System-Level Testing in Operational Environments

I. Dennis Rapport, Gerald W. Balkcom, Charles R. Stirrat, and Ronald L. Wilson

As typically implemented, systems engineering and operational test and evaluation are processes for ensuring that successful systems are delivered to users. Often, little consideration is given to supporting the test and evaluation of deployed systems. Some complex and mission-critical systems, however, require continuing test and evaluation programs, after deployment, to provide understanding of the systems' ability to respond to the complete range of environmental and operational parameters. Increasingly, developed systems require a means of planning to support modifications, life-cycle extensions, and responses to changing threats and environments. We describe the types of systems for which system-level testing in operational environments is appropriate, examine the unique requirements for evaluating deployed systems, and describe some test and evaluation programs designed by the Applied Physics Laboratory for these systems.

INTRODUCTION

Systems engineering is frequently viewed as the discipline necessary for the successful development of complex systems. The processes implicit in systems engineering are equally essential for the successful deployment and operation of these systems. This article addresses the contribution of system test and evaluation (T&E) to deployed systems and describes some programs where the approach has been successful.

System T&E, as embodied in major DoD programs, emphasizes product development. Considerably less focus is placed on the test and evaluation of deployed systems, and DoD policies have no explicit requirement for assessing these systems for a sustained period after deployment. The T&E programs in major acquisitions are designed to support milestone decisions, that is, whether acquisition should be continued, abandoned, or modified. Developmental tests address technical design issues, and operational tests answer questions concerning the system's ultimate employment by end users (e.g., military forces) rather than by designers. The traditional operational test program consists of early operational assessments and a final series of tests supporting the initial operational capability and deployment decision.

Figure 1 presents a time line of a system's life cycle. During the early phases of development, system capability increases steadily toward more refined performance. Developmental and operational test programs specified for the acquisition phase ensure that performance goals are being met and provide data for correcting problems or responding to changing needs. For most fielded systems, no such formalized T&E program exists, and changes in capability may go undetected. Test programs for fielded systems yield performance measures that identify degraded capability and support decisions regarding alternative uses, employment in response to changing threats, and upgrades. Although all deployed systems are tested at some level, there are

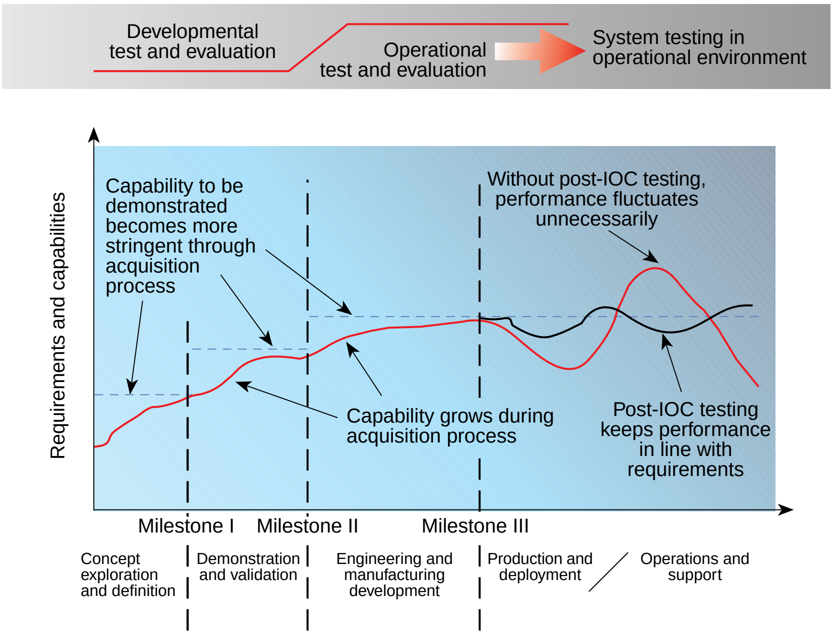

Figure 1. Testing during the acquisition process ensures that the system capability being demonstrated will meet requirements (IOC = initial operational capability).

those high-value systems that require a dedicated, integrated system T&E program that will continue throughout the life cycle.

SYSTEMS REQUIRING T&E AFTER DEPLOYMENT

After being fielded, military systems may be assessed through developer surveillance programs that track component reliability or through full-system tests conducted by end users in the operational environment. More frequently, however, testing conducted after deployment is intended to evaluate personnel and support training. The focus of such testing is on compliance with doctrine. Testing is frequently undertaken as part of the workup before overseas deployments (regularly scheduled rotations), but again, the focus is on equipment readiness and personnel training.

