United States General Accounting Office

GAO

Report to the Honorable Edward J. Markey, House of Representatives

February 2002

MISSILE DEFENSE

Review of Results and Limitations of an Early National Missile Defense Flight Test

GAO

Accountability * Integrity * Reliability

Contents

Tables

GAO
Accountability * Integrity * Reliability

United States General Accounting Office
Washington, DC 20548

February 28, 2002

The Honorable Edward J. Markey
House of Representatives

For a number of years, the Department of Defense has been researching and developing defenses against ballistic missile attacks on the United States, its deployed forces, friends, and allies. In 1990, the Department awarded research and development contracts to three contractors to develop and test exoatmospheric kill vehicles.[1] The Department planned to use the best of the three vehicles in a follow-on missile defense program. One of the contractors, Rockwell International, subcontracted a portion of its kill vehicle design work to TRW. TRW was tasked with developing software that could operate on a computer onboard the kill vehicle. The software was to analyze data collected in flight by the kill vehicle's sensor (which collects real-time information about threat objects), enabling the kill vehicle to distinguish an enemy warhead from accompanying decoys.[2]

The three contractors proceeded with development of the kill vehicle designs and built and tested key subsystems (such as the sensor) until 1994. In 1994, the Department of Defense eliminated Martin Marietta from the competition. Both Rockwell—portions of which in December 1996 became Boeing North American—and Hughes—now Raytheon—continued designing and testing their kill vehicles. In 1997 and 1998, the National Missile Defense Joint Program Office[3] conducted tests, in space, of the sensors being developed by the contractors for their competing kill vehicles. Boeing's sensor was tested in June 1997 (Integrated Flight Test 1A) and Raytheon's sensor was tested in January 1998 (Integrated Flight Test 2). Program officials said these tests were not meant to demonstrate that the sensor met performance requirements, nor were they intended to be the basis for any contract award decisions. Rather, they were early research and development tests that the program office considered

[1] An exoatmospheric kill vehicle is the part of a defensive missile that is designed to hit and destroy an incoming enemy warhead above the earth's atmosphere.

[2] In some instances, the system may also use ground radar data.

[3] The National Missile Defense Joint Program Office reports to the Ballistic Missile Defense Organization within the Department of Defense. The National Missile Defense program is now known as the Ground-based Midcourse Missile Defense Program and the Ballistic Missile Defense Organization is now the Missile Defense Agency.

experiments to primarily reduce risk in future flight tests. Specifically, the tests were designed to determine if the sensor could operate in space; to examine the extent to which the sensor could detect small differences in infrared emissions; to determine if the sensor was accurately calibrated; and to collect target signature[4] data for post-mission discrimination analysis.

After the two sensor tests, the program office planned another 19 flight tests from 1999 through 2005 in which the kill vehicle would attempt to intercept a mock warhead. Initially, Boeing's kill vehicle was scheduled for testing in Integrated Flight Test 3 and Raytheon's in Integrated Flight Test 4. However, Boeing became the Lead System Integrator for the National Missile Defense Program in April 1998 and, before the third flight test was conducted, selected Raytheon as the primary kill vehicle developer.[5]

Meanwhile, in September 1995, TRW had hired a senior staff engineer, Dr. Nira Schwartz, to work on various projects, including the company's effort to develop the exoatmospheric kill vehicle's discrimination software. The engineer helped evaluate some facets of a technology known as the Extended Kalman Filter Feature Extractor,[6] which TRW planned to add as an enhancement to its discrimination software. The engineer reported to TRW in February 1996 that tests revealed that the Filter could not extract the key characteristics, or features, from various target objects that an enemy missile might deploy and demanded that the company inform Rockwell and the Department of Defense. TRW fired the engineer in March 1996. In April 1996, the engineer filed a lawsuit under the False Claims Act[7] alleging that TRW[8] falsely reported or hid information to make the National Missile Defense Joint Program Office believe that the Extended Kalman Filter Feature Extractor met the

[4] A target object's signature is the set of infrared signals emitted by the target.

[5] The Department of Defense continued funding the Boeing kill vehicle at a reduced level as a backup to Raytheon's kill vehicle. In mid-2000, the Department terminated all funding for Boeing's kill vehicle, ending TRW's involvement in development of the kill vehicle's discrimination software.

[6] The Kalman Filter is a mathematical model commonly used in real time data processing to estimate a variable of interest, such as an object's position or velocity. The Extended Kalman Filter Feature Extractor is used to extract features, which are used to perform discrimination.

[7] 31 USC 3729-3733.

[8] Rockwell, now Boeing North American, was later added to the lawsuit.

Department’s technical requirements. The engineer has amended the lawsuit several times, including adding allegations that TRW misled the Department of Defense about the ability of its discrimination software to distinguish a warhead from decoys and that TRW's test reports on Integrated Flight Test 1A falsely represented the discrimination software’s performance.

The False Claims Act allows a person to bring a lawsuit on behalf of the U.S. government if he or she has knowledge that a person or company has made a false or fraudulent claim against the government. If the suit is successful, the person bringing the lawsuit may share in any money recovered. The Department of Justice reviews all lawsuits filed under the act before deciding whether to join them. If it does, it becomes primarily responsible for prosecuting the case.

To determine whether it should join the engineer's lawsuit against TRW, Justice asked the Defense Criminal Investigative Service, a unit within the Department of Defense Inspector General’s office,[9] to examine the allegations. The engineer cooperated with the Investigative Service for more than 2 years. During the course of the Department of Defense’s investigation into the allegations of contractor fraud, two groups examined the former employee’s specific allegations regarding the performance of TRW’s basic discrimination software and performed limited evaluations of the Extended Kalman Filter Feature Extractor. The first was Nichols Research Corporation, a contractor providing technical assistance to the Ground Based Interceptor Project Management Office for its oversight of the exoatmospheric kill vehicle contracts. (This office within the National Missile Defense Joint Program Office is responsible for the exoatmospheric kill vehicle contracts.) Because an investigator for the Defense Criminal Investigative Service was concerned about the ability of Nichols to provide a truly objective assessment, the National Missile Defense Joint Program Office asked an existing advisory group, known as

[9] Department of Justice officials told us that they often use other agencies’ investigative units to investigate contractor fraud cases.

the Phase One Engineering Team,[10] to undertake another review of the specific allegations of fraud with respect to the software. This group is comprised of scientists from Federally Funded Research and Development Centers who were selected for the review team because of their knowledge of the National Missile Defense system. In addition, both Nichols and the Phase One Engineering Team assessed the feasibility of using the Extended Kalman Filter Feature Extractor to extract additional features from target objects that an enemy missile might deploy.[11]

The Department of Justice and the Defense Criminal Investigative Service investigated the engineer's allegations until March 1999. At that time, the Department of Justice decided not to intervene in the lawsuit. The engineer has continued to pursue her lawsuit without Justice's intervention.

When a Massachusetts Institute of Technology professor, Dr. Theodore Postol, learned of the engineer's claims, he conducted his own analysis of Integrated Flight Test 1A. In May 2000, the professor wrote to the White House alleging that Boeing North American and TRW misrepresented the results of the test.

The professor claimed that his analysis of Integrated Flight Test 1A showed that the system can be defeated by the simplest of decoys and that the National Missile Defense Joint Program Office and its contractors attempted to hide this fact by tampering with the flight test data and altering their analysis of the sensor's discrimination capabilities. The professor also alleged that objects deployed as part of Integrated Flight Test 1A displayed no distinguishable differences that Boeing's infrared

[10] The Phase One Engineering Team, according to its director, was established in 1988 by the Strategic Defense Initiative Organization—later known as the Ballistic Missile Defense Organization—as an umbrella mechanism to obtain technical and engineering support from Federally Funded Research and Development Centers. To ensure that the scientists who work on each review undertaken by the Phase One Engineering Team have the requisite expertise in the subjects they are asked to review, the membership on each review team varies with each assignment. The team assembled to review TRW's software included two individuals from the Massachusetts Institute of Technology's Lincoln Laboratory, two from Lawrence Livermore National Laboratory, and one from the Aerospace Corporation.

[11] In October 1996, TRW removed the Extended Kalman Filter Feature Extractor from its discrimination software. According to company officials, the Filter required computer speed and memory resources that were not available in the kill vehicle's onboard processor. In addition, the officials said that the basic discrimination software would perform adequately even without the Filter.

sensor could use to identify the mock warhead from decoys and that the program office hid the sensor's weaknesses by reducing the number of decoys planned for future tests. Further, the professor claimed that the Phase One Engineering Team's analysis was faulty.

