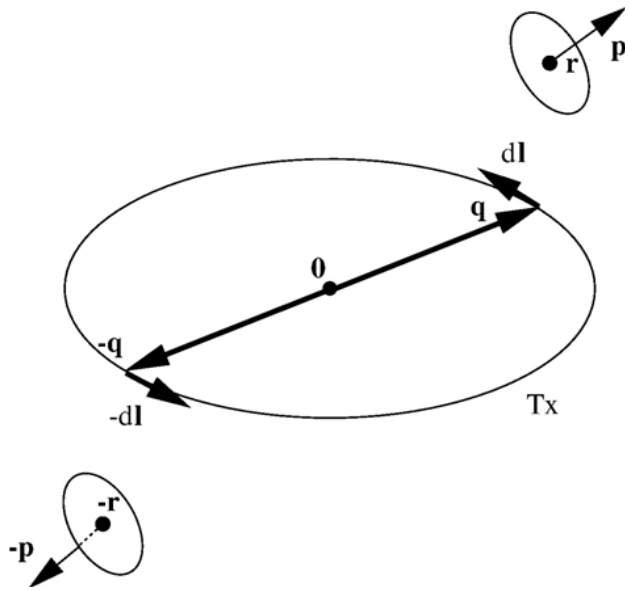


Fig. 1: Geometry of symmetric receiver loop pair, one receiver loop centered at ***r*** with axis along *p*, and the other centered at −***r*** with axis along −*p*.

Note that the combined magnetic field shown above is symmetric with respect to the change of sign of ***r***. This suggests that identical magnetic fields are induced at two points ***r*** and –***r***, which are mirror images with respect to the transmitter coil. This insight has been successfully implemented in the design of the Berkeley UXO Discrimminator (BUD) developed at LBNL, where three independent transmitter coil-pairs are arranged to have magnetic fields that are linearly independent. We have developed our analysis for the BUD system, although the methodology is general.

The BUD sensor system is a prototype EMI system developed for detecting and characterizing UXOs [2]. The sensor system consists of two pairs of orthogonal vertical loop transmitters (Tx and Ty in Fig. 2(c)) and a pair of horizontal loop transmitters (Tz) spaced apart vertically by 26" with a 39" x 39" footprint. The vertical coils are separated by 6" and are 45.5" x 23.5". The vertical coils are mounted on the diagonals between the horizontal loop coils (see Fig. 2).

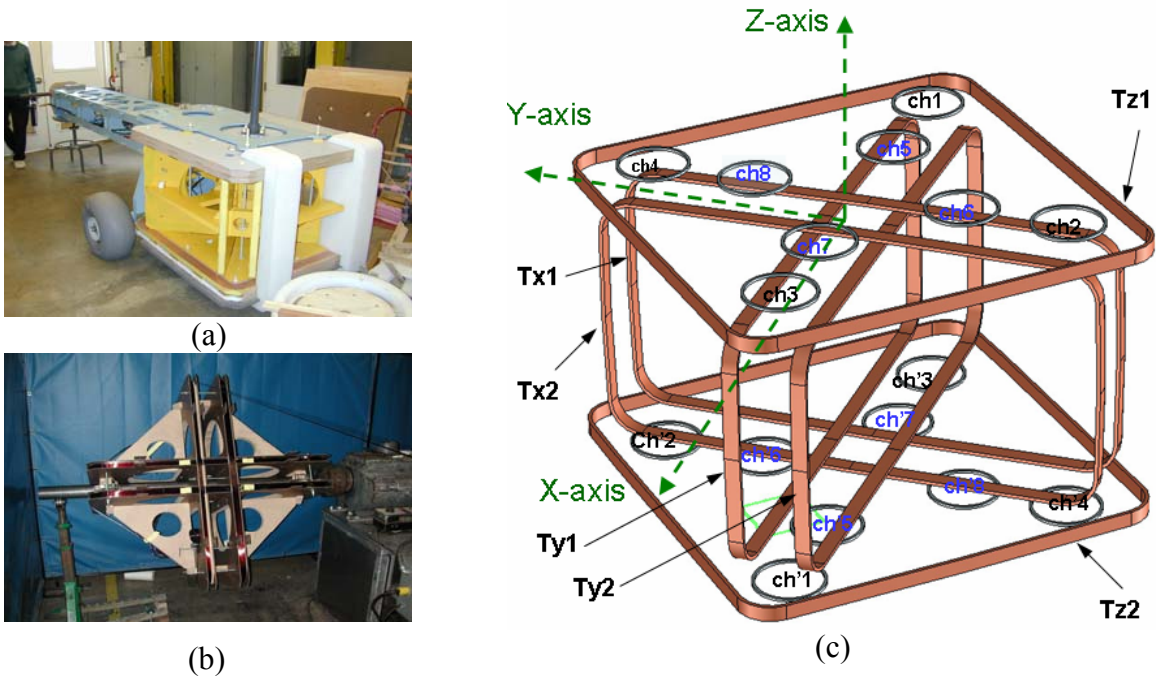


Figure 2. (a) Cart assembly with X, Y and Z transmitter coils, (b) Transmitter coils being wound, (c) Physical structure of the BUD sensor system.

Eight vertical field receivers (ch1 to ch8) are deployed in the upper and lower plane of the two horizontal loops (Tz1 and Tz2) and are arranged in pairs to measure offset vertical gradients of the fields. By design, the offset vertical field measured by the receivers are null-coupled to all three transmitted magnetic fields. The location and orientation of the three principal polarizabilities of a target can be recovered from a single position of the transmitter-receiver system. The system employs a bipolar half sine pulse train current waveform and the receivers are dB/dt induction coils designed to minimize the transient response of the primary field pulse. The whole sensor system is mounted on a cart, as shown in Fig. 2(a).

(b) Electromagnetic Induction Model

SIG has developed a forward model that simulates the time-domain response received by each of eight receiver pairs, for each of three transmitter coil excitations.