The Ship Antisubmarine Warfare Readiness/Effectiveness Measuring Program is one example of a comprehensive testing approach that includes data collection and analysis of special exercises, maintenance of a computerized database, and support of studies based on cumulative experience. However, the emphasis on training and system employment is clear; in particular, program goals include developing solutions for tactical problems connected with the application of surface antisubmarine warfare (ASW) systems, producing recommendations for effective use of weapons and sensors, and making improvements to tactics.

Other postdeployment test programs have a similar focus. For example, a review of the instructions for the tactical readiness evaluation required of each submarine periodically shows a clear intention to evaluate the crew's ability to effectively use their sonar and tactical weapon systems.

When tests for an operational system stress familiarization and proficiency training, little consideration is given to the system's capability and performance. Limited data collection is planned, and even less evaluation is performed. For highly integrated systems that are tested in this way, the lack of control over system variables can make it impossible to estimate tactical performance by extrapolating from the test environment. Depending on the system's complexity, the only realistic test option may be to design a dedicated T&E program that focuses on assessing system capability in varying environments, under changing concepts of operations, and with changing system capabilities.

Many systems may not require extensive T&E programs. If a system does not exploit risky or advanced technologies, if it does not play a critical role in the success of a military mission or campaign (because of redundancy with other systems or units), and if it is not highly integrated with other forces and systems, little may be gained from an expensive and extensive T&E program during the deployed life of the system.

Figure 2 illustrates how T&E scope and sophistication are correlated with system complexity and the importance of the mission. In this figure, "Opportunity T&E" describes the testing that is accomplished on the simplest of systems.

At the other end of the testing requirements spectrum are systems that involve a high degree of technical risk, are critical to the success of a military objective, and have complex components and interoperability requirements. Moreover, the performance of these systems may depend on a variety of environmental or operational parameters that can never be fully tested in all relevant combinations. For these systems, a dedicated program, augmented by high-fidelity models and simulations as well as subsystem testing and experimentation, is often required. In addition, there may be few other options to answering critical issues in support of strategic planning and making decisions about system employment changes, modifications, and upgrades.

Although our emphasis is on complex, high-value systems, a need for T&E focused on other deployed systems also exists. As budgets tighten, fewer completely new systems will be developed. Evolutionary change

