SECURITY STUDIES PROGRAM
Massachusetts Institute of Technology

Attachment D
Letter to John Podesta
White House Chief of Staff

**POET Study 1998-5**

# Independent Review of TRW Discrimination Techniques Final Report

**(Title UNCLASSIFIED)**


**M-J. Tsai, MIT Lincoln Laboratory**

**Larry Ng, Lawrence Livermore National Laboratory**

**Glenn Light, Aerospace Corporation**

**Frank Handler, POET/Lawrence Livermore National Laboratory**

**Charles Meins, MIT Lincoln Laboratory**

# Unclassified Draft

## EXECUTIVE SUMMARY

[U] The exoatmospheric kill vehicle (EKV) is the most technically challenging aspect of National Missile Defense (NMD) development. The EKV must acquire and track the target complex, discriminate and select the reentry vehicle (RV), maneuver to the RV, and kill it by kinetic impact. Target discrimination and selection must be done with or without handover information about the target from the rest of the NMD system.

[U] TRW has developed an onboard autonomous discrimination architecture and a set of algorithms for the Boeing EKV. These algorithms have been evaluated against threats described in the *Technical Requirements Document* (TRD) using simulated data, as well as against the target suite flown on integrated flight test IFT-1A using data collected by the Boeing EKV. During the past two years, various technical concerns have been raised regarding the concept, implementation, and performance of TRW's discrimination baseline algorithms (BLA).

[U] Recently a POET team was tasked by BMDO JN/I to conduct an independent review of TRW's baseline discrimination techniques. With cooperation from the Ground-Based Interceptor Program Office (GBI PO), TRW, NRC, and DCIS, the POET team has completed three major tasks: (1) a review of the scientific, mathematical, and engineering principles of TRW's BLA; (2) a review of the BLA's performance against the IFT-1A data as reported by TRW to the Government; and (3) a determination of the projected performance of the BLA against possible IFT-3 scenarios.

[U] In fulfillment of the first task, the POET team reviewed the concept, design, implementation, and operational performance analyses of the BLA, primarily by studying technical documents provided by TRW, JN/I, the GBI PO, and DCIS. The POET team's major findings may be summarized as follows:

[U] 1. Overall, the BLA are well designed and work properly, with only some refinement or redesign required to increase the robustness of the overall discrimination function. (Specific comments on redesign are listed in items 2 and 3 below; items 4 and 5 pertain to robustness).

[U] 2. The gap-filling algorithm (GFA), which was designed to fill data gaps induced by low signal-to-noise ratio (SNR) signals early in the engagement timeline, has not yet been demonstrated to be effective. Its use may actually hurt rather than help with baseline discrimination, as demonstrated in the IFT-1A postmission data analyses.

[U] 3. Among the eight features included in the BLA, only a few can be used in combination for the discrimination function. Because they are derived from only two independent radiant-intensity sequences as measured by the EKV sensor, the eight features are highly correlated. One of the consequences of this particular design is that the covariance matrices of certain feature subsets may become noninvertible.

[U] 4. The performance of the discrimination architecture may be fragile. Training data sets that closely match actual threats are crucial to the performance of the multimodal Bayesian quadratic classifier (MBQC) for the fine (i.e., precision) discrimination function. Performance may degrade significantly if incorrect prior knowledge regarding target signatures is used in the classifier database. The target signatures are heavily influenced by threat type, target characteristics, sensor-to-target geometry, and engagement timeline. Therefore, unexpected targets in the threat may challenge the performance of the MBQC. It would be desirable to expand the current discrimination architecture to make it more robust by including, for example, real-time adaptation capability and certain threat-type–based system-level discrimination.

...nsitivit... TRW

organizations.

[U] 6. The target to be intercepted is selected on the basis of an "importance" measure, which is an ad hoc combination of two different methods of inference. One of the inputs, the probability of being a lethal target ($P_{AT}$), is defined within the probability theory, yet its combination with another input (an ad hoc "confidence" factor) for the "importance" calculation is done using fuzzy logic. As a result, it is impossible to calculate the probability of selection ($P_{select}$) analytically, and Monte Carlo simulation becomes the only way for statistical evaluation of the discrimination performance.

[U] In fulfillment of the second task, the POET team investigated the purported discrepancies among TRW's IFT-1A data analysis reports and analyzed flight-test data in order to validate important algorithms and to check on the accuracy of key discrimination results reported by TRW. The POET's results may be summarized as follows:

[U] 1. The discrepancies in feature ellipses among the 45-day, 60-day, and revised 60-day reports result from the fact that in the 60-day report the size of the ellipses was mislabeled and for the 45-day report an improperly constituted GFA was used. Further, the POET team determined that TRW's decision to turn off the GFA for the IFT-3 mission was technically justified.

[U] 3. Outputs of the important discrimination algorithms reported by TRW have been verified.The fading memory filter was properly designed and tuned, the features were correctly extracted, and the classifier database included in the mission data load (MDL) was consistent with the reported feature ellipses.

[U] 5. The Kalman filter that was originally developed by TRW to provide enhanced discrimination features (but was later removed) could be tuned to perform properly for the IFT-1A MRV data and a set of noisy simulated data. But a robust procedure for filter initialization must be developed to make this a generally useful algorithm.

[U] In fulfillment of the third task (which was to be performed if time permitted), the POET team predicted the performance of the BLA against IFT-3 scenarios based on experience gained from IFT-1A. According to the GBI PO, five target suite options are currently under consideration for IFT-3. The most stressing target suite for EKV discrimination contains the same ten targets as those deployed in IFT-1A: an MRV, a multiple service launch system (MSLS), and eight penaids. Regardless of which target suite is selected, signature data collected on the IFT-1A targets will be extremely useful for constructing a representative MDL classifier database for IFT-3. If targets are deployed nominally in IFT-3, it is the POET team's assessment that the TRW BLA will successfully select the MRV as the intercept target, even for the most stressing target suite.

[U] The issue has been raised that TRW's discrimination concept and its implementation will not and cannot function to meet the TRD requirements. At the request of Mr. Keith Englander (JN/I), the POET team has made point-by-point responses to the issues raised in a 16-page document authored by DCIS. Those responses are included in Appendix B of this report. Overall, the issues raised serve to emphasize that discrimination is a highly complex and difficult function. Though it is not the intention of the POET team to interpret the TRD requirements, it sees no evidence that suggests any obvious noncompliance with the TRD.

[U] One of the most important criticisms raised by DCIS is that during the IFT-1A postmission analyses TRW unjustifiably shifted the preflight feature distribution ellipses between the 45-day report and the 60-day report in order to increase the BLA's discrimination performance. As previously noted, TRW's explanation that an improperly constituted GFA caused the discrepancies appears to be correct. The decision not to use the GFA for IFT-3 also seems reasonable. Another important issue concerns the "improper" concept of ranking probability used by TRW for selecting an intercept target. It is clear to the POET team that all points raised concerning this issue were based on misunderstanding of Bayes rule and misinterpretation of TRW's approach.

[U] It is a legitimate concern that the robustness of the BLA's discrimination function could be compromised by the assumption that threat-typing information would be available for use in the MDL database and that target characteristics, motion parameters, and engagement geometry would be known for a given threat system. Reliable prior knowledge on many of the threat types included in the TRD cannot be obtained. The POET study did not evaluate to what extent the BLA may be used to meet the general requirements described in the TRD beyond the ability to determine which object among a threat cloud is lethal (given that the threat type is known). However, the POET team's concerns regarding the potential lack of robustness of the current BLA have been indicated earlier. The POET team also recognizes that the concept of threat typing can have significant impact on EKV discrimination performance and that such impact should be examined carefully in the context of overall NMD architecture design and threat assessment. Within the EKV scenario it may be important to structure the discrimination architecture to take advantage of threat-typing information while at the same time reducing potential risk.

# Unclassified Draft

## ACKNOWLEDGMENTS

(U) We would like to thank the many people from BMDO JN/I, DCIS, the GBI Program Office, TRW, Boeing, and NRC who have made available the documents, data, and briefing materials upon which the results and conclusions presented in this report are based. Special thanks go to Dr. N. Schwartz, who identified a number of potential issues for us to focus on, and to Mr. A. J. Rienecke, who helped us acquire documents and arrange briefings.

# Unclassified Draft

## TABLE OF CONTENTS

# Unclassified Draft

## TABLE OF CONTENTS (Continued)

# Unclassified Draft

## LIST OF ILLUSTRATIONS





Unclassified Draft

# 1. INTRODUCTION

[U] The exoatmospheric kill vehicle (EKV) is the most technically challenging aspect of National Missile Defense (NMD) development. The EKV operates by acquiring a threat cluster from information supplied to it by surveillance sensors through the battle management, command, control, and communications ($BM/C^3$) element. The EKV must fuse this information with the scene that its onboard seeker sees, select the reentry vehicle (RV), and maneuver to the RV, destroying it by force of impact. If insufficient threat information is available from the rest of the NMD system, the EKV must be able to determine the lethal object using onboard discrimination and target selection.

[U] TRW has developed an onboard discrimination architecture and a set of algorithms for the Boeing EKV. These algorithms have been evaluated using data simulated according to the threat described in the *Technical Requirements Document* (TRD), as well as data collected by the EKV sensor on the integrated flight test IFT-1A mission. Over the past two years various questions have been raised concerning the concept, implementation, and performance of TRW's baseline discrimination design.

[U] A POET team was recently tasked by BMDO JN/I to independently evaluate TRW's baseline discrimination algorithms (BLA): the team began its work on 1 June 1998. The principal team members were Dr. Ming Tsai (MIT Lincoln Laboratory, task leader), Dr. Frank Handler (Lawrence Livermore National Laboratory/POET, task manager), Dr. Larry Ng (Lawrence Livermore National Laboratory), Mr. Glenn Light (Aerospace Corporation), and Dr. Charles Meins (MIT Lincoln Laboratory). As outlined in the Statement of Work (Appendix A), the POET team's major tasks included (1) a review of the scientific, mathematical, and engineering principles of TRW's BLA and (2) a review of the BLA's performance against the IFT-1A data, as reported by TRW to the Government. Time permitting, the POET team would also examine the BLA's performance for the proposed IFT-3 scenarios.

[U] For the POET team to complete the study in about three months, the help of the Ground-Based Interceptor Program Office (GBI PO), TRW, NRC, and DCIS was required. Appropriate briefings, documents, and flight-test data were identified and requested from these organizations through JN/I. The POET team obtained a total of 21 documents and 2 data tapes: [1-2] from JN/I, [3-16] from TRW, [17] and [23] from the GBI PO, and [18-22] from DCIS.

[U] The POET team also received a one-day briefing from DCIS (16 June 1998) and a two-day briefing from Boeing/TRW and NRC (21 and 22 July 1998). In the first briefing DCIS presented an evaluation of TRW's baseline discrimination technology, claiming it contained a number of inconsistencies and inaccuracies that could have caused noncompliance with the requirements given in the TRD [21]. Results documented in the "Threat Typing Sensitivity Study Report (TTSSR)" [6], "Kalman Filter Performance" [7], "IFT-1A 45-Day Report," and "IFT-1A 60-Day

# Unclassified Draft

Report" [10-12] were cited. In the second briefing TRW personnel presented an in-depth review of their EKV baseline discrimination technology, including concept, algorithm designs, operational performance analyses, and IFT-1A data analysis [16]. NRC personnel also presented the results of their independent testing.

[U] The POET team concluded that, with these documents and briefings, they have had adequate resources with which to carry out their tasks. The team reached their conclusions, which are documented in this report, primarily by studying the issues raised, reviewing the available documents, and independently analyzing the flight-test data.

[U] In Section 2 of this report, TRW's baseline discrimination technology is reviewed and evaluated. In Section 3 the POET team's assessment of the baseline discrimination performance against the IFT-1A data is presented. Projected performance for IFT-3 is given in Section 4. A preliminary evaluation of the performance of the Kalman filter is discussed in Section 5. Sections 2, 3, and 4 directly address the major tasks that were outlined in the Statement of Work, which, for completeness, is included as Appendix A. Appendix B was compiled as the result of a later request by Mr. Keith Englander of IN/I; it gives the POET team's point-by-point responses to the issues raised by DCIS.

Unclassified Draft

# 2. REVIEW OF TRW'S BASELINE DISCRIMINATION ALGORTHMS

[U] The POET study team's first task was to review the scientific, mathematical, and engineering principles of TRW's BLA. This task was carried out primarily by studying the briefings and documents provided by TRW. As reported in this section, the POET team's review includes the discrimination concept, algorithm design, and operational performance analysis against the TRD threats.

## 2.1 DISCRIMINATION CONCEPT

### 2.1.1 Overview

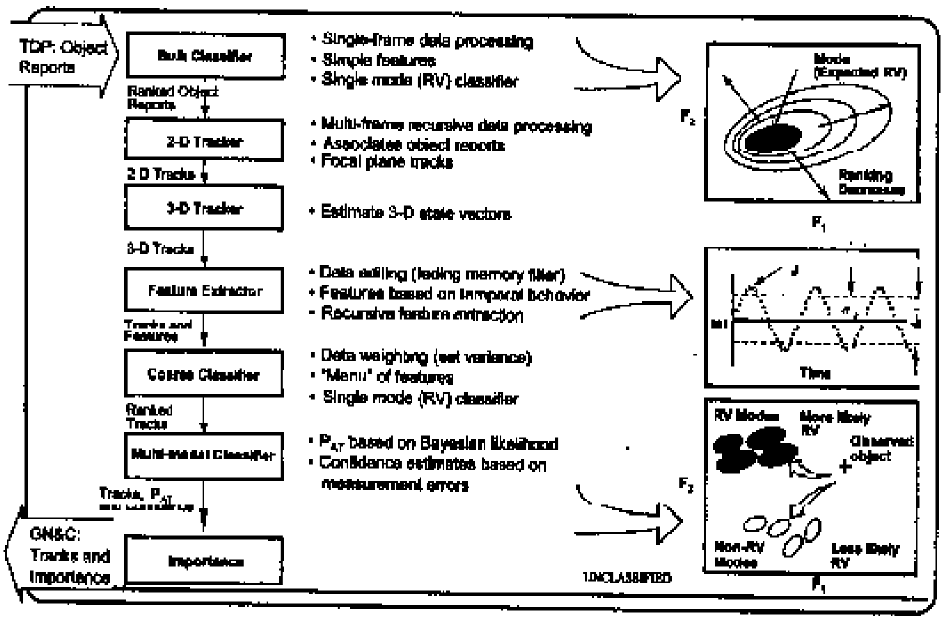
[U] As described in [8] and [16] and illustrated in Figure 1, the EKV discrimination function is highly integrated with the guidance, navigation, and control (GN&C) logic. The discrimination function continually provides target-ranking information to the GN&C in a tight coupling loop so that the seeker can, within its onboard processing power, keep the most likely lethal target in its focal plane throughout the homing phase and eventually select the correct target to intercept in the endgame. Discrimination is performed in three phases (bulk, coarse, and fine), which employ algorithms of increasing complexity and demand increasing processing power. The bulk discrimination function is mainly for quick rejection of easily identified nonlethal targets; the fine (i.e., precision) discrimination function is for final target selection.

[U] Fine discrimination is actually a classical multimodal Bayesian quadratic Gaussian classifier (MBQC). Such a classifier assumes a set of target models (or hypotheses) and is defined by a database that consists of prior probabilities of all target models combined with the associated mean vectors and covariance matrices of a set of features. The classifier provides, for each feature vector of a track file, what is called the probability of being the assigned target ($P_{AT}$) as an input to the mission director software for target selection.

[U] The MBQC database includes feature data from as many missile systems as specified in the TRD [1]. The database can be reduced using system-typing information; i.e., if the system type of the incoming threat is known prelaunch, only data pertinent to that missile type is provided to the classifier, using what is called the mission data load (MDL).

