SAND2016-11999J

Linear Models to Perform Treaty Verification Tasks for Enhanced Information Security

Christopher J. MacGahan[a,b,*], Matthew A. Kupinski[a], Erik M. Brubaker[b], Nathan R. Hilton[b], Peter A. Marleau[b]

[a] *College of Optical Sciences, The University of Arizona, 1630 E. University Blvd, Tucson, AZ, 85721, USA*
[b] *Sandia National Laboratories, Livermore, CA 94551*

Abstract

Linear mathematical models were applied to binary-discrimination tasks relevant to arms control verification measurements in which a host party wishes to convince a monitoring party that an item is or is not treaty accountable. These models process data in list-mode format and can compensate for the presence of variability in the source, such as uncertain object orientation and location. The Hotelling observer applies an optimal set of weights to binned detector data, yielding a test statistic that is thresholded to make a decision. The channelized Hotelling observer applies a channelizing matrix to the vectorized data, resulting in a lower dimensional vector available to the monitor to make decisions. We demonstrate how incorporating additional terms in this channelizing-matrix optimization offers benefits for treaty verification. We present two methods to increase shared information and trust between the host and monitor. The first method penalizes individual channel performance in order to maximize the information available to the monitor while maintaining optimal performance. Second, we present a method that penalizes predefined sensitive information while maintaining the capability to discriminate between binary choices. Data used in this study was generated using Monte Carlo simulations for fission neutrons, accomplished with the GEANT4 toolkit. Custom models for plutonium inspection objects were measured in simulation by a radiation imaging system. Model performance was evaluated and presented using the area under the receiver operating characteristic curve.

*Corresponding author
Email address: cmacgahan@optic.arizona.edu (Christopher J. MacGahan)

Preprint submitted to Nuclear Instrumentation and Methods A *October 30, 2016*

Keywords: arms control treaty verification; information barrier; discrimination algorithms; neutron imaging;

1. Introduction

Radiation imaging systems for measuring treaty accountable items (TAIs) have been proposed as a component of nuclear arms control treaty verification between host and monitoring countries. The host country desires to convince the monitoring country that the imaged item is a TAI without revealing any information they deem sensitive about the TAIs. A disadvantage of imaging techniques is that the monitoring party could use knowledge of the detector data and imaging system to reconstruct sensitive geometrical information on the host country's TAIs. In order to prevent disclosure of sensitive information to the monitor, the host could choose to implement a physical or software information barrier (IB). Examples of IB implementations are the CIVET system developed by Brookhaven National Laboratory [1] and Sandia National Laboratories' TRIS and TRADS systems [2–4]. The IBs use jointly approved data processing and analysis to output a decision for the desired task, viewable by the monitor. The monitor can authenticate the device, but the host must certify that the monitor cannot access the sensitive measurements used in the decision process. This makes the monitoring procedure more expensive, and an IB generally reduces confidence in the verification results.

Our approach is to develop mathematical models (called observer models in this work) to classify unverified test objects. The observer models are built on acquired calibration data from a trusted TAI and identify tested sources using raw projection data as opposed to a reconstructed image. The models ultimately arrive at a test statistic that is thresholded to make a decision. The observer models presented in this paper are able to process testing data in list-mode (LM) format. For each detected event, the relevant data is used to update a test statistic, and the LM data is then disposed of. Therefore, a decision is made without aggregating projection data or reconstructing spatial or spectral information on the object. To enforce LM processing in practice with a detector, an electronic board or system could be connected to the digital outputs that would update a test statistic with the incoming signals. The host would need to certify that the data vector is never aggregated in the system. This system is out of the scope of this article but is discussed in more detail in the future work section.

Alternative methods for reducing or altering the IB have been developed by groups at Princeton University and Pacific Northwest National Lab. Glaser, Barak and Goldston [5] have developed a zero-knowledge protocol (ZKP) method that preloads a measurement system with the complement of a measured data set from a trusted item. As data from an object is read in, the measurement data is updated. At the end of some predetermined acquisition time, if the aggregated output reaches some prestated value, that is evidence that the tested item was the same as the trusted item. Gilbert, Miller and White [6] are working towards a hardware solution to the problem, using a single pixel gamma-ray camera (SPGC) behind a coded aperture that changes dynamically with time, encoding the measurement data in time. Without knowing the mask sequence, the monitor could not reconstruct any sensitive details, though the measurement could be used to compare the data sets for a trusted item and tested item, performing a verification measurement. Neither of these methods has addressed the security of sensitive information in the presence of source-term variability, such as orientation and count-rate variability, that affect task performance but are not of interest to the task itself. The SPGC signal will change when the flux from the source changes, and the ZKP output will not be equal to the desired value, yielding sensitive information to the monitor. We refer to the variability caused by unknowns in the source term as nuisance parameters. Examples include object orientation and object orientation, as the imaging configuration may lack precision, or material age, which would affect the detected gamma-ray and neutron spectra and count rates. The methods discussed in this paper can be trained to account for this variability.

To the best of our knowledge, there is no model that can perform null hypothesis tests while utilizing list-mode information [7]. As such, we have decided to focus on binary discrimination tasks. This is an important point; warhead verification techniques generally seek to verify the presence of a warhead by matching it to a known template. The approach described in this paper could be used to distinguish between two hypotheses—a warhead and a well defined alternative, such as a collection of spoofs. As another example, after a warhead has been confirmed and designated for dismantlement, the monitor desires to verify that the explosive has been removed. The methods developed here are well suited to answer the question of whether the explosive has been removed or not.

In past work [8], we have demonstrated how the Bayesian ideal observer (BIO) [9], configured to process LM data, could be used to perform binary-discrimination tasks. In that model, the test statistic is equal to the likelihood ratio of measuring a data set given two hypotheses. The BIO thresholds the likelihood ratio and offers optimal performance. Nuisance parameters were incorporated in the model by integrating over them when evaluating the likelihoods. For optimal performance, this method required storage of the binned detector data for many different realizations of the nuisance parameters. This storage, whether source-term variability was present or absent, would necessitate an IB as the monitor could potentially reconstruct sensitive details of the TAI. As such, only the host country would have access to the calibration data and model (requiring an IB). Authentication would be a challenge with this method—the monitor would only be able to view the final test statistic, and testing the model on non TAI inspection items would be of limited utility due to the immense dimensionality reduction involved.

While we consider this a useful result, the desire remains to develop a model that only contains nonsensitive information on the TAIs yet still effectively discriminates them. This method would allow the host country to share the model with the monitor, allowing the monitor to use the system to test their own items and gain some measure of confidence that the system is working as stated. The sensitivity of stored information must be balanced against performance;

performance declines as less information is stored. The purpose of this paper is threefold:

1. Describe the Hotelling observer [10] (HO), which is the optimal linear discriminant and equivalent to the BIO when the statistics of the data are Gaussian. The Hotelling weights are applied to binned testing data to yield a test statistic which is then thresholded to make a decision. This model stores less sensitive information than the BIO and accounts for nuisance parameters.
2. Explore the advantages gained through utilization of the channelized Hotelling observer [11] (CHO). With this method, a series of weight vectors are applied to the image data, resulting in a lower-dimensional set of channelized values that are accessible to the monitor; the weighted sum of these values gives the test statistic. This method essentially gives the monitor access to multiple nonsensitive test statistics.
3. Study the addition of penalty terms to the CHO's optimization routine, either to maximize the number of channels available to the monitor while maintaining optimal performance or to create non-optimal channels that the monitor could access.

We first describe the methods to be used in this paper. These methods are not detector dependent. Therefore, the detector description is decoupled from the tasks and objects. Finally, the simulation studies we performed are outlined and results shown.

2. Theory

2.1. Definitions

We now briefly introduce the notation for LM data. Much of this is taken from work developed by Barrett, Parra, and Caucci [12, 13]. The data can be expressed in terms of N, the number of photons and neutrons that interact with the detector, which is Poisson distributed, and A_n, the LM data attributes estimated for each detected event. The detectable information for the n^{th} observed particle is contained in A_n. For an imaging detector, this data can be defined as

$$A_n = \{\text{particle type, pixel number, energy deposited}\}, \tag{1}$$

where n goes from 1 to N. Though we use a specific imager in our experiments, the methods developed here should apply to any measurement system that can process LM data.