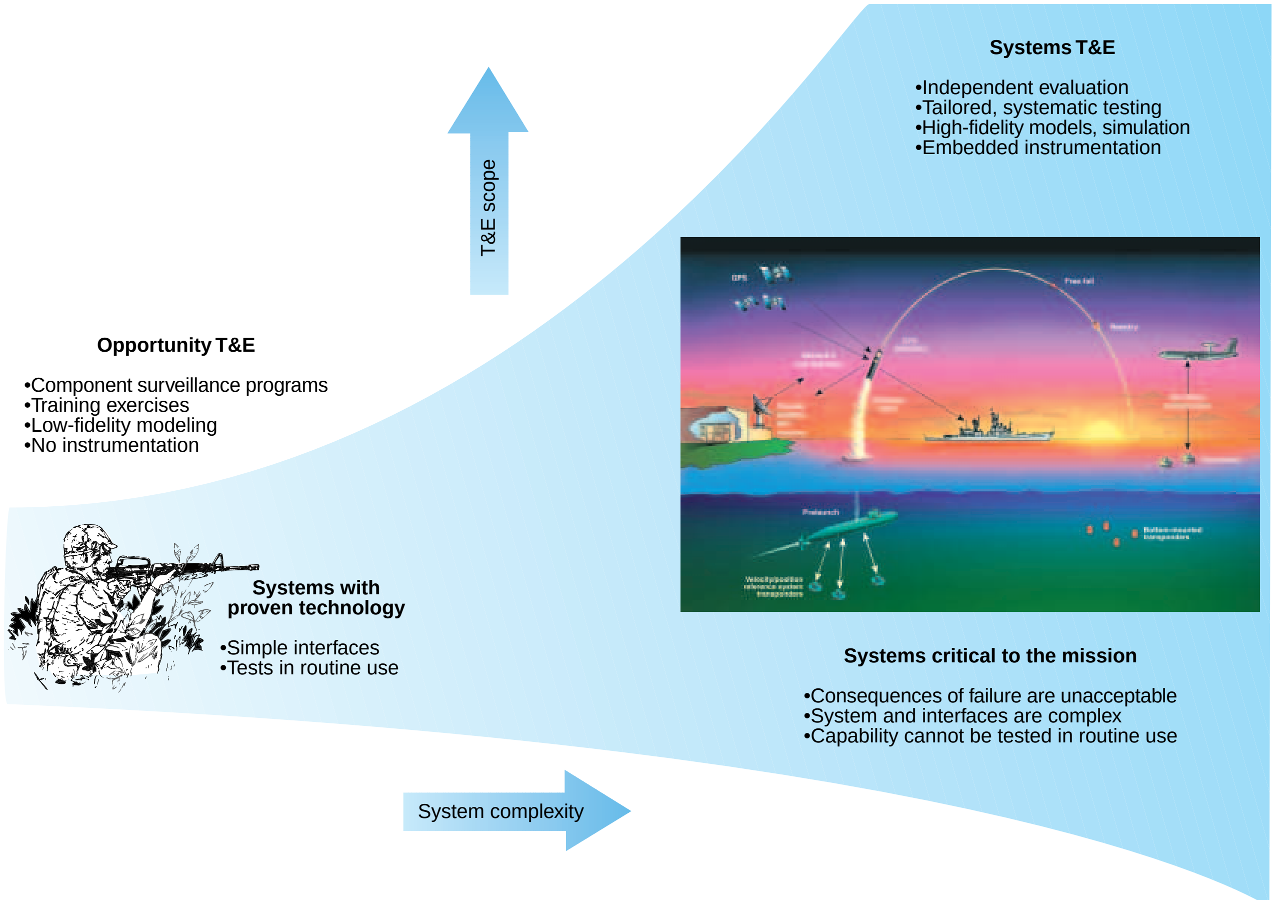

Figure 2. Fielded systems that are complex, are untested in routine use, and have unacceptably high consequences of failure are candidates for a systems T&E approach. (An expanded view of the inset is shown in Fig. 4.)

in response to new technologies or emerging threats is likely to be the rule. One inexpensive response to a new threat is to be able to state with confidence that the systems already deployed can meet that threat with only a change in concept of operations or minor modifications. Results of a T&E effort that support such an assessment will provide one of the only clear means to make this judgment.

SPECIAL REQUIREMENTS FOR TESTING FIELDED SYSTEMS

Test and evaluation provides the basis for making decisions about a system. The more important the question, the more confidence in our information we need to make a decision. In some sense, decisions required after deployment are more complex than those made during acquisition. During acquisition we ask if the system has attained technical performance specifications, and if it is operationally effective and suitable for intended use. After deployment we also want to know how well the system will work against a different threat and in a different environment as well as if it is cost-effective to modify the system for a new threat or environment.

In a major acquisition program, the purpose of an operational T&E undertaking is to determine whether a system meets the threshold of specific requirements against specific threats. These requirements are usually static in nature, although an increasing demand exists for reevaluating them throughout system acquisition. In contrast, for deployed systems, the critical issues are often poorly defined, and the answers are predictions of future performance in a changing environment against an uncertain threat. In either case, a credible T&E program is necessary to provide answers with confidence. For deployed systems, however, the greatest uncertainty in knowledge of system capability occurs when T&E capability is at its lowest.

Figure 3 shows the elements of a T&E program applicable to a deployed system. Although the indicated flow of information is generally left to right, in practice there are feedback paths among almost all of the processes. For deployed systems, all of the blocks in Fig. 3 undergo continuous change. For example, if the threat and mission concepts for a system change, the assessment of system effectiveness may change, independent of the collection of new test data. Similarly, as the interpretation of test data changes our underlying knowledge of system performance, we will revise our

