The ITEA Journal

December 2008
Volume 29, Number 4

Autonomous and Cognitive Systems

Published quarterly by
the International
Test and Evaluation
Association

CALCULEX

Electronics, Systems, Software and Support

Remember...

...the rotary phone?

...the TRS-80™?

...the Cabbage Patch Kids®?

...Disco?

...tape recorders?

... to move up to the most reliable, fastest and affordable solid state recorder technology...

MONSSTR®

info@calculex.com www.calculex.com (575) 525-0131

CONTENTS

TECHNICAL ARTICLES

DEPARTMENTS

ON THE COVER: Robots have become staples of assembly lines, and simple autonomous systems will vacuum for you; photos of sophisticated unmanned aircraft deployed to Afghanistan and Iraq have appeared in all of the news media. As such capabilities evolve, future civilian and military systems will become networked autonomous systems, homogeneous then heterogeneous, and finally networked cognitive systems. Reconfigurable and cognitive networks of sensors also offer the possibility of adaptable range instrumentation. Testing networks of autonomous and cognitive systems will severely challenge existing infrastructure, processes, and even safety practices.

■ ITEA Headquarters: 4400 Fair Lakes Court, Suite 104, Fairfax, Virginia 22033-3899; Tel: (703) 631-6220; Fax: (703) 631-6221, E-mail: *itea@itea.org;* Web site: *http://www.itea.org.*

■ ITEA is a not-for-profit international association founded in 1980 to further the development and exchange of technical information in the field of test and evaluation.

■ *The ITEA Journal* (ISSN 1054-0229) is published quarterly by the International Test and Evaluation Association at 4400 Lakes Court, Suite 104, Fairfax, Virginia 22033-3899. Single issue cover price for *The ITEA Journal* is $20. ITEA membership dues are $45 for individuals, $25 for full-time students, and $745 for corporations. Annual dues include a one-year subscription to *The ITEA Journal*. The annual subscription rate for libraries and other organizations providing timely reference material to groups is $60. All overseas mail (air mail or AOA) requires an additional $20. *The ITEA Journal* serves its readers as a forum for the presentation and discussion of issues related to test and evaluation. All articles reflect the individual views of the authors and not official points of view adopted by ITEA or the organizations with which the authors are affiliated.

POSTMASTER: Send address changes to: ITEA, 4400 Fair Lakes Court, Suite 104, Fairfax, Virginia 22033-3899.

The ITEA Journal
December 2008
Volume 29, Number 4

ITEA Journal 2008; 29: 323–324

As 2008 comes to a close, the ITEA Events Committee is gearing up for a new year focused on meeting the evolving needs of the test and evaluation (T&E) community. I would like to formally thank the many volunteers who worked tirelessly to host or support events this year. I certainly believe it is worth reiterating that the success of this organization lies with the chapters and their volunteers who work collaboratively to make the ITEA vision a reality. As a volunteer organization, our members are our greatest asset. If you have not been involved in local or national events and have an interest in doing so, I urge you to contact a member of the leadership (locally, regionally, or nationally), myself included. I am confident that you will find the volunteer experience rewarding, both professionally and personally.

We made some changes to our 2008 programming to meet the strategic goals of the organization—to attract more non-U.S. Department of Defense (DoD) participation and to provide forums on topics that reflect the current T&E hot topics. The first was the Test and Evaluation in an Age of Asymmetric Warfare workshop hosted by the ITEA Huachuca Chapter in March. The unique workshop format encouraged attendees to roll up their sleeves and address the multitude of issues the Age of Asymmetric Warfare brings to the T&E community. Specifically, panels examined a variety of test methods, resources, infrastructure, and new technologies required to support the T&E of rapidly evolving systems supporting our warfighters. Topics addressed included the T&E of computer network attack (CNA) applications; unconventional attacks (e.g., the chemical, biological, radiological, and nuclear environment, and T&E of barrier systems being employed in a Homeland Security role); and counter-improvised explosive devices (C-IED), infrastructure, and specialized instrumentation requirements at the test ranges.