The linear models developed in this paper require vectorized data. It is therefore necessary to bin the LM data into a data vector $\mathbf{g}$,

$$g_m = \sum_{n=1}^{N} f_m(A_n). \tag{2}$$

Here, f_m is the binning function for the m^{th} bin of $\mathbf{g}$, which has a total of M elements. A linear observer then applies a set of weights $\mathbf{W_g}$ to this $\mathbf{g}$ to return a scalar test statistic t as such,

$$t = \mathbf{W_g}^{\dagger}\mathbf{g} = \sum_{n=1}^{N}\sum_{m=1}^{M} W_m f_m(A_n). \tag{3}$$

This equation has been rewritten to put the sum over n out in front to emphasize the LM processing. The resulting test statistic is then thresholded to decide which hypothesis is true. Though $\mathbf{g}$ is not LM data, t can be updated by LM processing. The value of t would be initialized to zero, and the corresponding weight for each detected event would be added to the test statistic. That particle's data and the previous t value would then be deleted from memory.

There is always Poisson noise in an imaging system, but $\mathbf{g}$ becomes doubly stochastic when nuisance parameters that cause variability in the source term are present. Nuisance parameters such as object orientation or location could apply to all objects being imaged. Other nuisance parameters are object dependent, such as variations in gamma-ray or neutron intensities or energy distributions, which may be caused by variable material compositions or ages. We refer to the set γ_j of nuisance parameters, e.g.,

$$\gamma_j = \{\text{object orientation, object location, source age}\}, \tag{4}$$

where the index j is used to denote the object type. The index j can be either 1 or 2 since we are considering binary-discrimination tasks between two object classes. In this paper, hypothesis 1 corresponds to the TAI and 2 to the alternative hypothesis.

2.2. Hotelling Observer

The HO is the linear observer that best separates the test-statistic distributions for the two hypotheses [9]. As a reminder, one hypothesis would be that the item is a warhead of a certain type, and the second that the imaged item is a certain spoof or belongs to a collection of spoofs. It is also equivalent to the BIO when the data is normally distributed. The Hotelling weights $\mathbf{W_H}$ are defined as,

$$\begin{aligned} \mathbf{W_H} &= \mathbf{K_g}^{-1}\Delta\overline{\mathbf{g}} \\ \mathbf{K_g} &= \frac{\mathbf{K}_1 + \mathbf{K}_2}{2} \\ \Delta\overline{\mathbf{g}} &= \overline{\mathbf{g}_2} - \overline{\mathbf{g}_1}. \end{aligned} \tag{5}$$

Here $\mathbf{K}_j$ represents the covariance matrix and $\overline{\mathbf{g}_j}$ the mean data vector for source j. The averages are over the Poisson noise in the imaging system and, when present, nuisance parameters. The presence of nuisance parameters causes $\mathbf{K_g}$ to generally be dense. A discussion on the evaluation of the dense $\mathbf{K_g}^{-1}$ can be found in Appendix A. If the nuisance parameters are properly characterized by the host, false negatives will be kept to a minimum. Any deviation of the variability in these parameters between the objects the model is trained on compared to the objects the model is tested on will result in a higher percentage of false negatives.

The differences between the HO and BIO are critical in regards to information storage. While the BIO stores the detector data for each measurement in the covered nuisance parameter space, the HO simply contains a product of first and second order statistics over this space. As an example that is explored in Appendix B, if orientation is considered to be a nuisance parameter, the spatial and spectral information is averaged over this nuisance parameter density, yielding a set of weights that do not have an analogue to a physical pair of objects. The Hotelling weights $\mathbf{W_H}$ in (5) are analogous to a second imaging system that only observes the differences between the two TAIs. If the monitor gained access to $\mathbf{W_H}$, and tried to reconstruct $\mathbf{g}_{rec} = \mathbf{W_H}^{-1}t$ (where $\mathbf{W_H}^{-1}$ is a pseudoinverse), the resulting data set would look like the Hotelling weights, not the highly sensitive detector data. All other information in $\mathbf{g}$ is in the null space of $\mathbf{W_H}$. However, $\mathbf{W_H}$ is still useful information to the monitor, as it is related to the difference between geometries or material composition of the objects. This is information the host may want to keep hidden. For that reason, we will treat the Hotelling weights as sensitive through the rest of this paper.

In practice, the host would acquire calibration data and determine the weights. If the weights were deemed sensitive and stored behind an IB, authentication would again be a challenge as when using the BIO. The monitor would only be able to access the test statistic. There is a benefit to the test statistic being outside an IB as opposed to a standard red light/green light decision—the monitor would have some ability to differentiate objects other than the TAI and the alternate that the model is trained on.

2.3. Channelized Hotelling Observer

Use of the CHO has become widespread in the field of medical imaging as a cheap alternative to a human observer in image quality studies [14, 15]. Here we use it for a much different purpose—information security. The data vector $\mathbf{g}$, which has a size of thousands to millions or higher depending on the detector, is channelized by an operator $\mathbf{T}$ into a much smaller vector $\mathbf{v}$, with length L that can be as low as the user desires. Each row of $\mathbf{T}$ contains a template of the same size as $\mathbf{W_H}$. Using calibration data, a set of optimal channelized weights $\mathbf{W_v}$ are then found, and applied to channelized testing data to make decisions

$$\begin{aligned} \mathbf{v} &= \mathbf{Tg} \\ t &= \mathbf{W_v}^{\dagger}\mathbf{v}, \end{aligned} \tag{6}$$

where the weights that best separate the resulting test-statistic distributions are,

$$\mathbf{W_v} = \mathbf{K_v}^{-1}\Delta\overline{\mathbf{v}}. \tag{7}$$

$\mathbf{K_v}^{-1}$ and $\Delta\overline{\mathbf{v}}$ are analogous to the terms in (5). The CHO can also be represented by the weights for $\mathbf{g}$ by,

$$\mathbf{W}_{\mathbf{g}}^{\dagger} = \mathbf{W_v}^{\dagger}\mathbf{T} \tag{8}$$

Representing the model this way provides a point of comparison between non-optimal $\mathbf{T}$s for the CHO and the Hotelling weights. The performance of the CHO depends on the $\mathbf{T}$ chosen. To achieve performance equivalent to the HO, it is necessary to find the matrix that best separates the multivariate normal distributions on $\mathbf{v}_j$, or the signal-to-noise ratio,

$$SNR^2(\mathbf{T}) = \Delta\overline{\mathbf{v}(\mathbf{T})}^{\dagger}\mathbf{K_v}(\mathbf{T})^{-1}\Delta\overline{\mathbf{v}(\mathbf{T})}. \tag{9}$$

The $\mathbf{T}$ that optimizes this function is found through a gradient descent optimization routine with backtracking [16]. Matrix calculus [17] is required to take the derivative. With an optimal $\mathbf{T}$, performance equivalent to the HO is achieved.

We have therefore reduced storage from the sensitive $\mathbf{W_H}$ to a channelizing matrix $\mathbf{T}$ and a set of channelized weights $\mathbf{W_v}$. However, for an optimal $\mathbf{T}$,

$$\mathbf{W_H}^{\dagger} \approx \mathbf{W_v}^{\dagger}\mathbf{T}. \tag{10}$$

As we believe $\mathbf{W_H}$ would be deemed to be sensitive information, the host cannot reveal both $\mathbf{W_v}$ and $\mathbf{T}$ to the monitor. The monitor should only be given access to $\mathbf{T}$ or $\mathbf{W_v}$, or a subset of the two, but not the entirety of both. However, as is shown in the results section, the optimization of $\mathbf{T}$ often results in sensitive channels that resemble the Hotelling weights. Therefore, a standard implementation of the CHO in a treaty verification setting would require the host to gather calibration data and determine the channelizing matrix. The monitor would only have access to $\mathbf{W_v}$ and be allowed to observe the channelized $\mathbf{v}$ for each measured item.

In addition, because the SNR^2 is being optimized, the monitor is again unable to reconstruct $\mathbf{g}$, even if they were to cheat and gain access to $\mathbf{T}$. Similar to the HO discussion, $\mathbf{T}^{-1}\mathbf{v}$ (where $\mathbf{T}^{-1}$ is a pseudoinverse) yields a dataset that looks like the Hotelling weights.

The CHO does have some added benefits over the HO:

- Training of the CHO is far more computationally practical than the HO. Rather than needing M samples to generate an invertible covariance matrix, the host can get by with (at a minimum) L. $\mathbf{K_g}$ and $\overline{\mathbf{g}}$ can be found through the L samples, then the optimization occurs using $\mathbf{K_v} = \mathbf{T}^{\dagger}\mathbf{K_g}\mathbf{T}$ and $\mathbf{v} = \mathbf{T}\overline{\mathbf{g}}$. This is a fundamental advantage for the CHO.