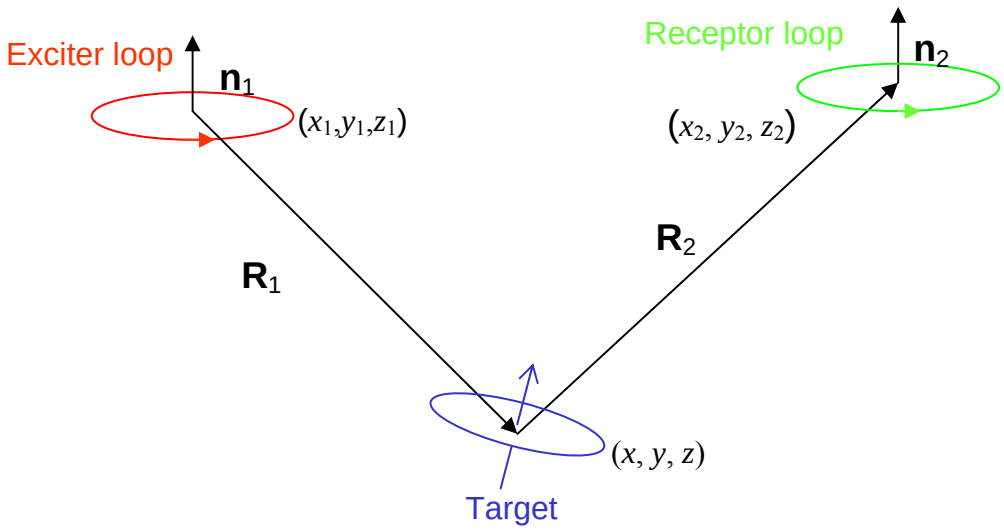


Figure 3. Schematic of an induction sensor interrogating a subsurface UXO

Figure 3 shows the schematic of an active EMI-based system for detection of buried metallic objects. The transmitter is represented here as a horizontal loop at location (x_1,y_1,z_1) (note that BUD sensor array consists of three transmitter loops along three mutually orthogonal directions). The target is located at (x,y,z), at a distance $\mathbf{R}_1$ from the center of the exciter loop. The receiver is represented by another horizontal loop at a distance $\mathbf{R}_2$ from the target center. The time-varying magnetic field produced by the exciter induces current on the target, which in turn develops a secondary magnetic field detected by the receptor coil. For an unit-strength target dipole (a target is modeled as a single dipole, assuming the excitation wavelength is much larger than the physical dimensions of the target), the magnetic field induced by the presence of an UXO is modeled [3] as

$$\mathbf{B} = (\frac{\mu_0}{4\pi})^2 \frac{IA}{R_1^3 R_2^3} \{9(\mathbf{n}_1 \cdot \mathbf{r}_1)(\mathbf{r}_1^T \mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{r}_2)\mathbf{r}_2 - 3(\mathbf{n}_1^T \mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{r}_2)\mathbf{r}_2 - 3(\mathbf{n}_1 \cdot \mathbf{r}_1)\mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{r}_1 + \mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{n}_1\}$$

where $\mathbf{M} = \begin{bmatrix} M_x \omega_x e^{-\omega_x t} & 0 & 0 \\ 0 & M_x \omega_x e^{-\omega_x t} & 0 \\ 0 & 0 & M_z \omega_z e^{-\omega_z t} \end{bmatrix}$ and

$$\mathbf{U}_1 = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \mathbf{U}_2 = \begin{bmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{U} = \mathbf{U}_1\mathbf{U}_2$$

The variables M_x and M_z represent magnetic moments along x and z direction, where z represents the direction along the central axis of an UXO. A buried metallic object, modeled as a single dipole, is fully characterized by the following parameters [4]:

1) Relative location of the dipole (*x,y,z*) with respect to the sensor system
2) Strength or Dipole Moment (M_x and M_z). Note that UXOs are assumed to bisectionally symmetric target. Hence M_x and M_y are assumed to be equal.
3) Orientation of the target (azimuth ϕ, and inclination θ)
4) Resonant frequencies (ω_x, ω_z)

The BUD sensor system may be treated as a combination of three linearly independent exciter loop pairs, along with eight receiver loop pairs. The time-domain response collected by a BUD sensor system excited by the presence of a UXO may be represented as

$$\mathbf{B}_{total} = \sum_i (\frac{\mu_0}{4\pi})^2 \frac{IA}{R_1^3 R_2^3} \left\{ 9(\mathbf{n}_1^i \cdot \mathbf{r}_1^i)(\mathbf{r}_1^{i^T} \mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{r}_2^i)\mathbf{r}_2^i - 3(\mathbf{n}_1^{i^T} \mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{r}_2^i)\mathbf{r}_2^i - 3(\mathbf{n}_1^i \cdot \mathbf{r}_1^i)\mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{r}_1^i + \mathbf{U}^T \mathbf{M}\mathbf{U}\mathbf{n}_1^i \right\}$$

We have used the above equation to simulate the time-domain data received by the receiver loops for any location and orientation of a buried object (the coil dimensions are modeled rigorously – no dipole assumption). In the next section, we discuss the optimal sequential search strategy that will be employed to detect the approximate target (dipole) location (*x,y,z*) and the associated dipole parameters (M_x,M_z,ω_x,ω_z).

II. Adaptive EMI Sensing of Buried Objects

The main objective of the project is to develop a systematic approach for detection and identification (or classification) of buried UXOs over a wide area, by optimally exploiting the capabilities of next-generation EMI systems. This approach is developed to mitigate an important limitation of powerful new systems such as BUD: while the sensor has significant capability, it is far more complex (*e.g.*, time consuming) to deploy, due to the complexity and size of the system. The new adaptive algorithms developed here allow one to retain the sophistication of such systems, but the system is deployed only on a specific set of locations such that the corresponding measurements are essential for an efficient detection and classification of buried objects. This is an important new paradigm: previously one made tradeoffs in sensor design when constructing the hardware; while this leads to more-efficient (faster) deployment, one also sacrifices potential sensor performance. In the approach developed here, one retains the full sensor sophistication in the hardware (*e.g.*, in the BUD system), and practical deployment is manifested algorithmically, with the algorithms developed here defining which measurements are essential for detection and classification (the data-collection choices are made adaptively, in the field, not at the hardware-development stage).