At your request, we reviewed the professor's allegations. Specifically, as discussed with your office, we addressed the following questions:

1. Did Boeing and TRW disclose the key results and limitations of the flight test to the National Missile Defense Joint Program Office?
2. How did the Ground Based Interceptor Project Management Office oversee Boeing's and TRW's technical performance?
3. Did the flight test show whether each object deployed in space by an attacking missile exhibits distinguishable features?
4. Why did the National Missile Defense Joint Program Office reduce the complexity of later flight tests?
5. What were the methodology, findings, and limitations of the evaluation conducted by the Phase One Engineering Team of TRW's discrimination software?

You also asked us to determine whether the Department of Defense misused the security classification process to stifle public discussion of possible problems with the National Missile Defense system. We addressed this question in a separate report, dated June 12, 2001.[12]

Disclosure of Key Results and Limitations

Boeing and TRW disclosed the key results and limitations of Integrated Flight Test 1A in written reports released between August 13, 1997, and April 1, 1998. The contractors explained in a report issued 60 days after the June 1997 test that the test achieved its primary objectives, but that some sensor abnormalities were noted.[13] For example, while the report explained that the sensor detected the deployed targets and collected

[12] *DOD Officials Acted in Accordance With Executive Order for Addressing Security Classification Concerns* (GAO-01-737R, June 12, 2001).

[13] Appendix V includes selected requirements that Boeing established before the flight test to evaluate sensor performance and the actual sensor performance characteristics that Boeing and TRW discussed in the report.

some usable target signals, the report also stated that some sensor components did not operate as desired and the sensor often detected targets where there were none. In December 1997, the contractors documented other test anomalies. According to briefing charts prepared for a December meeting, the Boeing sensor tested in Integrated Flight Test 1A had a low probability of detection; the sensor's software was not always confident that it had correctly identified some target objects; the software significantly increased the rank of one target object toward the end of the flight; and in-flight calibration of the sensor was inconsistent. Additionally, on April 1, 1998, the contractors submitted an addendum to an earlier report that noted two more problems. In this addendum, the contractors disclosed that their claim that TRW's software successfully distinguished a mock warhead from decoys during a post-flight analysis was based on tests of the software using about one-third of the target signals collected during Integrated Flight Test 1A. The contractors also noted that TRW reduced the software's reference data[14] so that it would correspond to the collected target signals being analyzed. Project office and Nichols Research officials said that in late August 1997, the contractors orally communicated to them all problems and limitations that were subsequently described in the December 1997 briefing and the April 1998 addendum. However, neither project officials nor contractors could provide us with documentation of these communications.

Although the contractors reported the test's key results and limitations, they described the results using some terms that were not defined. For example, one written report characterized the test as a "success" and the sensor's performance as "excellent." We found that the information in the contractors' reports, in total, enabled officials in the Ground Based Interceptor Project Management Office and Nichols Research to understand the key results and limitations of the test. However, because such terms are qualitative and subjective rather than quantitative and objective, their use increased the likelihood that test results would be interpreted in different ways and might even be misunderstood. As part of our ongoing review of missile defense testing, we are examining the need for improvements in test reporting.

Appendix I provides details on the test and the information disclosed.

[14] Reference data are a collection of predicted characteristics, or features, that target objects are expected to display during flight. The software identifies the warhead from the decoys by comparing the features displayed by the different target objects to the reference data.

Project Office Reliance on Various Sources for Contractor Oversight

The Ground Based Interceptor Project Management Office relied on an on-site engineer and Nichols Research Corporation to provide insight into Boeing's work. The project office also relied on Boeing to oversee the performance of its subcontractor, TRW. Oversight was limited by the ongoing competition between Boeing and another contractor competing for the exoatmospheric kill vehicle contract because the Ground Based Interceptor Project Management Office and its support contractors had to be careful not to affect competition by assisting one contractor more than another. Project officials said that they relied more on "insight" into the contractors' work rather than oversight of that work. Nichols gained program insight by attending technical meetings, assessing test reports, and sometimes evaluating technologies proposed by Boeing and TRW.

For more information on how the project office exercised oversight over its contractors' technical performance, see appendix II.

Distinguishable Differences in Objects Deployed in Space

Boeing and TRW reported that post-flight testing and analysis of data collected during Integrated Flight Test 1A showed that deployed target objects displayed distinguishable features when observed by an infrared sensor. The contractors reported the test also showed that Boeing's exoatmospheric kill vehicle sensor could collect target signals from which TRW's software could extract distinguishable features and that the software could identify the mock warhead from other objects by comparing the extracted features to the features that it had been told to expect each object to display. However, there has been no independent verification of these claims.

We talked with Dr. Mike Munn, who was, during the 1980s, the Chief Scientist for missile defense programs at Lockheed Missiles and Space Company. He agreed that a warhead and decoys deployed in the exoatmosphere likely display distinguishable differences in the infrared spectrum. However, the differences may not be fully understood or there may not presently be methods to predict the differences. Dr. Munn added that the key was in the ability to make both accurate and precise measurements and also to predict signatures accurately. He emphasized that robust discrimination depends on the ability to predict signatures and then to match in-space measurements with those predictions. The Phase One Engineering Team and Nichols Research Corporation have noted that TRW's software used prior knowledge of warhead and decoy differences, to the maximum extent available, to discriminate one object from the other and cautioned such knowledge may not always be available in the real world.

Decoy Reduction in Later Tests

National Missile Defense program officials said that after considerable debate among themselves and contractors, the program manager reduced the number of decoys planned for intercept flight tests in response to a recommendation by an independent panel, known as the Welch Panel.[15] The panel, established to reduce risk in ballistic missile defense flight test programs, viewed a successful hit-to-kill engagement as a difficult task that should not be further complicated in early tests by the addition of decoys. After contemplating the advice of the Welch panel and considering the opinions of program officials and contractors who disagreed over the number and complexity of decoys that should be deployed in future tests, the program manager decided that early tests should include only one decoy, a large balloon.

See appendix III for more information on the reduction of decoys in later tests.

Evaluation of TRW's Discrimination Software

The Phase One Engineering Team was tasked by the National Missile Defense Joint Program Office to assess the performance of TRW's software and to complete the assessment within 2 months using available data. The team's methodology included determining if TRW's software was based on sound mathematical, engineering, and scientific principles and testing the software's critical modules using data from Integrated Flight Test 1A.

The team reported that although the software had weaknesses, it was well designed and worked properly, with only some changes needed to increase the robustness of the discrimination function. Further, the team reported that the results of its test of the software using Integrated Flight Test 1A data produced essentially the same results as those reported by TRW. Based on its analysis, team members predicted that the software would perform successfully in a future intercept test if target objects deployed as expected.

Because the Phase One Engineering Team did not process the raw data from Integrated Flight Test 1A or develop its own reference data, the team cannot be said to have definitively proved or disproved TRW's claim that

[15] The Welch Panel was chaired by Larry Welch, President of the Institute for Defense Analyses, and included 15 other members, some of whom were retired flag officers and former Department of Defense officials.

its software successfully discriminated the mock warhead from decoys using data collected from Integrated Flight Test 1A. A team member told us its use of Boeing- and TRW-provided data was appropriate because the former TRW employee had not alleged that the contractors tampered with the raw test data or used inappropriate reference data.

Appendix IV provides additional details on the Phase One Engineering Team evaluation.

Agency Comments and Our Evaluation

In commenting on a draft of this report, the Department of Defense concurred with our findings. It also suggested technical changes, which we incorporated as appropriate. The Department's comments are reprinted in appendix VII.

We conducted our review from August 2000 through February 2002 in accordance with generally accepted government auditing standards. Appendix VI provides details on our scope and methodology. The National Missile Defense Joint Program Office’s process for releasing documents significantly slowed our work. For example, the program office took approximately 4 months to release key documents such as the Phase One Engineering Team’s response to the professor’s allegations. We requested these and other documents on September 14, 2000, and received them on January 9, 2001.

As arranged with your staff, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days from its issue date. At that time, we plan to provide copies of this report to the Chairmen and Ranking Minority Members of the Senate Committee on Armed Services; the Senate Committee on Appropriations, Subcommittee on Defense; the House Committee on Armed Services; and the House Committee on Appropriations, Subcommittee on Defense; and the Secretary of Defense; and the Director, Missile Defense Agency. We will make copies available to others upon request.