Unclassified Draft

UNCLASSIFIED

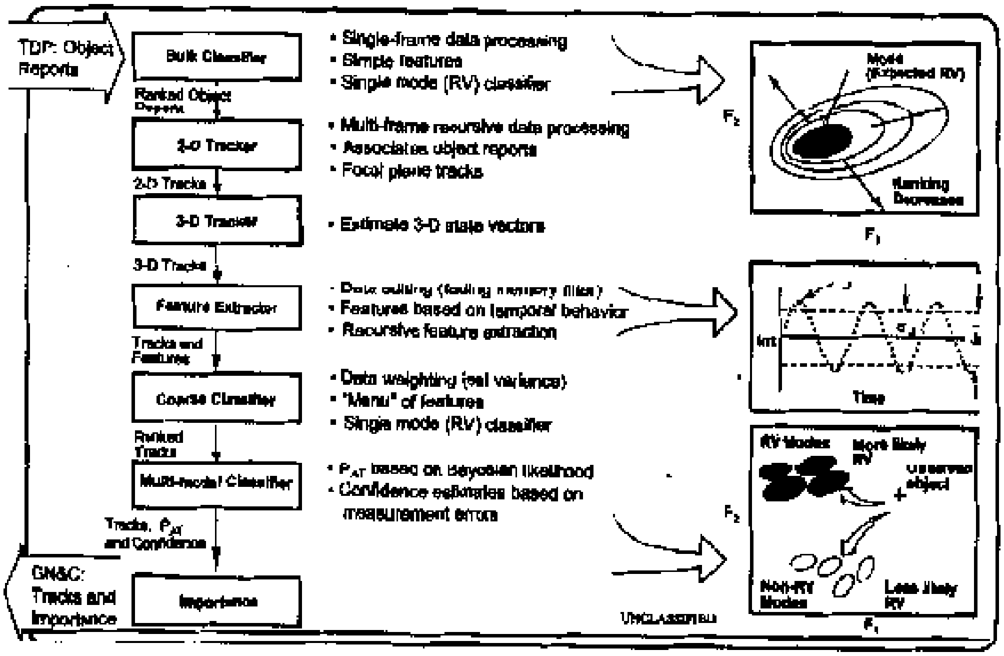



*[U] Figure 1. TRW/Boeing EKV discrimination concept (courtesy of TRW).*

### 2.1.1 POET Review

[U] Tight coupling between the discrimination and GN&C functions is a valid concept. In fact, a design based on this concept is virtually unavoidable given that the EKV seeker is required to provide early acquisition (approximately 60 s before intercept) and to operate even without target object mapping handovers from the Ground-Based Radar (GBR)

[U] The MBQC, as adopted for the fine discrimination function, has several attractive properties. It is understandable, easy to train (if representative data are available), easy to use, and expandable to more complex threat scenarios. Other missile defense elements, such as the Theater High-Altitude Area Defense (THAAD) radar and the NMD GBR, use the same classifier design.




*[U] Figure 1. TRW/Boeing EKV discrimination concept (courtesy of TRW).*

### 2.1.1 POET Review

[U] Tight coupling between the discrimination and GN&C functions is a valid concept. In fact, a design based on this concept is virtually unavoidable given that the EKV seeker is required to provide early acquisition (approximately 60 s before intercept) and to operate even without target object mapping handovers from the Ground-Based Radar (GBR).

[U] The MBQC, as adopted for the fine discrimination function, has several attractive properties. It is understandable, easy to train (if representative data are available), easy to use, and expandable to more complex threat scenarios. Other missile defense elements, such as the Theater High-Altitude Area Defense (THAAD) radar and the NMD GBR, use the same classifier design.

The MBQC has a significant inherent drawback, however, in that it depends very much on the training data. If the actual threat and the training data are closely matched, the classifier is expected to perform very well; if there are meaningful mismatches, the performance may degrade considerably. It may be difficult for this kind of classifier to handle unexpected targets in the threat successfully.

[U] The concept of $P_{AT}$ may need clarification. When the MBQC is executed for a given feature vector, $X$, the a posteriori probability of the feature vector being from one of the hypotheses defined in the classifier database, $p(H_i|X)$, is calculated. Suppose there are $N$ hypotheses. Then

$$p(H_i|X) = [p(H_i)p(X|H_i)]/\left[\sum_{i=1}^{N} p(H_i)p(X|H_i)\right] \quad ,$$

where $p(H_i)$ is the a priori probability of $H_i$ and $p(X|H_i)$ is the likelihood function of $X$. It is obvious from the above equation that

$$\sum_{i=1}^{N} p(H_i|X) = 1 \quad .$$

Note that $P_{AT}$ as defined by TRW is, in fact, the $p(H_i|X)$ of a given $X$ when $H_i$ is the RV. Suppose there are $M$ targets under track at a given time by the EKV sensor. For each target and the corresponding feature vector, $X_j$, $P_{AT}(X_j) = p(RV|X_j)$ is calculated by the MBQC. There is no reason to assert that the summation of the $P_{AT}$ of all the targets in the threat complex has to be 1. That is, usually

$$\sum^{M} P_{AT}(X_j) \neq 1 \quad .$$

## 2.2 ALGORITHM DESIGNS

### 2.2.1 Overview

[U] As described in [5] and [16], the BLA include the bulk classifier, gap-filling algorithm (GFA), feature extractor, coarse classifier, fine discrimination function, and "importance" calculation. These algorithms interact with other elements within the track, fusion, and discrimination (TFD) module, including the two- and three-dimensional trackers.

lk classifie m TFD f
nd

. so:
ıne MBQ

### 2.2.2 POET Review

[U] In general, the BLA are well designed, though there are a few areas where refinement or redesign is necessary in order to increase the robustness of the overall discrimination function.

trated to be u

suitability f

[U] The GFA fills the OR gaps with values equal to $k \cdot Th$, where $k$ is a preset constant between 0 and 1, and $Th$ is the detection threshold. Because of data gaps, the mean and standard deviation estimates of a radiant-intensity signal are biased; gap-filling is meant to reduce such biases. However, if $k$ is not correctly chosen, another kind of bias may be created. It is the POET team's opinion that the GFA should be redesigned. One option is to incorporate into the gap-filling process knowledge about the signal, such as the limit of the bandwidth.

:ach tr

nd are m(

[U] The mean vectors and sample covariance matrices of the MDL classifier database are obtained premission out of many simulation runs using feature vectors calculated at a fixed time close to the endgame. In real-time operation, for each target a feature vector and its associated estimation covariance are calculated frame by frame, and the feature vector alone is used to calculate $P_{AT}$ in the MBQC. This approach introduces a mismatch between training (MDL) and testing data, because the latter usually have larger dispersion than the former (due to lower SNR). A proper approach is to include the covariance of feature extraction to compensate for the SNR-induced mismatch.

baseline fea

use features

led the
rgt f track
n tl : to an
ig t m, the
pai rence.

[U] The "importance" measure, the basis upon which the intercept target is selected, is the outcome of an odd combination of two different methods of inference. One of the inputs, $P_{AT}$, is defined within the probability theory, yet its combination with the confidence factor for the importance calculation is done using fuzzy logic. As a result, it is impossible to calculate the probability of selection analytically, and Monte Carlo simulation becomes the only way for statistical evaluation of the discrimination performance.

## 2.3 OPERATIONAL PERFORMANCE ANALYSIS

### 2.3.1 Overview

[U] TRW has been using a software package, DISCSIM, to evaluate the performance of the baseline discrimination architecture. DISCSIM includes tools for signature generation (OPTISIG and OSC), feature extraction (MODEL), feature selection (BEST), overall performance simulation (VICTIM), and data visualization. The most important performance index, what is called $P_{select}$, is the fraction of VICTIM Monte Carlo runs for which the RV is selected as the intercept target.

[U] DISCSIM has been applied by TRW to evaluate the baseline discrimination function against the TRD near-term threat [8] and the NSEN C1 threats [16], as well as to study the threat-typing sensitivity [6]. Results show that, in all cases, the baseline discrimination architecture and BLA met the required performance levels. NRC repeated the same tests for the TRD near-term threat and basically confirmed TRW's results [17].

### 2.3.2 POET Review

[U] The best feature subset is selected within BEST according to what is called the probability of identification ($P_{ID}$), which is calculated using numerical integration. $P_{ID}$ is different than $P_{select}$, which is the output of VICTIM Monte Carlo runs. In each of these Monte Carlo runs the MBQC is actually executed using the feature subset as selected by BEST. Given that $P_{ID}$ and $P_{select}$ are two different performance indices, the best feature set determined by BEST may not produce the best performance by VICTIM. In its independent assessment work [17] NRC reported that the BEST feature set produced a $P_{select}$ that was below the required performance levels in a few cases and sometimes caused the simulation tools to crash. A potential solution to this problem is to select the best features according to the $P_{AT}$ of the MBQC in a large number of Monte Carlo runs.

[U] In TRW's performance testing, test samples were generated in VICTIM under exactly the same conditions as were used to generate the MDL classifier database in MODEL. Both modules assumed the same target characteristics, motion parameters, and sensor models, and both used the same values for the variables used for the BLA (track length, filter constants, flags, etc.). This perfect match may at least partly explain the excellent test results. A question remains as to what the performance will be in real-world situations, where threats are not exactly known and sensors may deviate from nominal operating conditions. It will be interesting to test the MDL designed for the TRD near-term threat against data collected by intelligence sensors and/or data simulated by independent organizations.

# Unclassified Draft

## 3. REVIEW OF REPORTED PERFORMANCE FOR IFT-1A

[U] The POET team's second task was to review the performance of the BLA technology against the IFT-1A flight-test data as reported by TRW to the Government. TRW's analysis results were first documented in the 45-day report [10], then in the 60-day report [11, 12], and finally in the revised 60-day report [13]. The data reported in the 60-day report were provided to the POET team [14, 15]. A number of issues were resolved by analyzing these data. In this section the discrepancies among TRW's data analysis reports are discussed first. The results of the POET's analyses are then presented. The emphasis is on discrimination issues.

### 3.1 DISCREPANCIES AMONG DATA ANALYSIS REPORTS

[U] There are two discrepancies regarding discrimination performance among the IFT-1A 45-day report, the 60-day report, and the revised 60-day report. The first discrepancy is that the 1-σ ellipses that represent the data dispersion of the MDL target models change substantially in size, shape, and location from the 45-day to the 60-day report (cf. Figures 4.4.2.5.4.3-3 and 4.4.2.5.4.3-7 of [13]). As a result, the $P_{AT}$ of the MRV increases greatly (cf. Figures 4.4.2.5.4.3-9 and 4.4.2.5.4.3-6 of [13]). The second discrepancy is that the sizes of the 1-σ ellipses decrease a great deal from the 60-day report to the revised 60-day report (cf. Figure 4.4.2.5.4.3-1 of [12] and Figure 4.4.2.5.4.3-1 of [13]), but the $P_{AT}$ does not change (cf. Figure 4.4.2.5.4.3-5 of [12] and Figure 4.4.2.5.4.3-5 of [13]).

[U] TRW's explanation for the second discrepancy was that the ellipses shown in the 60-day report were mislabeled, and that they were actually 2.8-σ (rather than 1-σ) ellipses. The POET team was able to verify this point by calculating the 1-σ ellipses using the 200 MDL feature vectors per target provided by TRW [14].

[U] The explanation that TRW offered for these discrepancies appears to be reasonable. Their decision to turn off the GFA for IFT-3 was also reasonable. At the same time, these problems highlight the issue of robustness of the BLA. As demonstrated in the IFT-1A postmission analyses, to obtain "good" training data is not an easy task even for a domestic flight test.

## 3.2 INDEPENDENT DATA ANALYSIS RESULTS

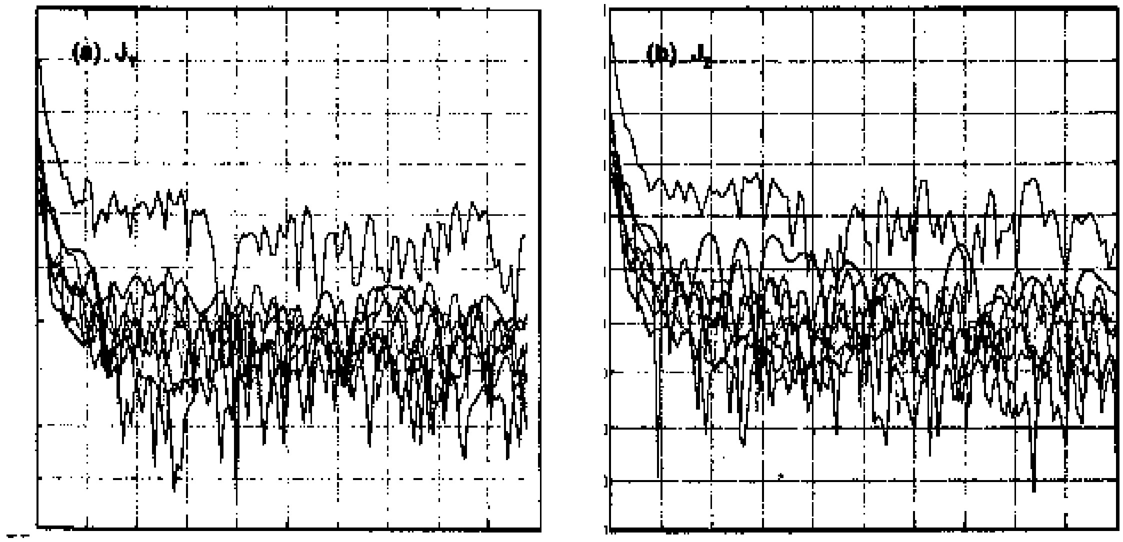
[U] The key results of TRW's IFT-1A discrimination data analysis were presented in Figures 4.4.2.5.4.3-1 through 4.4.2.5.4.3-6 of the revised 60-day report [13]. Those results were checked for accuracy by the POET team using data provided by TRW [14,15]. In addition, the POET team tried to verify the outputs of some important algorithms.

d and used
; results pre

### 3.2.1 Fading Memory Filter

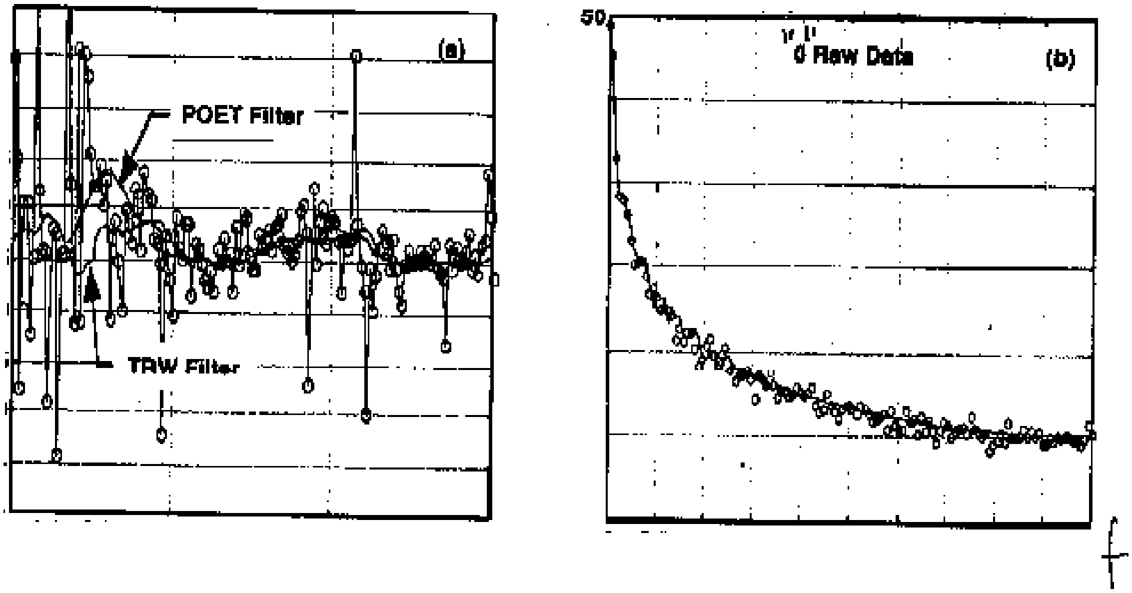
[U] The measured radiant-intensity signals in each of the two color bands, $J1$ and $J2$, are filtered using a fading memory filter before they are combined and used for feature extraction. The filter time constant was set to 1 s for the IFT-1A postmission data analyses. Both raw and filtered radiant intensity signals were provided to the POET team. The consistency of these data with the tuning of the fading memory filter was then analyzed by the POET team.

ectral nine
respec :low
luctio ove
abling




*[U] Figure 2. Power spectral densities of the raw intensity signals of the nine targets observed for IFT-1A for J1 (a) and J2 (b).*

[U] Figure 3a shows the results of applying the fading memory filter and a second-order Butterworth low-pass filter to the J1 sequence of the MRV. The fading memory filter is also a low-pass filter, but it is first-order and the filter coefficients are time-variant. The time constant of the fading memory filter was set to 1 s, and the cutoff frequency of the Butterworth filter was tuned to 0.3 Hz. The outputs of the two filters are essentially identical after the initial 4-s transient response. Further demonstration of the similarity between the two filtered signals is their power spectral density functions, as shown in Figure 3b. Based on these results, the POET team concluded that the fading memory filter was properly designed, implemented, and tuned.