Another hot topic forum was the T&E of Unmanned Aircraft Workshop hosted by the ITEA Southern Maryland Chapter in cooperation with the Patuxent River Chapter of the Society of Flight Test Engineers. Unmanned Aerial Vehicles are increasingly becoming significant assets for DoD, Homeland Security, Border Patrol, and local Police and Fire Departments. The workshop had an overwhelming response including participation by many non-traditional ITEA attendees. We hope to continue to provide programming that captures the interest of a broader T&E community.


John Smith

Our Annual Symposium in November, Advancing T&E in the Global Community was phenomenal. The symposium agenda was geared towards increasing interest from non-Department of Defense participants and attendees. The expansion focus this year was the Federal Aviation Administration and Department of Homeland Security. The program included VIP speakers and panel chairs from these organizations. In order to address one of the objectives of ITEA, the symposium consisted of a panel which discussed the challenge of attracting younger members of our T&E community.

In 2009, we are continuing the press for organizational growth and to remain a viable mechanism for moving T&E into the future starting with the Live-Virtual-Construction (LVC) workshop in January. The White Sands Chapter has hosted a highly successful Modeling and Simulation (M&S) Conference for the past thirteen years. The 2009 workshop will contain the traditional components of modeling and simulation, but the revamped LVC theme will allow the introduction of new, relevant topics of interest in line with initiatives geared towards Testing in a Joint Environment.

The 2009 Annual Symposium will also be a mechanism for change and growth. As a new approach, we partnered with Leading Edge Events and Media of London and their showcase exposition, AeroTest America, to enable more opportunities for growth, especially in the international community, and to increase the number of visitors for our exhibitors. The 2009 Committee is excited about the prospect of developing a program designed to increase participation and networking opportunities. I hope that you join me in looking forward to watching this unique symposium unfold.

ITEA has had an exciting 2008, and I personally am looking forward to continued momentum in 2009 with workshops, symposium, and educational opportunities geared towards the evolving requirements of our important community. To ensure that ITEA remains effective in meeting the needs of its members and participants, the Board of Directors will be focusing considerable effort towards effective strategic planning. We are currently exploring the utilization of a strategic planning consulting firm to guide us through an evaluation of our existing Strategic Plan with regard to its relevancy in moving ITEA towards our desired future. Through a step-by-step process, we will revise the plan and its implementation components, as necessary, to ensure that the future of ITEA is in fact aligned to meet the needs of the T&E profession and its technology advancements in the years to come. Stay tuned for more information as we initiate this important process. I thank each and every one of you for your continued support of our organization. See you in 2009!

ITEA Journal 2008; 29: 325–327

Testing Unmanned Systems

C. David Brown, Ph.D., PE
Consultant
Havre de Grace, MD

The testing of unmanned, meaning remotely piloted or autonomous, aerial and ground systems is taking us into the world of testing artificial intelligence and testing operators as they interface with this artificial intelligence. These unmanned systems may not look very different from traditional combat systems, and often they support missions very similar to those of manned systems. Often, however, the missions, control, and operator interface are drastically different. These differences may dictate very different developmental and operational testing than that required for traditional manned systems.

C. David Brown

There seem to be two very different schools of thought on how to test unmanned systems. One view is that unmanned systems are usually designed to replace or augment manned systems, so they should simply be tested in the same manner to see if they can perform as their manned counterparts. The other view is that unmanned systems are very different "animals", as the operator is displaced or even replaced, and often these systems are deployed very differently than their manned counterparts. I strongly subscribe to the latter view. Therefore, I believe that we testers and evaluators have much work to do to modify and develop test procedures, develop test facilities, and develop evaluation methods and criteria to address the unique characteristics, operation, and missions of unmanned systems.