The objective of the first phase of our algorithm is to efficiently estimate the parameters of the target model. As discussed above, a target is modeled as a single dipole (it can be extended to a multi-dipole model) with an unknown location, orientation, dipole strength, and resonant frequencies. Although we assume that the structure of dipole model is known, the inverse model, developed at SIG to estimate the target parameters from the time-domain data, is sensitive to sensor noise and has many local minima. The idea is to make robust and reliable estimation of these parameters using as few sensing measurements as possible (reducing deployment costs for the new sensor, while retaining overall capability). The proposed approach develops a fundamental information-theoretic framework to adaptively and sequentially identify sensing locations in order to minimize the uncertainty on the target-parameter estimation (note that this is distinct from laboriously and potentially redundantly collecting data on a fixed grid).

The objective of this phase of our research is to efficiently estimate the model parameters θ. We have developed an active-learning based search strategy to achieve this goal with a minimum number of sensing actions. Let p_n be the sensor parameters (location and orientation for both transmitter and receiver coils), and O_n is the time-domain data associated with the nth measurement (sensing action). Assuming no prior knowledge, sensing starts at any random location (p_1 is chosen randomly within the search area) and takes a set of measurements (O_1). Based on (p_1, O_1), one may estimate the dipole model parameters $\hat{\Theta}_1$ utilizing the inverse model. The goal is to choose the sensor parameters for the next measurement, denoted as p_2, to improve the estimate of Θ. In general, after N measurements are performed, from which $\hat{\Theta}_N$ is determined, the objective is to choose p_{n+1}, in order to maximally improve the estimation of the target properties, $\hat{\Theta}_{N+1}$.

The search strategy assumes that the Nth measurement is represented as $O_N(p_N, \Theta) = B_{total}(\Theta, p_n) + G_N$, where $B_{total}(\Theta, p_n)$ is the noise-free target response and G_N is the additive white Gaussian noise. The search strategy is based on choosing measurement parameter p_{N+1} that minimize the Cramer-Rao bound (CRB) [5] computed as the optimal variance of the unbiased estimate of Θ_{N+1}. Assuming white Gaussian noise, the maximum-likelihood estimation of Θ, based on measurements $\{p_n, O_n\}_{n=1:N}$, reduces to a least-square (LS) fit [5]

$$\hat{\Theta} = \arg\min_{\Theta} g(\Theta) = \arg\min_{\Theta} \sum_{n=1}^{N} \left| B_{total}(\Theta, p_n) - O_n(p_n) \right|^2 \qquad (1)$$

where Θ = [x, y, z, M_x, M_z, ϕ, θ ,ω_x ,ω_z] represent the target parameters and p_N represents the sensor parameters (location and orientation). An important issue concerning the above computation involves the existence of multiple local minima.

Assuming β represents the inverse of the noise variance, the likelihood of the measured observations O can be written as

$$p(O;p,\Theta) = \frac{\beta}{2\pi}\exp\left(-\frac{1}{2}\beta\left|O - B_{total}(\Theta, p_n)\right|^2\right)$$

Let the Fisher information matrix [6] be denoted as J, and its $(i,k)^{\text{th}}$ element J_{ik} can be evaluated as

$$J_{ik} = \int p(e,\Theta)\frac{\partial \ln p(e,\Theta)}{\partial \Theta_i}\frac{\partial \ln p(e,\Theta)}{\partial \Theta_k} de_R de_I \text{ , where } O - B_{total}(\Theta, p_n) = e_R + je_I$$

Hence our objective is to determine optimal sensor parameters p_{N+1} for $N+1^{th}$ measurement, where the quality of p_{N+1} is based on the Fisher information matrix [6], evaluated as

$$J = \beta\sum_{n=1}^{N}\mathrm{Re}\left\{\left[\nabla_\Theta B_{total}(\Theta, p_n)\right]\left[\nabla_\Theta B_{total}(\Theta, p_n)\right]^H\right\} \tag{2}$$

Where ∇_Θ represents the gradient evaluated with respect to the target parameters Θ and superscript H represents the complex transpose. The above equation is evaluated at $\Theta = \hat{\Theta}_N$, assuming the model parameter estimate is correct after N measurements. The objective in selecting sensor parameters p_{N+1} is to reduce the uncertainty in the estimated target parameters, characterized through the Cramer-Rao bound $C = J^{-1}$. We define the Fisher information measure q of a measurement sequence $\{p_n, O_n\}_{n=1:N}$ as

$$q(\{p_1,....,p_N\}) = \left|J(p_1,....,p_N)\right| = \left|\sum_{n=1}^{N} J^n(p_n)\right| ,$$

where $J^n = \beta\,\mathrm{Re}\left\{\left[\nabla_\Theta B_{total}(\Theta, p_n)\right]\left[\nabla_\Theta B_{total}(\Theta, p_n)\right]^H\right\}$. Note that J^n is a function of all n prior measurements, via the estimated target parameters $\hat{\Theta}_n$.

Considering a new sensor parameter p_{N+1} , one can show that

$$\begin{aligned} q(\{p_1,....,p_N, p_{N+1}\}) &= \left|\sum_{n=1}^{N} J^n(p_n) + \beta F(p_{N+1})F^T(p_{N+1})\right| \\ &= \left|\sum_{n=1}^{N} J^n(p_n)\right|\left|I + \beta F^T\left(\sum_{n=1}^{N} J^n(p_n)\right)^{-1} F\right| \end{aligned}$$

where I is a 2x2 identity matrix and F is a Kx2 matrix (assuming K target parameters), $F = [F_R, F_I] = \left[\mathrm{Re}\{\nabla_\Theta B_{total}(\Theta, p_n)\}, \mathrm{Im}\{\nabla_\Theta B_{total}(\Theta, p_n)\}\right]$. The logarithmic increase of the Fisher information measure is

$$\delta(p) = \ln q(\{p_1, ..., p_N, p_{N+1}\}) - \ln q(\{p_1, ..., p_N\}) = \ln\left|I + \beta F^T B_N^{-1} F\right| \tag{3}$$

where $B_N = \sum_{n=1}^{N} J^n$ is the Fisher information matrix computed using the first N sensor parameters $\{p_n\}_{n=1:N}$, based on the latest estimate of the model parameters $\hat{\Theta}_N$. Therefore, the sensor parameters p_{N+1} for the $(N+1)^{\text{th}}$ measurement are selected at the point where the model "error bars" $\beta F^T B_N^{-1} F$ are largest. Since our objective is to achieve the maximum information gain, we define the optimal sampling point p_{N+1} as

$$p_{N+1} = \arg\max_p \delta(p) = \arg\max_p \ln\left|I + \beta F^T B_N^{-1} F\right| \tag{4}$$