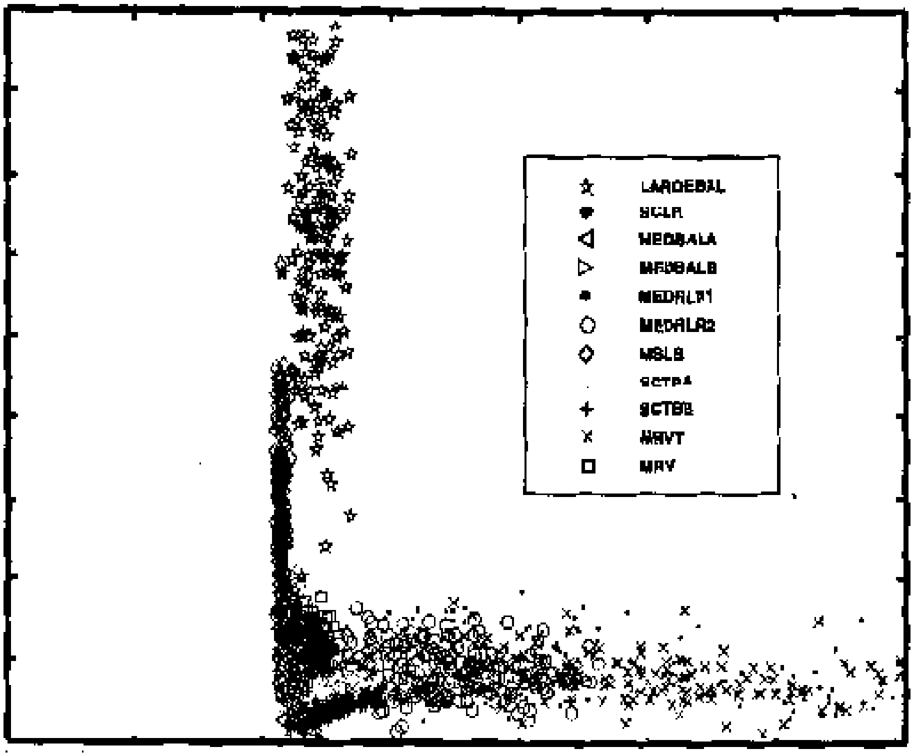



*[U] Figure 3. Outputs in the time domain (a) and frequency domain (b) of the TRW fading memory filter (1 s time constant) and a second-order Butterworth low-pass filter (0.3-Hz cutoff frequency) when applied to the J1 sequence of the MRV.*

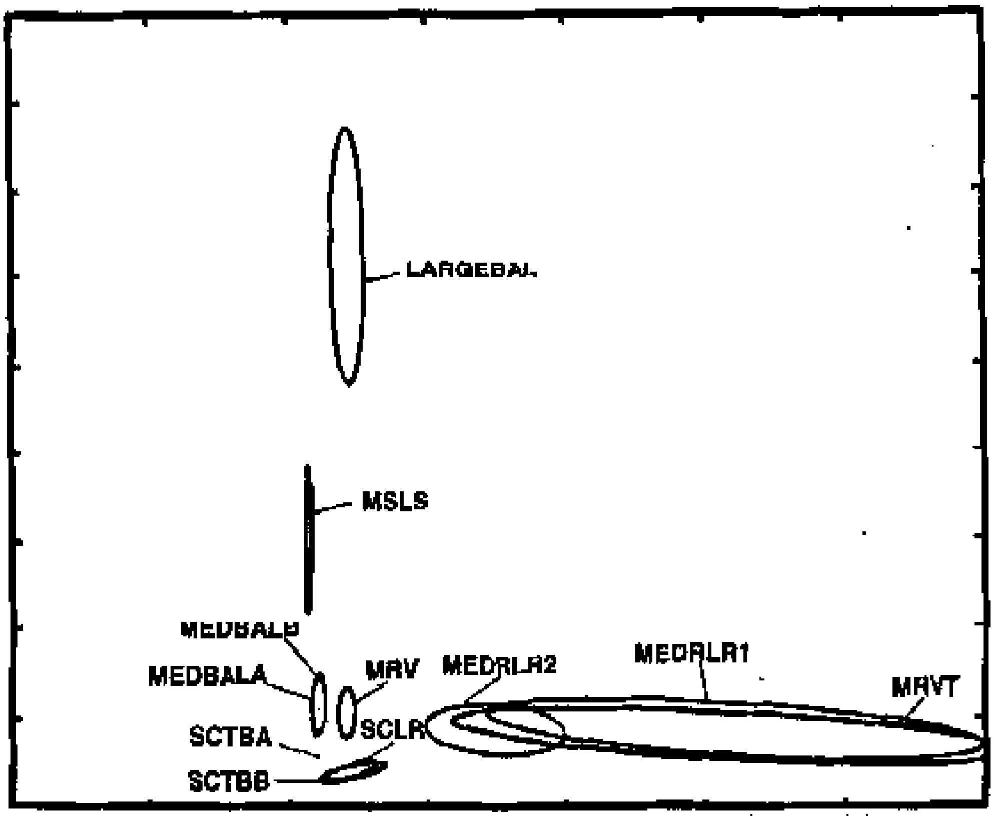
#### 3.2.2 MDL Data Distribution

[U] Figure 4 shows the scatter plots of the feature vectors generated and used by TRW in the 60-day MDL. Figure 4 confirms the scatter plot shown in Figure 4.4.2.5.4.2-1 of the 60-day report [12, 13]. There are a total of 11 target models (or hypotheses) in the database. One target model is constructed for each target in the mission complex, except that the MRV is represented by two models: one spinning (MRV) and one tumbling (MRVT).

[U] Figure 5 shows the 1-σ ellipses computed from 200 feature vectors. (Although a set of 1000 feature vectors per target was actually used to generate the MDL, only 200 of them were provided to the POET team.) The center of each ellipse corresponds to the sample mean of the 200 feature vectors. The probability of samples falling within a 1-σ ellipse is 39.3%. Figure 5 confirms that the ellipses shown in Figure 4.4.2.5.4.3-1 of the 60-day report [12] are 2.8-σ rather than 1-σ ellipses




*(U) Figure 4. Scatter plots of the feature vectors (200 per target) generated and used in the IFT-1A 60-day MDL.*




*[U] Figure 5. One-σ ellipses computed from the feature vectors shown in Figure 4.*

### 3.2.3 Statistical Performance

ve ecifically,
M
nt
(M
e);

[U] The POET team's estimate of the MDL's statistical performance was obtained according to the following procedures:

[U] 1. Calculate for each of the 11 target models the mean vector and the covariance matrix using the given 200 feature vectors.

[U] 2. Assign a prior probability of 0.1 to each of the 11 target models, except the MRV and MRVT. Priors for the MRV and MRVT are set to 3/4 and 1/4 of 0.1, respectively. (This partition is arbitrary.)

[U] 3. An MBQC is constructed using results obtained from (1) and (2).

[U] 4. For each target, test each of the same 200 feature vectors against the classifier designed in (3). For each feature vector under test, calculate the a posteriori probability of being each one of the 11 target models. Assign the feature vector to the target model that has the largest a posteriori probability. Tally the assignments of the 200 feature vectors.

[U] 5. Construct a confusion matrix by combining the tallies obtained in (4) for all targets.

[U] 6. Estimate the probability of leakage ($P_L$), the probability of false alarm ($P_{fa}$), and the probability of error ($P_e$) based on the confusion matrix.

# TABLE 1

## Confusion Matrix for Object-Oriented Discrimination [U]

- **Filter A**
- **200 feature vectors per target model**
- **Identifical training and testing data sets**

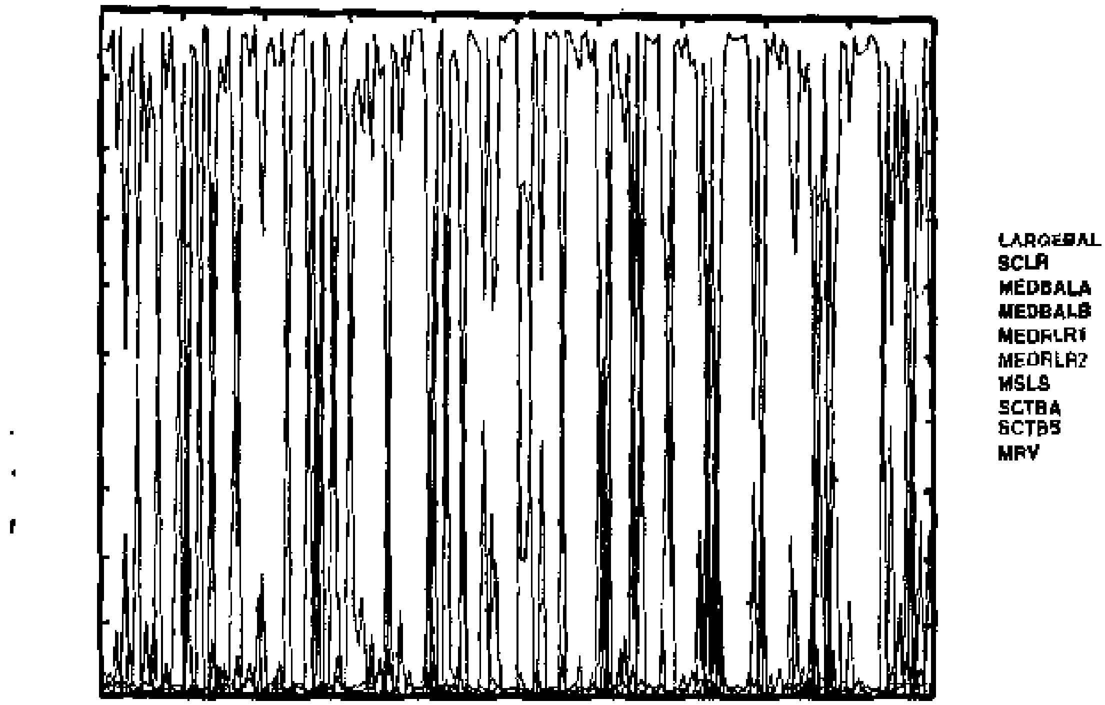
| True Class | Apriori Probability | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (MRV) | 0.1 x 0.75 | 162 | 0 | 0 | 4 | 29 | 0 | 2 | 3 | 0 | 0 | 0 |
| (MALR1) | 0.1 | 0 | 119 | 81 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| (MALR2) | 0.1 | 1 | 11 | 174 | 3 | 7 | 4 | 0 | 0 | 0 | 0 | 0 |
| (SCLR) | 0.1 | 0 | 0 | 3 | 117 | 24 | 56 | 0 | 0 | 0 | 0 | 0 |
| (SCB-A) | 0.1 | 6 | 0 | 6 | 14 | 144 | 20 | 2 | 8 | 0 | 0 | 0 |
| (SCB-B) | 0.1 | 0 | 0 | 7 | 84 | 14 | 95 | 0 | 0 | 0 | 0 | 0 |
| (MB-A) | 0.1 | 7 | 0 | 0 | 0 | 7 | 1 | [illegible] | [illegible] | 0 | [illegible] | 0 |
| (MB-B) | 0.1 | 5 | 0 | 0 | 1 | 8 | 1 | 128 | 55 | 0 | 2 | 0 |
| (LB) | 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 199 | 0 | 0 |
| (MSLS) | 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 2 | 2 | 193 | 0 |
| [illegible] | 0.1 x [illegible] | 4 | [illegible] | [illegible] | 1 | [illegible] | 0 | [illegible] | [illegible] | 0 | 0 | 0 |

[U] The statistical performance shown in Table 1 is based on the so-called object-oriented discrimination, in which the discrimination decision is made object by object. But what is employed by TRW is the so-called complex-oriented discrimination, in which the lethal target is selected by examining the whole threat complex observed and tracked by the EKV sensor. The complex-oriented discrimination is expected to provide better performance than the object-oriented discrimination. Because the IFT-1A data were provided to the POET team on an object-by-object basis, the statistical performance of the complex-oriented discrimination cannot be obtained in this case.

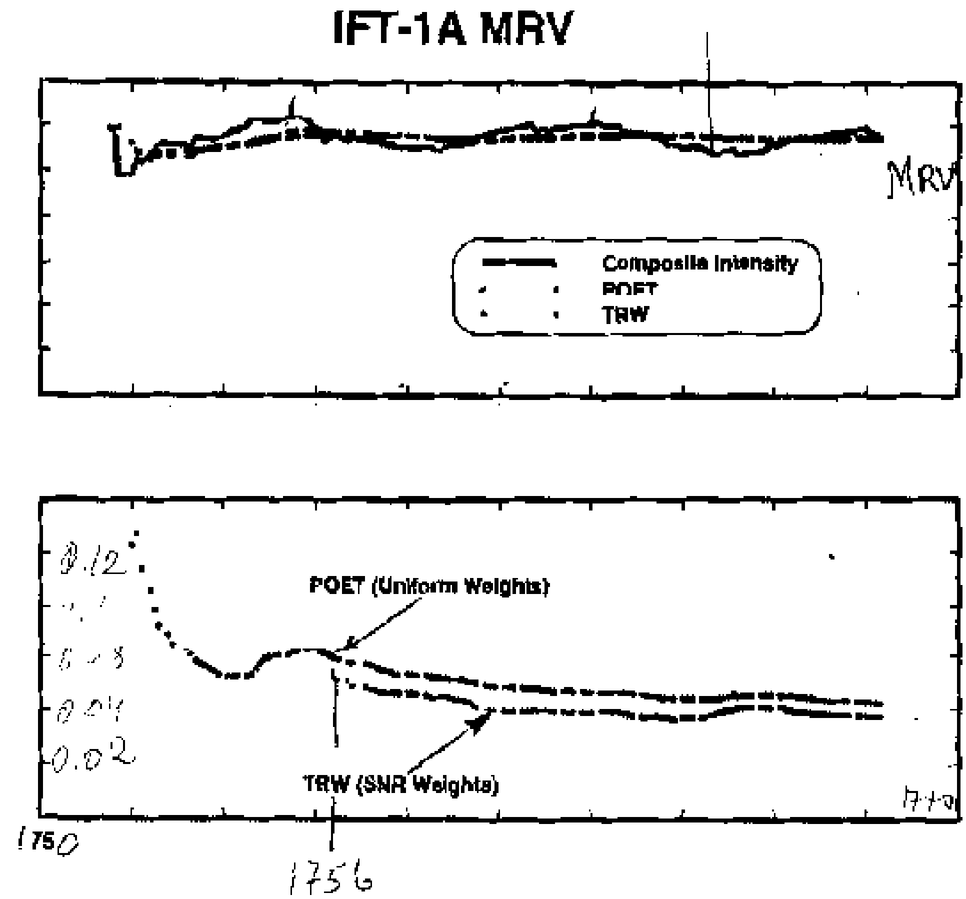
### 3.2.4 IFT-1A Feature Extraction

[U] Discrimination features were extracted recursively at each frame time using radiant intensity measurements of the two color bands. Appropriate weights were assigned to the measurements. For illustration, the sequences of features #1 and #5 of the MRV and the MRLR2 as extracted by TRW postmission are shown in Figures 7 and 8, respectively. It can be seen that these features converge in about one oscillation period; values of feature #5 are zeroed out in the first 5 s after acquisition; and, for unknown reasons, feature data provided by TRW have dropouts.