If you or your staff have any questions concerning this report, please contact Bob Levin, Director, Acquisition and Sourcing Management, on (202) 512-4841; Jack Brock, Managing Director, on (202) 512-4841; or Keith Rhodes, Chief Technologist, on (202) 512-6412. Major contributors to this report are listed in appendix VIII.

Sincerely yours,

Jack L. Brock, Jr.
Managing Director
Acquisition and Sourcing Management

Keith Rhodes
Chief Technologist
Applied Research and Methods

Appendix I: Disclosure of Flight Test's Key Results and Limitations

Boeing and TRW disclosed the key results and limitations of an early sensor flight test, known as Integrated Flight Test 1A, to the Ground Based Interceptor Project Management Office. The contractors included some key results and limitations in written reports submitted soon after the June 1997 test, but others were not included in written reports until December 1997 or April 1998. However, according to project office and Nichols officials, all problems and limitations included in the written reports were communicated orally to the project management office in late August 1997. The deputy project office manager said his office did not report these verbal communications to others within the Program Office or the Department of Defense because the project office was the office within the Department responsible for the Boeing contract.

One problem that was included in initial reports to program officials was a malfunctioning cooling mechanism that did not lower the sensor's temperature to the desired level. Boeing characterized the mechanism's performance as somewhat below expectations but functioning well enough for the sensor's operation. We hired experts to determine the extent to which the problem could affect the sensor's performance. The experts found that the cooling problem degraded the sensor's performance in a number of ways, but would not likely result in extreme performance degradation. The experts studied only how increased noise[1] affected the sensor's performance regarding comparative strengths of the target signals and the noise (signal to noise ratio). The experts did not evaluate discrimination performance, which is dependent on the measurement accuracy of the collected infrared signals. The experts' findings are discussed in more detail later in this appendix.

The Test

Integrated Flight Test 1A, conducted in June 1997, was a test of the Boeing sensor—a highly sensitive, compact, infrared device, consisting of an array of silicon detectors, that is normally mounted on the exoatmospheric kill vehicle. However, in this test, a surrogate launch vehicle carried the sensor above the earth's atmosphere to view a cluster of target objects that included a mock warhead and various decoys. When the sensor detected the target cluster, its silicon detectors began to make precise measurements of the infrared radiation emitted by the target objects. Over the tens of seconds that the target objects were within its field of view, the sensor continuously converted the infrared radiation into an electrical

[1] Noise is undesirable electronic energy from sources other than the target objects.

current, or signal, proportional to the amount of energy collected by the detectors. The sensor then digitized the signal (converted the signals into numerical values), completed a preliminary part of the planned signal processing, and formatted the signal so that it could be transmitted via a data link to a recorder on the ground. After the test, Boeing processed the signals further[2] and formatted them so that TRW could input the signals into its discrimination software to assess its capability to distinguish the mock warhead from decoys. In post-flight ground testing, the software analyzed the processed data and identified the key characteristics, or features, of each signal. The software then compared the features it extracted to the expected features of various types of target objects. Based on this comparison, the software ranked each item according to its likelihood of being the mock warhead. TRW reported that the highest-ranked object was the mock warhead.

The primary objective of Integrated Flight Test 1A was to reduce risk in future flight tests. Specifically, the test was designed to determine if the sensor could operate in space; to examine the extent to which the sensor could detect small differences in infrared emissions; to determine if the sensor was accurately calibrated; and to collect target signature[3] data for post-mission discrimination analysis. In addition, Boeing established quantitative requirements for the test.[4] For example, the sensor was expected to acquire the target objects at a specified distance. According to a Nichols' engineer, Boeing established these requirements to ensure that its exoatmospheric kill vehicle, when fully developed, could destroy a warhead with the single shot precision (expressed as a probability) required by the Ground Based Interceptor Project Management Office. The engineer said that in Integrated Flight Test 1A, Boeing planned to measure its sensor's performance against these lower-level requirements so that Boeing engineers could determine which sensor elements, including the software, required further refinement. However, the engineer told us that because of the various sensor problems, of which the contractor and project office were aware, Boeing determined before the test that it would not use most of these requirements to judge the sensor's performance. (Although Boeing did not judge the performance of its sensor against the

[2] The signal processing that Boeing completed after the test will be completed onboard the exoatmospheric kill vehicle in an operational system.

[3] A target object's signature is the set of infrared signals emitted by the target.

[4] These requirements were established by the contractor and were not imposed by the government.

requirements as it originally planned, Boeing did, in some cases, report the sensor's performance in terms of these requirements. For a summary of selected test requirements and the sensor's performance as reported by Boeing and TRW in their August 22, 1997, report, see app. V.)

Reported Key Results and Limitations

Table 1 provides details on the key results and limitations of Integrated Flight Test 1A that contractors disclosed in various written reports and briefing charts.

Table 1: What and When Key Results and Limitations Were Included in Contractors' Written Reports

August 13, 1997, Report	August 22, 1997, Report	December 11, 1997, Briefing	April 1, 1998, Report
Detected deployed targets	Detected deployed targets	High false alarm rate	Failure of gap-filling module[a]
Target signals collected	Target signals collected	Sensor did not cool to desired temperature	Target signals collected during selected portion of the flight timeline used in assessment of discrimination software
Discrimination software distinguished mock warhead from decoys	Discrimination software distinguished mock warhead from decoys	Software confidence factor remained small for two target objects	Selected reference data used in assessment of discrimination software
	Excellent performance of sensor payload	Sensor had a lower than expected probability of detection	
	Power supply caused noisy target signals	Software significantly increased rank of one target object toward the end of the flight	
	Sensor did not cool to desired temperature	In-flight calibration of sensor was inconsistent	
	High false alarm rate		
	Slow turn-around of launch vehicle caused data loss		

[a]TRW designed a gap-filling module for its discrimination software to replace missing or noisy portions of collected and simulated target signals.

Although the contractors disclosed the key results and limitations of the flight test in written reports and in discussions, the written reports described the results using some terms that were not defined. For example, in their August 22, 1997, report, Boeing and TRW described Integrated Flight Test 1A as a "success" and the performance of the Boeing sensor as "excellent." We asked the contractors to explain their use of these terms. We asked Boeing, for example, why it characterized its sensor's performance as "excellent" when the sensor's silicon detector array did not cool to the desired temperature, the sensor's power supply created excess noise, and the sensor detected numerous false targets. Boeing said that even though the silicon detector array operated at temperatures 20 to 30 percent higher than desired, the sensor produced useful data. Officials said they knew of no other sensor that would be capable of producing any useful data under those conditions. Boeing officials went on to say that the sensor continuously produced usable, and, much of the time, excellent data in "real-time" during flight. In addition, officials said the sensor component responsible for suppressing background noise in the silicon detector array performed perfectly in space and the silicon detectors collected data in more than one wave band. Boeing concluded that the sensor's performance allowed the test to meet all mission objectives.

Based on our review of the reports and discussions with officials in the Ground Based Interceptor Project Management Office and Nichols Research, we found that the contractors' reports, in total, contained information for those officials to understand the key results and limitations of the test. However, because terms such as "success" and "excellent" are qualitative and subjective rather than quantitative and objective, we believe their use increases the likelihood that test results would be interpreted in different ways and could even be misunderstood. As part of our ongoing review of missile defense testing, we are examining the need for improvements in test reporting.

The August 13 Report

This report, sometimes referred to as the 45-day report, was a series of briefing charts. In it, contractors reported that Integrated Flight Test 1A achieved its principal objectives of reducing risks for subsequent flight tests, demonstrating the performance of the exoatmospheric kill vehicle's sensor, and collecting target signature data. In addition, the report stated that TRW's software successfully distinguished a mock warhead from accompanying decoys.[5]

The August 22 Report

The August 22 report, known as the 60-day report, was a lengthy document that disclosed much more than the August 13 report. As discussed in more detail below, the report explained that some sensor abnormalities were observed during the test, that some signals collected from the target objects were degraded, that the launch vehicle carrying the sensor into

[5] Boeing and TRW reported that the original test objectives did not include a test of TRW's discrimination software. However, program officials decided immediately prior to the test that it offered an excellent opportunity to assess the software's capability even though post-processing tools needed to assess the software were not yet available and would need rapid development after Integrated Flight Test 1A.

space adversely affected the sensor's ability to collect target signals, and that the sensor sometimes detected targets where there were none. These problems were all noted in the body of the report, but the report summary stated that review and analysis subsequent to the test confirmed the "excellent" performance and nominal operation of all sensor subsystems.