Unmanned systems can generally be classified into three categories—remotely operated, fully autonomous, and leader-follower. There are obviously combinations of the above, for instance many remotely piloted systems have some degree of autonomy to assist the operator and leader-follower systems have autonomous capability to handle the differences in the leader and follower situations. Generally, however, these three major bins are different enough that very different test and evaluation methods, facilities, and measures need to be developed and applied to each.

Let's first examine remotely piloted systems. This seems pretty basic, just relocate the operator station, and replace mechanical linkages with servo motors and wiring with radio links. Give the operator a video picture of what can be seen from inside the system, and we are good to go. The mechanics and electronics are now fairly mature, but there are three areas that are now just beginning to be examined. First, the environment and inputs to the operator are very different in the remote station than aboard the system. Even though the operator station may be the same, the operator's interface with the system now goes through a transfer function both ways (control from operator to the system and feedback from system and system's environment to the operator). This drives very significant changes in operation, and thus operator training; the test instrumentation must be more extensive and very different; and finally, the developmental testing must include significantly more human factors testing to get the operator interface right. Second, the system characteristics can change significantly when they don't have to accommodate an on-board operator, for instance, they can be much smaller. Thus, these systems function and respond very differently from their manned counterparts. They likely will have very different or a much wider range of operational environments, thus dictating a wider range of test scenarios and environments. Third, the "operatorless" systems can be employed in very different situations and missions because of their system characteristics (e.g. smaller size) and no need to keep an operator safe. Thus, the tactics, techniques, and procedures; force development testing; and operational testing will all change substantially. Finally, throughout all testing, there are range safety issues with systems that could become disconnected from the operator, especially if these systems are weaponized.

Fully autonomous systems pose all of the above testing challenges, except perhaps that of the real-time operator interface. However, an operator must still tell the system what to do through a set of waypoints or other method entered in advance of the mission. The technical complexity of the navigation sensors and decision system increases significantly over remotely piloted systems, and thus, increases the complexity of the technical testing to support development of the system and evaluation of the operation of the navigation and guidance systems. This represents a new frontier in sensor, software, interface, and control sub-system testing. The missions for autonomous systems are even more varied and challenging, and thus, so is the operational testing. Range safety issues while testing now become paramount as there is no real-time operator, and these safety issues can become very significant for weaponized autonomous systems.

Leader-follower systems present all of the above challenges, as well as another set of testing challenges. A semi-autonomous system must follow a leader, which is a manned or another following system, all of which are performing a convoy type mission. The unmanned system receives or derives control commands from the leader system, and also performs some autonomous control to properly navigate given the difference between its situation and that of its leader. Again, we have a testing challenge. Somewhere in the chain is a manned leader system, and the operator of this system is essentially controlling an unconnected train of systems over or through an uncontrolled environment. Developmental testing needs to address how well the follower systems follow, but also must address how well the leader system leads. In fact, just like the typical dilemma of whether the problem is hardware or software, we may now have a leader versus follower performance issue. Operational testing must address the human factors issues involved with the operator essentially driving vehicles that he or she is not really driving. In addition, the above described range safety issues apply to the operatorless follower vehicles.

Unmanned vehicles are thrusting us into the realm of artificial intelligence on the battlefield that futurists have been describing for many years. Until recently, robotic systems have been confined to very structured environments such as the factory floor or warehouse, and even outer space. Now, robotic systems are taking the leap from these highly structured environments to the most unstructured environments we have, the natural environment clouded with the chaos of combat. Testing is all about discovering "unknown unknowns." We do this by exercising structured missions while immersing a system under test into a test environment that is as close as we can create to the natural combat environment. Therefore, we are testing a system with structured intelligence as it exercises a structured mission in a somewhat structured environment. It seems to me that there may be a flaw in there somewhere. We do the same with manned systems, but we depend on human intelligence and "Yankee ingenuity" to adapt to the unknown and untested challenges of the actual natural combat environment. Unmanned systems lack this adaptability to untested challenges to some extent, as remotely piloted and leader-follower systems certainly hamper or degrade the interface between the human and the operational environment, and completely lack the adaptability in the case of fully autonomous systems.

