GAO
Accountability * Integrity * Reliability

United States Government Accountability Office
Washington, DC 20548

June 1, 2006

The Honorable Jeff Sessions
Chairman
The Honorable Bill Nelson
Ranking Minority Member
Subcommittee on Strategic Forces
Committee on Armed Services
United States Senate

Subject: *Defense Acquisitions: Space System Acquisition Risks and Keys to Addressing Them*

On April 6, 2006, we testified before the subcommittee on the Department of Defense's (DOD) space acquisitions. In fiscal year 2007, DOD expects to spend nearly $7 billion to acquire space-based capabilities to support current military and other government operations as well as to enable DOD to transform the way it collects and disseminates information, gathers data on its adversaries, and attacks targets. Despite its growing investment in space, however, DOD's space system acquisitions have experienced problems over the past several decades that have driven up costs by hundreds of millions, even billions, of dollars; stretched schedules by years; and increased performance risks. In some cases, capabilities have not been delivered to the warfighter after decades of development.

Within this context, you requested that we provide additional comments regarding the need for better program management, space acquisition policy, and DOD's Space Radar and Transformational Satellite Communications System acquisitions. Your specific questions and our answers are discussed below.

Question: What are the top obstacles to achieving program success from the point of view of program managers?

As part of a 2005 review[1] on program management best practices, we surveyed DOD's major weapon program managers, including some managing space programs, who cited the following as "top" obstacles to achieving successful outcomes in an open ended question:

- funding instability (about 36 percent),
- requirements instability (13 percent),

[1] GAO, *Best Practices: Better Support of Weapon System Program Managers Needed to Improve Outcomes*, GAO-06-110 (Washington, D.C.: Nov. 30, 2005).

- staffing problems (8 percent),
- excessive oversight (7 percent), and
- inexperienced leadership (7 percent).

Although the majority of respondents to our survey believed that the initial baselines of their programs were reasonable, a significant group, about 24 percent, responded that their program parameters were not reasonable at the start, and 45 program managers responded that their program had been rebaselined one or more times for cost and schedule increases. In addition, 18 percent said one or more key technologies fell below best practice standards for maturity.

Our reviews of space programs are consistent with these views—we have found technologies to be immature at program start for major space programs. Further, in delving deeper into the root causes behind these problems, we have found that competition for funding has incentivized programs to produce optimistic cost and schedule estimates, over promise on capability, suppress bad news, and forsake the opportunity to identify potentially better alternatives. In addition, because DOD starts more weapons programs than it can afford, it invariably finds itself in the position of having to shift funds to sustain programs—often to the point of undermining well-performing programs to pay for poorly performing ones. We also have found that DOD starts its space programs too early, that is, before it has assurance that the capabilities it is pursuing can be achieved within available resources (time, money technology, people, etc.) and time constraints, and it allows new requirements to be added well into the acquisition phase, a course of action that can further stretch technology challenges. This is encouraged by the funding process, as acquisition programs tend to attract the majority of research, development, test, and evaluation (RDT&E) dollars. Many officials working within the space community agreed that these were key underlying causes of acquisition problems during a review we conducted last year.[2] In addition, officials we spoke with also cited pressures resulting from having a diverse array of officials and organizations involved with the space acquisition process, tensions between the science and technology (S&T) and acquisition communities as to who is better suited to translate technology concepts into reality, pressures resulting from short tenures among staff critical to achieving acquisition success, and difficulties in overseeing contractors.

Question: Do you believe that the Air Force is addressing these obstacles?

The Air Force has recently taken steps to put is Transformational Satellite Communications System (TSAT) program on a more executable track by reducing its expectations in the level of sophistication for the first two satellites so that it can meet its schedule goals. It is also holding off entering product development of the first increment until critical technologies are proven. If the Air Force adheres to this commitment for TSAT and applies it to Space Radar, as it has also informally committed to do, then it would be addressing some of the obstacles noted above. For example, it would reduce the risk of funding instability since cost estimates would be more realistic. In addition, the Air Force has committed to estimating cost and funding new acquisitions to an 80-percent confidence level, strengthening systems

[2] GAO, *Defense Acquisitions: Incentives and Pressures That Drive Problems Affecting Satellite and Related Acquisitions*, GAO-05-570R (Washington, D.C.: June 23, 2005).

engineering, and strengthening the acquisition workforce. And for some specific programs, the Air Force has applied additional mechanisms to regulate requirements. These actions could also remove obstacles, if effectively implemented.