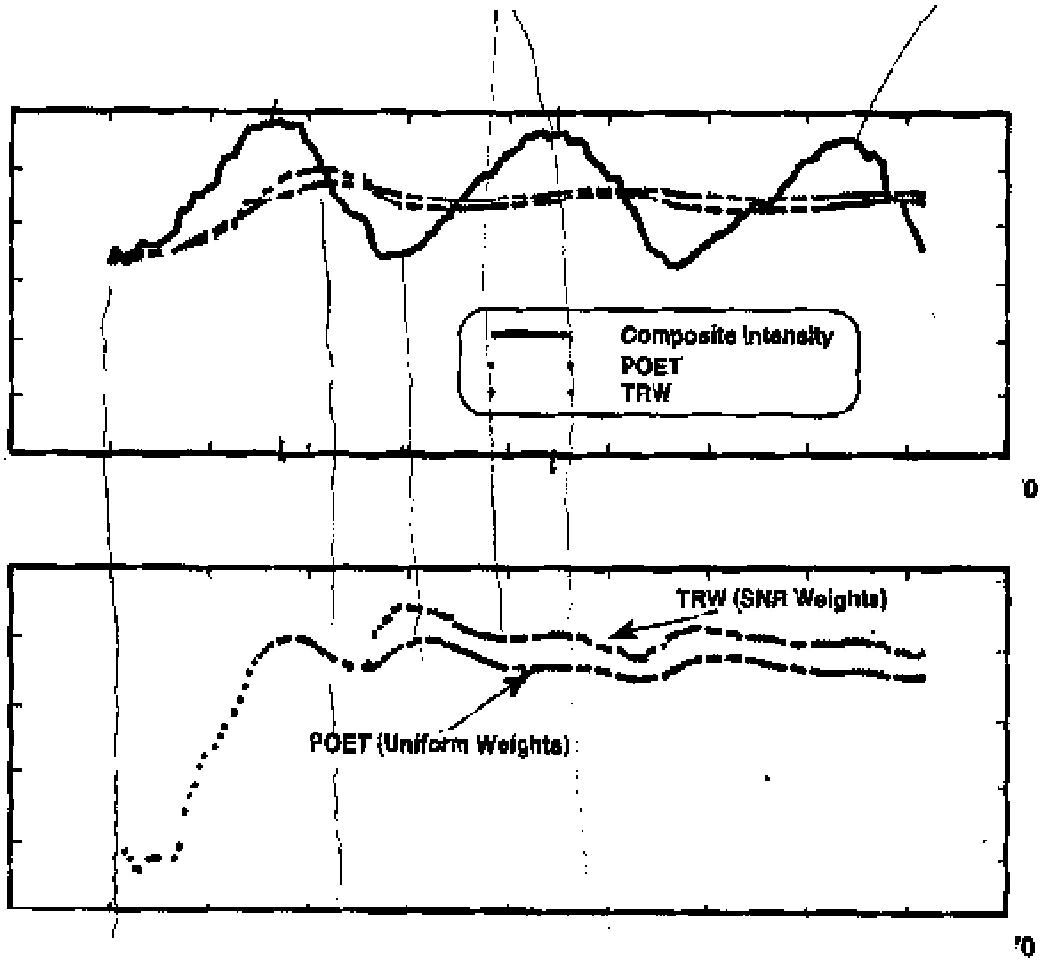
[U] The POET team independently extracted these features using the filtered radiant intensity measurements [14]. Because measurement weights were not included in the data package, uniform weights were assumed. The feature sequences thus extracted are also shown in Figures 7 and 8 for comparison with TRW's results. Obviously, the estimates for feature #1 are in good agreement, and the trends for feature #5 are similar, although the actual estimates are somewhat different. Taking into account the difference in using the measurement weights, the POET team concludes that TRW's feature extraction algorithm has been validated.




*[U] Figure 6. A posteriori probabilities of being each of the target models for each of the 200 MRV feature vectors; they add up to 1.*




*[U] Figure 7. Comparison of TRW's and POET's feature extraction for the MRV.*




*[U] Figure 8. Comparison of TRW's and POET's feature extraction for the MRLR2.*

### 3.2.5 IFT-1A $P_{AT}$ Calculations

[U] According to the TFD architecture, in an interval immediately before the endgame the feature vector derived from each track file is evaluated against the MDL classifier, and the a posteriori probability of being one of the 11 target models as defined in the MDL classifier database is calculated. The a posteriori probability of being classified as an RV is referred to as the $P_{AT}$. A particular target among all targets in the threat complex is selected for intercept based on the $P_{AT}$ and what is called the confidence factor. If the confidence factor is the same for all targets, the target with the largest $P_{AT}$ is selected.

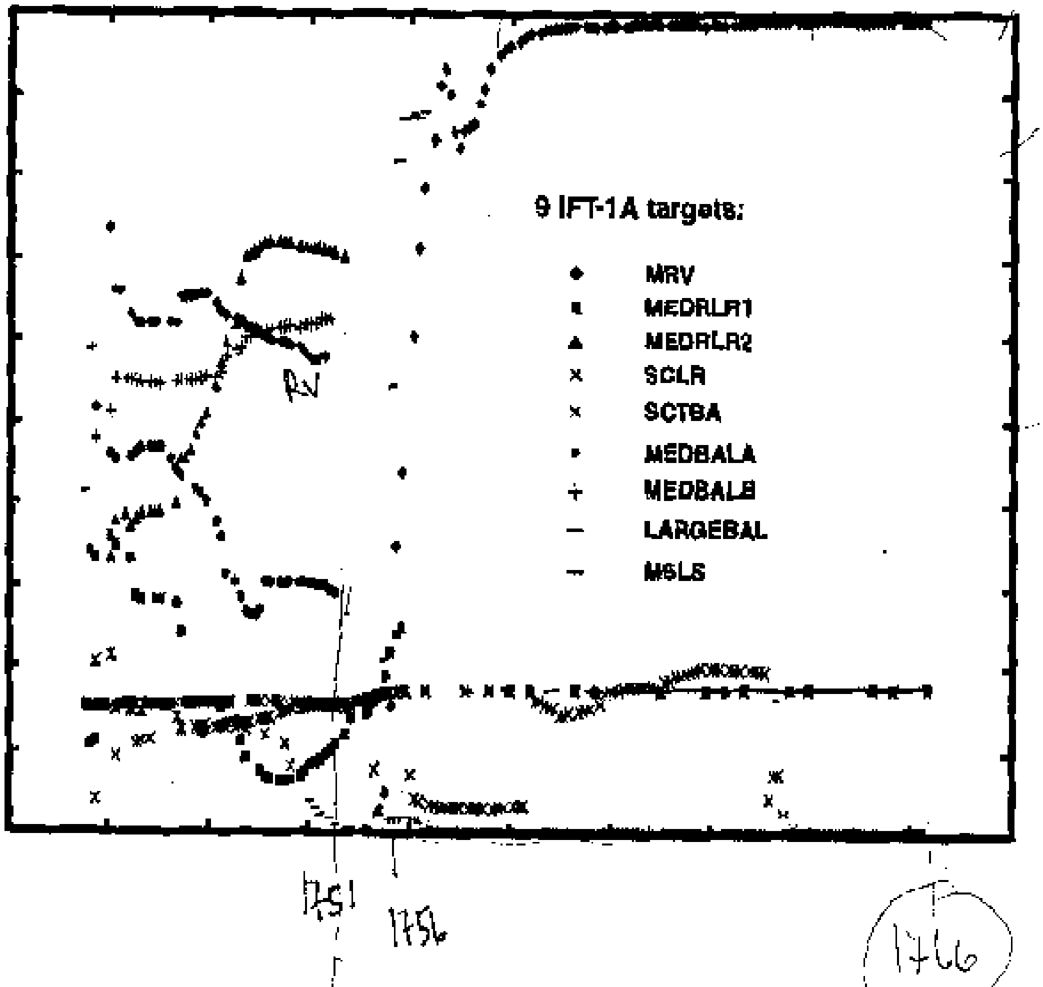
[U] The values of $P_{AT}$ for all targets observed by the EKV in IFT-1A were calculated postmission. Results for filter A were presented in the 60-day report (Figure 4.4.2.5.4.3-5 in [13]) and are repeated here in Figure 9. Only two features (#1 and #2) were used in the first 5 s after acquisition; two additional features (#5 and #7) were used in the next 12 s.

no: 'uplicate
wh #5. The
re # ed later.

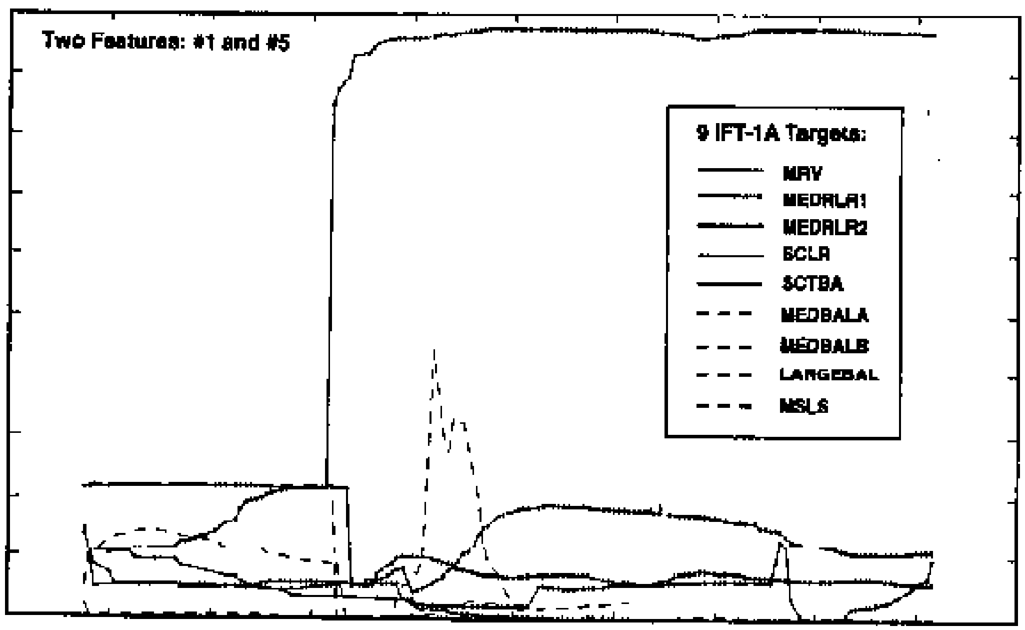
al (: among
. co re, it is
t ta

s an ture #1
fide

ne P ecomes
the can be




*[U] Figure 9. Values of $P_{AT}$ for all targets observed by the EKV as calculated by TRW.*

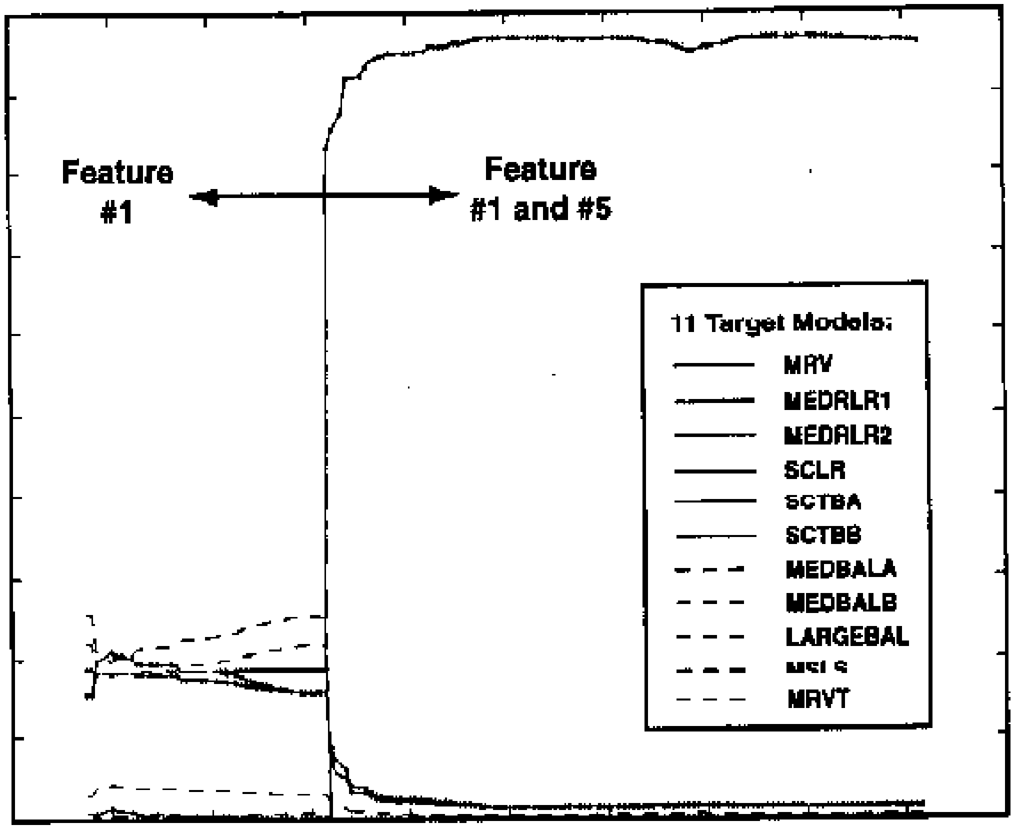



*[U] Figure 10. Values of $P_{AT}$ for all targets observed by the EKV as calculated by POET.*

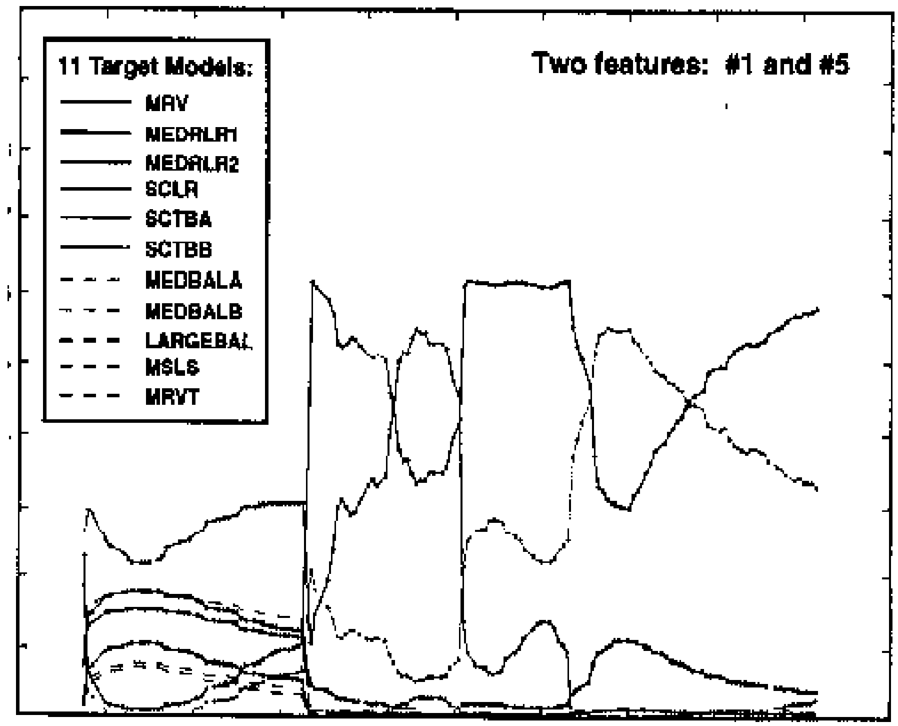
...esult, may

present pro...