The search for the next sensing location p_{N+1} is performed in a two-dimensional space, corresponding to sensor position (x_s, y_s). The target parameter estimate is updated ($\hat{\Theta}_{N+1}$) based on $\{p_n, O_n\}_{n=1:N+1}$. It is important to note that B_N is only invertible after performing a sufficient number of measurements. Such a limitation is handled by adding a "diagonal loading" to the matrix B_N (*i.e.*, replace B_N by $B_N+\lambda I$) for first few measurements. This sequential process of choosing sensor locations for measurements is terminated when the Fisher information gain is below a threshold, yielding a stable estimate of the approximate target location, orientation and model parameters. Once this is achieved, we enter the second phase of our work, where the objective is to *identify* the buried object.

III. Optimal Sensing using POMDP for UXO Classification

We have developed a partially observable Markov decision process (POMDP) based autonomous decision making system for UXO identification, assuming the approximate location of the buried object is known (*e.g.*, using the technique in the previous section, or based on other information that may be available, for example from a magnetometer). The algorithm has been successfully developed and tested on data simulated to replicate the BUD sensor.

Given an area where different types of UXO and clutter are buried, our objective is to identify the buried objects through sensing, without costly excavation. We use the same BUD sensor system as before, but unlike the previous approach, we incorporate the cost of sensing in our sequential decision making process. In other words, the adaptive search strategy discussed in the previous section used all three transmitters and eight receivers to obtain measurements, which in turn were utilized for target parameter inversion (executing all of these measurements may be unnecessary/redundant and wasteful, and now we seek to address this issue). The strategy in the previous section also ignored the cost of moving the sensor array from one location to the next, which is incorporated in the policy design in the POMDP-based approach discussed next.

We develop a policy that evaluates the next best action to take at any time, based on the data measured thus far. The policy is optimal in the sense that it tries to maximize the *long-term* ("non-myopic") discounted reward through its sequential choice of actions. For example, the policy decides whether it should declare the "ID" of the buried object, and when it needs to sense more (using a risk analysis). We assume that correct and incorrect declarations also have corresponding reward/penalty – this, along with sensing costs, are provided by the policy maker prior to training the POMDP-based optimal sensing and declaration policy.

We have developed a partially observable Markov decision process (POMDP) [7] based policy design that that answers the following questions.

1. *How to optimally choose sensing positions, so as to use as few sensing locations as possible to identify the buried object correctly (minimizing the number of times the sensor must be moved.*
2. *How to optimally choose a sensor from the sensor array at each sensing location.*
3. *When to stop sensing and make a declaration with regard to the target ID (UXO vs. clutter).*

Policy design proceeds in two phases. The first phase involves *model training,* consisting of (a) designing the target/clutter model, which involves estimation of the model parameters based on previously collected training data; (b) training of the optimal policy given the specified model. There exist many policy-learning algorithms for given POMDP models. We have employed the state-of-the-art point-based value iteration (PBVI) algorithm [8] here.

After model learning and policy design (done off-line), the second phase involves *model testing*. The testing phase is executed in real time to make decisions to sense or declare the ID of the buried object, as dictated by the trained policy. We briefly discuss the salient features of a POMDP model, followed by the specifics of the model as designed for UXO classification.

a) Partially Observed Markov Decision Processes (POMDP)

A POMDP is a model of an agent interacting synchronously with its environment. The agent starts with an initial estimate (belief) of the underlying unobservable states. A belief vector is a probability distribution over the states of the model. The agent takes an action dictated by the policy. This produces an observation and a reward from the environment. The agent updates its belief based on the observation, and takes the next action based on the updated belief. The agent keeps the entire history of the past {action, observation, reward} sequence compressed in the form of an updated belief vector over the unobservable states (the belief vector is a "sufficient statistic"). This process continues until the agent takes one of the terminal actions (*e.g.*, the declaration action in the current problem).

A POMDP model is defined by the tuple $\{S, A, T, R, \Omega, O\}$ [7], where S is a finite set of discrete states of the environment, A is a finite set of discrete actions, and Ω is a finite set of discrete observations providing noisy state information. In the current problem, the states represent the area on the ground divided into square grids (more details are presented below in the results section). Note that states S in a POMDP are hidden. In the

current problem, the hidden part is the class of the buried object (UXO or clutter), not the physical location of the buried object. The actions in our problem consist of three different sensing actions (corresponding to three mutually orthogonal transmitters), along with five moving actions (moving the sensor system east, west, north, and south by a fixed distance for the next measurement; or possibly no-movement, corresponding to taking another measurement at the same point). In addition, we incorporate three declaration actions: UXO, clutter and clean region. These declaration actions serve as terminal actions for a POMDP agent, by which it makes its final declaration about the class of the buried object before moving to a new location.

The state transition probability is represented by matrix T where

$$T: S \times A \rightarrow \Pi(S) \text{, where } \quad T(s,a,s') = \Pr(S_{t+1} = s' \mid S_t = s, A_t = a)$$

represents the probability of transitioning from state s to s' upon taking action a. The observation function is defined as

$$O: S \times A \rightarrow \Pi(\Omega), \text{ where } \quad O(a,s',o) = \Pr(O_{t+1} = o \mid A_t = a, S_{t+1} = s')$$

represents the probability of receiving observation o after taking action a, and transiting to state s'. The reward structure is represented as R: $S \times A \rightarrow \Re$, where $R(s, a)$ is the expected reward (cost) received by taking action a in state s.