- The L channel values, assuming $L << M$, are nonsensitive. The more channels present, the more information the monitor can use to discriminate possible spoofs from the TAIs other than those designated alternatives the model is trained on. The individual channel values can be aggregated as more sources are tested. The monitor could compare the test statistic value returned when measuring an unknown item to the test statistic distribution that results when measuring the objects that the model is trained on.

- There are an infinite number of channelizing matrices that maximize the SNR^2. Ideally, the monitor could use this fact to gain confidence that the algorithm is not outputting a result that is predetermined by the host, possibly by randomly picking the channelization matrix from many different optimizations. This would increase confidence that the host is not placing a spoof in front of the detector, as the host would not know in advance what the channels are, and therefore cannot design a spoof to return the same value. This could be combined with techniques to be discussed in the upcoming sections on generating a nonsensitive channelizing matrix.

2.4. Inclusion of Penalty Terms in Optimization Routine

While the standard optimization of $\mathbf{T}$ discriminates on sensitive information for certain tasks, the addition of a penalty term to the objective function in the form of

$$f_{obj}(\mathbf{T}) = SNR^2(\mathbf{T}) - f_{pen}(\mathbf{T}) \tag{11}$$

offers some possibilities in circumventing the IB requirement. In this section we discuss two different methods that have been developed to hide or explicitly penalize out sensitive information.

2.4.1. Channel Performance Penalty to Create Nonsensitive Channels

Here, we treat $\mathbf{W_H}$ as the sensitive information that the monitor cannot access. We attempt to maximize the amount of nonsensitive information available to the monitor while still maintaining optimal performance. Nonsensitive channels were created using a penalty term that reduces the performance of each individual channel,

$$f_{pen}(\mathbf{T}) = \eta \sum_{l=1}^{L} SNR^2(\mathbf{T}_{i^{th} channel}). \tag{12}$$

This penalizes the ability of each channel to distinguish the TAI from the spoof. Despite this, it is possible to maintain optimal performance, as the optimization routine now focuses on the relationships between the channels rather than the channels themselves.

As the results section shows, this method is not a perfect answer for the information security problem. While the individual channels are nonsensitive, if the monitor was given the entire $\mathbf{T}$, a singular value decomposition [18] could lead them to the Hotelling weights. However, it does allow the host to give the monitor a large number of rows of $\mathbf{T}$. This would be helpful to verify the channelization process is working properly.

2.4.2. Penalizing Declared Sensitive Information

The first method treats $\mathbf{W_H}$ as the sensitive information that the host cannot share with the monitor, but that is not strictly true. If the host is able to explicitly declare what information in the TAI is sensitive (such as mass or isotopic composition), the optimization of the channelization matrix can be used to penalize the model's ability to discriminate along that sensitive parameter. If the host does not want the monitor to know the parameter p, which takes on a value p_0, within a tolerance Δp, the host can create the following objective function,

$$\begin{aligned} f_{pen}(\mathbf{T}) =& SNR^2_{1-2}(\mathbf{T}) \\ &- \eta SNR^2_{(1,p=p_0)-(1,p=p_0+\Delta p)}(\mathbf{T}), \end{aligned} \tag{13}$$

where object 1 is the TAI and 2 is the spoof hypothesis. This objective function finds a $\mathbf{T}$ that maximizes the distance between the channelized-data distributions for the TAI and spoof while minimizing the distance between the TAI constructed with $p = p_0$ and TAI constructed with $p = p_0 + \Delta p$. The sum of the optimally weighted channels $\mathbf{W_v}^\dagger\mathbf{T}$ then no longer contains information about the differences in measured data when the TAI changes along a sensitive parameter. In essence, this routine puts that sensitive information into the null space of the channelizing matrix. As stated in the introduction, a mathematical model such as this one, which can't be used to determine the value of sensitive parameters of the TAI, could be shared with the monitor. The only IB required in this setup would be in the measurement of the trusted items and implementation of the channelization procedure. The monitor would be able to access the L channelized values and be given $\mathbf{T}$. The monitor's ability to measure and analyze their own items using the same model would increase their confidence that a useful measurement is being performed during verification.

There is a legitimate concern that the second term in eq. (13) could be altered to include a spoof that the monitor is not aware of, yielding a model that cannot differentiate the chosen spoof. Ultimately, because this model leads to a nonsensitive $\mathbf{T}$, the monitor could use the model to test for simulated spoofs in order to investigate if the host has cheated.

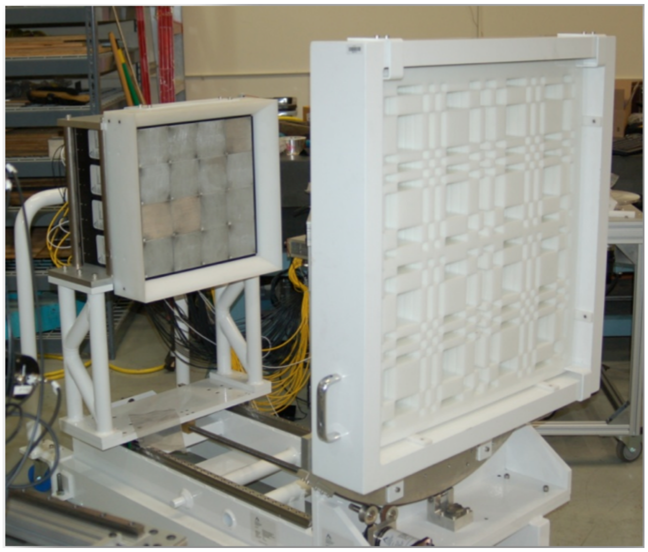

Figure 1: Fast-neutron coded-aperture imaging system. The imager uses a polyethylene coded aperture and a 4×4 array of liquid-scintillator detectors, each consisting of 10×10 $(1 \text{ cm})^2$ pixels. A quarter-inch lead plate (not pictured) is positioned in front of the pixelated detectors.

3. Detector Description

The simulated detector is the fast-neutron coded-aperture imager (see Figure 1), developed by Oak Ridge National Laboratory and Sandia National Laboratories. More details can be found in reference [19]. In the simulations performed here, a rank 19 modified uniformly redundant array [20] mask is used. Other detector configuration parameters are set such that a 50 cm $\times$ 50 cm field of view is achieved. The raw projection data discussed in this paper consists of counts or count rates in 40 $\times$ 40 detector pixels after nominal event-selection cuts. The detector system had a source-to-mask distance of 70.5 cm, mask-to-detector distance of 60 cm, mask element size of 1.21 cm and mask thickness of 6.95 cm.

4. Task Description and Simulation Data

The models were trained on data simulated from a source placed at two different locations, with the data set from one of those locations being treated as the TAI. The model was then asked to decide the location that an independently simulated data set was taken from. The relevant details necessary to replicate the simulations and the simulated data are also discussed in this section. No nuisance parameters were considered in this task—for an example incorporating nuisance parameters, see Appendix B.

4.1. Sources

Throughout this paper, a single discrimination task is used. The BeRP ball [21], a 3.79 cm radius solid sphere, was simulated at (0 cm, 0 cm) and (2 cm, 2 cm). The source at the origin was treated as the TAI and the off-center source was treated as the spoof we desire to discriminate from the TAI. the models were asked to determine the location of independently simulated data based on the neutron image.

4.2. Forward Model in GEANT4

We developed an application that uses the GEANT4 toolkit [22, 23] to model neutron transport through Monte Carlo methods. Particles emitted by the objects were transported through the object geometry to the detector. A detector-response code records the detected energy and the pixel-dependent light output (determined through experimental calibration data), applies an energy smear specific to the detector, and bins the event into a mean pixel location. A perfect pulse-shape discrimination between gamma-rays and neutrons was assumed. Visualization of our GEANT4 simulation is shown in Figure 2.