While artificial intelligence systems cannot adapt to situations that they have not encountered before, they can be programmed to change parameters in response to past operator or environmental occurrences, and therefore accomplish some elementary "learning." A simple example is the current breed of automobile automatic transmission controllers that adjust to operator driving habits to shift gears more effectively. Undoubtedly, unmanned systems will incorporate such features to help assist the operator interface and improve control. These features create a new set of testing challenges, in that a system under test is essentially changing throughout the testing. Issues of repeatability, test repetitions, configuration control, and what and when to evaluate will certainly arise.

One could read the previous two paragraphs to say that we cannot effectively test unmanned systems. I am not saying that, but I do believe that we must take extra care to test unmanned systems through a full spectrum of challenges. The old adage of testing to find unknowns is far more important here. With no operator, the system only "knows" how to react to the knowns, those challenges envisioned by the designer of the artificial intelligence subsystem. We, of course, need to test to these knowns to assess and document the technical performance. However, tests which are structured to emphasize designed in abilities can be deceiving. For instance, recent operational testing of a small unmanned ground vehicle was conducted in an urban like environment where the buildings had no doors. Human scouts can open doors and traverse other such "obstacles," where the unmanned scout can't. Does this mean it is not useful? Not necessarily, but it does mean that its missions may be limited, or at least different than those of manned systems. More important, the test designers must do their best to envision the "unknowns," as the likelihood of these systems to be able to handle those unknowns is greatly diminished compared to traditional manned systems. While we rarely have the luxury of doing more testing, we certainly, now more than ever, must test more "smartly."

In this editorial, I have been talking about unmanned aerial and ground systems that replace or augment manned systems. I also have an Army perspective of ground and close in air support of ground operations. I realize that missiles have essentially been remotely operated and autonomous since their creation. I also realize that the Air Force and Navy have their high fliers such as Predator and Global Hawk. Although their operational environment, human interface, missions, and acquisition climate are very different from ground support aerial and ground systems; we can perhaps apply some lessons learned, M&S, test design, instrumentation, procedures, and analysis methods to the testing of these relatively new unmanned systems.

The testing of unmanned systems will challenge test designers, test facilities, and evaluators. Human factors testing must be taken to new heights as we increase and significantly change the operator interface and workload with these systems. Operational testing will change significantly when the military participants are in some cases more part of the test environment than they are the direct operators. It is a paradox that as we try to remove the human from the systems through autonomy and remote control, the human interface and workload become so much more complex. We now must instrument, test, evaluate, and understand the operator more than ever. While I admit that these new systems are "smarter" in that they have more artificial intelligence, they lack to some extent, the "smarts" of an operator. I therefore, contend that they are ultimately far less "smart" than their manned counterparts. In summary, these less "smart" systems dictate "smarter" testing. ❑

Dr. C. David Brown is a consulting engineer in the areas of systems engineering, acquisition, program management, and test and evaluation. He spent 34 years working in defense acquisition as an Army officer, civil servant, and senior executive. His most recent responsibility for the DOD was as the executive director for test and evaluation for the Army's Future Combat Systems Program. He has a BS, MS and Ph.D. in Electrical Engineering from the University of Delaware and a masters in National Security Policy from the National Defense University. He has two patents, one of which is for an autonomous land navigation subsystem, and has authored numerous technical papers on the subject. Email: David.Brown10@us.army.mil

Invited Article

ITEA Journal 2008; 29: 329–331

Some Musings on Test and Evaluation of Unmanned Ground Vehicles

Ellen M. Purdy

Joint Ground Robotics Enterprise,

Office of the Under Secretary for Defense (Acquisition, Technology, and Logistics), Washington, D.C.