However, as we testified, such actions should be accompanied by an investment strategy for space, and ultimately DOD's entire weapons portfolio, to separate wants from needs and to alleviate long-standing pressures associated with competition within DOD to win funding. DOD could also instill the best practices it is now embracing into its space acquisition policy. In addition, we have recommended that DOD, as a whole, take steps to hold people and programs accountable when best practices are not pursued. This will require DOD to empower program managers to make decisions related to funding, staffing, and moving into subsequent phases and to match program manager tenure with delivery of a product. It may also require DOD to tailor career paths and performance management systems to provide incentives for longer tenures. By embracing a model that incorporates all these elements, DOD can achieve better outcomes for its space programs. By not doing so, there will still be incentives and allowances to overpromise capability, underestimate cost and schedule, and to start programs prematurely, which, in turn, can eventually undo other improvement efforts.

Question: DOD starts more space and weapons programs than it can afford, which, according to GAO, "pressures programs to underestimate costs and over promise capabilities." Can you provide a few examples of this problem in space programs and [say] if and how the problem is being addressed?

Actual costs for nearly every major space acquisition we review each year as part of our annual weapon system assessment have greatly exceeded earlier estimates--a clear indication that programs consistently underestimate costs. For example, the Space Based Infrared System (SBIRS)-High cost estimate climbed from about $4 billion as of October 1996 to over $10 billion in September 2005, and costs are expected to increase further. Estimated costs for the Evolved Expendable Launch Vehicle (EELV) program have climbed from about $15 billion in October 1998 to $27 billion in August 2005 with 43 fewer launches to be purchased than anticipated. Estimated costs for the Advanced Extremely High Frequency Satellite program (AEHF) increased from $5.6 billion as of October 2001 to $6.2 billion as of August 2005, with quantities decreasing from five to three satellites. Estimated costs for the National Polar-orbiting Operational Environmental Satellite System (NPOESS) grew from $5.9 billion in August 2002 to nearly $8 billion in September 2005. Our past reports have also identified cases where programs have overpromised capabilities. For example, the Space Based Infrared System (SBIRS)-Low program started under the assumption that the satellites would be able to detect and track multiple objects and differentiate a threatening warhead from decoys, even though that technology challenge was exceedingly high. In fact, the program was never able to achieve this capability. It was eventually shut down in the face of cost and schedule overruns that came with addressing technology challenges. The SBIRS-High program began with the assumption that there would be four satellites in geosynchronous orbit, but more than 10 years later, DOD plans to reduce the number of satellites it will procure and still does not have the assurance it needs that the missile detection capability can be achieved in time to replace the existing detection system. In addition, DOD has initiated efforts to develop a parallel competing capability with the SBIRS-High program. Similarly, the NPOESS program is now considering dropping some of its planned capability because of technology and design-related challenges.

DOD has been taking actions to improve cost estimating and we are in the process of assessing these actions. As mentioned above, for example, it has committed to estimating cost and funding new acquisitions to an 80-percent confidence level. In addition, the Air Force is requiring the use of independent cost estimates—rather than estimates produced by a program office or a contractor. It is also committed to strengthening its cost-estimating capabilities—in terms of people, methodologies, and tools. In regard to the issue of overpromising capability, the Air Force has deferred pursuing some of its more ambitious capabilities on its TSAT program, so that the program can be better positioned to meet its schedule. We do not know at this point whether it will be doing the same for its new Space Radar program. As we underscored in our testimony, it is important that these and other individual actions be made within a framework of broader, systemic improvements to DOD's overall acquisition process, the acquisition workforce, and an overall investment strategy.

Question: The second problem is that DOD "starts its space programs too early, that is, before it is sure that the capabilities it is pursuing can be achieved within available resources and time constraints." Can you provide a few examples of this problem in space programs and [indicate] if and how the problem is being addressed?

Many of our annual reviews of major space acquisitions show that programs have started with relatively low levels of technology maturity—meaning DOD does not have assurance that the technologies can work as intended. This includes, AEHF, NPOESS, SBIRS-High, and SBIRS-Low—now known as the Space Tracking and Surveillance System. Exceptions include the Navy's Mobile User Objective System, or MUOS (though the program later added two additional technologies that did not meet best practices standards for maturity) and the Global Positioning System Block IIF. At times, we have found that key sensors to be included in new satellites were not fully tested, or even prototyped, before being included in a program. In other cases, technologies used to support the health of the overall satellite, such as cooling systems, were immature. And in other cases, software needs were vastly underestimated. In the case of AEHF, technical resources to support security needs were underestimated.