4.3. Data

As the difference in neutron images (Figure 3) was expected to be the greatest difference between the data sets, the neutron detector data was summed over energy. Gamma-rays were ignored. No neutron background was considered, and

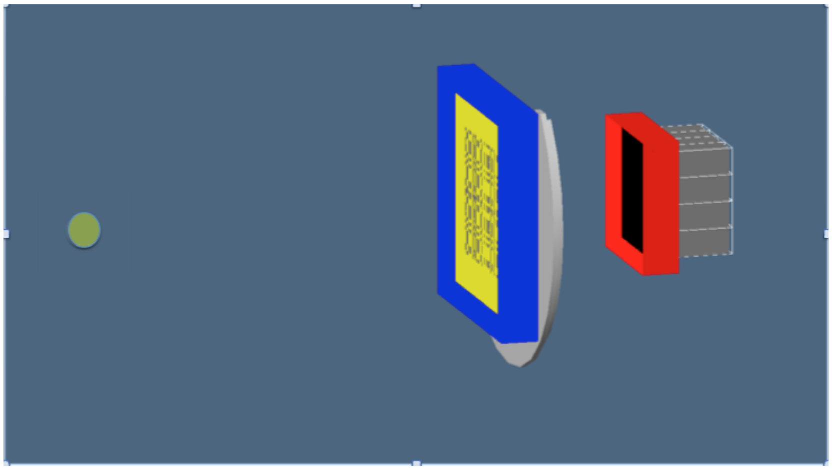

Figure 2: Geant4 model of system. The BeRP ball is shown on the left in green. The polyethylene mask, shown in yellow, is in the center and surrounded by a mask holder shown in blue. On the right is the pixelated detector.

as such studies distinguishing between different BeRP ball locations use only the source data.

4.4. Experimental Outline

For each object, two simulations were executed. One simulation was used for calibration to generate the observer model, and the second set of data was used to test the model. This was done to prevent overestimating performance.

4.5. Evaluating Performance

The receiver operating characteristic (ROC) curve [24] plots the true positive fraction (the fraction of times the model correctly declares the object is the second type) against the false positive fraction (the fraction of times the model incorrectly declares the object is the second type) for a range of thresholds across the test statistic t. The metric chosen to evaluate the models was the area under the ROC curve (AUC). We chose this metric because we are not immediately concerned with where to set the test-statistic threshold in the observer studies due to unknown costs associated with incorrect outcomes, and

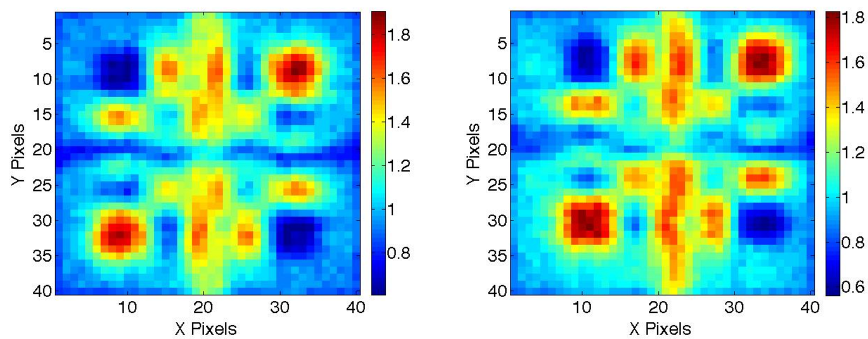

Figure 3: The figure on the left corresponds to a BeRP ball imaged at (0 cm, 0 cm) while on the right it was imaged at (2 cm, 2 cm). The colorbar is in units of detected counts per second. There were 5.5 million total detected counts (corresponding to an acquisition time of about 23 minutes) that went into each of these images.

the AUC represents performance integrated over all thresholds. Generally, the AUC increases along with the acquisition time.

We used the two-alternative forced-choice test [25] to calculate the AUC metric. The observer is presented with a series of pairs of testing datasets. In each pair, one dataset is from a measurement of source 1, and the other is from a measurement of source 2. For each dataset, the observer calculates a test statistic that is intended to have a higher value for source 2 than for source 1. The AUC is equivalent to the fraction of pairs that have a larger test statistic using the source 2 data.

5. Experiments and Results

Calibration data for the two imaged BeRP ball locations was simulated and the HO and CHO were built from that data, then tested on independent data. All results presented here were done with the same parameters in the gradient descent optimization routine. The backtracking β parameter was set to 0.5. The optimization was ended when the difference in function value between steps was 1e-6 of the function magnitude or when the step size of $\mathbf{T}$ reached 1e-5 or the

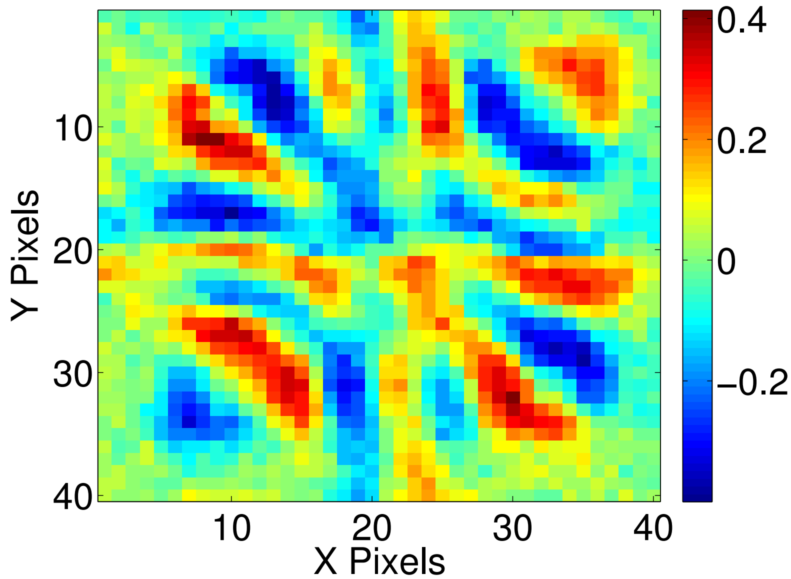

Figure 4: The Hotelling weights when the objects are BeRP balls at two different positions.

magnitude of the gradient reached 1e-5.

5.1. Hotelling Weights

The Hotelling weights (Figure 4) correspond to a scaled version of the image shift in Figure 3. Here, the BeRP ball at (2 cm, 2 cm) was source 2 in (5). The Hotelling weights in this example relate directly to the differences in projection data between the two objects.

5.2. Channelization Examples and Observer Performance

A four channel optimization (9) of the channelizing matrix was performed. The results in Figure 5 show multiple channels with strong performance that look similar to the Hotelling weights. In this task, the channels themselves would constitute sensitive information. This is further evidenced by analyzing the performance of the models and the best performing channel. We see that when the SNR^2 is optimized, the HO and CHO have equivalent performance and the best channel shows near equal performance to these optimal observers.

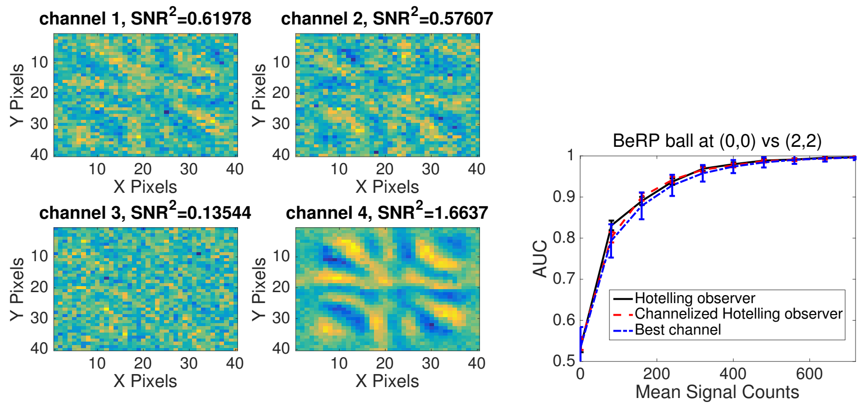

Figure 5: The left plot shows an example optimization of the channelizing matrix for the BeRP ball location study. Multiple channels show strong performance. Not shown here is the optimally weighted sum of channels, which results in an image very similar to the Hotelling weights for this task with an SNR^2 of 1.74. On the right is the performance plot for the BeRP ball location study. Many optimizations of the channelizing matrix were done for each point on the AUC curve.

Because the individual channels themselves are sensitive, the host would by necessity have to treat $\mathbf{T}$ as sensitive.

5.3. Models with Penalty Terms

This section presents implementations of the models incorporating penalty terms in the channelization matrix optimization routine.

5.3.1. Channel Performance Penalty to Create Nonsensitive Channels

We begin by showing the effect that the penalty term in eq. (12) has on the maximum channel SNR^2. For this study, we chose an acquisition time corresponding to high performance in Figure 5, so that on average 400 signal counts were observed. As Table 1 shows, when η is increased up to a value of one, the maximum channel performance continues to drop. An example output of the channel optimization when η has a value of 1 is shown in Figure 6. We see that the individual channels themselves are nonsensitive–each performs very poorly in discriminating the two BeRP ball locations.