Some Sensor Abnormalities Were Observed During the Test

Boeing disclosed in the report that sensor abnormalities were observed during the test and that the sensor experienced a higher than expected false alarm rate. These abnormalities were (1) a cooling mechanism that did not bring the sensor's silicon detectors to the intended operating temperature, (2) a power supply unit[6] that created excess noise, and (3) software that did not function as designed because of the slow turnaround of the surrogate launch vehicle.

In the report's summary, Boeing characterized the cooling mechanism's performance as somewhat below expectations but functioning well enough for the sensor's operation. In the body of the report, Boeing said that the fluctuations in temperature could lead to an apparent decrease in sensor performance. Additionally, Boeing engineers told us that the cooling mechanism's failure to bring the silicon detector array to the required temperature caused the detectors to be noisy. Because the discrimination software identifies objects as a warhead or a decoy by comparing the features of a target's signal with those it expects a warhead or decoy to display, a noisy signal may confuse the software. Boeing and TRW engineers said that they and program office officials were aware that there was a problem with the sensor's cooling mechanism before the test was conducted. However, Boeing believed that the sensor would perform adequately at higher temperatures. According to contractor documents, the sensor did not perform as well as expected, and some target signals were degraded more than anticipated. Boeing disclosed in the report that sensor abnormalities were observed during the test and that the sensor experienced a higher than expected false alarm rate. These abnormalities were (1) a cooling mechanism that did not bring the sensor's silicon detectors to the intended operating temperature, (2) a power supply unit that created excess noise, and (3) software that did not function as designed because of the slow turnaround of the surrogate launch vehicle.

In the report's summary, Boeing characterized the cooling mechanism's performance as somewhat below expectations but functioning well

[6] The power supply unit is designed to power the sensor's electronic components.

enough for the sensor's operation. In the body of the report, Boeing said that the fluctuations in temperature could lead to an apparent decrease in sensor performance. Additionally, Boeing engineers told us that the cooling mechanism's failure to bring the silicon detector array to the required temperature caused the detectors to be noisy. Because the discrimination software identifies objects as a warhead or a decoy by comparing the features of a target's signal with those it expects a warhead or decoy to display, a noisy signal may confuse the software. Boeing and TRW engineers said that they and program office officials were aware that there was a problem with the sensor's cooling mechanism before the test was conducted. However, Boeing believed that the sensor would perform adequately at higher temperatures. According to contractor documents, the sensor did not perform as well as expected, and some target signals were degraded more than anticipated.

Power Supply Creates Noise

The report also referred to a problem with the sensor's power supply unit and its effect on target signals. An expert we hired to evaluate the sensor's performance at higher than expected temperatures found that the power supply, rather than the temperature, was the primary cause of excess noise early in the sensor's flight. Boeing engineers told us that they were aware that the power supply was noisy before the test, but, as shown by the test, it was worse than expected.

Payload Launch Vehicle Affected Software's Ability to Remove Background Noise

The report explained that, as expected before the flight, the slow turnaround of the massive launch vehicle on which the sensor was mounted in Integrated Flight Test 1A caused the loss of some target signals. Engineers explained to us that the sensor would eventually be mounted on the lighter, more agile exoatmospheric kill vehicle, which would move back and forth to detect objects that did not initially appear in the sensor's field of view. The engineers said that Boeing designed software that takes into account the kill vehicle's normal motion to remove the background noise, but the software's effectiveness depended on the fast movement of the kill vehicle. Boeing engineers told us that, because of the slow turnaround of the launch vehicle used in the test, the target signals detected during the turnaround were particularly noisy and the software sometimes removed not only the noise but the entire signal as well.

Sensor Sometimes Detected False Targets

The report mentioned that the sensor experienced more false alarms than expected. A false alarm is a detection of a target that is not there. According to the experts we hired, during Integrated Flight Test 1A, the Boeing sensor often mistakenly identified noise produced by the power supply as signals from actual target objects. In a fully automated discrimination software program, a high false alarm rate could overwhelm the tracking software. Because the post-flight processing tools were not fully developed at the time of the August 13 and August 22, 1997, reports, Boeing did not rely upon a fully automated tracking system when it processed the Integrated Flight Test 1A data. Instead, a Boeing engineer manually tracked the target objects. The contractors realized, and reported to the Ground Based Interceptor Project Management Office, that numerous false alarms could cause problems in future flight tests, and they identified software changes to reduce their occurrence.

December 11 Briefing

On December 11, 1997, Boeing and TRW briefed officials from the Ground Based Interceptor Project Management Office and one of its support contractors on various anomalies observed during Integrated Flight Test 1A. The contractors' briefing charts explained the effect the anomalies could have on Integrated Flight Test 3, the first planned intercept test for the Boeing exoatmospheric kill vehicle, identified potential causes of the anomalies, and summarized the solutions to mitigate their effect. While some of the anomalies included in the December 11 briefing charts were referred to in the August 13 and August 22 reports, others were being reported in writing for the first time.

The anomalies referenced in the briefing charts included the sensor's high false alarm rate, the silicon detector array's higher-than-expected temperature, the software's low confidence factor that it had correctly identified two target objects correctly, the sensor's lower than expected probability of detection, and the software's elevation in rank of one target object toward the end of the test. In addition, the charts showed that an in-flight attempt to calibrate the sensor was inconsistent. According to the charts, actions to prevent similar anomalies from occurring or impacting Integrated Flight Test 3 had in most cases already been implemented or were under way.

Contractors Report Further on False Alarms

The contractors again recognized that a large number of false alarms occurred during Integrated Flight Test 1A. According to the briefing charts, false alarms occurred during the slow turnarounds of the surrogate launch vehicle. Additionally, the contractors hypothesized that some false alarms resulted from space-ionizing events. By December 11, engineers had identified solutions to reduce the number of false alarms in future tests.

Briefing Charts Include Observations on Higher Detector Array Temperature

As they had in the August 22, 1997, report, the contractors recognized that the silicon detector array did not cool properly during Integrated Flight Test 1A. The contractors reported that higher silicon detector array temperatures could cause noisy signals that would adversely impact the detector array's ability to estimate the infrared intensity of observed objects. Efforts to eliminate the impact of the higher temperatures, should they occur in future tests, were on-going at the time of the briefing.

Some Software Confidence Factors Lower Than Expected

Contractors observed that the confidence factor produced by the software was small for two target objects. The software equation that makes a determination as to how confident the software should be to identify a target object correctly, did not work properly for the large balloon or multiple-service launch vehicle. Corrections to the equation had been made by the time of the briefing.

Sensor's Probability of Detection Is Lower Than Expected

The charts state that the Integrated Flight Test 1A sensor had a lower than anticipated probability of detection and a high false alarm rate. Because a part of the tracking, fusion, and discrimination software was designed for a sensor with a high probability of detection and a low false alarm rate, the software did not function optimally and needed revision. Changes to prevent this from happening in future flight tests were under way.

Software Increases the Rank of One Object Near Test's End

The briefing charts showed that TRW's software significantly increased the rank of one target object just before target objects began to leave the sensor's field of view. Although a later Integrated Flight Test 1A report stated the mock warhead was consistently ranked as the most likely target, the charts show that if in Integrated Flight Test 3 the same object's rank began to increase, the software could select the object as the intercept target. In the briefing charts, the contractors reported that TRW made a software change in the model that is used to generate reference data. When reference data was generated with the software change, the importance of the mock warhead increased, and it was selected as the target. Tests of the software change were in progress as of December 11.

In-Flight Calibration Was Inconsistent

The Boeing sensor measures the infrared emissions of target objects by converting the collected signals into intensity with the help of calibration data obtained from the sensor prior to flight. However, the sensor was not calibrated at the higher temperature range that was experienced during Integrated Flight Test 1A. To remedy the problem, the sensor viewed a star with known infrared emissions. The measurement of the star's intensity was to have helped fill the gaps in calibration data that was essential to making accurate measurements of the target object signals. Boeing disclosed that the corrections based on the star calibration were inconsistent and did not improve the match of calculated and measured target signatures. Boeing subsequently told us that the star calibration corrections were effective for one of the wavelength bands, but not for another, and that the inconsistency referred to in the briefing charts was in how these bands behaved at temperatures above the intended operating range. Efforts to find and implement solutions were in progress.