Since the state is not observed directly, a belief state b is introduced. The belief state is a probability distribution over all states, representing the agent's probability of being in each of the states based on past actions and observations. The belief state is updated by Bayes' rule after each action and observation, based on the previous belief state.

$$b_t(s') = \frac{O(a_t, s', o_t) \sum_{s \in S} T(s, a_t, s') b_{t-1}(s)}{\Pr(o_t \mid a_t, b_{t-1})}, \quad \forall s' \in S$$

$$= \frac{O(a_t, s', o_t) \sum_{s \in S} T(s, a_t, s') b_{t-1}(s)}{\sum_{s' \in S} O(a_t, s', o_t) \sum_{s \in S} T(s, a_t, s') b_{t-1}(s)}$$

A POMDP policy is a mapping from belief states to actions, telling the agent which action to take based on the current belief state. The goal of the POMDP is to find an optimal policy by maximizing the expected discounted reward $V = E[\sum_t \gamma^t R(s_t, a_t)]$, which is accrued over a finite or infinite horizon. The discount factor $\gamma \in (0,1]$ describes the degree to which future rewards are discounted relative to immediate rewards. When the agent (sensor system in this case) in belief state *b*, the maximum expected discounted reward is given by

$$V^*(b) = \max_{a \in A}\left[R(b,a) + \gamma \sum_{b' \in B} T(b,a,b')V^*(b') \right]$$

where *R(b,a)* is the immediate reward and $\gamma \sum_{b' \in B} T(b,a,b')V^*(b')$ is the discounted future reward over an infinite horizon. For a finite-horizon case, *V**(*b*) has been shown to be

$$V^*(b) = \max_i b^T \alpha_i^* = \max_i \sum_s b(s)\alpha_i^*(s)$$

and is piecewise linear and convex in belief state, and represented by a set of |*S*|-dimensional a vector $\{\alpha_1^*, ..., \alpha_M^*\}$. The objective is to estimate these α vectors. There exist many algorithms to achieve this, of which we prefer the PBVI algorithm [8] that has been shown to outperform others for models with large state space.

b) Point Based Value Iteration (PBVI)

The objective of the *point-based value iteration* (PBVI) algorithm [8] is to solve a POMDP for a finite set of representative belief points $B=\{b_1,\ldots,b_N\}$, rather than for the entire belief space. These belief points are chosen carefully using stochastic trajectories, and by maintaining one hyperplane (α vector) per belief point it solves problems with large state space. Assume that we have a carefully chosen *N* belief points. The algorithm starts with an initial estimate of the α vectors, one for each belief point. Without any prior knowledge, each such α vector is initialized to $\alpha_0 = \frac{R_{\max}}{1-\gamma}$. Each such alpha vector is updated iteratively using value backup (explained in the following paragraph). The

complete PBVI algorithm is designed as an *anytime* algorithm, interleaving steps of value iteration and steps of belief set expansion. It starts with an initial set of belief points for which it applies a first series of backup operations. It then grows the set of belief points, and finds a new solution for the expanded set. We briefly describe how PBVI performs value backups and choose the representative belief points.

i) Point-based value backup

The exact value backup for POMDP is given by

$$V(b) = \max_{a \in A} \left[\sum_{s \in S} R(s,a) b(s) + \gamma \sum_{o \in O} \max_{\alpha \in V} \sum_{s} \sum_{s'} T(s,a,s') \Omega(o,s',a) \alpha(s') b(s) \right]$$

In the PBVI algorithm, the exact value backup is modified such that only one α-vector per belief point is maintained. For a point-based update, PBVI creates projections as

$$\begin{aligned} \Gamma^{a,*} &\leftarrow \alpha^{a,*}(s) &= R(s,a) \\ \Gamma^{a,o} &\leftarrow \alpha_i^{a,o}(s) &= \gamma \sum_{s \in S} T(s,a,s') \Omega(o,s',a) \alpha_i^{'}(s') \end{aligned}$$

Now the best action for each belief point b is evaluated as

$$\begin{aligned} V &\leftarrow \arg\max_{\Gamma_b^a, \forall a \in A} (\Gamma_b^a \cdot b), \quad \forall b \in B \quad \textit{where} \\ \Gamma_b^a &= \Gamma^{a,*} + \sum_{o \in O} \arg\max_{\alpha \in \Gamma^{a,o}} (\alpha \cdot b) \end{aligned}$$

ii) Belief point set expansion

PBVI focuses its planning on relevant beliefs. It has been proven that PBVI performs best when its belief set is uniformly dense in the set of reachable beliefs. Consequently, we start with a small, randomly initialized belief set B, and greedily expand it to capture reachable belief points. For a given belief point $b \in B$, PBVI stochastically simulates a

single-step forward trajectory using each action to produce a set of new beliefs $\{b_{a_1},.....,b_{a_N}\}$, one for each possible action $a \in A$. Finally, it keeps only the belief b_{a_i} that is furthest from the starting belief *b*. So at every step of belief-set expansion, the number of belief points essentially doubles and the performance of the algorithm converges after a few expansion steps. Since expansion phases are interleaved with value iteration, PBVI is an *anytime* solution.

c) POMDP Model Design for UXO Identification

It is assumed in this phase of the project that the approximate location of a buried object is known *a priori*, although their identification (UXO or clutter) is unknown to the agent (sensor system in this case). The objective is to classify the buried object through a careful choice of sensing actions around the buried object so as to facilitate classification without costly excavation. The agent can be in one of three possible (hidden) situations, where the underlying buried object is an UXO, clutter, or the subsurface is relatively clean. Each such situation is designated as a "world" in the POMDP, while all worlds constitute the "universe". The agent can move between states of a world, but no transition between the worlds is permitted. This is intuitive in the sense that the sensor system can move from one location to another around the buried object, while the nature of the buried object does not change. Since the class of the buried object identifies the world, no state transition is permitted between worlds. Out of three possible worlds, one of them is the true world, but the agent does not possess this information (it is "hidden").

We model a given region via nine state, where each state represents a .5m x .5m square area. A "clean" world is modeled as a single state. Assuming the object location is approximately known, the agent starts in state s5 of world 1 (UXO) or 2 (Clutter), or s1 of world 3 (Clean). The idea is to identify which world the agent is in. The sequential sensing process terminates when the agent makes a declaration about the class of the buried object.