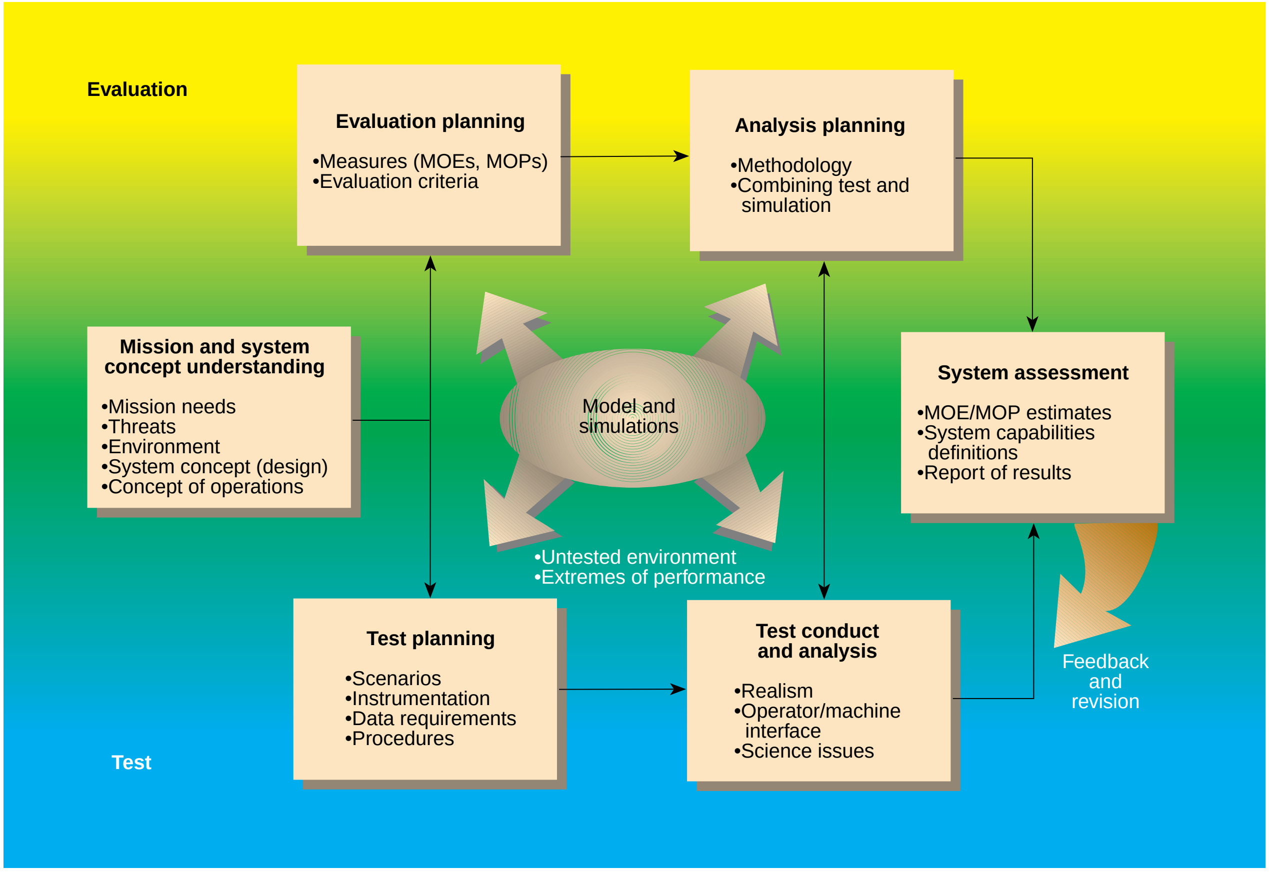

Figure 3. Systematic deployed system assessment is characterized by testing that is both representative of tactical situations and responsive to science issues. Testing is driven by evaluation programs that produce useful measures of performance and effectiveness (MOPs and MOEs, respectively).

predictions of system effectiveness in untested environments and situations.

ESSENTIAL FEATURES OF FULL-SYSTEM EVALUATION IN THE OPERATIONAL ENVIRONMENT

Effective system-level testing in the operational environment involves four key features.

1. An integrated mixture of testing (equipment and subsystem surveillance, representative free play with complete systems, environmental measurements, and component-level engineering data collection) supported by instrumentation, an evaluation process, and a mission-based assessment
2. Validated models and simulations to leverage test results into assessments of system performance in untested or infrequently tested areas
3. Instrumentation that is built into the system design
4. A fundamental understanding of the system, the environment in which it operates, and the interactions among components, subsystems, and the environment

These features strengthen the view that postdeployment testing should support more than pass/fail decisions. We can reach an appropriate level of understanding of system performance that allows us to extrapolate into untested areas or respond to changing plans and emerging threats. The T&E process is a continuum of increasing understanding of the system, in which detailed analysis of system success is as important as fault isolation after system failure.

To illustrate the advantages of this systems engineering approach to the testing and evaluation of deployed systems, we describe the T&E programs of two Navy systems—the Fleet Ballistic Missile/Strategic Weapon System and the passive sonar suite of SSBNs (nuclear-powered fleet ballistic missile submarines).

The Submarine-Launched Ballistic Missile (SLBM) Strategic Weapon System

The Joint Chiefs of Staff requires the Navy to provide credible, quantitative information for strategic target planning. The Navy developed, with APL providing primary support, a T&E program that continues throughout the life cycle of the weapon system.

Instrumentation is integrated in all subsystems. Annual flight test programs are planned, and instrumented missiles are flown from test ranges with special impact areas to evaluate in-flight performance and accuracy. This test program includes significant hallmarks.