April 1, 1998, Report

On April 1, 1998, Boeing submitted a revised addendum to replace an addendum that had accompanied the August 22, 1997, report. This revised addendum was prepared in response to comments and questions submitted by officials from the Ground Based Interceptor Project Management Office, Nichols Research Corporation, and the Defense Criminal Investigative Service concerning the August 22 report. In this addendum, the contractors referred in writing to three problems and limitations that had not been addressed in earlier written test reports or the December 11 briefing. Contractors noted that a gap-filling module, which was designed to replace noisy or missing signals, did not operate as designed. They also disclosed that TRW's analysis of its discrimination software used target signals collected during a selected portion of the flight timeline and used a portion of the Integrated Flight Test 1A reference data that corresponded to this same timeline.

Gap-Filling Software Module Did Not Perform As Designed

The April 1 addendum reported that a gap-filling module that was designed to replace portions of noisy or missing target signals with expected signal values did not operate as designed. TRW officials told us that the module's replacement values were too conservative and resulted in a poor match between collected signals and the signals the software expected the target objects to display.

Assessment Uses Selected Target Signals

The April 1, 1998, addendum also disclosed that the August 13 and August 22 reports, in which TRW conveyed that its software successfully distinguished the mock warhead from decoys, were based on tests of the software using about one-third of the target signals collected during Integrated Flight Test 1A. We talked to TRW officials who told us that Boeing provided several data sets to TRW, including the full data set. The officials said that Boeing provided target signals from the entire timeline to a TRW office that was developing a prototype version of the exoatmospheric kill vehicle's tracking, fusion, and discrimination software,[7] which was not yet operational. However, TRW representatives said that the test bed version of the software that TRW was using so that it could submit its analysis within 60 days of Integrated Flight Test 1A could not process the full data set. The officials said that shortly before the August 22 report was issued, the prototype version of the tracking, fusion, and discrimination software became functional and engineers were able to use the software to assess the expanded set of target signals. According to the officials, this assessment also resulted in the software's selecting the mock warhead as the most likely target. In our review of the August 22 report, we found no analysis of the expanded set of target signals. The April 1, 1998, report, did include an analysis of a few additional seconds of data collected near the end of Integrated Flight Test 1A, but did not include an analysis of target signals collected at the beginning of the flight.

Most of the signals that were excluded from TRW's discrimination analysis were collected during the early part of the flight, when the sensor's temperature was fluctuating. TRW told us that their software was designed to drop a target object's track if the tracking portion of the software received no data updates for a defined period. This design feature was meant to reduce false tracks that the software might establish if the sensor detected targets where there were none. In Integrated Flight Test 1A, the fluctuation of the sensor's temperature caused the loss of target signals. TRW engineers said that Boeing recognized that this interruption would cause TRW's software to stop tracking all target objects and restart the discrimination process. Therefore, Boeing focused its efforts on

[7] The purpose of TRW's tracking, fusion, and discrimination software, which was being designed to operate on-board Boeing's exoatmospheric kill vehicle, was to record the positions of the target objects as they moved through space, fuse information about the objects collected by ground-based radar with data collected by the kill vehicle's infrared sensor, and discriminate the warhead from decoys. The software's tracking function was not operational when the project office asked the contractors to determine the software's ability to discriminate. As a result, Boeing hand-tracked the target objects so that TRW could use test bed discrimination software, which is almost identical to the discrimination portion of the operational version of the tracking, fusion, and discrimination software, to assess the discrimination capability.

processing those target signals that were collected after the sensor's temperature stabilized and signals were collected continuously.[8]

Some signals collected during the last seconds of the sensor's flight were also excluded. The former TRW employee alleged that these latter signals were excluded because during this time a decoy was selected as the target. The Phase One Engineering Team cited one explanation for the exclusion of the signals. The team said that TRW stopped using data when objects began leaving the sensor's field of view. Our review did not confirm this explanation. We reviewed the target intensities derived from the infrared frames covering that period and found that several seconds of data were excluded before objects began to leave the field of view. Boeing officials gave us another explanation. They said that target signals collected during the last few seconds of the flight were streaking, or blurring, because the sensor was viewing the target objects as it flew by them. Boeing told us that streaking would not occur in an intercept flight because the kill vehicle would have continued to approach the target objects. We could not confirm that the test of TRW's discrimination software, as explained in the August 22, 1997, report, included all target signals that did not streak. We noted that the April 1, 1998, addendum shows that TRW analyzed several more seconds of target signals than is shown in the August 22, 1997, report. It was in these additional seconds that the software began to increase the rank of one decoy as it assessed which target object was most likely the mock warhead. However, the April 1, 1998, addendum also shows that even though the decoy's rank increased the software continued to rank the mock warhead as the most likely target. But, because not all of the Integrated Flight Test 1A timeline was presented in the April 1 addendum, we could not determine whether any portion of the excluded timeline might have been useful data and if there were additional seconds of useful data whether a target object other than the mock warhead might have been ranked as the most likely target.

Corresponding Portions of Reference Data Excluded

The April 1 addendum also documented that portions of the reference data developed for Integrated Flight Test 1A were also excluded from the

[8] When the Ground Based Interceptor Project Management Office asked Boeing to assess the discrimination capability of its sensor's software, TRW's prototype tracking, fusion, and discrimination software was not operational. To perform the requested assessment, TRW used test-bed discrimination software that was almost identical to the discrimination software that TRW engineers designed for the prototype tracking, fusion, and discrimination software. Because the test-bed software did not have the ability to track targets, Boeing performed the tracking function and provided the tracked signals to TRW.

discrimination analysis. Nichols and project office officials told us the software identifies the various target objects by comparing the target signals collected from each object at a given point in their flight to the target signals it expects each object to display at that same point in the flight. Therefore, when target signals collected during a portion of the flight timeline are excluded, reference data developed for the same portion of the timeline must be excluded.

Information Provided Verbally to Project Office

Officials in the National Missile Defense Joint Program Office's Ground Based Interceptor Project Management Office and Nichols Research told us that soon after Integrated Flight Test 1A the contractors orally disclosed all of the problems and limitations cited in the December 11, 1997, briefing and the April 1, 1998, addendum. Contractors made these disclosures to project office and Nichols Research officials during meetings that were held to review Integrated Flight Test 1A results sometime in late August 1997. The project office and contractors could not, however, provide us with documentation of these disclosures.

The current Ground Based Interceptor Project Management Office deputy manager said that the problems that contractors discussed with his office were not specifically communicated to others within the Department of Defense because his office was the office within the Department responsible for the Boeing contract. The project office's assessment was that these problems did not compromise the reported success of the mission, were similar in nature to problems normally found in initial developmental tests, and could be easily corrected.

Effect of Cooling Failure on Sensor's Performance

Because we questioned whether Boeing's sensor could collect any usable target signals if the silicon detector array was not cooled to the desired temperature, we hired sensor experts at Utah State University's Space Dynamics Laboratory to determine the extent to which the sub-optimal cooling degraded the sensor's performance. These experts concluded that the higher temperature of the silicon detectors degraded the sensor's performance in a number of ways, but did not result in extreme degradation. For example, the experts said the higher temperature reduced by approximately 7 percent the distance at which the sensor could detect targets. The experts also said that the rapid temperature fluctuation at the beginning and at the end of data acquisition contributed to the number of times that the sensor detected a false target. However, the experts said the major cause of the false alarms was the power supply noise that contaminated the electrical signals generated by the sensor in response to the infrared energy. When the sensor signals were processed after Integrated Flight Test 1A, the noise appeared as objects, but they were actually false alarms.

Additionally, the experts said that the precision with which the sensor could estimate the infrared energy emanating from an object based on the electrical signal produced by the energy was especially degraded in one of the sensor's two infrared wave bands. In their report, the experts said that the Massachusetts Institute of Technology's Lincoln Laboratory analyzed the precision with which the Boeing sensor could measure infrared radiation and found large errors in measurement accuracy. The Utah State experts said that their determination that the sensor's measurement capability was degraded in one infrared wave band might partially explain the errors found by Lincoln Laboratory.

Although Boeing's sensor did not cool to the desired temperature during Integrated Flight Test 1A, the experts found that an obstruction in gas flow rather than the sensor's design was at fault. These experts said the sensor's cooling mechanism was properly designed and Boeing's sensor design was sound.

Appendix II: Project Office Reliance on Various Sources for Contractor Oversight

The Ground Based Interceptor Project Management Office used several sources to monitor the contractors' technical performance, but oversight activities were limited by the ongoing exoatmospheric kill vehicle contract competition between Boeing and Raytheon. Specifically, the project office relied on an engineer and a System Engineering and Technical Analysis contractor, Nichols Research Corporation, to provide insight into Boeing's work. The project office also relied on Boeing to oversee TRW's performance.