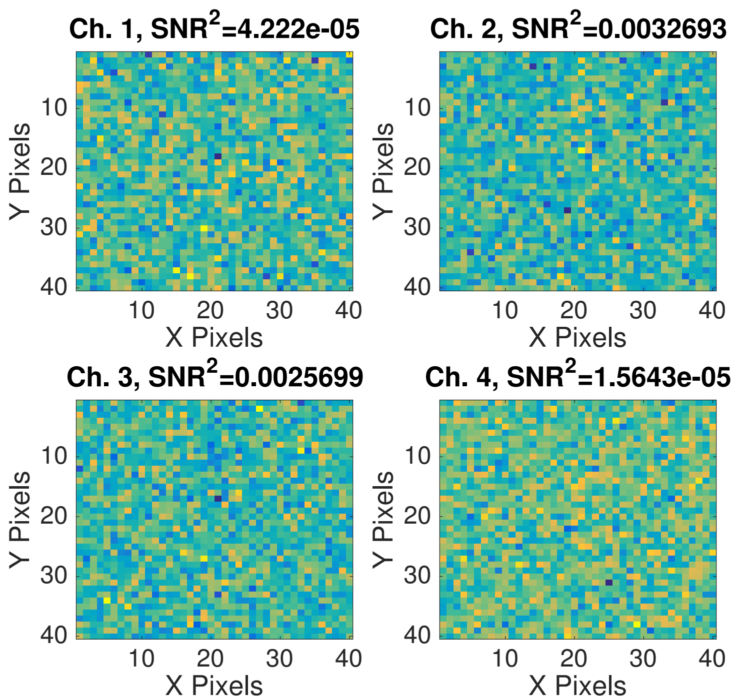

Figure 6: An example optimization of the channelizing matrix when the channel performance penalty coefficient η=1. Each channel appears to be random, but when properly weighted, maximum performance is obtained.

η	Mean SNR^2 for max channel	SNR^2 All Channels	Percent failed (total $SNR^2 < 0.1$)
1e-4	8.13	8.63	0
1e-3	6.66	8.63	0
1e-2	1.17	8.63	0
1e-1	0.106	8.63	0
1	0.0042	8.63	0
1.1	0.0025	3.88	55

Table 1: For each row in this table, 100 optimizations were performed. The acquisition time for this study was set so an average of 400 mean signal counts were read in. This corresponds to an optimal SNR^2 of 8.63 between the test-statistic distributions. As η increases, the best channel performance drops. When η rises above 1, the optimizations fail increasingly often.

Though this routine can effectively generate nonsensitive channels, a singular value decomposition [18] of $\mathbf{T}$, as shown in Figure 7, reveals that the singular vector with the lowest singular value looks like the sensitive $\mathbf{W_H}$. This result emphasizes the fact that a $\mathbf{T}$ that optimizes the SNR^2 necessarily contains sensitive information on the objects.

To this point, the channel penalty has not resulted in the monitor being able to access any more information, as $\mathbf{T}$ is still sensitive. However, due to the importance this optimization routine places on the relationship between channels, removing a single channel or small number of channels from the resulting $\mathbf{T}$ causes a debilitating effect on performance (see Table 2)—regardless of how large L is, the performance of a large number $L_{mon} < L$ of the channels is very poor. This creates an interesting application for treaty verification. The host can give the monitor L_{mon} channels of $\mathbf{T}$, the channel weights $\mathbf{W_v}$ and in testing the host can access all of the channelized data $\mathbf{v}$. The host would only keep a small number of channels, which could be chosen at random. When 90% of channels are given, performance is poor, but there are still occasional

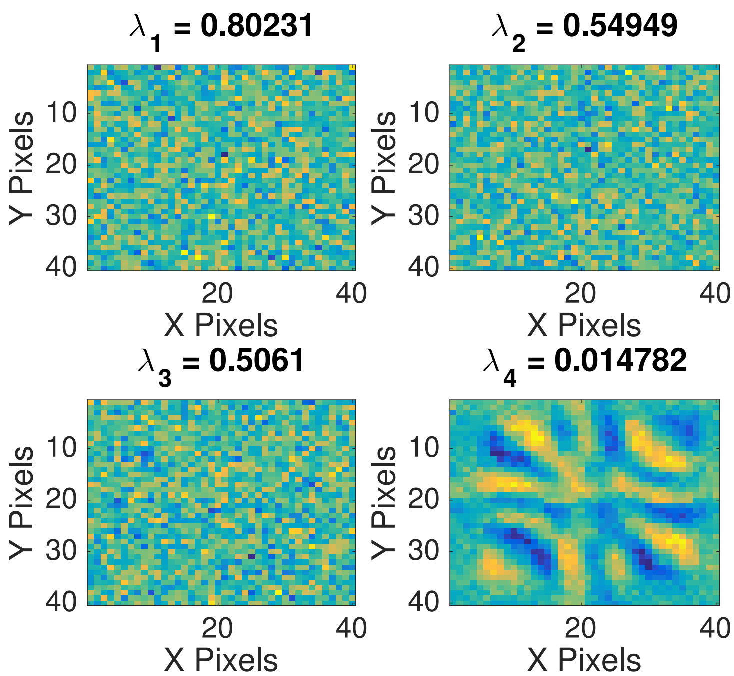

Figure 7: Singular value decomposition of the channelizing matrix shown in Figure 6. The singular vector with the lowest singular value contains the Hotelling weights.

L	L_{mon}	SNR^2 for L_{mon} channels	Percent failed (total $SNR^2 < 0.1$)
4	3	0.269	88
10	9	0.7105	50
10	7	0.0195	98
25	24	0.975	18
25	22	0.291	46
25	18	0.0314	98

Table 2: 50 optimizations of the channelizing matrix with L channels were performed in each row in this table. The second column L_{mon} is how many the monitor would have access to. The right two columns shows how L_{mon} channels would perform at the task, and what percentage of the optimizations effectively minimized the SNR^2 of the remaining channels. The acquisition time for this study was set so an average of 400 mean signal counts were read in. This corresponds to an optimal SNR^2 of 8.63.

optimizations where the monitor could gain useful information. When only 75% of channels are given to the monitor, the resulting performance is very poor, and task performance is near the guessing observer.

Overall, this method presents two advantages over the standard implementation of the CHO. Because the host can share L_{mon} channels, the monitor could hypothetically image a known nonsensitive test object, see the nonsensitive $\mathbf{v}$ and verify that the algorithm is working properly on the shared channels. Because they have access to the channels, the monitor could determine what alternative spoofs could be used to fool these channels and replicate the channelized value distributions for the trusted items.

5.3.2. Penalizing Declared Sensitive Information

To carry out this task, we created a toy problem (see Figure 8) based on the BeRP ball location data. We treated differences in the image data due to $\hat{x}$ location, from a range of 0 mm to 20 mm, as sensitive and differences in data due to the $\hat{y}$ location as nonsensitive. The BeRP ball was simulated at (20 mm, 0 mm) to penalize $\hat{x}$ information within this tolerance. This was accomplished using the following objective function,

$$\begin{aligned} f_{obj}(\mathbf{T}) =& SNR^2_{(0,0),(20,20)}(\mathbf{T}) \\ & - \eta SNR^2_{(0,0),(20,0)}(\mathbf{T}). \end{aligned} \tag{14}$$

Figure 9 shows the result of this optimization routine. Choosing a high value for the penalty coefficient, $\eta = 50$, the CHO is no longer able to distinguish the pair of sources that only differ in their $\hat{x}$ coordinate. A plot of $\mathbf{W_v T}$ is shown in Figure 10. Note that the result corresponds to a simple $\hat{y}$ shift in the count maps. The effectiveness of this method is demonstrated further in a performance study (see Figure 11). When $\eta = 0$, the optimization routine only maximizes the SNR^2 of the two different objects, and the performance matches Figure 5. When $\eta = 50$, performance in discriminating the objects we chose to optimize is still very good while discrimination between the (0 mm, 0 mm) and (20 mm, 0 mm) objects is near the guessing observer. We are also able to distinguish a BeRP ball at (0 mm, 20 mm) from (0 mm, 0 mm) with this method, which is expected because the $\hat{y}$ differences between the objects were not penalized. Finally, this particular study carries the added benefit that a tested source inside the tolerance at (10 mm, 0 mm) cannot be distinguished from (0 mm, 0 mm).