[U] Figure 11 shows, for each given MRV feature vector, the a posteriori probability of being one of the 11 target models (hypotheses) defined in the MDL. It can be seen that the a posteriori probability of being the MRV is the largest, except in the first 5 s of the engagement timeline. Similarly, Figure 12 shows the a posteriori probabilities for the first MB (MB-A). In this case, the largest a posteriori probability belongs to target models (SCB-A, SCLR, or SCB-B) other than the MB-A at all times. This is not a surprise because the MB-A was misdeployed. These figures demonstrate once again that the a posteriori probabilities of all target models add up to one for a given target feature vector.




*] Figure 11. A posteriori probabilities of all hypotheses given MRV feature vectors.*




*[U] Figure 12. A posteriori probabilities of all hypotheses given MB-A feature vectors.*

## 4. PROJECTED PERFORMANCE FOR IFT-3

[U] The Boeing/TRW EKV is scheduled to participate in the IFT-3 flight test. It is desirable to project its discrimination performance in such a flight test based on experience gained from IFT-1A. According to TRW's postmission analysis reports [12,13], the BLA can successfully discriminate IFT-1A targets.

[U] According to the GBI PO, five options of target suite are currently under consideration for IFT-3. The target deployment conditions are also yet to be specified. As shown in Figure 13, the last and current option of target suite is identical to that for IFT-1A. This option, which contains a total of ten targets [an MRV, a multiple service launch system (MSLS), and eight penaids], represents the most stressing target suite for the EKV discrimination. On the other hand, discrimination is not an issue at all for the first option because the target suite only includes an MRV and MSLS. Subsets of the eight penaids are inserted in an increasing order of complexity into target suites for options 2, 3, and 4.

[U] Regardless of which target suite will be selected for IFT-3, the targets to be flown have been observed by the EKV sensor in IFT-1A. Signature data collected on these targets will be extremely useful for constructing a representative MDL classifier database for IFT-3. If targets are deployed nominally in IFT-3, it is POET team's assessment that TRW's BLA will successfully select the MRV as the intercept target, even for the most stressing scenario (option 5). However, there are concerns when the deployment is nonnominal. Nonnominal deployment conditions have been observed in previous flight tests; for example, the medium balloons in IFT-1A and IFT-2 were only partially inflated.

[U] TRW studied the impact of anomalous deployments of penaids on discrimination in an IFT-3 scenario [16]. The MDL included a small number of misdeployed penaid models in the database. VICTIM simulation results showed that discrimination performance was largely insensitive to nonnominal deployment of penaids.

a tumbl

. is a ma

**UNCLASSIFIED**

***OPTION 1***

- **Medium Reentry Vehicle (MRV)**

***OPTION 2***

- **Same as Option 1 Except:**
  - **Add 2 Canisterized Small Balloons (CSB) and 1 Large Balloon (LB)**

***OPTION 3***

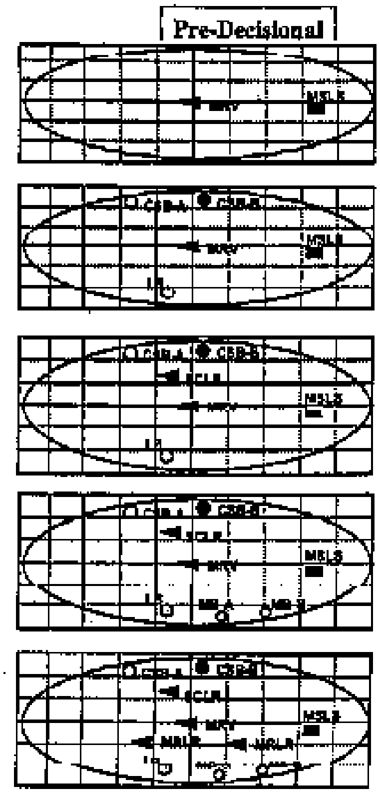
- **Same as Option 2 Except:**
  - **Add Small Canisterized Light Replica (SCLR)**

***OPTION 4***

- **Same as Option 3 Except:**
  - **Add 2 Medium Balloons (MB)**

***Current***

- **Same as Option 4 Except:**
  - **Add 2 Medium Rigid Light Replica (MRLR)**




*[U] Figure 13. Target suite options for IFT-3.*

## 5. REVIEW OF THE EXTENDED KALMAN FILTER

[U] A Kalman filter (KF) was originally developed as a performance enhancement to the TFD function. The KF can provide a set of six additional discrimination features. A detailed study of the KF and its applications to discrimination was conducted in November 1995 [7]. The KF was removed from the TFD in June 1996 because, according to TRW, the EKV's onboard processors could not support its operations in real time.

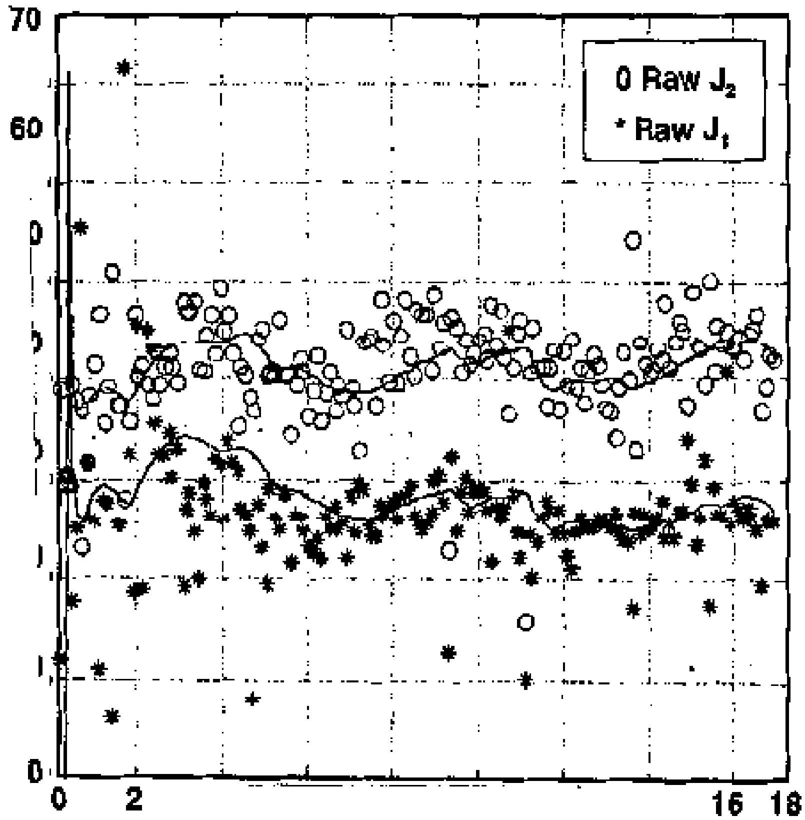
[U] The KF was originally studied by DCIS [19-21]; the POET team also reviewed the KF, even though this task was not included in the team's original Statement of Work. Because of time constraints, the team only looked into the feasibility of using the KF for feature extraction.

The signal model employed in the TRW KF is given as

In such a model the time sequences of $J1$ and $J2$ are described as the sum of two components: a mean and an oscillation. The model includes a state vector of six elements: mean of $J1$ ($\overline{J1}$), mean of $J2$ ($\overline{J2}$), amplitude ($b$), frequency ($\omega$) and phase ($\phi$) of target intensity oscillation, and tilt angle ($\theta$) between the two color bands.

[U] The KF estimates the six states (or parameters) recursively. In essence, it is intended to solve a nonlinear parameter estimation problem. The observability matrix of the signal model is invertible, meaning that all of the six states are observable. Given that the model is nonlinear, an extended Kalman filter (EKF) was needed. The EKF derived its linear model for estimation at each iteration based on prior best estimates of the state vector. For a poor state estimate and its companion covariance matrix, the EKF could easily diverge.

[U] The POET team emulated an EKF according to [7] using MatLab software. This filter was tested using data from simulation as well as from IFT-1A field measurements. In those exercises the initial state vector and covariance were obtained using visual estimation, and the process noise of the filter was set equal to 1% of the state vector. Figure 14 shows the outputs of the EKF when applied to raw $J1$ and $J2$ signals of the IFT-1A MRV. The EKF appears to track the signals reasonably well. Compared with the results obtained from the fading memory filter (Figure 3), the EKF seems to provide a faster track convergence: the convergence time is roughly 1 s for the EKF and 4 s for the fading memory filter.

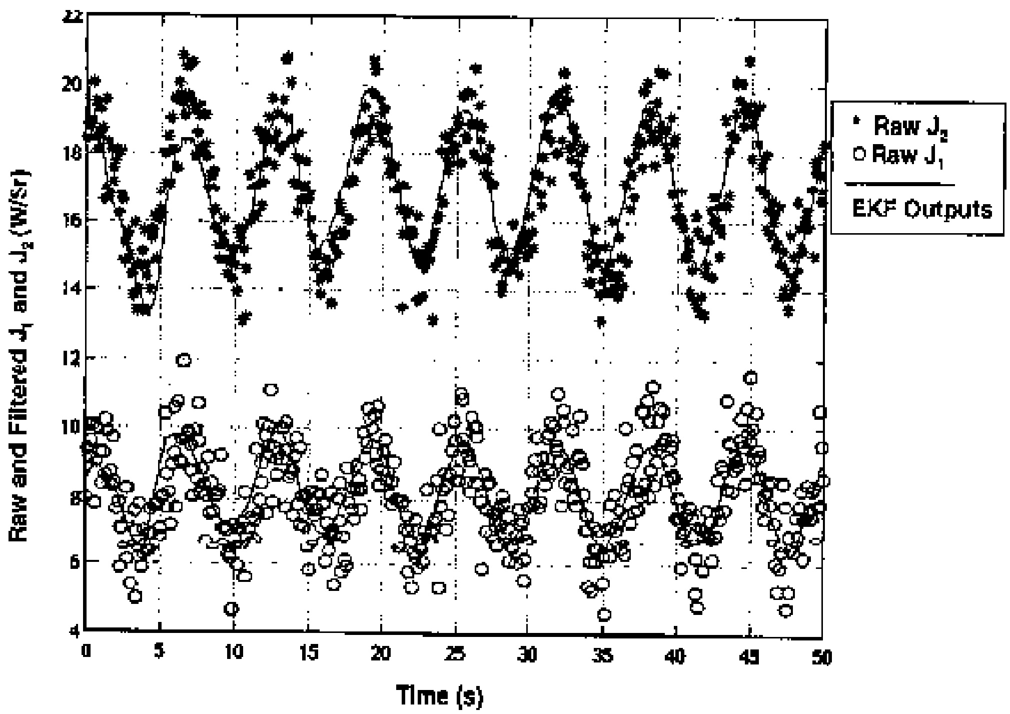



*[U] Figure 14. Outputs of the extended Kalman filter as applied to raw J1 and J2 signals of the IFT-1A MRV.*

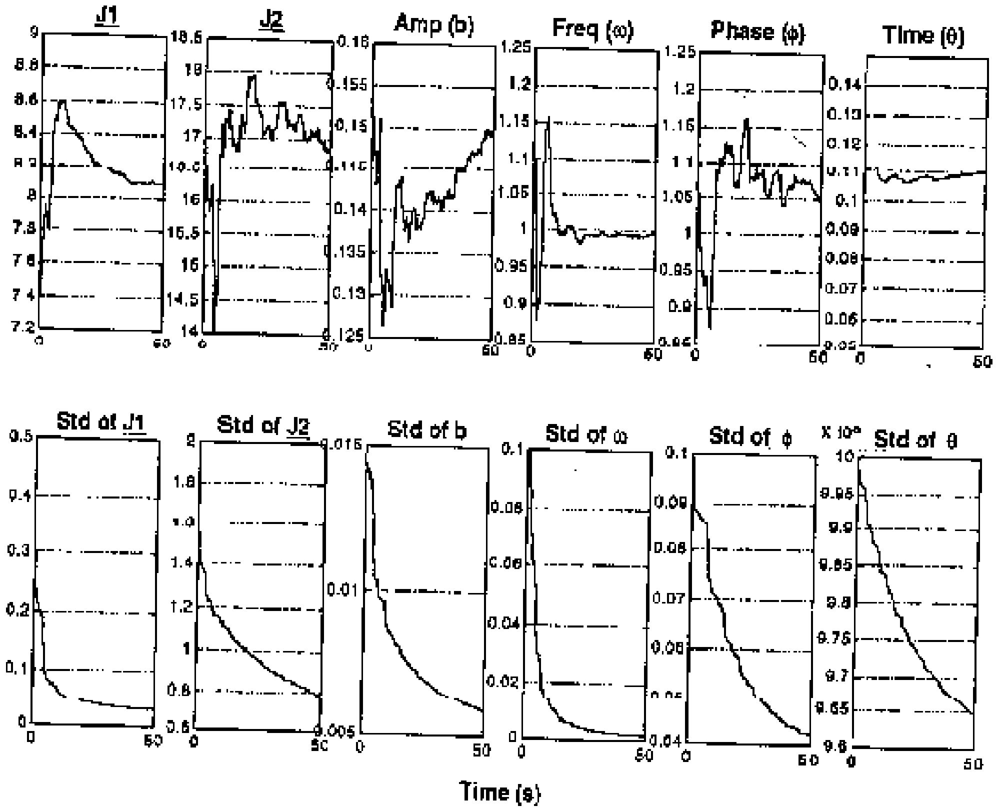
[U] To further demonstrate the convergence of the EKF, Figure 15 shows the results from a computer simulation. J1 and J2 were simulated using a particular set of values for the six states. Each signal was then perturbed using an additive zero-mean white Gaussian noise with standard deviation equal to 10% of the signal. Figure 16 correspondingly shows the converging time histories of the six states and the companion standard deviations. These results show that the EKF can be tuned to behave properly in a noisy environment.

Unclassified Draft

UNCLASSIFIED




*[U] Figure 15. Demonstration of the convergence of the extended Kalman filter using simulated noisy two-band intensity signals: inputs and outputs of the filter.*




*[U] Figure 16. Demonstration of the convergence of the extended Kalman filter using simulated noisy two-band intensity signals: time histories of the state vector convergence.*

obtaine ere por
how to rob e th
nce sen ; are ains
he impl tion ture
ate with nati (φ
ion beca n na ncy
at can o inat e fr
target o ion he (
lone m scin RV

# 6. SUMMARY

[U] A POET team was tasked by BMDO JN/I to study TRW's baseline discrimination algorithms (BLA) for the Boeing EKV. The study included three major tasks: (1) a review of the scientific, mathematical, and engineering principles of the BLA, (2) a review of performance against the integrated flight test IFT-1A data as reported by TRW to the Government, and (3) a projection of the BLA's performance for IFT-3. The scope is basically limited to the BLA and the scenarios where threat types are known. The POET team carried out the tasks mainly by reviewing the technical documents provided by multiple organizations and by analyzing flight-test data.