The deputy manager of the Ground Based Interceptor Project Management Office told us that competition between Boeing and Raytheon limited oversight to some extent. He said that because of the ongoing competition, the project office monitored the two contractors' progress but was careful not to affect the competition by assisting one contractor more than the other. The project office primarily ensured that the contractors abided by their contractual requirements. The project office deputy manager told us that his office relied on "insight" into the contractors' work rather than oversight of that work.

The project office gained insight by placing an engineer on-site at Boeing and tasking Nichols Research Corporation to attend technical meetings, assess test reports, and, in some cases, evaluate Boeing's and TRW's technologies. The on-site engineer was responsible for observing the performance of Boeing and TRW and relaying any problems back to the project office. He did not have authority to provide technical direction to the contractors. According to the Ground Based Interceptor Project Management Office deputy manager, Nichols essentially "looked over the shoulder" of Boeing and TRW. We observed evidence of Nichols' insight in memorandums that Nichols' engineers submitted to the project office suggesting questions that should be asked of the contractors, memorandums documenting engineer's comments on various contractor reports, and trip reports recorded by the engineers after various technical meetings.

Boeing said its oversight of TRW's work complied with contract requirements. The contract between the Department of Defense and Boeing required Boeing to declare that "to the best of its knowledge and belief, the technical data delivered is complete, accurate, and complies with all requirements of the contract." With regard to Integrated Flight Test 1A, Boeing officials said that they complied with this provision by selecting a qualified subcontractor, TRW, to develop the discrimination concepts, software, and system design in support of the flight tests, and by holding weekly team meetings with subcontractor and project office officials. Boeing officials stated that they were not required to verify the validity of their subcontractor's flight test analyses; rather, they were only required to verify that the analyses seemed reasonable. According to Boeing officials, both they and the project office shared the belief that TRW possessed the necessary technical expertise in threat phenomenology modeling, discrimination, and target tracking, and both relied on TRW's expertise.

Appendix III: Reduced Test Complexity

National Missile Defense Joint Program Office officials said that they reduced the number of decoys planned for intercept flight tests in response to a recommendation by an independent panel, known as the Welch Panel. The panel, established to reduce risk in ballistic missile defense flight test programs, viewed a successful hit-to-kill engagement as a difficult task that should not be further complicated in early tests by the addition of decoys. In contemplating the panel's advice, the program manager discussed various target options with other program officials and the contractors competing to develop and produce the system's exoatmospheric kill vehicle. The officials disagreed on the number of decoys that should be deployed in the first intercept flight tests. Some recommended using the same target set deployed in Integrated Flight Test 1A and 2, while others wanted to eliminate some decoys. After considering the differing viewpoints, the program manager decided to deploy only one decoy—a large balloon—in early intercept tests.

Decoys in Early Intercept Tests

As flight tests began in 1997, the National Missile Defense Joint Program Office was planning two sensor tests—Integrated Flight Test 1A and 2—and 19 intercept tests. The primary objective of the sensor flight tests was to reduce risk in future flight tests. Specifically the tests were designed to determine if the sensor could operate in space; to examine the extent to which the sensor could detect small differences in infrared emissions; to determine if the sensor was accurately calibrated; and to collect target signature[1] data for post-mission discrimination analysis.

Initially, the next two flight tests were to demonstrate the ability of the competing kill vehicles to intercept a mock warhead. Integrated Flight Test 3 was to test the Boeing kill vehicle and Integrated Flight Test 4 was to test the Raytheon kill vehicle. Table 1 shows the number of target objects deployed in the two sensor tests, the number of objects originally planned to be deployed in the first two intercept attempts, and the number of objects actually deployed in the intercept attempts.

[1] A target object's signature is the set of infrared signals emitted by the target.

Table 2: Planned and Actual Targets for Initial Flight Tests

Target suite	Actual targets in integrated flight tests 1A and 2	Initial plan for integrated flight tests 3 and 4	Actual targets deployed for integrated flight tests 3 and 4
Mock warhead[a]	1	1	1
Medium rigid light replica[b]	2	2	0
Small canisterized[c] light replica	1	1	0
Canisterized small balloon	2	2	0
Large balloon	1	1	1
Medium balloon	2	2	0
Total objects	9	9	2

[a]The mock warhead, also known as the medium reentry vehicle, is the test target. Not included in this table is the multi-service launch system, which carries the mock warhead and all of the decoys into space. The launch system will likely become an object in the field of view of the exoatmospheric kill vehicle, like the mock warhead and decoys, and must be discriminated.

[b]This is a replica of the warhead.

[c]Decoys can be stored in canisters and released in flight.

Source: GAO generated from Department of Defense information.

By the time Integrated Flight Tests 3 and 4 were actually conducted, Boeing had become the National Missile Defense Lead System Integrator and had selected Raytheon's exoatmospheric kill vehicle for use in the National Missile Defense system. Boeing conducted Integrated Flight Test 3 (in October 1999) and Integrated Flight Test 4 (in January 2000) with the Raytheon kill vehicle. However, both of these flight tests used only the mock warhead and one large balloon, rather than the nine objects originally planned. Integrated Flight Test 5 (flown in July 2000) also used only the mock warhead and one large balloon.

Program officials told us that the National Missile Defense Program Manager decided to reduce the number of decoys used in Integrated Flight Tests 3, 4, and 5, based on the findings of an expert panel. This panel, known as the Welch Panel, reviewed the flight test programs of several Ballistic Missile Defense Organization programs, including the National Missile Defense program. The resulting report,[2] which was released shortly

[2] *Report of the Panel on Reducing Risk in Ballistic Missile Defense Flight Test Programs*, February 27, 1998.

after Integrated Flight Test 2, found that U.S. ballistic missile defense programs, including the National Missile Defense program, had not yet demonstrated that they could reliably intercept a ballistic missile warhead using the technology known as “hit-to-kill.” Numerous failures had occurred for several of these programs and the Welch Panel concluded that the National Missile Defense program (as well as other programs using "hit-to-kill" technology) needed to demonstrate that it could reliably intercept simple targets before it attempted to demonstrate that it could hit a target accompanied by decoys. The panel reported again 1 month after Integrated Flight Test 3[3] and came to the same conclusion.

The Director of the Ballistic Missile Defense Organization testified[4] at a congressional hearing that the Welch Panel advocated removing all decoys from the initial flight tests, but that the Ballistic Missile Defense Organization opted to include a limited discrimination requirement with the use of one decoy. Nevertheless, he said that the primary purpose of the tests was to demonstrate the system’s “hit-to-kill” capability.

Opinions on Decoys

Program officials said there was disagreement within the Joint Program Office and among the key contractors as to how many targets to use in the early intercept flight tests. Raytheon and one high-ranking program official wanted Integrated Flight Tests 3, 4, and 5 to include target objects identical to those deployed in the sensor flight tests. Boeing and other program officials wanted to deploy fewer target objects. After considering all options, the Joint Program Office decided to deploy a mock warhead and one decoy—a large balloon.

Raytheon officials told us that they discussed the number of objects to be deployed in Integrated Flight Tests 3, 4, and 5 with program officials and recommended using the same target set as deployed in Integrated Flight Tests 1A and 2. Raytheon believed that this approach would be less risky because it would not require revisions to be made to the kill vehicle’s software. Raytheon and program officials told us that Raytheon was confident that it could successfully identify and intercept the mock warhead even with this larger target set.

[3] *National Missile Defense Review*, November 1999.

[4] *Statement of Lieutenant General Ronald T. Kadish, USAF, Director, Ballistic Missile Defense Organization, Before the House Armed Services Committee, Subcommittee on Military Research & Development*, June 14, 2001.

One high-ranking program official said that she objected to reducing the number of decoys used in Integrated Flight Test 3, because there was a need to more completely test the system. However, other program officials lobbied for a smaller target set. One program official said that his position was based on the Welch Panel's findings and on the fact that the program office was not concerned at that time about discrimination capability. He added that the National Missile Defense program was responding to the threat of "nations of concern," which could only develop simple targets, rather than major nuclear powers, which were more likely to be able to deploy decoys.