6. Conclusion

Using the linear HO and CHO models, we have developed methods that can be used to reduce storage of sensitive information while maintaining task performance. Treating the Hotelling weights as the sensitive information, a penalty term on individual channel performances would allow the host and monitor to share some channels while maintaining optimal task performance. If the host can define precisely what parameters on their object are sensitive, they could penalize out that specific information in the model, yielding test-statistic distributions with equal means for all objects that differ within the range of parameter values defined as sensitive.

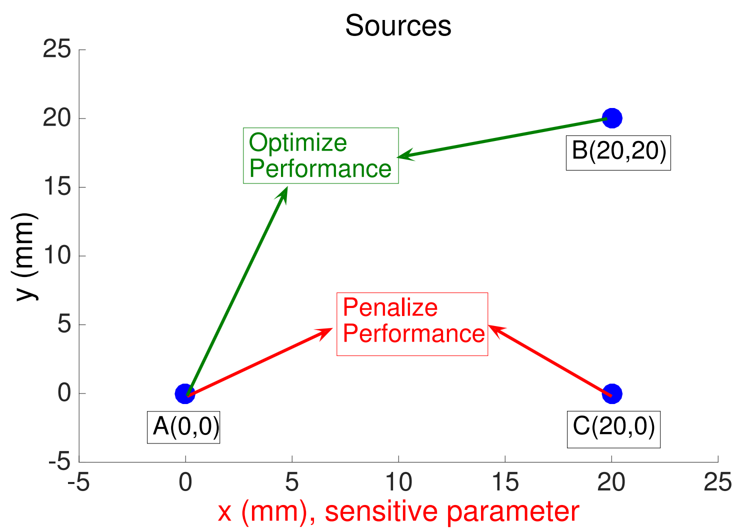

Figure 8: Diagram of BeRP ball locations (note that the BeRP ball size is larger than the difference between locations shown here). Performance between (0 mm, 0 mm) and (20 mm, 20 mm) is optimized while (0 mm, 0 mm) and (20 mm, 0 mm) is penalized in this task.

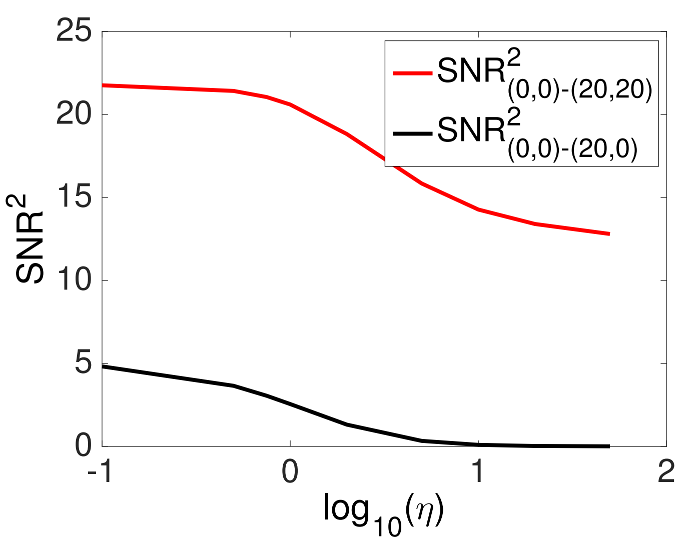

Figure 9: The SNR^2 for the BeRP ball at (0 mm, 0 mm) and (20 mm, 20 mm) vs (0 mm, 0 mm) and (20 mm, 0 mm), for an acquisition time corresponding to an average of 1000 signal counts observed. As η increases, the ability to discriminate the penalized pair of sources drops to zero while the SNR^2 for the optimized sources drops by approximately a factor of 1/3.

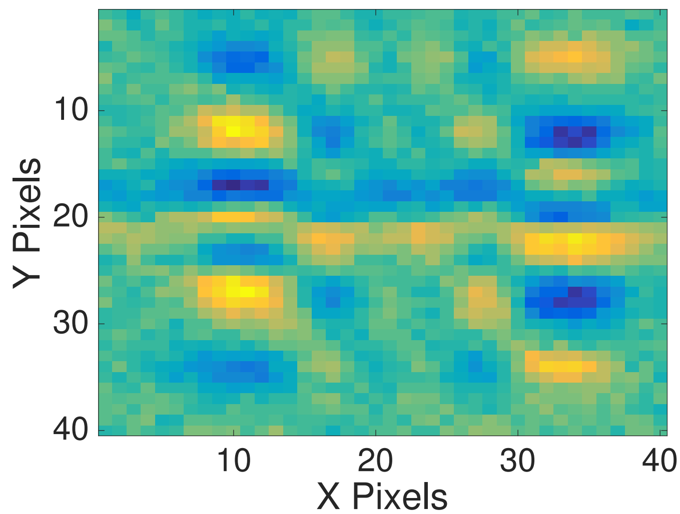

Figure 10: A plot of $\mathbf{W_v T}$ when the $\hat{x}$ information has been penalized with $\eta = 50$.

We consider this a very encouraging result, but it is a first step. In the BeRP ball localization problem, a single penalty term can be used to prevent the model from differentiating objects within the sensitive $\hat{x}$ range, but additional studies have proven this is not always true. Multiple penalty terms would likely need to be incorporated for a given penalized parameter. Furthermore, in practice multiple parameters would likely be deemed sensitive by the host, and that may require an additional penalty term with a distinct η in eq. (13). The inclusion of nuisance parameters adds another layer of difficulty to this problem, as we would need to know the effect of nuisance parameters on both objects in the penalty term. Lastly, to adapt this method for real-life verification, the host would likely need to simulate the items that differ along sensitive parameters, as construction of these items would be too expensive.

7. Future Work

While we believe these results are strong, the penalty terms discussed in this paper may not be ideal. For example, the penalization of the SNR^2 in eq. (13)

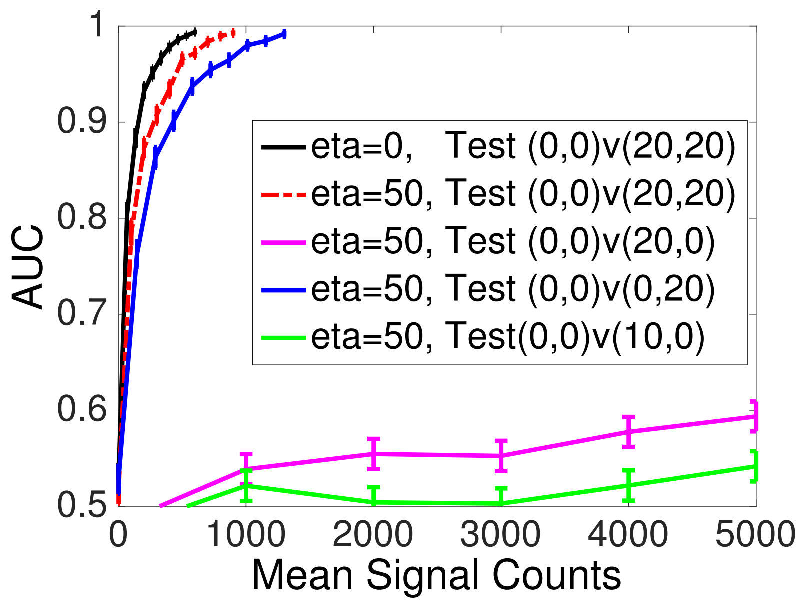

Figure 11: Performance of the CHO with channelization matrix optimized by eq. (14). The black line shows the performance of a standard optimization without a penalty term. Including the penalty in the optimization of $\mathbf{T}$, good performance is maintained when classifying sources that differ in $\hat{y}$, while near guessing observer performance is seen when classifying sources that differ in $\hat{x}$.

only necessarily equalizes the means of the test-statistic distributions for the two objects. If the monitor was given the test statistics for each tested item, they could experiment with various object configurations, trying to arrive at an object geometry that matched the variance of t and not just the mean. A more appropriate penalty term could penalize the distance between distributions using metrics described in [7].

A physical implementation of the nonsensitive channelization matrix is the next step for this project. One possibility is an electronic board that would read in the pulse characteristics (to determine particle type), integrated charge (to determine energy) and PMT ratios (to determine pixel ID) for each event, find the weight corresponding to these values, and update the test statistic. Another possibility is an attenuating plate placed between the mask and the detector plane, where the attenuator thickness varies over the detector plane to create indistinguishable data sets for objects that vary along a sensitive parameter. Further investigation would be required to determine whether this would be physically realizable.