- To supplement system-level testing, a comprehensive test program is conducted at the subsystem level to bolster statistical confidence in the performance of critical elements.
- Modeling and simulation is a major contributor to test design, instrumentation specification, and extrapolation of test results to untested environments.
- Instrumentation is designed into the system; all aspects of the weapon system contain built-in instrumentation that is deployed with the strategic assets. This approach provides an early understanding of system performance. Supported by detailed system understanding, the need for follow-on testing is reduced.
- The evaluation requirements drive the test program—test objectives, test design, and test sizing are specified with respect to integrated performance measures.

Instrumentation for the newest SLBM weapon system (Trident II) records on board real-time data, flight telemetry/Global Positioning System data, and impact area sonobuoy detections. Figure 4 shows a schematic of the instrumentation support for a Trident II missile launch. Data from a test flight can be used to identify contributions to miss distance down to the subsystem level.

The Trident II Accuracy Model is the analytical tool that combines subsystem and system data from flight tests to validate fundamental system accuracy parameters and provide estimates of overall system accuracy. During system development, developers and evaluators collaborated under the Navy's Strategic Systems Programs to model Trident accuracy to the component level and to develop a propagation model to combine the elements. This model is now used routinely in system accuracy assessments and is the preeminent tool for identifying the accuracy of system design or employment changes. Our reliance on this model continues to increase as the numbers of missile test ranges as well as trajectory options available to the Navy for direct systems accuracy tests have decreased. Figure 5 shows how the Accuracy Model is used to "systems engineer" the results of flight tests by decomposing missile reentry body impact results from a test flight to component accuracy contributors.