The Boeing/TRW team also wanted to reduce the number of decoys used in the first intercept tests. In a December 1997 study, the companies recommended that Integrated Flight Test 3 be conducted with a total of four objects—the mock warhead, the two small balloons, and the large balloon. (The multi-service launch system was not counted as one of the objects.) The study cited concerns about the inclusion of decoys that were not part of the initially expected threat and about the need to reduce risk. Boeing said that the risk increased significantly that the exoatmospheric kill vehicle would not intercept the mock warhead if the target objects did not deploy from the test missile as expected.

According to Boeing/TRW, as the types and number of target objects increased, the potential risk that the target objects would be different in some way from what was expected also increased. Specifically, the December 1997 study noted that the medium balloons had been in inventory for some time and had not deployed as expected in other tests, including Integrated Flight Test 1A. In that test, one medium balloon only partially inflated and was not positioned within the target cluster as expected. The study also found that the medium rigid light replicas are the easiest to misdeploy and the small canisterized light replica moved differently than expected during Integrated Flight Test 1A.

Appendix IV: Phase One Engineering Team's Evaluation of TRW's Software

In 1998, the National Missile Defense Joint Program Office asked the Phase One Engineering Team to conduct an assessment, using available data, of TRW's discrimination software even though Nichols Research Corporation had already concluded that it met the requirements established by Boeing.[1] The program office asked for the second evaluation because the Defense Criminal Investigative Service lead investigator was concerned about the ability of Nichols to provide a truly objective evaluation.

The Phase One Engineering Team developed a methodology to (1) determine if TRW's software was consistent with scientific, mathematical, and engineering principles; (2) determine whether TRW accurately reported that its software successfully discriminated a mock warhead from decoys using data collected during Integrated Flight Test 1A; and (3) predict the performance of TRW's basic discrimination software against Integrated Flight Test 3 scenarios. The key results of the team's evaluation were that the software was well designed; the contractors accurately reported the results of Integrated Flight Test 1A; and the software would likely perform successfully in Integrated Flight Test 3. The primary limitation was that the team used Boeing- and TRW-processed target data and TRW-developed reference data in determining the accuracy of TRW reports for Integrated Flight Test 1A.

Phase One Engineering Team's Methodology

The team began its work by assuring itself that TRW's discrimination software was based on sound scientific, engineering, and mathematical principles and that those principles had been correctly implemented. It did this primarily by studying technical documents provided by the contractors and the program office. Next, the team began to look at the software's performance using Integrated Flight Test 1A data. The team studied TRW's August 13 and August 22, 1997, test reports to learn more about discrepancies that the Defense Criminal Investigative Service said it found in these reports. Team members also received briefings from the

[1] The Ground Based Interceptor Project Management Office identified the precision (expressed as a probability) with which the exoatmospheric kill vehicle is expected to destroy a warhead with a single shot. To ensure that the kill vehicle would meet this requirement, Boeing established lower-level requirements for each function that affects the kill vehicle's performance, including the discrimination function. Nichols compared the contractor-established software discrimination performance requirement to the software's performance in simulated scenarios.

Defense Criminal Investigative Service, Boeing, TRW, and Nichols Research Corporation.

Team members told us that they did not replicate TRW's software in total. Instead, the team emulated critical functions of TRW's discrimination software and tested those functions using data collected during Integrated Flight Test 1A. To test the ability of TRW's software to extract the features of each target object's signal, the team designed a software routine that mirrored TRW's feature-extraction design. The team received Integrated Flight Test 1A target signals that had been processed by Boeing and then further processed by TRW. These signals represented about one-third of the collected signals. Team members input the TRW-supplied target signals into the team's feature-extraction software routine and extracted two features from each target signal. The team then compared the extracted features to TRW's reports on these same features and concluded that TRW's software-extraction process worked as reported by TRW. Next, the team acquired the results of 200 of the 1,000 simulations that TRW had run to determine the features that target objects deployed in Integrated Flight Test 1A would likely display.[2] Using these results, team members developed reference data that the software could compare to the features extracted from Integrated Flight Test 1A target signals. Finally, the team wrote software that ranked the different observed target objects in terms of the probability that each was the mock warhead. The results produced by the team's software were then compared to TRW's reported results.

The team did not perform any additional analysis to predict the performance of the Boeing sensor and its software in Integrated Flight Test 3. Instead, the team used the knowledge that it gained from its assessment of the software's performance using Integrated Flight Test 1A data to estimate the software's performance in the third flight test.

[2] The Phase One Engineering Team reported that TRW ran 1,000 simulations to determine the reference data for Integrated flight Test 1A, but the Team received the results of only 200 simulations. TRW engineers said this was most likely to save time. Also, the engineers said that the only effect of developing reference data from 200 simulations rather than 1,000 simulations is that confidence in the reference data drops from 98 percent to approximately 96 percent.

The Phase One Engineering Team's Key Results

In its report published on January 25, 1999, the Phase One Engineering Team reported that even though it noted some weaknesses, TRW's discrimination software was well designed and worked properly, with only some refinement or redesign needed to increase the robustness of the discrimination function. In addition, the team reported that its test of the software using data from Integrated Flight Test 1A produced essentially the same results as those reported by TRW. The team also predicted that the Boeing sensor and its software would perform well in Integrated Flight Test 3 if target objects deployed as expected.

Weaknesses in TRW's Software

The team's assessment identified some software weaknesses. First, the team reported that TRW's use of a software module to replace missing or noisy target signals was not effective and could actually hurt rather than help the performance of the discrimination software. Second, the Phase One Engineering Team pointed out that while TRW proposed extracting several features from each target-object signal, only a few of the features could be used.

The Phase One Engineering Team also reported that it found TRW's software to be fragile because the software was unlikely to operate effectively if the reference data—or expected target signals—did not closely match the signals that the sensor collected from deployed target objects. The team warned that the software's performance could degrade significantly if incorrect reference data were loaded into the software. Because developing good reference data is dependent upon having the correct information about target characteristics, sensor-to-target geometry, and engagement timelines, unexpected targets might challenge the software. The team suggested that very good knowledge about all of these parameters might not always be available.

Accuracy of Contractors' Integrated Flight Test 1A Reports

The Phase One Engineering Team reported that the results of its evaluation using Integrated Flight Test 1A data supported TRW's claim that in post-flight analysis its software accurately distinguished a mock warhead from decoys. The report stated that TRW explained why there were differences in the discrimination analysis included in the August 13, 1997, Integrated Flight Test 1A test report and that included in the August 22, 1997, report. According to the report, one difference was that TRW mislabeled a chart in the August 22 report. Another difference was that the August 22 discrimination analysis was based on target signals collected over a shorter period of time (see app. I for more information regarding

TRW's explanation of report differences). Team members said that they found TRW's explanations reasonable.

Predicted Success in Integrated Flight Test 3

The Phase One Engineering Team predicted that if the targets deployed in Integrated Flight Test 3 performed as expected, TRW's discrimination software would successfully identify the warhead as the target. The team observed that the targets proposed for the flight test had been viewed by Boeing's sensor in Integrated Flight Test 1A and that target-object features collected by the sensor would be extremely useful in constructing reference data for the third flight test. The team concluded that given this prior knowledge, TRW's discrimination software would successfully select the correct target even in the most stressing Integrated Flight Test 3 scenario being considered, if all target objects deployed as expected. However, the team expressed concern about the software's capabilities if objects deployed differently, as had happened in previous flight tests.

Limitations of the Team's Evaluation

The Phase One Engineering Team's conclusion that TRW's software successfully discriminated is based on the assumption that Boeing's and TRW's input data were accurate. The team did not process the raw data collected by the sensor's silicon detector array during Integrated Flight Test 1A or develop their own reference data by running hundreds of simulations. Instead, the team used target signature data extracted by Boeing and TRW and developed reference data from a portion of the simulations that TRW ran for its own post-flight analysis. Because it did not process the raw data from Integrated Flight Test 1A or develop its own reference data, the team cannot be said to have definitively proved or disproved TRW's claim that its software successfully discriminated the mock warhead from decoys using data collected from Integrated Flight Test 1A. A team member told us its use of Boeing- and TRW-provided data was appropriate because the former TRW employee had not alleged that the contractors tampered with the raw test data or used inappropriate reference data.

Appendix V: Boeing Integrated Flight Test 1A Requirements and Actual Performance as Reported by Boeing and TRW

The table below includes selected requirements that Boeing established before the flight test to evaluate sensor performance and the actual sensor performance characteristics that Boeing and TRW discussed in the August 22 report.