8. Acknowledgements

This work is supported by the Office of Defense Nuclear Nonproliferation Research and Development, Nuclear Weapon and Material Security Team. Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. Internal report number SAND2016-XXXXX.

9. Appendix A

This appendix discusses calculation of the inverse of a dense covariance matrix in the case that there is additional variability in the object, background, or imaging system. In cases where the object is known exactly (signal-known-exactly, or SKE), the only randomness in the data $\mathbf{g}$ is due to Poisson noise, and the covariance matrices would be diagonal with values equal to the mean. In the SKE case, taking the inverse of the covariance matrix is trivial.

Upon incorporation of nuisance parameters, the averages become doubly stochastic or worse and the problem becomes more difficult. The terms that comprise the Hotelling weights are

$$\begin{aligned} \mathbf{K}_j &= \left\langle \left\langle (\mathbf{g}_j - \overline{\overline{\mathbf{g}_j}})^\dagger (\mathbf{g}_j - \overline{\overline{\mathbf{g}_j}}) \right\rangle_{g_j|\gamma_j} \right\rangle_{\gamma_j} \\ \overline{\overline{\mathbf{g}_j}} &= \left\langle \left\langle \mathbf{g}_j \right\rangle_{g_j|\gamma_j} \right\rangle_{\gamma_j} . \end{aligned} \tag{15}$$

The first average is over Poisson noise while the second is over the nuisance parameter distributions. This can cause the covariance to become dense and the inverse unwieldy; in a detector with 1600 pixels, where 64 energy bins are used there are approximately 1e5 elements in $\mathbf{g}$ with 1e10 elements in $\mathbf{K_g}$. The large size of $\mathbf{g}$ presents issues when using the HO in practice. A realistic number of measurements would be far less than the size of the data vector, leading to a non-invertible covariance matrix. Certain estimation techniques can be used to overcome this limitation [26, 27], but require often unrealistic assumptions.

In simulation, this computational barrier can also be overcome using the Matrix Inversion Lemma [28]. To use this lemma, the covariance matrix must be able to be represented in the form,

$$\mathbf{K} = \mathbf{A} + \mathbf{BCD}, \tag{16}$$

where A must be diagonal. To put the covariance matrix in this form, one can add and subtract $\overline{\mathbf{g}_j}$, the noise-averaged data from inside each parenthesis in the covariance matrix equation of eq. (15), so each term is $((\mathbf{g}_j - \overline{\mathbf{g}_j}) + (\overline{\mathbf{g}_j} - \overline{\overline{\mathbf{g}_j}}))$. When taking the statistical average, the cross terms average to zero and the covariance matrix can be expressed as,

$$\mathbf{K}_j = \left\langle \mathbf{K}_{\mathbf{g}_j|\gamma_j} \right\rangle_{\gamma_j} + \mathbf{K}_{\gamma_j} = Diag(\overline{\overline{\mathbf{g}_j}}) + \mathbf{K}_{\gamma_j} \tag{17}$$

The left matrix is diagonal with values equal to the mean of the detector data, averaging over Poisson randomness and all nuisance parameter densities. The right term is dense.

Treating orientation as a nuisance parameter, a vector $\boldsymbol{\theta}$ can be defined which corresponds to the orientation of the object being imaged, where [1 0 0 ....0] and [0 1 0 ....0] correspond to different object orientations. Then a system response matrix $\mathbf{H}$ can be defined and resulting detector data $\mathbf{g} = \mathbf{H}\boldsymbol{\theta}$. The covariance matrix for source j, $\mathbf{K}_{\gamma_j}$, can then be represented by the detector response function and the covariance matrix for the $\boldsymbol{\theta}$ vector,

$$\mathbf{K}_{\gamma_j} = \mathbf{H}\mathbf{K}_{\boldsymbol{\theta}_j}\mathbf{H}^{\dagger}. \tag{18}$$

In simulation we have knowledge of the nuisance parameter distributions, which are used to find $\mathbf{K}_{\boldsymbol{\theta}_j}$. By assuming the GEANT4 data is the "true" system response to each object we find the covariance matrix $\mathbf{K}_{\gamma_j}$. This technique allows us to apply the matrix inversion lemma, reducing the problem from a $M \times M$ inverse, where M can be thousands to millions, to a $P \times P$ inverse, where P is the number of orientations chosen to average over.

10. Appendix B

This appendix demonstrates the effect of the incorporation of nuisance parameters in the Hotelling observer model. In this study, the observer models were applied to discriminate between a pair of Idaho National Laboratory (INL) inspection objects labeled 8 and 9 [29], which are referred to as IO8 and IO9 (see Figure 12).

10.1. Objects and Imager

IO8 and IO9 differ in their shielding material, causing a disparity in the gamma-ray spectra in the image data. Multiple orientations of each object were considered in this task in order to understand how nuisance parameters affect the HO. The fast-neutron coded aperture was once again used to measure the objects. Though the detector is designed to observe neutrons, it also serves as a

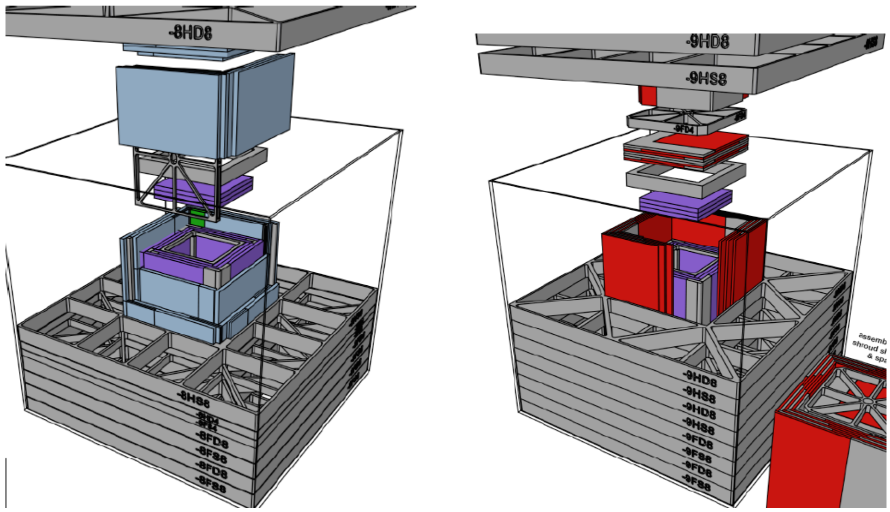

Figure 12: IO8 and IO9 developed by INL. IO8 is plutonium shielded by depleted uranium (DU) while IO9 is plutonium shielded by highly-enriched uranium (HEU). Both assemblies are supported by an aluminum framework inside an 8" × 8" × 8" aluminum box that is 1" thick.

gamma-ray imager and low-resolution spectrometer. Gamma-rays were binned into 50keV energy bins.

10.2. GEANT4 model

Due to similarities in the geometry between the two objects, neutrons were not incorporated in this study as the spectral and spatial differences were expected to be minimal. A linear energy bias, as well as a low-energy cutoff of 100 keV were used to make the simulations computationally feasible. Due to the quarter inch lead plate in front of the detector pixels, the cutoff has a negligible effect on the output. A generic gamma-ray radiation background spectrum was generated using the Gamma-ray Detector Response and Analysis Software (GADRAS) [30], and was added to all pixels.

To simulate multiple object orientations, we used a method developed by Arvo [31] that can be used to rotate the items into evenly spaced orientations. The method uses three numbers between zero and one to generate rotations of

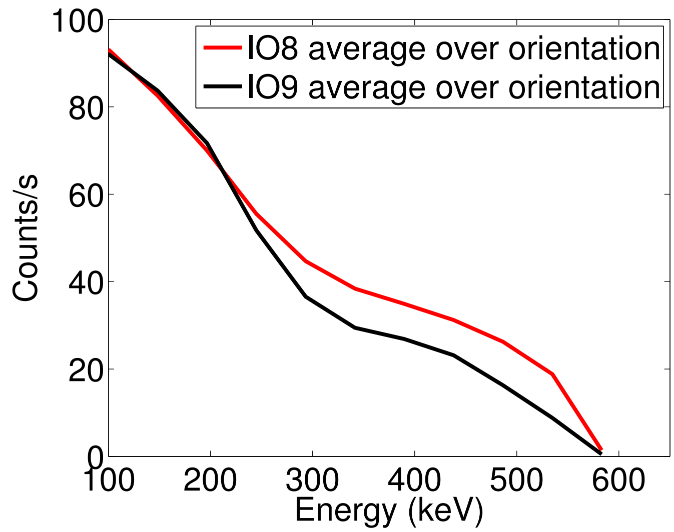

Figure 13: Comparison of IO8 and IO9 gamma-ray spectra after averaging over all simulated orientations. DU inside IO8 is the cause of the increased counts at higher energies.

an object. The object is first rotated around the $\hat{z}$ axis using the first number; then the $\hat{z}$ (vertical) axis is rotated to a certain location in ϕ, θ space using the last two. The model was trained on sixty total evenly spaced orientations. Three initial rotations around $\hat{z}$ were chosen. Then $\hat{z}$ was rotated into twenty different points (five in ϕ, four in θ).