Many programs we have studied felt the need to start the acquisition process before such needs were better understood because acquisition programs tend to attract more funding than science and technology efforts. In addition, in the case of space, programs have historically attempted to satisfy all requirements in a single step, regardless of the design challenge or the maturity of the technologies to achieve the full capability. While this is partly attributable to a desire to speed delivery of capability, it has perversely slowed down programs, since programs were at increased risk of facing costly and disruptive technical and design problems.

As noted previously, DOD has committed to delay the development of one new major space program—TSAT—until technology needs are better understood. It has also committed to deliver new space-based capabilities in an incremental fashion so that acquisition efforts can be more executable and the science and technology base can be more engaged in major space programs. It has not taken such action yet on other new programs, notably Space Radar, though it has informally committed to. In addition, DOD's space acquisition policy still allows major acquisitions to begin without demonstrating that technology can work as intended.

Question: A third issue is that DOD has "allowed new requirements to be added well into the acquisition phase." I would also add that sometimes the original requirements may be unrealistic or unaffordable and that this may be part of the problem. Can you provide a few examples of the requirements problem in space programs and [indicate] if and how the problem is being addressed?

Our past reports have pointed to requirements setting problems in the AEHF, NPOESS, and SBIRS-High programs. In the case of SBIRS-High, we pointed to problems related to not adequately defining requirements up front. These were further detailed in subsequent DOD studies, including those by the SBIRS-High Independent Review Team and the Defense Science Board. Both noted that the acquisition approach the Air Force was following, known as Total System Performance Responsibility, placed too much responsibility on the part of the contractor to negotiate requirements, and that the process eventually broke down. In the case of NPOESS, we reported in the early phases of the program that the Air Force and the National Oceanic and Atmospheric Administration had difficulty resolving diverging requirements. In the case of AEHF, we reported that DOD substantially and frequently altered requirements and design in the early phases of the program. While considered necessary, some changes increased costs by hundreds of millions of dollars and caused scheduling delays on a program that DOD was trying to accelerate in order to address a potential capability gap. DOD has since rejected the acquisition approaches that led to requirements-setting problems on both SBIRS-High and AEHF. It has also instituted control mechanisms to regulate requirements on SBIRS-High. In our testimony, we noted that DOD could take further steps to strengthen requirements setting by implementing processes and policies, as needed, which stabilize requirements for acquisitions, like NPOESS, that are being shared with other agencies.

We have also reported on programs that took on unrealistic or potentially unaffordable requirements. The SBIRS-Low program's pursuit of discrimination capability is an older example of such a program. More recently, we have pointed to affordability and feasibility issues related to Space Radar and the TSAT programs, which together, have been preliminarily estimated to cost about $40 billion. Specifically, we have stated that DOD was planning to start these acquisitions even when many of their critical technologies were still immature, and it was pursuing a highly ambitious path in terms of the technology push. Given that these systems were among the most complex programs ever undertaken for space, they were being counted on to enable wider DOD transformation efforts, and DOD was already contending with highly problematic space efforts, we believed that DOD could not afford to pursue such risky approaches for TSAT and Space Radar. As noted earlier, DOD has taken steps to ensure it is pursuing realistic requirements for TSAT, and it has informally committed to do the same for Space Radar.

Question: Is there a clear definition of each Technology Readiness Level (TRL) that all of you agree on (GAO and DOD) and that exists in writing and that clearly applies to space programs?

The National Aeronautics and Space Administration (NASA) developed the original ranking and definitions of technology maturity levels. GAO and DOD agree on the TRL definitions—in its reports, GAO continues to reference the TRL scale for assessing critical technologies from DOD's *Interim Defense Acquisition Guidebook* (app 6, dated October 30, 2002). However, for space system acquisitions, GAO and DOD have disagreements on what the TRLs should be at major decision points. According to our work on best practices, product development should be initiated after critical technologies have been incorporated into a system prototype and tested in an operational environment—meaning the cold-radiated vacuum of space. Our prior reports have recognized that space systems are uniquely difficult to test in a true operational environment. However, DOD has found ways to test sensors and other critical technologies on experimental satellites. Nonetheless, DOD continues to stand up formal space system acquisitions too early—before critical technologies have been tested in operational or relevant environments—that is, before DOD has assurance that the capabilities it is pursuing can be achieved. This causes DOD to extend technology invention to its acquisitions, which have reverberating effects and require large amounts of time and money to fix. In these cases, DOD points to its National Security Space Acquisition Policy, which allows it to take such an approach—unlike DOD's acquisition policy for non-space acquisitions, where TRL 7 (testing in an operational environment) is preferred before product development is initiated (TRL 6 is required). As long as GAO continues to base its reviews of space programs on best practices and DOD continues to use the wide leeway afforded in its space acquisition policy regarding critical technologies and their maturity levels to initiate product development, GAO and DOD will continue to have disagreements in this area.