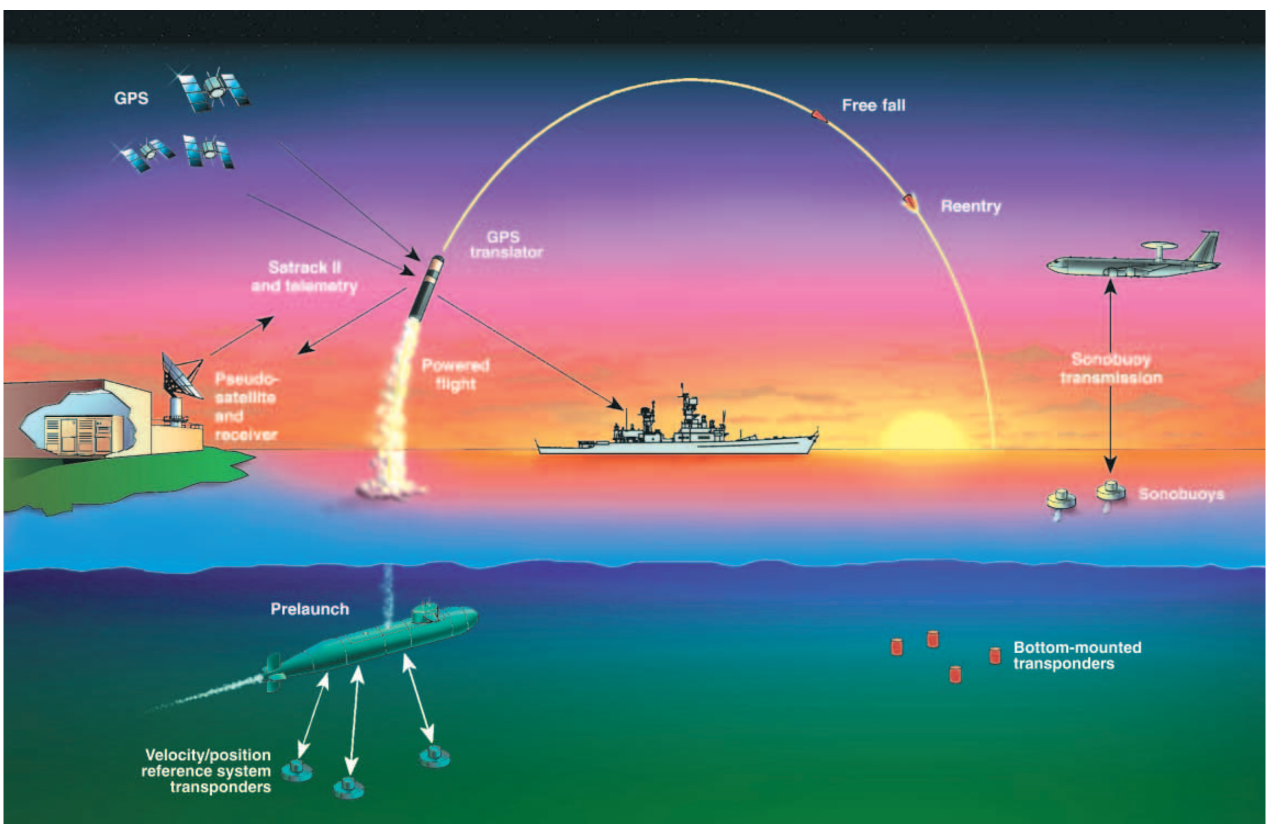

Figure 4. Evaluating the capability of the Strategic Weapon System requires instrumentation covering launch, boost, free fall, and reentry, complemented by sophisticated processing and evaluation methods. Onboard real-time recording includes data on navigation, missile launch-tube environment, missile/guidance initialization, major interfaces, system events, and voice communications. (GPS = Global Positioning System; white lines and arrows denote acoustics, black represent telemetry, and yellow show trajectory.)

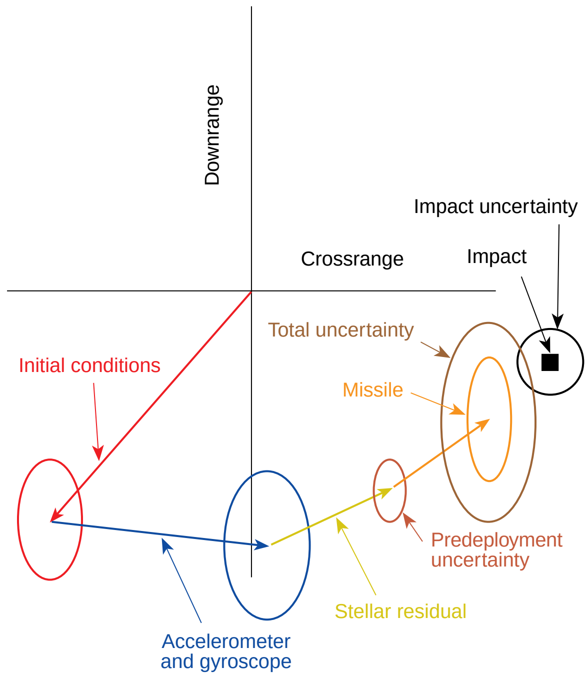

Figure 5. Identification of sources of miss distance, even for successful tests, allows identification of compensating errors that could adversely affect other scenarios (total uncertainty includes modeled reentry aerodynamics uncertainty).

The attention to fundamental knowledge of the phenomenology of system performance is central to a successful T&E program. For the Navy's Strategic Weapon System Program, this knowledge is gained through an integrated test plan that combines the results of an assortment of tests to provide statistical significance to system assessments that complement the solid understanding of the underlying physics of system performance. These assessments range from end-to-end system tests (including missile firing) to a range of system and subsystem tests that collect data on various components of the weapons system and evaluate important interfaces. Most of these tests serve a dual role, accommodating the requirements of the integrated test program while providing training for systems operators.

The data provided for validation of models and simulations constitute an important aspect of the integrated test program. Modeling and simulation has often been heralded as the primary complement to live testing for estimating a system's performance. During the acquisition process, for example, it is cheaper to explore design alternatives by computer simulation than by building and testing prototypes. Without an adequate body of test data to support model validation and verification, however, there can be little credibility in predictions based solely on these evaluation tools. For that reason, modeling and simulation plays the central role in coupling tests of live systems with evaluations of the performance capabilities of those systems.

In the Strategic Weapon System Program, many of the tests intended to support estimations of systems capability are augmented by test phases whose sole purpose is to validate simulations or to provide data for constructing credible models. For example, consider the automatic depth control system whose purpose is to maintain the submarine at a nearly constant depth during a ripple launch. The ability to launch 24 missiles without ship control delay cannot be demonstrated, so a ship control simulation provides the basis for establishing this capability. Tests to support these simulations include "closed-loop" launches in which the depth control system operates automatically to transfer ballast water into and out of the submarine. In addition, some launches are conducted in "open-loop" mode with sea valves closed. These tests provide data to estimate launch forces on the submarine to be used as inputs to depth control simulations. Although the open-loop launches are not representative of actual system application, they provide the best data for estimating the complex forces acting on the submarine as the missile is ejected and seawater floods back into the vacated launch tube (Fig. 6). A physics-based submarine ship control simulation is used to estimate these forces.