The gamma-ray detection rates and energy spectra present the most significant difference between the two sources (see Figure 13). The spectral disparity is due to the difference in shielding material, as HEU has a highly active emission line at 186 keV and DU a moderate activity at 1001 keV. Both the detected spectra and overall count rate are dependent on the orientation chosen, though the count rate varies more significantly with orientation.

10.3. Hotelling Weights

For the INL inspection objects, 60 orientations of each object were measured in simulation, the models were developed from that data, and then they were asked to discriminate based on an independently simulated single data set from

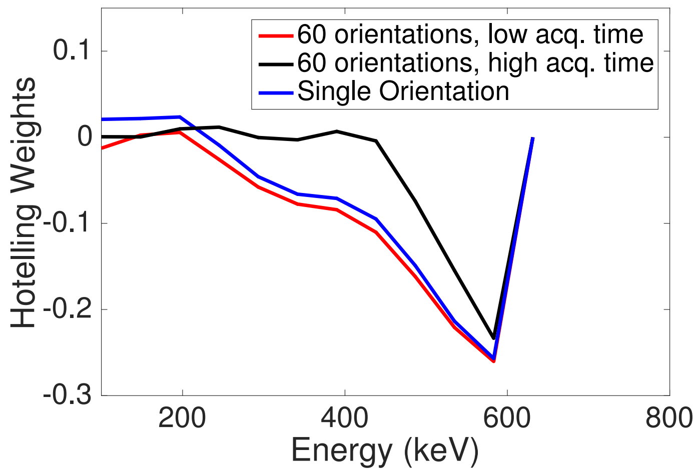

Figure 14: The Hotelling weights when discriminating IO8 and IO9. The orientation-known-exactly observer (blue line) directly relates to the observed spectra for that orientations. Their change upon incorporation of nuisance parameters is shown in the black and red lines.

one of those 60 orientations for a tested object. The Hotelling weights for the two inspection objects (Figure 14) relate directly back to Figure 13. When the orientation nuisance parameter is not present, the Hotelling weights are found by taking the difference between the counts per second in each bin of the spectra for each object (here IO9 was treated as source 2 and IO8 as source 1 in eq. (5)) and scaling by the inverse of the average of the count rates. When nuisance parameters are accounted for, the covariance matrix is dense as in eqs. (17) and (18) and there is a shift in the Hotelling weights. Furthermore, $\mathbf{W_H}$ is time dependent, as the covariance matrix in (17) that averages over nuisance parameters dominates at high acquisition times.

References

[1] P. B. Zuhoski, J. P. Indusi, P. E. Vanier, Building a dedicated information barrier system for warhead and sensitive item verification, Tech. Rep. BNL-66214, Brookhaven National Lab (1999).

[2] K. Seager, D. Mitchell, T. Laub, K. Tolk, R. Lucero, K. Insch, Trusted radiation identification system, in: Proceedings of the 42nd Annual INMM Meeting, Indian Wells, CA, 2001.

[3] D. Mitchell, K. Tolk, Trusted radiation attribution demonstration, in: Proceedings of the 41st Annual INMM Meeting, New Orleans, LA, 2000.

[4] B. Geelhood, B. Bartos, R. Comerford, D. Lee, J. Mullens, J. Wolford, Information barrier working group evaluation of trads with particular attention to its authentication merits, Tech. Rep. PNNL-13259 (August 2000).

[5] A. Glaser, B. Barak, R. J. Goldston, A Zero-Knowledge Protocol for Nuclear Warhead Verification, Nature 510 (7506) (2014) 497–502.

[6] A. J. Gilbert, B. W. Miller, T. A. White, Information barriers for imaging: The single-pixel gamma camera, in: Institute of Nuclear Materials Management (INMM) 57th Annual Meeting, 2016.

[7] S. H. Cha, Comprehensive survey on distance/similarity measures between probability density functions, International Journal of Mathematical Models and Methods in Applied Sciences 1 (4).

[8] C. J. MacGahan, M. A. Kupinski, N. R. Hilton, E. M. Brubaker, W. C. Johnson, Development of an ideal observer that incorporates nuisance parameters and processes list mode data, JOSA A 33 (4) (2016) 689–697.

[9] H. H. Barrett, K. J. Myers, Foundations of image science, Foundations of Image Science, by Harrison H. Barrett, Kyle J. Myers, pp. 1584. ISBN 0-471-15300-1. Wiley-VCH, October 2003. 1.

[10] H. Hotelling, The generalization of student's ratio, Ann. Math. Statist. 2 (3) (1931) 360–378. `doi:10.1214/aoms/1177732979`.
URL `http://dx.doi.org/10.1214/aoms/1177732979`

[11] H. H. Barrett, J. Yao, J. P. Rolland, K. J. Myers, Model observers for assessment of image quality, Proceedings of the National Academy of Sciences 90 (21) (1993) 9758–9765.

[12] H. H. Barrett, T. White, L. C. Parra, List-mode likelihood, JOSA A 14 (11) (1997) 2914–2923.

[13] L. Caucci, H. H. Barrett, Objective assessment of image quality. v. photon-counting detectors and list-mode data, JOSA A 29 (6) (2012) 1003–1016.

[14] J. Yao, H. H. Barrett, Predicting human performance by a channelized hotelling observer model, in: San Diego'92, International Society for Optics and Photonics, 1992, pp. 161–168.

[15] S. Wollenweber, B. Tsui, D. Lalush, E. Frey, K. LaCroix, G. Gullberg, Comparison of hotelling observer models and human observers in defect detection from myocardial spect imaging, Nuclear Science, IEEE Transactions on 46 (6) (1999) 2098–2103.

[16] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge university press, 2004.

[17] E. Bodewig, Matrix calculus, Elsevier, 2014.

[18] G. H. Golub, C. Reinsch, Singular value decomposition and least squares solutions, Numerische mathematik 14 (5) (1970) 403–420.

[19] P. Hausladen, M. A. Blackston, E. Brubaker, D. Chichester, P. Marleau, R. J. Newby, Fast-neutron coded-aperture imaging of special nuclear material configurations, in: 53rd Annual Meeting of the INMM, Orlando, FL, USA, 2012.

[20] S. R. Gottesman, E. Fenimore, New family of binary arrays for coded aperture imaging, Applied optics 28 (20) (1989) 4344–4352.

[21] J. Mattingly, Polyethylene-reflected plutonium metal sphere: Subcritical neutron and gamma measurements, SAND2009-5804 Revision 2.

[22] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, et al., Geant4—a simulation

toolkit, Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (3) (2003) 250–303.

[23] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. A. Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek, et al., Geant4 developments and applications, Nuclear Science, IEEE Transactions on 53 (1) (2006) 270–278.

[24] J. A. Hanley, B. J. McNeil, The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve., Radiology 143 (1) (1982) 29–36.

[25] G. T. Fechner, E. G. Boring, D. H. Howes, H. E. Adler, Elements of Psychophysics. Translated by Helmut E. Adler. Edited by Davis H. Howes And Edwin G. Boring, With an Introd. by Edwin G. Boring, Holt, Rinehart and Winston, 1966.

[26] P. J. Bickel, E. Levina, Regularized estimation of large covariance matrices, The Annals of Statistics (2008) 199–227.

[27] J. Bai, S. Shi, Estimating high dimensional covariance matrices and its applications.

[28] M. A. Woodbury, Inverting modified matrices, Memorandum report 42 (1950) 106.

[29] R. Neibert, J. Zabriskie, C. Knight, J. L. Jones, Passive and active radiation measurements capability at the inl zero power physics reactor (zppr) facility, Tech. rep., Idaho National Laboratory (INL) (2010).

[30] D. J. Mitchell, Gamma detector response and analysis software (gadras), Tech. Rep. SAND88-2519, Sandia National Laboratories (1988).

[31] J. Arvo, Fast random rotation matrices, in: Graphics Gems III, Academic Press, 1992, pp. 117–120.