Question: What is the difference between TRL 6 and 7 and what is the advantage or disadvantage of being at level 6 or 7 at the [Critical] Design Review?

The main difference between TRL 6 and 7 is the testing environment. For TRL 6, the testing environment would be a laboratory or a simulated operational environment, and for TRL 7, the testing environment would be an operational environment—meaning in space. According to GAO's work on best practices, achieving a high level of technology maturity at program start is an important indicator of whether available resources in terms of knowledge, time, money and capacity match the customer's requirements. In addition, the key measure for a successful critical design review (CDR) is when 90-percent of the design drawings have been submitted to manufacturing. When space programs reach CDR and TRLs are below 6, it is unlikely that a high percentage of design drawings would have been released to manufacturing, thereby increasing program risk at this juncture. Another key point to remember is that CDR is the point at which programs begin ordering long-lead parts to build the first few satellites. This investment in hardware is at risk if the technologies do not prove out to work as intended. Achieving TRL 6 or 7 by CDR is a matter of risk—if the critical technologies in question are supremely important and have no space-based heritage, then it is warranted to test the technologies in space before proceeding through CDR. For TSAT, some critical technologies have a heritage of being tested or operated in space, and they are all slated to be at TRL 6 at the time of CDR—an approach that GAO did not fault.

Question: The Transformational Communications Satellite program, though still very early in the process, appears to have begun to adopt some of the recommendations of the GAO as well as the Young Panel and is focusing on technology maturity. Integration of the satellite appears to be the next difficult step for the TSAT program. What plans are in place to ensure successful integration?

The TSAT program is taking several steps to ensure its integration efforts are successful. First, according to program officials, the plan is to demonstrate critical technologies at TRL 6 when key integration tests are conducted in fiscal year 2007. Second, the program plans to use the results of its first round of integration tests to refine the testing to be conducted during a second round of more comprehensive integration testing. Third, the program is conducting a series of independent tests to verify results of contractor testing as it incrementally builds toward the two main integration tests facing the program—tests of the Next Generation Processor Router and Optical Standards Validation Suite. The program office plans to have knowledge on how these two major subcomponents work to reduce risk by uncovering technical problems before awarding the space segment contract for the design and assembly of the satellites. Finally, the TSAT program also plans to assess the results of the main integration tests before making a decision to enter the production development phase.

Question: What actions would you recommend to the programs managers to ensure successful integration?

According to GAO's prior work on best practices, leading firms ensure that (1) the right validation events—tests, simulations, and other means for demonstrating product maturity—occur at the right times, (2) each validation event produces quality results, and (3) the knowledge gained from an event is used to improve the product. Fully disclosing the results of tests (from low-level brass board tests to the main integration tests) and documenting the actions taken to address shortcomings further validates product knowledge. It is imperative that problems are fully addressed before rushing efforts to begin the next round of testing. It is also important that program managers use the test and evaluation parameters originally established, and any changes should be fully disclosed along with the reasons for doing so. Finally, the program manager needs assurance that all testing that has been done is reflective of the capabilities that the program is trying to deliver. Rigorous and sophisticated testing early and often will uncover problems when they are relatively easy and inexpensive to fix. Waiting too long to fully stress and test components will put the program in a risky position.

In preparing answers to your questions, we relied on our prior work on DOD's space acquisition policy, best practices in weapon system acquisitions, and our reviews of specific space acquisitions as well as DOD studies. In addition, for specific space systems development and cost growth, we relied on our annual assessment of selected major weapon programs. Because we relied on previously issued work, we did not obtain comments from DOD on a draft of this letter. We conducted our work from April 2006 through May 2006 in accordance with generally accepted government auditing standards.

We are sending copies of this letter to the Secretaries of Defense and the Air Force and interested congressional committees. We will also make copies available to others upon request. In addition, the report will be available at no charge on the GAO web site at http://www.gao.gov.

If you or your staff have any questions concerning these comments, please contact me at (202) 512-4841.

Sincerely yours,

Cristina Chaplain
Acting Director
Acquisition and Sourcing Management

cc: cc list

(120561)