[U] In fulfillment of the first task, the POET team reviewed the concept, design, implementation, and operational performance analyses of the BLA. The major findings include the following: (1) most of the BLA are well designed and work properly, except for a few, for which refinement or redesign is necessary in order to increase the robustness of the overall discrimination function; (2) the performance of the MBQC may degrade significantly if incorrect prior knowledge about the threat is used in the classifier database; (3) the GFA, which was designed to fill the data gaps induced by low-SNR signals early in the engagement timeline, has not yet been demonstrated to be useful; and (4) the "importance" measure, the basis upon which the intercept target is selected, is the outcome of an ad hoc combination of probability theory and fuzzy logic.

[U] In fulfillment of the second task, the POET team looked at several IFT-1A data analysis reports and analyzed flight-test data to verify the key discrimination results reported by TRW. One of the major discrepancies between the 45-day and the 60-day reports involves the shifts of feature ellipses. The POET team agrees that an improperly constituted GFA was the source for this discrepancy and considers the decision to turn it off for IFT-3 reasonable. POET's IFT-1A data analyses have validated the fading memory filter and the feature extractor. The target-ranking probabilities obtained by TRW and POET are in good agreement, with an exception that during feature transition TRW's results show that the $P_{AT}$ of the MRV drops to almost 0 before increasing to close to 1; the POET's results show a smooth transition. Statistical results indicate that performance of the TRW BLA within the allowable scenarios of IFT-1A is less than perfect.

f the the PC erfo
on in ience ı In II
conta ts depi gets
is PC that th ly se
even cenari al de
hen th ibling ! risk
cluded

# Unclassified Draft

[U] It has been a concern that the discrimination concept and its implementation will not and cannot function to meet TRD requirements; however, it is the POET team's consensus that the majority of the specific technical arguments to support this concern are unwarranted.

[U] Another concern was raised that in the generation and use of the classifier database it is assumed that threat-typing information is available and, for a given threat system, target characteristics, motion parameters, and engagement geometry are known, even though no reliable prior knowledge on threat type can be obtained. This issue is actually outside the scope of the POET study; however, the POET team recognizes that prior knowledge of the threat type can improve discrimination performance if it is correct and degrade performance if it is incorrect. How to provide a correct threat type is an overall NMD system design issue. Also, within the EKV it may be important to structure the discrimination architecture to take advantage of threat-typing information and avoid the potential risk at the same time.

[U] The POET also took an excursion from its statement of work to investigate the KF, which was originally developed to provide enhanced discrimination features and was later removed from the BLA. A KF was emulated and run against both simulated and IFT-1A field data, and the results showed that the filter could be tuned to perform properly, but that a robust initialization procedure has yet to be developed.

## APPENDIX A
## POET STATEMENT OF WORK

### A.1 SCOPE

[U] The POET will provide technical evaluations of TRW's discrimination techniques developed for the BOEING Exoatmospheric Kill Vehicle (EKV). The POET evaluation will address the technical performance of the techniques as indicated by available data and as reported by TRW.

#### A.1.1 Background

[U] The Exoatmospheric Kill Vehicle (EKV) or front end of the interceptor is the most technically challenging aspect of NMD [National Missile Defense] interceptor development. The EKV operates by acquiring a threat cluster from information supplied to it by surveillance sensors through the Battle Management, Command, Control, and Communications (BM/C$^3$) element. The EKV must fuse this information with the scene its on-board seeker sees, select the RV [reentry vehicle], and maneuver to the RV, destroying it by force of impact. If insufficient threat information is available from the rest of the NMD system, the EKV must be able to determine the lethal object through on-board discrimination and target selection. To do this, the EKV must process and fuse on board sensor multiband data, perform target correlation and fusion with data from off board sensors (GBS), perform feature extraction, and classify targets according to lethal versus non-lethal objects.

#### A.1.2 Objective

[U] The overall objective of this effort is to assess the performance of the EKV discrimination algorithms implemented in the BOEING EKV by TRW, and to complete the assessment within about two months. The assessment will evaluate the performance of the algorithms using available data from Integrated Flight Test 1A and other information from TRW, Nichols Research Corporation, the GBI Program Office, Dr. Nira Schwartz, and the Department of Defense Inspector General. The POET study will address two primary issues:

[U] Is the set of discrimination algorithms, software implementation, and associated data developed by TRW for use in the BOEING EKV consistent and correct in its scientific, mathematical and engineering principles?

[U] What performance do the algorithms, software implementation, and associated data provide using the data provided by IFT 1A?

# Unclassified Draft

[U] It may also be possible within the 2-month study duration to estimate the performance the algorithms, implementation and associated data would provide given the expected sensor data for IFT 3. If evaluation of IFT3 potential performance is not possible in the 2-month study duration, such work could be done by extending the study as needed if desired by JN/I.

## A.2 TASKS

[U] The work will be divided into two tasks corresponding to the issues to be addressed. Every effort will be made to complete the evaluation by August 1, 1998, but both tasks will be done in cooperation with TRW, Nichols Research Corporation [NRC], the GBI Program Office, Dr. Nira Schwartz, and the Department of Defense Inspector General. Full and timely cooperation from these individuals is required for completion of the POET study in 2 months as desired. BMDO-JN/I will provide liaison to the above organizations and assure full and timely cooperation. POET will provide an interim progress and status report to JN/I at the end of the first month of activity.

[U] Task 1: Review of discrimination algorithms, software implementation, and associated data developed by TRW for use in the BOEING EKV for consistency and correctness in its scientific, mathematical and engineering principles

[U] This task examines the principles and assumptions on which the algorithms are based and the manner in which they are implemented to determine that they are fundamentally consistent and correct. In particular, the statistical and mathematical assumptions, techniques and applications will be reviewed for accuracy, sufficiency for the purposes intended and appropriateness for the application. The scientific principles used in feature definition, measurement, and characterization will be reviewed for consistency with the best available information on the relevant phenomenology. The engineering approach used to implement the algorithms will be reviewed for consistency with the algorithm foundations and for consistency with current best engineering practice.

[U] Task 2: Review of performance against IFT-1A data

[U] This task examines the performance of the algorithms and implementation when used to discriminate specific data. In particular, this task reviews the performance obtained with the data from IFT-1A and evaluates the performance reported by TRW to the government for accuracy.

## A.3 DELIVERABLES

[U] The deliverable will be an interim and final report describing the conclusions reached and delivered to JN/I. The final report will be produced and delivered to JN/I on completion of the study.

[U] Every effort will be made to complete the evaluation by August 1, 1998. However, as indicated above, the timely completion of the study requires the cooperation of TRW, Nichols Research Corporation, GBI Program Office, Dr. Nira Schwartz, and the Department of Defense Inspector General in providing data and information necessary for the POET evaluation. If it becomes apparent that completion by August 1, 1998 is not possible, the POET will inform JN/I promptly and indicate the data or information required for completion.

### A.5 PROGRAM MANAGEMENT

[U] JN/I will be responsible for overall task management with Mr. Keith Englander the point of contact. The POET task leader will be Dr. Ming-Jer Tsai. Dr. Frank Handler will be the POET task manager. Release of results or information generated by POET analysis will be only to JN/I unless otherwise specifically directed by JN/I.

### A.6 LEVEL OF EFFORT

Evaluation is intended to be for 2 months. During this time it is anticipated as many as 4 POET personnel will be participating at any one time, using approximately 4 staff months over the intensive period.

# Unclassified Draft

## APPENDIX B
## RESPONSES TO ISSUES RAISED BY DCIS

### B.1 REVIEW OF ISSUES

[U] In the course of its review of TRW's EKV discrimination work, DCIS has asserted that TRW's discrimination concept and its implementation will not and cannot function to meet the specifications of the *Technical Requirements Document* (TRD). Responding to these issues was not included in the original study plan (see Appendix A); the POET team was requested to do so after the study was completed by Mr. Keith Englander of JN/I.

[U] The most recent concerns are documented in [19-21]. Given that many of the concerns raised in these documents are similar and that [21] is the most complete statement of these concerns, the POET team believes it is appropriate to respond to the issues raised in [21]. Most of these issues have already been discussed in general terms in the main body of this report. In this section, the most critical issues are grouped into six points; the POET team's findings are then presented. Point-by-point responses are found in Section B.2.

#### B.1.1 Critical Issues

[U] Many issues (a total of 53 bullets) were raised in the 16-page briefing document [21]. The most critical may be stated as follows:

[U] 1. In their operational performance analysis TRW took a narrow, limited approach in assuming that threat-typing information would be available and that, for a given threat system, target characteristics, motion parameters, and engagement geometry would be known. In actuality, reliable prior knowledge on threat type cannot be obtained, and the use of such knowledge for discrimination is not permitted by the TRD.

[U] 2. Because improper feature extraction methods were employed, many anomalies were observed, including that features were correlated, that some baseline features (#3 and #6 in particular) were never used, that discrimination performance degraded with the addition of features, and that the best feature subset selected did not produce the best performance.

[U] 3. The concept of "probability of ranking" used for selecting an intercept target is improper. When applied to the IFT-1A postmission analyses, many problems were identified: normalizing the $P_{AT}$ to 1 resulted in noncompliance performance with the TRD; feature ellipses were not drawn in terms of 3-σ uncertainty; using the "distances to the centers of the assigned ellipses" for target ranking was misleading; and large overlapping area between ellipses indicated poor performance.

[U] 4. During the IFT-1A postmission analyses, the preflight feature distribution ellipses were shifted between the 45-day report and the 60-day report in order to increase the discrimination performance. The explanation that the gap-filling algorithm (GFA) is the reason for the shifts is not true.

[U] 5. The optical signatures contain spikes and dips due to natural physical characteristics of the targets. NRC test results show that when the signatures possess spikes no target is selected in many cases and there are a large number of false alarms; when dips occur the baseline algorithms (BLA) collapse.

[C] [redacted] (KF), which [redacted] at the suggestion [redacted]

Ex [redacted] use [redacted] ory and pr [redacted] r.

### B.1.2 POET Team's Findings

[U] It is the consensus of the POET team that the majority of these concerns lack technical justification. Overall, however, these concerns reflect the widely recognized complexity of the discrimination task. The POET team sees no evidence that suggests any obvious noncompliance of the BLA with the TRD.

[U] The POET team's findings concerning the six major issues are summarized in the following paragraphs. The team based these findings on careful review of all the available documents, independent analyses of flight-test data, and the experiences of the members of the POET team.

[U] (1) The TRD contains no requirements relative to single-threat system assignment; however, it is well recognized that prior knowledge of the threat type can improve discrimination performance if it is correct and can degrade the performance if it is incorrect. How to provide a correct threat type is an NMD system-design issue. Nonetheless, the EKV discrimination architecture can be structured to take advantage of threat-typing information and avoid risk at the same time. For example, at the expense of additional complexity, the performance degradation due to threat-typing error can be minimized by running in parallel multiple sets of the mission data load (MDL) and then accepting results from the particular MDL that produces the best system-level discrimination performance.

[U] (2) Most of the feature extraction algorithms described in [5, 16] appear to be sound in both concept and implementation, the GFA may be the only exception.

[S
in
in
n
I
tempe... e featu... ghly co...

[U] The fact that increasing the number of features after first improving performance eventually causes the performance to decrease is curious but not unprecedented. Possible explanations include (a) insufficient training data, (b) non-Gaussian feature distributions, (c) highly correlated features, and (d) mismatches between training and testing of the multimodal Bayesian quadratic classifier (MBQC).

[U] The best feature subset selected by BEST may not always produce the highest $P_{select}$ because BEST makes the feature selection based on $P_{ID}$, which is calculated numerically, and VICTIM obtains $P_{select}$ based on Monte Carlo runs. It may be a good idea that BEST also uses Monte Carlo runs to select features.

[U] (3) Clearly, the concept of "probability of ranking" was misunderstood and misinterpreted. Good descriptions of this concept can be found in [5,16].

[U] $P_{AT}$ is a probabilistic measure calculated according to the Bayes' rule for each target in the observed threat complex. For a target under track, $P_{AT}$ is calculated against all target models as defined in the MDL. It is actually the a posteriori probability of the target being an RV among all target models. $P_{AT}$ does not need to be normalized to 1 because the summation of $P_{AT}$ over all targets is not equal to 1. For example, given two target tracks with $P_{AT}$ of 0.9 and 0.5, the track with $P_{AT} = 0.9$ will be selected as the RV although the two $P_{AT}$ values do not add up to 1.

[U] To calculate the $P_{AT}$ of a target, the covariance matrix, which describes the distribution or dispersion of feature vectors of the target, is used. In a two-dimensional feature space, the covariance matrix is typically represented as an ellipsis of a certain size and orientation. It does not matter whether the ellipse is drawn for 3 or 1 $\sigma$ or for any other value, as long as it is correctly identified.

[U] In TRW's design the "distances to the centers of the assigned ellipses" are never used for target selection. An intercept target is selected using what is called the "importance" measure, which is calculated by combining $P_{AT}$ and a confidence factor. The Mahalanobis distance is used for bulk and coarse discriminations, which is the distance between a given feature vector and the center (or the feature mean vector) of a target feature distribution normalized by the associated covariance.

[U] TRW's fine (i.e., precision) discrimination function implicitly includes what is called object-oriented discrimination and complex-oriented discrimination. The former computes the a posteriori probability of belonging to each of the target models as defined in the MDL. The latter examines the a posteriori probabilities of being an RV (i.e., $P_{AT}$) for all targets in the observed threat complex and then selects the target with the largest $P_{AT}$ as the intercept target. When the RV and non-RV ellipses overlap heavily, the object-oriented discrimination function alone usually produces poor performance, but this performance can be improved significantly by the follow-on complex-oriented discrimination. This means that $P_{select}$ can be higher than what is perceived by inspecting the feature ellipses.

[U] (4) The feature ellipses changed in location, size, and shape from the 45-day report to the 60-day report. As a result, the $P_{AT}$ of the medium-size reentry vehicle (MRV) increased significantly. TRW's explanation, which hinges on the GFA, seems to be reasonable. In either case the MRV was selected as the lethal target.

8

ace
ully
the
tion

ic

[U] TRW admits that additional work on the GFA is necessary, and they appear to be proceeding with caution regarding the use of a GFA. The data to support its use have not yet been presented to the POET team.

[U] (5) The ACPS part of the CSS-4 and the reflecting appendages/threat wings as cited by DCIS do not physically produce spikes on optical signature data. The "spikes" referred to here are artifacts of an incorrect sensor noise model and improper logic in parts of VICTIM, which are not TRW's responsibility.

(standard deviation of the composite intensity normalized by the mean) and #7 (standard deviation of temperature). It is incorrect to assert that because dips occasioned by an improperly constituted GFA cause a failure of the system dips of some other origin, i.e., aspect-angle variation, will cause a similar result.

[S] ... ial F- ... and ...

I ... s and I

[U] TRW's report has shown that the extended Kalman filter (EKF) can be tuned to converge [7]. Independent simulation by the POET team confirmed these results. But neither study adequately addressed the issue of sensitivity to noise perturbation. In noisy conditions the EKF is known to be sensitive to initialization of the state vector and covariance matrix. A robust initialization procedure is needed to make this a generally useful algorithm. TRW has acknowledged that the KF, EKF, and Iterative EKF need additional development before they are ready to be implemented, but there is no contractual requirement to implement any of them.

### B.2 POINT BY POINT DISCUSSION

[U] This section presents the POET team's point-by-point responses to the issues by DCIS. At the request of JN/I, the issues given in [21] are presented page-by-page and bullet-by-bullet; the POET team's response to each bullet is inserted in red immediately under the original text.

**Page 1 - Non-Compliance without exception of Baseline Discrimination technology with the TRD**

[U] Title page.

**Page 2 - Kalman Filter (KF) family of Feature Extractors failed to comply without exception to the TRD/SOW**

- Jameison [ISEG] committee recommended spectrum analysis.

[U] There is no contractual requirement that the KF function should be included in the discrimination process.

- **Kalman filter was to extract additional large number of features.**
  - Kalman Filter April 1996 report pages 11 (Jameison) and page 120 (feature list).