The Trident Passive Sonar Suite

The Strategic Weapon System exemplifies a system for which a comprehensive life-cycle and valuation

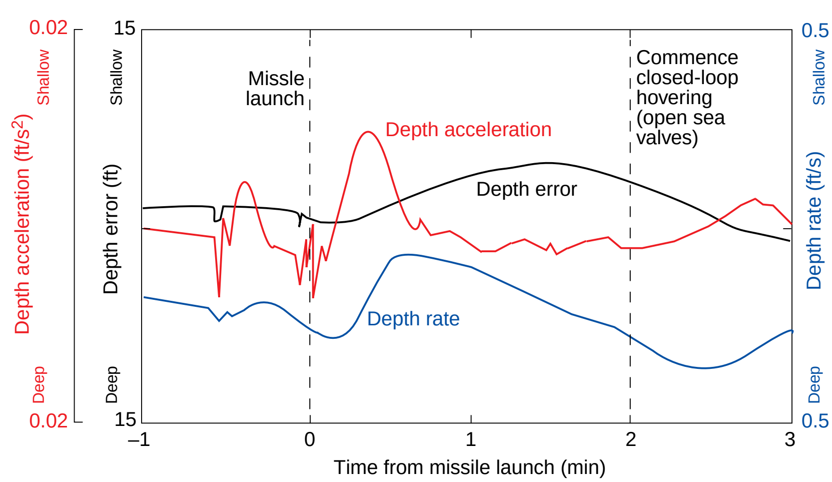

Figure 6. To validate depth control system models, the missile is launched from a submarine with the depth control computer operating but sea valves closed. The computer output, shown here, provides data for deriving launch forces on the submarine and for validating the complete depth-keeping simulation.

program was envisioned during the design phase. The test program for the SSBN passive sonar suite was established after the system had been deployed. Nevertheless, it is an ambitious program involving many test types and a sophisticated instrumentation suite supported by elaborate postprocessing capability and evaluation methods.

The Navy and APL have fitted selected submarine sonar systems with instrumentation to record raw and processed hydrophone data for postpatrol processing and analysis. The recording system interfaces with the tactical sonar suite and records up to 1300 sonar acoustic inputs (6.25 MB/s) continuously during an entire deployment. These data are then processed at the Laboratory on state-of-the-art sonar processors (Fig. 7), which provide flexible beamforming as well as user-selectable displays.

We use a mix of tests and integrate the results into a complete assessment of sonar capability. These tests include an evaluation of sonar performance against selected contacts of opportunity chosen from those available during routine patrol operations. We also use data from security exercises, during which submarines are ordered to operate together in an ASW scenario in which one plays the role of the pursuer and the other the role of the evader. The test program is rounded out with sonar demonstration and shakedown operation exercises in which the exercise geometries may be controlled to provide data for special investigations. Some parts of the controlled exercises are usually reserved for collecting data for model validation (e.g., to establish a baseline for modeling acoustic propagation).

The test program for deployed Trident sonars typifies the way in which developing a thorough understanding of system capability supports planning and decision making when faced with a changing threat. Twenty years ago, when the program was conceived, the threat to U.S. strategic assets consisted of Soviet SSNs (nuclear-powered attack submarines) that were generally noisy and inefficient compared with our submarines. Since then, the successors to those Soviet SSNs are the submarines of the modern Russian Navy. Despite the collapse of the Soviet Union, these new submarines are highly capable, with quieting levels that rival some U.S. submarines. Our understanding of the capability of U.S. strategic assets to tackle a threat as capable as these Russian submarines comes from detailed system knowledge, stemming from a test program that concentrates on quantifying performance factors throughout the system's life cycle rather than establishing whether it has passed preestablished test criteria.

CONCLUSION

The role of T&E during the acquisition cycle is well established. We have identified a class of defense systems that require a credible understanding of performance and a continuing test program to support planning, system modifications, and responses to changing threats. Because we are interested in more than test thresholds, the evaluation of test data concentrates on understanding system performance at the subsystem or even component level. Accurate assessments of the system's interaction with its environment are also necessary. Finally, the test program should support model development and validation.

A full-scope integrated test program is not appropriate for all systems. For some critical systems, a continuing evaluation of system capability ultimately saves the cost of planning military campaigns in the face of unnecessary uncertainty, or developing follow-on or replacement systems without knowing how current systems might perform in new or modified roles. Increasingly, however, other developed systems require planning to support system modifications, life extensions, and responses to changing threats. The core components of this form of T&E provide another potential tool in the systems engineering process of developing, fielding, and modifying these systems.