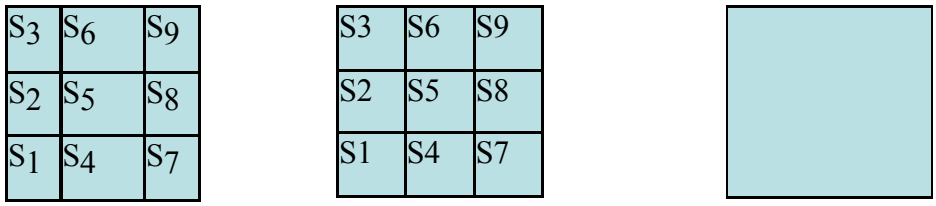


Figure 4. (a) World -1:UXO (b) World-2:Clutter (c) World-3:Clean

We have modeled five moving actions {stay, go-east, go-west, go-north, go-south} coupled with three sensing actions {sense with transmitter coil Tx, Ty or Tz}, or declaration action {declare-UXO, declare-clutter, or declare-clean}. Also, note that we have modeled the transitions between states as determinimistic, assuming that after the algorithm identifies a moving actions as the next best action to take, there would be outside help to move the sensor system to the new location. Hence we can concentrate on the policy training for the underlying worlds, without being involved in the stochastic motion modeling. According to the model described above, the agent's motion would be constrained within the boundaries of the world. This is implemented within the state transition model in the following way: If the agent is in state S_9 of world 1, moving action towards north or right would not move the agent, whereas moving action towards west or south would lead the agent to state S_6 or S_8 respectively, with probability one. Although a POMDP is capable of handling stochastic transitions, we have used deterministic transitions (with probability 0 or 1) in our problem.

The POMDP model is based on discrete observations. For each sensing action, the forward model (which possesses the knowledge of the target location and other parameters) simulates a set of eight time-domain signals corresponding to eight receiver coil pairs. Given the set of continuous time-domain signals, we perform a model inversion to estimate the underlying target parameters. Note that any such model inversion technique leads to multiple local optimas. Hence one needs to estimate the distribution of the inverted parameter vectors prior to POMDP training. In order to achieve the above goal, we first generate a large set of simulated UXO, Clutter, and clean area responses, from each of the 19 probable states of the universe. Given a set of received time-domain signals, we ran model-inversion algorithm with many random

seeds to generate a large set inverted target parameters. We discretized this large sample set using vector quantization (VQ) to develop a codebook. The codebook serves as the set of possible discrete observations for the POMDP. We choose a codebook size of 15 for discretizing the parameter space for the entire universe. The POMDP model requires a discrete observation probability distribution. We approximate this distribution by the relative frequency of observing each codebook element within each of 19 states in the universe.

IV. Performance Analysis

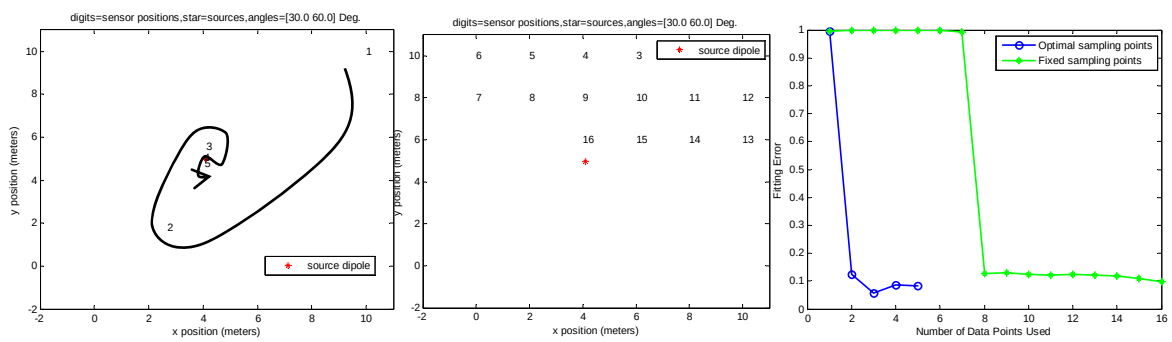
We analyzed the performance of the proposed algorithms based on simulated data modeled on the active electromagnetic BUD sensor system developed at LBNL. The forward model that emulates the BUD sensor system is capable to illuminating the target with one of three orthogonal transmitter coils and receive secondary magnetic field in all eight horizontal receivers. The parameters of the targets used in these simulations were based on inverting *measured* data from the BUD system.

a) Phase 1: Adaptive EMI Sensing

The first phase of our research was detection of buried objects, where our principal aim is to identify the approximate location and parameters (shape, size, dipole moments etc). We have designed an active-learning based information-theoretic technique that efficiently chooses a sequence of sensing actions to minimize the uncertainty on the unknown model parameters. The algorithm starts with a randomly picked starting location, from where its makes its first sensing action. In this phase, a sensing action involves all three transmitters and eight receivers. Based on the set of eight time-domain signals, the algorithm estimates the approximate target location and the uncertainty on each of these estimates. The algorithm then evaluates the Fisher information matrix based on the current estimate of the model parameters and evaluates the next best location to sense that would maximally reduce the uncertainty on the model parameters. This greedy

search strategy is continued until the reduction in uncertainty is lower than a predefined threshold, leading to a stable estimate of the target location.

Figure 5(a) represent the sequence of sensing locations as evaluated by the adaptive strategy. As shown in the figure, the target is located near the middle of the search area (marked by a red dot), while the sensor system starts sensing from an arbitrarily chosen location (marked as "1"). Based on the observation it receives, the target parameter Θ is estimated (using an inverse model developed at SIG). The Fisher information matrix [6] is evaluated as shown in Eq. 2, assuming the current parameter estimate $\hat{\Theta}_1$ is correct. The next best location to sense is evaluated (marked as "2") that maximizes the gain in Fisher information (as defined in Eq. 3).



(a) (b) (c)

Figure 5: (a) Variation of the BUD sensor location p_n as a function of measure n; p_n is determined adaptively; (b) Variation of the BUD sensor location p_n as a function of measure n, using a fixed grid of sensor position; (c) Fitting error for the estimation of the UXO parameters Θ for the two search strategies.