[S] extract ...ency and
ty osc
he ha

repl small

- **Statement of Work (SOW) requires system design robustness to future modifications. KF and baseline discrimination failed to comply with the TRD or SOW.**

[U] There is no evidence that the system design is not robust enough to accommodate a well-functioning KF if that should be required in the future.

- **The KF, extended Kalman filter (EKF), and iterative extended Kalman filter (IEKF) all failed NRC test on November 1996.**

[U] NRC's test results are not available for review. But TRW's report [7] has shown that the EKF can be tuned to converge. Independent simulation by the POET team confirmed these results (see Sections 5 and B.1).

- **Entire KF family oscillated/collapsed under small percent of noise perturbation.**

[U] Neither TRW's study [7] nor POET's independent simulation adequately address the issue of sensitivity to noise perturbation. But the EKF is known to be sensitive in noisy conditions to the initialization of the state vector and covariance matrix. What is needed is a robust initialization procedure.

- **In November 1996 the KF was removed from the EKV**

[U] TRW has acknowledged that the KF, EKF, and IEKF need additional development before they are ready to be implemented, but there is no contractual requirement to implement any of them.

- **TRW excused the KF removal due to the excess memory it required. However, the KF failed with all the memory it needed to operate properly.**

[U] The requirement is for TRW to meet the $P_{select}$ requirements levied by Boeing in the *Prime Item Development* (PID) document. They can do this with any combination of algorithms that succeeds. TRW claimed that the baseline discrimination alone is adequate for the near-term threat specified by the TRD and the IFT-3 mission.

**Page 3 - Non-Compliance without Exception with the TRD**

- It is impossible to predict/assign, just prior to the attack, a single threat system.
  - Sources:
  - TTSSR 1995 and 1996
  - 45-day, 60-day report and Addendum 1 to the 60-day report
  - NRC test reports, Final and prior dated 12/2/98
  - TRD
  - KF April 1996 test report.

[U] The TRD specifies no requirements relative to single-threat system assignment, and this issue is outside the scope of the POET study. The POET's study did not evaluate to what extent the BLA may be used to meet the general requirements described in the TRD, beyond the ability to determine which object within a threat cloud is lethal (given that the threat type is known); however, the POET team recognizes that the concept of threat typing has significant impact on EKV discrimination performance and that this impact should be examined carefully in the context of overall NMD architecture planning and threat assessment.

[U] It is well recognized that prior knowledge of threat type can improve discrimination performance if it is correct and can degrade performance if it is incorrect. How to provide a correct threat type is an NMD system design issue. It is believed that the EKV discrimination architecture can be structured to take advantage of threat-typing information and avoid risk at the same time.

- Even if single threat system was assigned just prior to the attack then:
  - The Single Assigned Threat System Parameters (SATSP) needed for the Baseline discrimination are unavailable, and the probability of predicting their accuracy does not comply without exception to the TRD.
- Even if SATSP were available just prior to the attack then:
  - Feature extraction 'rules' are false. Dependent Features are falsely selected to bias ellipsis location to improve P-select.
  - Increasing number of features causes P-select decreases (failed TRD requirements).

[U] The fact that increasing the number of features after (as expected) first improving performance eventually causes $P_{select}$ to decrease is curious and warrants further investigation, but it is not a basis for claiming failure to meet the specifications of the TRD. The TRD (actually the

# Unclassified Draft

*Prime Item Development* document) requires a certain performance (expressed as $P_{select}$ ); it is of course sensible to optimize performance and not insist that all features must be used regardless of their effect on performance.

- Ellipsis location and feature extracted values are subject to unpredicted bias created by signature GAP to dramatically influence P-select (fails TRD requirements).

[U] The intensity sequences generated by MODEL for training purposes exhibited data gaps in the beginning of the engagement timeline when the SNR is low. TRW employed the GFA to fill the data gaps, which biased and rotated the feature distribution ellipses. In the IFT-1A 60-day report [12], TRW effectively eliminated the GFA problems by delaying the processing of intensity data until the SNR was sufficient.

[U] An improperly constituted GFA clearly would have a negative effect on discrimination, but this does not mean that a failure to meet the TRD requirements has occurred.

- Probability of Ranking does not comply with the SSPK as shown in TRD.

[U] Ranking according to probability of lethality is a sound statistical concept. These probabilities (referred to as $P_{AT}$ by TRW) for all targets in the threat complex do not add up to be unity. There are no reasons to suggest any conflict in principle between this ranking concept and the requirement to meet an overall SSPK.

**Page 4 - "Threat Typing Sensitivity Study Report" (TTSSR): A Narrow Limited Concept that does not comply without exception with the TRD**

- TTSSR 1995 page 5 (U): The battle manager can use surveillance data to provide threat typing information to the EKV system at launch. Figure 3.1 shows how detailed surveillance data can be used to narrow the threat system from any of the thirteen systems down to potentially a single system. TRW provided no validation of their selection.
- Narrowing the threat system down just Prior to the attack based on fuel type and launch location is grossly inaccurate. The TRD does not assign a threat type system to booster fuel type and launch location for a specific country (TTSSR 1995, page 12, TRD page A-72, 10).
  -This is incorrect since:
  - The enemy is known to switch Land and Sea missiles and their payloads (GBI on 12/2/98).
  - Just prior to the attack single threat system out of the entire threat library for a specific country and along the TRD can be not assigned.
  - The assigned threat system and all the RVAO[d]/[t]/[s-d]/[Dips]/[Gaps] and what

# Unclassified Draft

is needed for TRW to perform and to assign a target is not available (Page 123, addendum 60-day Report).

- Intelligence provides updated data every 2-6 months, not prior to the attack (GBI information supplied on 12/2/98).

- Intelligence/surveillance community assigned a threat system library per country, not single threat system target just prior to the attack.

- The accuracy/probability of the prediction is 1/N where N is the number of possibilities in the TRD library and it does not comply with the TRD (SSPK).

- What are unpredictable war conditions? (TRD page A-2; second para).

[U] The TRD gives no requirements relative to threat typing sensitivity, and the issue is outside the scope of this study.

[U] At the expense of additional complexity, the degradation of discrimination performance due to threat typing error may be minimized by running in parallel multiple sets of the MDL and then accepting results from the particular MDL that produces the best complexwide discrimination performance.

**Page 5 - TRD versus TRW Discrimination Concept**

[U] This page is a cartoon that illustrates the concept of using threat-typing information.

**Page 6 - Improper Feature extraction method that does not comply without exception to the TRD/SOW**

- BEST software malfunction is not the reason for the non-compliance of the Baseline technology with the TRD but improper/lack of feature extraction rules are the problem.

  - NRC hand extraction features/values assure high P-select and eliminate this excuse.

[U] The best feature subset selected by BEST may not always produce the highest $P_{select}$, because BEST makes its feature selection based on $P_{ID}$, which is calculated using numerical integration. VICTIM obtains $P_{select}$ based on Monte Carlo runs. It may be a good idea that BEST also use Monte Carlo runs to select features.

- Discrimination performance wrongly degrades with the addition of features.

  - Final test report by NRC, page 16, indicates that without feature #5 results improved "substantially" contradicts the TRD on page 6, 10 that request the use of all available data.

[U] Page 6 of the TRD states that the "KV shall integrate all available data to optimize target selection." Clearly the relevant phrase here is "to optimize target selection." The fact that

# Unclassified Draft

discrimination performance degrades with the addition of more features is curious and warrants further investigation, but the most important conclusion to draw is that such features should not be used. It is not reasonable to use all available data if that leads to degraded performance.

[U] The improvement in performance noted in the NRC test report [17] (which is associated with leaving out feature #5) is possibly related to the improper use of the GFA, as stated by NRC [17] and confirmed by TRW [16].

- Features #3 and #6 are never used since they cause the baseline system to collapse:
  - NRC Test Report dated 12/2/97 on page 28 of test and NRC presentation on 12/2/97.
  - While feature #3 is essential for the Pre-flight single threat system assignment and the ability to draw pre-flight ellipsis, however, there is never an in-flight match against this feature.
- Addendum 60-day and 60-day reports page 123, 124, and 125.
  - Just prior to the attack, neither the radar nor any one else can supply feature #3 values.
  - Even if pre-flight Feature #3 are not of any use, it is essential to match this feature onboard.
  - EKV measurements of feature #3 are not of any use. Which incorrectly leaves the essential pre-flight feature vector not in common with the in-flight feature vector.

[U] No substantial argument is being made here. Although feature #3 must be known or calculated as a part of the generation of an MDL, the fact that it is not then particularly useful in discrimination is largely irrelevant. The optimization process should specify the most useful features to employ in performing discrimination. The fact that the resulting set of features may not conform to intuitive guesses simply means that intuition is not a good guide in this case. It matters not at all that this parameter is needed to construct an MDL but then is not useful in actually doing discrimination.

[U] It should also be noted that features # 2 and # 3 are highly correlated. If feature #2 is selected (as in the MDL for IFT-1A) feature #3 will very likely be left out.

- Features are selected from the "dependent" group of features to falsely relocate ellipsis centers and falsely improves discrimination performance. This is an indication that the discrimination is subject to bias. NRC final test dated 12/2/97 page 18.

[U] The eight baseline features are indeed dependent because they are derived from just two independent sensor measurement sequences. This is a limitation of the EKV sensor measurements rather than intentional falsification of data.

**Page 7 - Improper Gap filling methods fail to comply without exception to the TRD/SOW**

# Unclassified Draft

- GAP filling algorithms are subject to bias and influence the feature/values extracted for discrimination.
  - NRC Final Test Report dated 12/2/97 on page 16, 17, 18.

[U] The GFA as implemented for the IFT-1A postmission data analysis in fact caused two problems: bias shift of ellipses and rotation of ellipses due to correlation between features.

- GAP sizes and locations are unpredictable and unpredictably bias the extracted feature values, which result in unpredictable non-compliance with the TRD.

[U] The improperly constituted GFA did affect the results significantly, as shown in the IFT-1A 45-day report [10]. But as shown in the 60-day report [12], the ill effects of the GFA can be overcome by delaying the feature extraction process until the SNR is high enough to avoid data gaps. NRC reported that when it modified the GFA to achieve the intended benefit of properly implemented GFA, performance did indeed improve.

- GAP filling algorithms can not be embedded into already predicted pre-flight signatures. Objects/GAP/ellipsis correlation are unknown and to be found and not assumed. But GAP filling was applied to pre-flight ellipsis and caused them to randomly shift location.
  - The addendum 60-day report page 125 dated April 1, 1998 indicates no gap filling was involved (only in the 45-day report).

[U] Actually, a new MDL was generated for the IFT-1A 60-day data analysis based on track intervals later and shorter than those used in the 45-day analysis.

- The drastic shift of pre-flight ellipsis center with the induced Gaps is an indication of the poor performance and the non-robust nature of the discrimination method to bias.

[U] The need to adjust the MDL from the 45-day analysis to the 60-day analysis arose mainly because the GFA was not properly implemented. The problem reflects the fact the discrimination algorithms are still under development and testing.

- While NRC stated that the feature extracted values are subject to Gap filling, Mr. Irv Balin in his letter to GBI on March 5, 1998 indicated that the new gap filling algorithm was integrated into the VICTIM. The algorithm is "statistically sound."

[U] The ill effects of the GFA can be avoided by properly weighing the data points.

- The Addendum 60-day report dated April 1, 1998 claims the need for Gap evaluation.

# Unclassified Draft

- **But there is no method to fill the Gaps so features extract values will not be subject to bias, and therefore the complete concept is wrong. (statistical Analysis with missing data).**

[U] To say that the GFA induces bias in the data and therefore that the "complete concept is wrong" ignores the fact that, even in the absence of gap filling, there is naturally and necessarily a bias in the data caused by the missing data below the detection threshold. While it is clear that an improperly constituted GFA would negatively impact the discrimination process, it is also quite possible that a properly constituted GFA will have substantial benefit to the process. Whether or not the use of a GFA will be beneficial can only be determined after substantial simulation and testing. It is not possible to say at this time whether a GFA will be beneficial or not. TRW appears to be proceeding with appropriate caution regarding the use of a GFA.

**Page 8. Improper "probability of ranking" method that fails to comply without exception with TRD/SOW**

- Normalizing the Probability of Assigned Target (PAT) to value one, as requested by the TRD page 10, results in non-compliance performance with the TRD.

[U] Nowhere in the TRD is the request is made to normalize $P_{AT}$ to one.

[U] $P_{AT}$ is a probabilistic measure calculated according to Bayes' rule for each target in the observed threat complex. It does not need to be normalized to 1 because the summation of $P_{AT}$ over all targets is not equal to 1. For example, given two target tracks with $P_{AT}$ of 0.9 and 0.5, the track with the $P_{AT}$ of 0.9 will be selected as the RV, even though the two $P_{AT}$ values do not add up to 1.

- "Probability of ranking" is incorrectly performed against one single pre-flight target assignment and not against the entire attacking country intelligence threat library.

[U] For a target under track, $P_{AT}$ is calculated against all target models as defined in the MDL. $P_{AT}$ is the a posteriori probability of the target being an RV among all target models. The MDL can be generated using target models from all threat systems or from a single system, depending on how the threat-typing information is used.

- P-select can not supersede the BEST calculated Prime Item Development (PID) value of P-select.
  - Addendum 60-day report page 123.
  - NRC test report dated 12/2/97 page 35.

[U] There appears to be some confusion here between the *Prime Item Development* (PID) specification and the probability of correct target identification, $P_{ID}$. The PID specifies the contractually required value of $P_{select}$ that the TRW discrimination functions must meet.

# Unclassified Draft

[U] It is clear from the description of $P_{ID}$ provided in [7, 8, 16] that there is no a priori reason why the value of $P_{select}$ cannot be larger or smaller than $P_{ID}$, which is the product of the off-line use of the algorithm BEST to find the best features to use in discrimination for a particular target suite, engagement, and illumination condition.

- **The ellipses should be drawn for three sigma since Boeing requested that the PID and P-select with not be less than a three sigma value.**
  **- NRC test report dated 12/2/97 page 35.**

[U] It makes no difference whether the ellipses are drawn for 3 σ, 1 σ, or for any other value, as long as they are correctly identified.

- **Large overlapping ellipsis areas as shown in the TRW 45-day, 60-day and addendum 1 to 60-day indicate non-compliance with the TRD (SSPK).**

[U] Large overlapping ellipses do not necessarily imply failure to comply with the TRD. The connection between feature ellipses and the TRD SSPK is rather complicated. The discrimination requirement is specified in terms of $P_{select}$, which is only a part of SSPK. $P_{select}$ is estimated through a number of Monte Carlo runs, in each of which the lethal target is selected according to $P_{AT}$

[U] In the flight data analysis reports, feature distributions are shown as ellipses in two-feature space. It is necessary to consider all features being used and their relative weights (covariance matrices) in calculating the value of $P_{AT}$ for each target. Values of $P_{AT}$ for all targets in the observed threat complex are then used in determining the value of $P_{select}$.

- **Distances to the center of the assigned ellipsis are subject to bias error.**

[U] This is certainly true. A good MBQC database design should exclude any bias in the feature distributions.

- **High parameter separation is not a guarantee for high discrimination performance, and measuring distances to the center of the assigned ellipsis are misleading.**

[U] The design for fine (i.e., precision) discrimination does not use "distances to the centers of the assigned ellipses" alone for target selection.

**Page 9 - IFT-1A versus IFT-3; both fail to comply without exception with TRD/SOW**

- **Flight test IFT-1a and IFT-3 use identical threat system objects/dynamics/etc. The same features extracted for IFT-1a must be identical for IFT-3. However, this is not the case.**
  **- NRC test report dated 12/2/98 page 217.**

# Unclassified Draft

[U] It has not yet been decided what the target suite shall be for IFT-3; therefore, it is premature to say that the extracted features for IFT-3 and IFT-1A must be identical. In addition to target characteristics, the extracted features also depend on sensor models, engagement geometry, and illumination conditions. It is expected that the seeker sensor planned for IFT-3 will be an improved version of that used in IFT-1A.