Table 3: Integrated Flight Test 1A Requirements Established by Boeing and Actual Performance

Capability Tested[a]	Requirement	Integrated Flight Test 1A performance reported by Boeing/TRW
Acquisition range[b]	The sensor subsystem shall acquire the target objects at a specified distance.	The performance exceeded the requirement.[c]
Probability of detection	The sensor shall detect target objects with a specified precision, which is expressed as a probability.	The performance satisfied the requirement.
False alarm rate	False alarms shall not exceed a specified level.	The performance did not satisfy the requirement. The false alarm rate exceeded Boeing's requirement by more than 200 to 1 because of problems with the power supply and the higher than expected temperature of the sensor.
Infrared radiation measurement precision	The sensor subsystem shall demonstrate a specified measurement precision at a specified range.	The contractor met the requirement in one infrared measurement band, but not in another.
Angular Measurement Precision (AMP)	Given specified conditions, the sensor subsystem shall determine the angular position of the targets with a specified angular measurement precision.	The performance was better than the requirement.
Closely spaced objects resolution	Resolution of closely spaced objects shall be satisfied at a specified range.	The closely spaced objects requirement could not be validated because the targets did not deploy with the required separation.
Silicon detector array cool-down time	The time to cool the silicon detector array to less than a desired temperature shall be less than or equal to a specified length of time.	The performance did not satisfy the requirement because the desired temperature was not reached. Nevertheless, the silicon detector operated as designed at the higher temperatures.
Hold time[d]	With a certain probability, the silicon detector array's temperature shall be held below a desired temperature for a specified minimum length of time.	Even though the detector array's temperature did not reach the desired temperature, the array was cooled to an acceptable operating temperature and held at that temperature for longer than required.

[a]The requirements displayed in the table were established by the contractor and were not imposed by the government. Additionally, because of various sensor problems recognized prior to the test, Boeing waived most of the requirements. Boeing established these requirements to ensure that its exoatmospheric kill vehicle, when fully developed, could destroy a warhead with the single shot precision (expressed as a probability) required by the Ground Based Interceptor Project Management Office.

[b]Boeing's acquisition range specification required that the specified range, detection probability, and false alarm rate be achieved simultaneously. Boeing's Chief Scientist said that because the range and target signals varied with time and the total observation time was sharply limited during Integrated Flight Test 1A, the probability of detection could not be accurately determined. As a result, the test was not a suitable means for assessing whether the sensor can attain the specified acquisition range.

[c]The revised 60-day report states that the sensor did not detect the target until approximately two-thirds of the nominal acquisition range. Boeing engineers told us that while this statement appears to contradict the claim that the target was acquired at 107 percent of the specified range, it does not. Boeing engineers said that the nominal acquisition range refers to the range at which a sensor that is performing as designed would acquire the target, which is a substantially greater range than the specified acquisition range. However, neither Boeing nor TRW could provide documentation of the nominal acquisition range so that we could verify that these statements are not contradictory.

[d]In the main body of the August 22 report, the contractor discussed "hold time." However, it is not mentioned in the appendix to the August 22 report that lists the performance characteristics against which Boeing planned to evaluate its sensor's performance. Rather, the appendix refers to a "minimum target object viewing" time, which has the same requirement as the hold time. Boeing reported that its sensor collected target signals over approximately 54 seconds.

Appendix VI: Scope and Methodology

We determined whether Boeing and TRW disclosed key results and limitations of Integrated Flight Test 1A to the National Missile Defense Joint Program Office by examining test reports submitted to the program office on August 13, 1997, August 22, 1997, and April 1, 1998, and by examining the December 11, 1997, briefing charts. We also held discussions with and examined various reports and documents prepared by Boeing North American, Anaheim, California; TRW Inc., Redondo Beach, California; the Raytheon Company, Tucson, Arizona; Nichols Research Corporation, Huntsville, Alabama; the Phase One Engineering Team, Washington, D.C.; the Massachusetts Institute of Technology/Lincoln Laboratory, Lexington, Massachusetts; the National Missile Defense Joint Program Office, Arlington, Virginia, and Huntsville, Alabama; the Office of the Director, Operational Test and Evaluation, Washington D.C.; the U.S. Army Space and Missile Defense Command, Huntsville, Alabama; the Defense Criminal Investigative Service, Mission Viejo, California, and Arlington, Virginia; and the Institute for Defense Analyses, Alexandria, Virginia.

We held discussions with and examined documents prepared by Dr. Theodore Postol, Massachusetts Institute of Technology, Cambridge, Massachusetts; Dr. Nira Schwartz, Torrance, California; Mr. Roy Danchick, Santa Monica, California; and Dr. Michael Munn, Benson, Arizona.

In addition, we hired the Utah State University Space Dynamics Laboratory, Logan, Utah, to examine the performance of the Boeing sensor because we needed to determine the effect the higher operating temperature had on the sensor's performance. We did not replicate TRW's assessment of its software using target signals that the Boeing sensor collected during the test. This would have required us to make engineers and computers available to verify TRW's software, format raw target signals for input into the software, develop reference data, and run the data through the software. We did not have these resources available, and we, therefore, cannot attest to the accuracy of TRW's discrimination claims.

We also examined the methodologies, findings, and limitations of the review conducted by the Phase One Engineering Team of TRW's discrimination software. To accomplish this task, we analyzed the Phase One Engineering Team's "Independent Review of TRW EKV Discrimination Techniques" dated January 1999. In addition, we held discussions with Phase One Engineering Team members, officials from the National Missile Defense Joint Program Office, and contractor officials.

We did not replicate the evaluations conducted by the Phase One Engineering Team and cannot attest to the accuracy of their reports.

We reviewed the decision by the National Missile Defense Joint Program Office to reduce the complexity of later flight tests by comparing actual flight test information with information in prior plans and by discussing these differences with program and contractor officials. We held discussions with and examined documents prepared by the National Missile Defense Joint Program Office, the Institute for Defense Analyses, Boeing North American, and the Raytheon Company.

Our work was conducted from August 2000 through February 2002 according to generally accepted government auditing standards. The length of time the National Missile Defense Joint Program Office required to release documents to us significantly slowed our review. For example, the Program Office required approximately 4 months to release key documents such as the Phase One Engineering Team's response to the professor's allegations. We requested these and other documents on September 14, 2000, and received them on January 9, 2001.

Appendix VII: Comments from the Department of Defense

OFFICE OF THE UNDER SECRETARY OF DEFENSE
3000 DEFENSE PENTAGON
WASHINGTON, DC 20301-3000

ACQUISITION,
TECHNOLOGY
AND LOGISTICS

December 20, 2001

Mr. Jack L. Brock
Managing Director, Acquisition and Sourcing Management
U.S. General Accounting Office
Washington, D.C. 20548

Dear Mr. Brock:

This is the Department of Defense (DoD) response to the General Accounting Office (GAO) draft report to The Honorable Edward J. Markey, House of Representatives, GAO-02-124, "MISSILE DEFENSE: Review of Results and Limitations of an Early National Missile Defense Flight Test," dated November 2001 (GAO Code 707541). The Department appreciates the opportunity to comment on the draft report.

The Department concurs with the comments contained in the draft report (GAO did not have any recommendations) and recommends minor changes to the draft which will downgrade the report classification from Secret to Unclassified.

Sincerely,

George R. Schneiter
Director
Strategic and Tactical Systems

Appendix VIII: Major Contributors

Acquisition and Sourcing Management

Bob Levin, Director
Barbara Haynes, Assistant Director
Cristina Chaplain, Assistant Director, Communications
David Hand, Analyst-in-charge
Subrata Ghoshroy, Technical Advisor
Stan Lipscomb, Senior Analyst
Terry Wyatt, Senior Analyst
William Petrick, Analyst

Applied Research and Methods

Nabajyoti Barkakati, Senior Level Technologist
Hai Tran, Senior Level Technologist

General Counsel

Stephanie May, Assistant General Counsel

GAO's Mission

The General Accounting Office, the investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily e-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
P.O. Box 37050
Washington, D.C. 20013

To order by Phone:	Voice:	(202) 512-6000
	TDD:	(202) 512-2537
	Fax:	(202) 512-6061

Visit GAO's Document Distribution Center

GAO Building
Room 1100, 700 4th Street, NW (corner of 4th and G Streets, NW)
Washington, D.C. 20013

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm,
E-mail: fraudnet@gao.gov, or
1-800-424-5454 or (202) 512-7470 (automated answering system).

Public Affairs

Jeff Nelligan, Managing Director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G. Street NW, Room 7149,
Washington, D.C. 20548