The next set of measurements are taken at location "2" and the target parameters $\hat{\Theta}_2$ is updated. It is important to note that the parameter estimate $\hat{\Theta}_2$ is based on both the observations collected from locations "1" and "2". This process is continued for five sensing actions, marked 1 to 5 on Fig. 5(a). Observe that the agent makes sensing actions around the target, while gradually moving towards the correct location of the target. Within five sensing actions, the sensor system is on top of the buried object and the

corresponding parameter estimate is close to the ground truth (comparisons are tabulated in table 1). In order to recognize the efficacy of the active learning technique, we compare its performance with a uniform sampling approach (shown in Fig. 5(b)), where sensing is performed on an uniform grid for sixteen measurements. Figure 5(c) compares the two approaches where the vertical axis represents the fitting error on the target parameters and horizontal axis represents the number of points used. One can easily recognize that the fitting error reduces drastically using only five sensing measurements using the adaptive strategy, whereas the uniform sampling approach takes eight sensing operations to achieve a similar performance.

The comparison between the model parameters estimated by the adaptive strategy and the fixed grid strategy is shown in Table 1. As noted above, $\{x,y,z\}$ correspond to the physical location of the buried target, $\{\theta,\phi\}$ correspond to the orientation, and $\{M_x, \omega_x, M_z, \omega_z\}$ correspond to the dipole moments. This clearly shows the benefit of the adaptive strategy in estimating the approximate target parameter with only a small number of sensing operations.

True parameters, parameters fitted adaptively, parameters fitted with fixed-grid									
	θ	ϕ	x	y	z	M_x	ω_x	M_z	ω_z
True	0.52	1.05	4.10	4.98	-0.37	20	90000	30	11000
fitted adaptively	0.55	0.97	4.08	4.99	-0.36	15.68	11036	62	4752
fitted with fixed-grid	0.38	0.91	4.10	4.93	-0.44	184.73	1543	113	4987

Table 1: Target parameters using two search

(b) Target Classification using POMDP-based sensor scheduling

Once the approximate location and model parameters of the buried object are obtained, we employ the second phase of the strategy, where a POMDP-based policy dictates how a buried objected needs to be illuminated by different transmitter coils in order to identify the "class" of the buried object. This model assumes the sensing cost with three individual transmitter are known *a priori*, along with the cost of declaring the "ID" of the

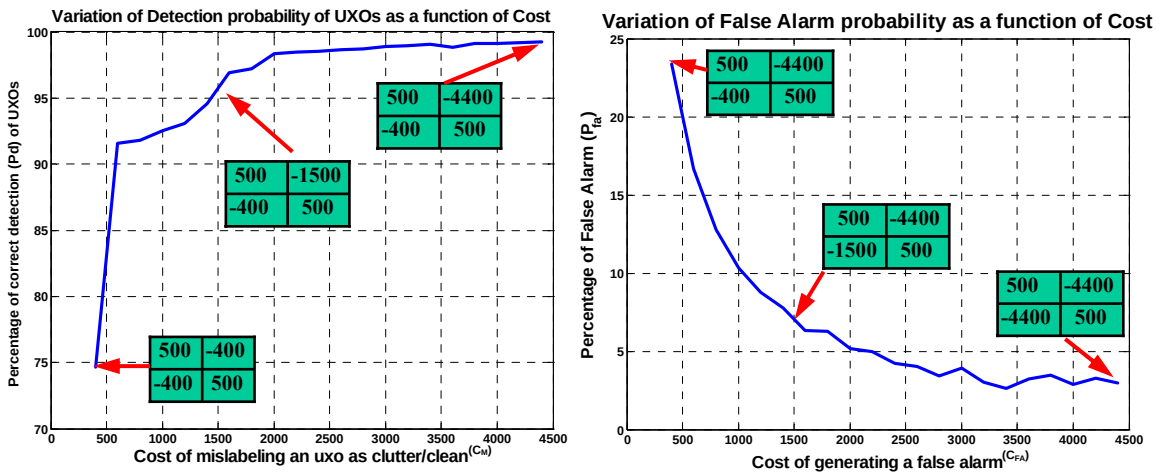
buried objects, both correctly and incorrectly. Note that a policy maker needs to carefully design these costs since they can significantly alter the policy. Suppose the cost structure for declaration is represented by a 2x2 matrix as

Reward for correct detection of an UXO	Reward(-cost) for missing an UXO
Reward (-cost) for generating a false alarm	Reward for correct labeling of a clutter

As described in Section III, the POMDP model needs to be trained prior to deploying the agent (the sensor system) on the field to make decisions on where to sense, which sensor to use for sensing, and when to stop further sensing to make declaration of the "class" of the buried object. The model consists of transition and observation probability matrices, along with cost/reward structure. These rewards/costs consist of cost of employing individual transmitters (which can be estimated based on their use of various resources like battery power etc), and the opportunity cost of mislabeling the underlying objects (described by the 2x2 table above). The trained model generates a set of α vectors, each associated with a discrete action from the action set. During the testing phase, the agent (sensor system) starts at the center of one of the underlying worlds and takes a measurement. Based on the output of the measurement, the trained policy decides the next best action, which could be taking another measurement using the same or different transmitter, or movement to another location and for further sensing. As the agent takes a sequence of actions and gathers a sequence of observations, it sequentially updates its belief over the entire universe. This iterative process terminates, when the agent is certain enough about the "class" of the underlying target (meaning the combined belief over all states of one world is close to one), at which it "declares" the class of the buried object.

Figure 6 shows the variation in classification performance as the declaration cost structure is varied. In Fig. 6(a), the reward for correct classification of an UXO or clutter is kept fixed at 500, while gradually increasing the cost of missing an UXO (or incorrectly declaring an UXO as clutter) from 400 to 4400. It is observed the probability of detection (p_d) also increases monotonically from 0.75 to 0.99. In the next set of

experiments (shown in Fig. 6(b)), we increased the cost of incorrectly declaring a clutter as an UXO from 400 to 4400. The consequence is the monotonic fall in the false alarm rate from 0.24 to 0.04. Both of these phenomena is expected and it demonstrates the direct effect of the cost structure on the probability of detection and false alarm. Figure 7 displays the variation of the average number of sensing actions taken by the sensor system before declaring the "ID" of the buried object as a function of the cost structure. As expected the number of sensing actions increase monotonically in order to achieve higher detectability of UXOs and lower false alarm rate.