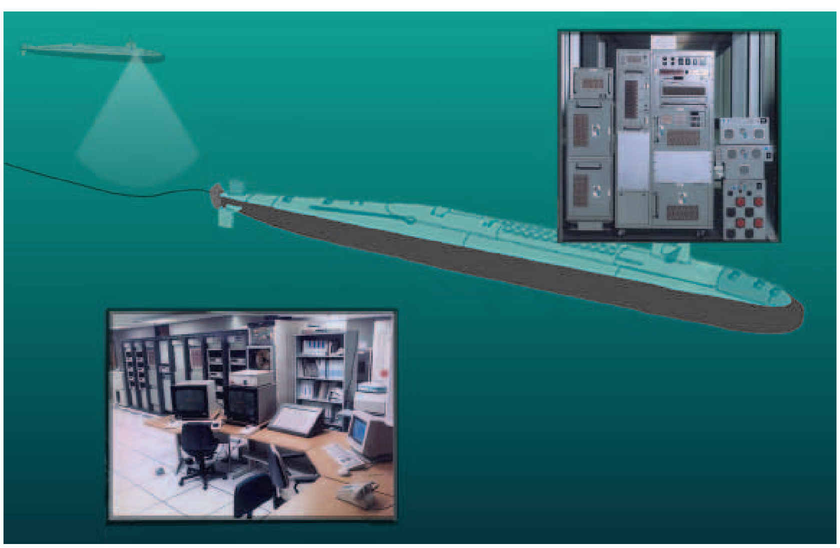

Figure 7. To provide an independent assessment of the capability of Trident sonars, hydrophone outputs are recorded throughout a patrol. These data are analyzed to establish baseline capabilities to compare with demonstrated performance (top right, acoustic recorder; bottom left, Signal Processing Laboratory).

ACKNOWLEDGMENTS: Kevin Curtiss supplied the information supporting the illustration of model validation for an open-loop missile launch from a submarine. Lee Simkins and James Huneycutt, Jr., supplied the information illustrating the use of models and instrumentation to identify sources of miss distance.

THE AUTHORS

I. DENNIS RAPPORT received a B.S. degree in systems and automatic control from Washington University in 1968 and an M.S. degree in system engineering from Case Western Reserve University in 1969. After a 3-year tour in the Navy, he joined APL's Strategic Systems Department. Mr. Rapport has been involved with the evaluation and simulation of submarine depth control systems, the development of targeting performance factors for submarine-launched ballistic missiles, the test and evaluation of submarine sonar systems, and the test and analysis design for mine countermeasures. He is a member of the Principal Professional Staff and is currently supervisor of the Undersea Systems Evaluation Group. His e-mail address is Ian.Rapport@jhuapl.edu.

GERALD W. BALKCOM is a member of the Principal Professional Staff of APL's Strategic Systems Department and supervisor of the Prelaunch Systems Group. He received a B.S. in physics from Oklahoma State University in 1967 and an M.S. in computer science from The Johns Hopkins University in 1981. He joined APL in 1973 after 2 years in the Army and 3 years with the Seismic Interpretive Center of Texaco, Inc. Since joining APL, he has developed performance analysis requirements for submarine-launched ballistic missile fire control systems and has conducted operational effectiveness evaluations for deployed systems as well as systems in development. He is a member of the International Council on Systems Engineering and the International Test and Evaluation Association. His e-mail address is Gerald.Balkcom@jhuapl.edu.

CHARLES R. STIRRAT is a member of the Principal Professional Staff of APL's Strategic Systems Department. He received a B.S. in physics and mathematics from the University of Wisconsin—River Falls in 1968 and an M.S. and Ph.D. in physics from Michigan State University in 1970 and 1974, respectively. After joining APL in 1974, he spent 16 years performing tests and evaluations of submarine-launched ballistic missile fire control systems as a member and supervisor of the Strategic Systems Department's Fire Control Group. From 1990 to 1995, he evaluated Strategic Defense Initiative Organization Brilliant Pebbles flight tests and supported the Ballistic Missile Defense Test and Evaluation Program. He is currently the supervisor of the Systems Studies and Analysis Group. Dr. Stirrat's e-mail address is Charles.Stirrat@jhuapl.edu.

RONALD L. WILSON is a member of the Principal Professional Staff of APL's Strategic Systems Department (SSD). He received his B.S.M.E. from Virginia Polytechnic Institute and State University in 1962. He joined the Laboratory in 1966 after completing a tour in the Air Force and working at the Goddard Space Flight Center. Mr. Wilson has served as a missile systems engineer in SSD, and previously managed the department's Polaris/Poseidon Program and the Fleet Ballistic Missile Patrol Program. He is currently the supervisor of the SSD Test Evaluation Branch. He has taught several technical management courses in The Johns Hopkins University G.W.C. Whiting School of Engineering since 1987. His e-mail address is Ronald.Wilson@jhuapl.edu.