- While IFT-1a had limited use for feature #1 and #2 and needed #5 to perform, the IFT-3 simulation used #1 and #2 and found #5 destroyed performance.
  - Addendum 1 60-day report page 125.
  - NRC on IFT-3 (case #45) used #1 and #2 and found that removing feature #5 cause results to improve substantially. (Final Test dated 12/2/98 on page 16, 44, 28, 135-141.

[U] The sensitivity of the discrimination performance to feature #5 has been stated in [17] to be due to the use of the GFA in its poorly constituted condition. When NRC modified the GFA, performance improved markedly. This appears to be a credible assertion by NRC and is a reasonable explanation for the concern raised.

- Flight IFT-1A and IFT-3 failed to perform with spikes that are integrated as part of the TRD requirements.
  - NRC baseline test report page 16: "no target selected for many cases" and "Large number of false alarms" when the signatures posses "Spikes." These cases were not integrated into the calculations for P-select/PAT (probability to select the RV). NRC test report and presentation on 12/3/98.

[U] The "spikes" referred to here are artifacts of an incorrect sensor noise model and improper logic in parts of VICTIM, which are not TRW's responsibility.

- But Spikes on signatures are not error in the noise modeling as NRC claims, but part of the TRD requirements, "several scattering centers" on:
  - the ACPS part of CSS_4 (please refer to TRD page A-32).
  - reflecting appendages/threat wings (please refer to TRD page A-49, A-50, SS-18-M5, SS-18-M4).
  - NRC test report dated 12/2/97 pages 224 balloon with strips.

[U] The "several scattering centers" on the ACPS of CSS-4 are believed to be radar scattering centers. The "reflecting appendages/threat wings" do not physically produce spikes on optical signature data. Artificial spikes may indeed be generated by the current OSC and OPTISIG codes because of problems with reflectance calculations. It is incorrect to suggest that the effect of these spikes, which are of completely artificial origin, carry any significance for the baseline discrimination function. The existence of reflective scattering centers on targets is presumably properly included in the MDL. Any special sensitivity to these spikes can be properly investigated by running test cases with these attributes on the targets.

**Page 10 - Lack of pre-flight threat targets assignment results in non-compliance with the TRD**

- GBI said on 12/3/97: "Lack of Target Assignment prior to onboard discrimination degraded the discrimination performance drastically . . . it does not comply with the TRD."

[U] It is true that discrimination performance degrades because of the lack of prior knowledge of the threat type. The degree of degradation depends on how well targets in the entire threat library can be separated. For a simple threat library such as the TRD near-term threats, the degradation is minimal, as shown in the TTSSR [6]. The TRD gives no requirements relative to single-threat system assignment, and this issue is outside the scope of this study.

- Discrimination performance is subject to bias.

**Page 11 - Threat Number (NRC Test Run 3)**

**Page 12 - Threat Number (NRC Test Run 3 cont'd)**

**Page 13 - Threat Number (NRC Test Run 9)**

[U] These three pages illustrate the effect of aspect-angle variation on radiant intensity measurements.

**Page 14 - Dips appears in Signatures with Tether, Appendages and single RV or a change in sensor Aspect Angle that comply with the TRD**

- Dips in Signatures are created by a tether/appendage attachment to the deployed RV.
- A change in the sensor aspect angle creates Dips in signature even with a single RV.

[S] Th... t

...normal ... ndard deviation ...

- Case run 3 does not have justified Dips in its composite signatures (Contradictory to NRC).
  - NRC final test page 51 dated 12/2/98.
  - TRD requires discrimination with aspect angle 0-180 degree (page A-80).

# Unclassified Draft

- TRW/NRC admits that Discrimination algorithm collapses when Signature has a Dip.
  - TRW letter to Mr. Roberson GBI dated March 5, 1998.
- Automatic total Discrimination with single RV and attachment of a tether/appendage is false since algorithm collapses with a Dip in composite signatures.
  - NRC final test dated 12/2/98 is false, since algorithm collapses when single RV has attachment or a tether.

[U] Tethered and appendaged RV cases have been run many times by NRC with VICTIM. No particular sensitivities have been reported. NRC has not reported that these cases lead to a "collapse" of the software. In the few cases that did not meet requirements, the failure was not attributable to lack of feature separation but rather to "faulty FOV [field of view] management logic and nonsensical sensor noise model effects" [17].

[U] It is incorrect to assert that because dips occasioned by an improperly constituted GFA cause a failure of the system, dips of some other origin, i.e., aspect-angle variation, will cause a similar result, particularly when this has been demonstrated to be false (see test cases shown on page 64 of NRC test report [17]).

**Page 15 Non compliance with the TRD for TRW 45 day, 60 day, Addendum 1 to 60-day, and NRC test Reports**

- TRW and NRC falsely claim that the sensor aspect angle remains constant during the flight.
  - NRC test report dated 12/2/97 page 70 "sensor aspect angle," Addendum 1 to 60-day report page 56.

[U] It is difficult to respond to this claim because neither of the cited references (NRC test report, page 70 [17] and classified addendum to the 60 -day report, page 56 [12]) refer to sensor aspect angle.

- But the Sensor Aspect angle does change and creates a signature Dip.
  - TRD page A-83 indicates a change of 17 degree over 20 seconds for SS-18 M5.

[U] See the discussion on the previous page.

- The ellipse in the 45-day report are different from the ellipse in the 60-day report and different from the ellipse in the Addendum 1 to 60-day report.
  - Addendum 1 to 60-day report pages 124 till 134 "one sigma ellipse for each class."
  - 60-day report "one sigma ellipse for each class pages 130-131.

# Unclassified Draft

[U] The differences in the ellipses between the 45-day and the 60-day reports have been satisfactorily explained as being the result of the changing impact of the GFA, which should indeed have a large influence on the shape and location of ellipses. The differences between the 60-day and the revised 60-day reports are simply as matter of mislabeling: the ellipses in the 60-day report should read 2.8-σ instead of 1-σ.

- Ellipsis change is not justified. The pre-flight single threat assignment was not changed.

[U] Actually, a new MDL was generated for the IFT-1A 60-day data analysis that was based on track intervals later and shorter than those used in the 45-day analysis.

- Ellipses on the 45-day and 60-day report are overlapping each other. The large overlapping area indicates non-compliance of the discrimination with the TRD.

[U] A large overlap of ellipses does not necessarily imply noncompliance with the TRD: however, it does indicate the likelihood of poor object-oriented discrimination performance. Also see responses to the fifth issue stated on page 8 of the DCIS document.

- TRW claims that the only change between the 45-day and 60-day reports is gap filling. This should not influence the shape of the ellipse.

[U] Actually, the change was in track intervals used in generating the MDLs. This change did influence the size, shape, and orientation of the feature ellipses. Please refer to Section 3.i of the main report for a detailed discussion.

- The ellipse should be drawn for three sigma since Boeing requested that the PID and P-select will be of value not less than three sigma.
  - NRC test report dated 12/2 /97 page 35.

[U] It is immaterial what multiplicative value of σ is used to draw the ellipsis, e.g., 1, 2, or 3 σ, provided that they are correctly identified. The ellipses are merely visual aids that reflect the statistical distribution of the feature values. Provided that the value of σ is invariant, drawing the ellipses at various σ values has no impact on any material result.

**Page 16 - Non-compliance with the TRD for TRW 45-day, 60-day, Addendum 1 to 60-day, and NRC test Reports**

- TRW claims that the only change in between the 45-day and 60-day reports is gap filling. This should not influence the shape of the ellipse.
- The ellipse should be drawn for three sigma since Boeing requested that the PID and P-select will be of value not less than three sigma.
  - NRC test report dated 12/2/97 page 35.

# Unclassified Draft

[U] The above two issues repeat the last two from page 15.

- How does TRW explain the "high ranking with 1000 test runs" when large areas of the RV ellipse overlaps with RVAOs?
  - Addendum 1 to 60-day report page 132: "when VICTIM is used to evaluate a single realization, as IFT-1a flight test, the output is the object ranking metric (probability that the object is the RV or assigned target) as a function of observation time."

[U] TRW's fine (i.e., precision) discrimination function implicitly includes what is called object-oriented discrimination and what is called complex-oriented discrimination. The former, given a target feature vector, computes the a posteriori probability of belonging to each of the target models defined in the MDL and then assigns the feature vector to the target model that is associated with the largest a posteriori probability. The latter function examines the a posteriori probabilities of being an RV (i.e., $P_{AT}$) for all targets in the observed threat complex and then selects the target with the largest $P_{AT}$ as the intercept target. When RV and non-RV ellipses overlap heavily, the object-oriented discrimination function alone usually produces a poor performance, but the performance can be improved significantly by the follow-on complex-oriented discrimination. This means that $P_{select}$ can be higher than what is perceived by only inspecting the feature ellipses.

[U] Figure B-1 illustrates this point; the figure shows a pair of RV and RV associated object (RVAO) feature vectors located within two overlapping ellipses. The RVAO feature vector is classified as an RVAO according to the object-oriented discrimination function, because its a posteriori probability of being an RVAO is higher than that of being an RV. The RV feature vector is classified as an RVAO for the same reason. But the RV is selected as the intercept target by the complex-oriented discrimination because the $P_{AT}$ of the RV feature vector is higher than that of the RVAO feature vector.

- Even when the target is assigned there are large numbers of false alarms and non-compliance with the TRD (about half the cases as informed by NRC on 12/2/97).

[U] It is difficult to understand this statement. In the target selection process, either an RV or a non-RV is selected. The probability of successful target selection and the probability of false alarm add up to 1. Therefore, when the RV is consistently assigned (or selected) the false alarm rate is zero.

- TRW addendum 1 to the 60-day report claims that to fully demonstrate the robustness of the algorithms will require a thousand more flight tests. This is false, the non-robustness, non-compliance of the algorithms without exception to the TRD can be shown by the current available reports.

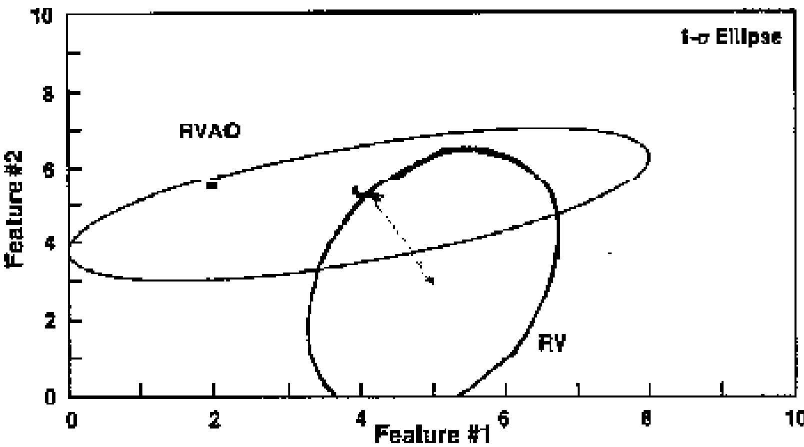
# Unclassified Draft

[U] The IFT-1A flight test is only one realization of the allowable test scenarios. Success in IFT-1A does not guarantee a 100% success rate if the same flight test is repeated many times within the allowable perturbation range. Given that the number of flight tests is quite limited, the performance in a statistical sense should be obtained through either analytical calculations or Monte Carlo simulations.

[U] The POET team disagrees with the generalized assessment that the algorithms are in noncompliance without exception, as shown in the current TRW IFT-1A data analysis reports.

# Unclassified Draft

**UNCLASSIFIED**




| | RV | RVA( |
|---|---|---|
| Mean | 5, 3 | 4, 5 |
| COV | 3 2<br>2 12 | 16 (<br>5 4 |

| | *Object Orientated Decision* | *Complex-Orientation Decision* |
|---|---|---|
| P (RV/RV■) = PAT(RV) =0.41<br>P (RVAO/RV■) = 0.59 | RV■ → RVAO | Select RV ■ |
| P (RV/RVAO ■) = PAT(RVAO) =0.12<br>P (RVAO/RVAO ■) = 0.88 | RVAO ■ → RVAO | |

*[U] Figure B-1. An illustration of object- and complex-oriented discrimination.*

# Unclassified Draft

## REFERENCES

1. "Technical Requirements Document (TRD) for the Ground Based Interceptor Experiment (GBI-X)" (U), U.S. Army PEO Missile Defense (28 September 1993), SECRET.

2. "Exoatmospheric Kill Vehicle (EKV) Critical Design Review (CDR) Book 2" (U), Rockwell Aerospace (1-2 August 1995), SECRET.

3. "Software Requirements Specification for the Tracking, Fusion & Discrimination (TFD) CSCI of the Exoatmospheric Kill Vehicle (EKV) Program, Rev C," TRW/Rockwell (13 March 1998), UNCLASSIFIED.

4. "Software Requirements Specification for the Tracking, Fusion & Discrimination (TFD) CSCI of the Exoatmospheric Kill Vehicle (EKV) Program, Rev C, Appendix A" (U), TRW/Rockwell (13 March 1998), SECRET.

5. "Detail Design for the Tracking, Fusion and Discrimination (TFD) of the Ground Based Interceptor (GBI) Program," Rockwell Aerospace (13 November 1996), UNCLASSIFIED.

6. W.D. Beitzel, J.W. Rapp, and D.E. Genereux, "Threat Typing Sensitivity Study Report, KV CDR Data Report SK1510" (U), TRW (23 July 1996), SECRET.

7. J.T. Tang, C. Frieler, and S. Cook, "Kalman filter performance" (U), TRW (April 1996), SECRET.

8. "GBI Discrimination Performance" (U), TRW/Boeing (18 November 1996), SECRET.

9. Hank Crowder et al., "GBI IFT 1A Post Flight Analysis" (U), TRW/Boeing (13 August 1997), SECRET.

10. "Sensor Flight Test Final (45) Day Report," Boeing (13 August 1997), UNCLASSIFIED

11. "Sensor Flight Test Final (60) Day Report," Boeing (22 August 1997), UNCLASSIFIED.

12. "Sensor Flight Test Final (60 Day) Report Classified Addendum" (U), Boeing (22 August 1997), SECRET.

13. "GBI IFT-1A Post Flight Analysis" (U), TRW (1 April 1998), SECRET.

14. "IFT-1A Ellipse Models and Measured Signatures from 45 and 60 Day Reports" (U), data tape, TRW (13 July 1998), SECRET.

15. "IFT-1A Features and SNR for POET" (U), data tape, TRW (6 August 1998), SECRET.

16. "POET Review of EKV Discrimination Algorithms" (U), TRW (21-22 July 1998), SECRET.

17. "TRW EKV Discrimination Algorithm Independent Testing Final Report" (U), NRC (2 December 1997), SECRET.

18. Robert A. Young, letter to Mr. Keith Englander, DCIS (22 December 1997), UNCLASSIFIED.

19. Robert A. Young, letter to Mr. Keith Englander, DCIS (25 March 1998), UNCLASSIFIED.

20. S. W. Reed, letter to Mr. Keith Englander, DCIS (29 May 1998), UNCLASSIFIED.

21. N. Schwartz, "Non-Compliance without Exception of Baseline Discrimination Technology with the TRD," presentation to the POET team (14 June 1998), UNCLASSIFIED.

# Unclassified Draft

22. S. W. Reed, letter to Mr. Keith Englander, DCIS (29 June 1998), UNCLASSIFIED.

23. S. J. Proffitt, *Ground Based Elements (GBE) response to the 25 March 1998 letter from R. A. Young (DCIS) to Keith Englander (BMDO/JPO)*, U.S. Army PEO Air and Missile Defense (8 June 1998), UNCLASSIFIED.