(a) (b)

Figure 6. (a) Variation of p_d as a function of the cost of missing an UXO; (b) Variation in p_{fa} as a function of the cost of labeling a clutter as an UXO;

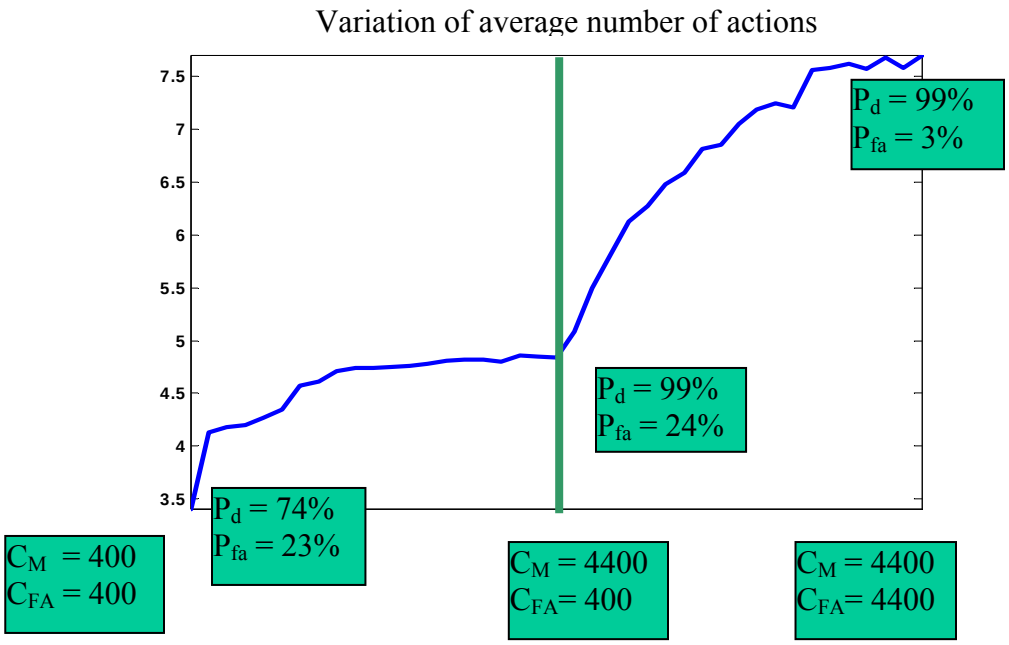


Figure 7. Variation in the number of sensing actions as a function of the declaration costs.

V. Conclusions

We have developed two algorithms with an ultimate goal of optimal sensor management for detection and classification of buried UXOs. The algorithms under development are designed to *efficiently* and *adaptively* exploit the full capabilities of next-generation EMI systems, such as the LBNL BUD system. The first algorithm is designed to approximately identify the location and model parameters of a buried object in a wide area. This approach adaptively identifies the sensing locations that minimize the uncertainty in the model estimates. The information-theoretic approach is shown to outperform the uniform sampling approach with better model estimates with less number of sensing actions. While this approach is effective in approximately identifying object locations, it does not incorporate cost of sensing or moving the sensor array from one location to the next. This approach also assumes that all the transmitter and receivers are employed for each sensing action, which might be inefficient and costly for wide area sensing. We are currently investigating the prospect of embedding these costs in the adaptive search algorithm. Once the approximate target location is identified, we employ our second algorithm that ensures optimal sequence of sensing actions to maximize target

identification, while minimizing the total sensing cost. The trained policy is optimized for a cost/reward structure provided by the policy maker. Although we have used only three transmitter coils, the scope of the policy can be expanded easily to include multi-modality where a sensor system includes multiple sensors with different sensing costs. We have restricted ourselves to only three sensing actions in this problem (corresponding to the choice of any one of the three transmitter coils for each sensing action), although it can be generalized to any possible combinations of three transmitters (six choices) and any combination of the eight receivers. Since the receivers are passive, we perceived their deployment as low-cost endeavor, hence we employed all eight receivers for each sensing action.

There are a few constraints of the proposed approaches. The adaptive search strategy assumes the sensor noise is white Gaussian which is often not completely true. The POMDP approach is based on the assumption that the approximate location of the target is known. This would not be true if the adaptive search strategy fails to locate the target. In this case the state space of the POMDP has to be increased. The POMDP model complexity grows exponentially with the size of the state space, hence it might not be practical if the adaptive search strategy fails. We are currently investigating POMDP training approaches that are capable of handling a larger state space and action space. In addition, the POMDP policy training employed here is an offline training scheme, where the policy needs to be trained every time a new sensor is added to the current set, or if the sensing/declaration costs (provided by the policy maker) change to accommodate a change in the environment. We are also investigating online POMDP training algorithms that perform concurrent exploration and exploitation to adapt to the changes in the environment in real-time and achieve the goal.

References

[1] H. P. Huang, B. SanFilipo, I. J. Won, "Planetary exploration using a small electromagnetic sensor," *IEEE Trans. Geoscience Remote Sensing*, vol. 43(7), pp. 1499–1506, July 2005.

[2] J. T. Smith, H. F. Morrison, L. R. Doolittle, H. Tseng, "Multi-transmitter multi-receiver null coupled systems for inductive detection and characterization of metallic objects", *Journal of applied Geophysics*, May 2006.

[3] N. Geng, C. E. Baum, and L. Carin, "On the low-frequency natural response of conducting and permeable targets," *IEEE Trans. Geoscience Remote Sensing*, vol. 37, pp. 347–359, Jan. 1999.

[4] Y. Zhang, L. Collins, H. Yu, C. Baum and L. Carin, "Sensing of unexploded ordnance with magnetometer and induction data: Theory and signal processing," *IEEE Trans. Geoscience Remote Sensing*, vol. 41, pp. 1005-1015, May 2003.

[5] Xuejun Liao and Lawrence Carin, "Application of the theory of optimal experiments to adaptive electromagnetic-induction sensing of buried targets", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 8, pp. 961 - 972, 2004.

[6] V.V. Fedorov, Theory of Optimal Experiments. Academic Press, 1972.

[7] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, pp. 99-134, 1998.

[8] J. Pineau, G. Gordon, and S. Thrun. "Anytime point-based approximations for large POMDPs," *Journal of Artificial Intelligence Research*, 2006.