The U.S. Department of Defense has come a long way in the last 5 years when it comes to fielding and employing unmanned ground vehicles. In 2003 with operations in Afghanistan, 3 Packbots were sent to the theater of operations to support cave reconnaissance. The systems worked so well that there were calls for more. Then came the idea of using ground robots to defeat improvised explosive devices. Experiences with these robots in the theatre proved such a success that today there are more than 6,000 unmanned ground vehicles in the inventory, the majority of them serving in Iraq and Afghanistan. No doubt about it, ground robotics are and forever more will be part of the materiel that enables our Service members to conduct their missions. With such a rapid adoption of this technology, most of which was acquired through rapid means, it is worth taking some time to consider whether the Department is fully prepared to test and evaluate the new robotic systems that will be called for in the future.

Key words: Autonomous tactical behavior; unmanned systems; hijacked system; rogue behavior; safety; reliability & mission duration; test metrics; trust; weight; speed.

Recently, Service experts with unmanned systems experience were engaged in an exercise to identify potential missions that could feasibly be conducted by unmanned systems during the next 25 years. For the ground domain, missions ranged from reconnaissance to casualty evacuation, range clearance, runway inspection, dirty bomb interrogation and disposal, tunnel exploration and mapping, and facility protection, to a myriad other missions. These experts also projected that in 25 years unmanned systems would likely advance from the teleoperation of today to fully autonomous tactical behaviors. They would likely be able to fully and autonomously team with manned systems and other unmanned systems across air, land, and sea.

Ellen M. Purdy

These technological feats are unprecedented in military history, so it begs the question of whether there is work to do when it comes to conducting appropriate and reliable test and evaluation (T&E) of these systems. The purpose of this article is to encourage dialog now, while unmanned systems are still in their infancy, particularly when it comes to full autonomy. Such dialog should take place not only among T&E personnel but also with users. As testers and evaluators wrestle with the difficult issues associated with the unique aspects of unmanned systems, they are in a position to assist users with sorting through appropriate measures of operational effectiveness.

Test considerations: T&E of unmanned systems is really no different for manned systems, or is it?

Unmanned systems, by their nature when they cross the boundary into fully autonomous behaviors, will be doing many of the perception and reasoning tasks that have up to this point always been done by people. Do we have all the right test methodologies to test these capabilities? Maybe, maybe not.

A case in point is a study published in January 2008 by the Massachusetts Institute of Technology (MIT) regarding the commonly adopted methodology for assessing the effectiveness of computer vision. According to MIT News:

> *"For years, scientists have been trying to teach computers how to see like humans, and recent research has seemed to show computers making progress in recognizing visual objects. A new*

MIT study, however, cautions that this apparent success may be misleading because the tests being used are inadvertently stacked in favor of computers." (Delude 2008)

When it comes to unmanned systems in combat situations, it is imperative we get our testing right so that we accurately characterize what these systems can and cannot do.

Given that T&E will have to mature right along with unmanned systems, or better yet at a faster pace so testing does not slow down technology development and system acquisition, with what exactly should testers and evaluators be concerned? The first thing that comes to mind is metrics and how they impact test infrastructure. The U.S. Department of Defense (DoD) test community is very good and very experienced at testing mobility of platforms on the ground. Consider this though: so far all the vehicles that have been tested regardless of speed and types of terrain have always been "governed" by the constraints associated with passengers. There are certain speeds and terrains these vehicles do not traverse because human bodies are not built for that kind of punishment. Our vehicle technology is capable of higher speeds and able to traverse harsher terrains as long as there is no human being to worry about getting dizzy, sick, or bruised. In our current test tracks, do we have the kinds of terrain roughness that unmanned ground vehicles (UGVs) will be able to handle across the next 25 years—at UGV speeds?

What about trust? What are our metrics and how do we measure a person's trust of unmanned systems? Twenty-five years from now there will be an entire generation that does not know what life is like without unmanned systems. Today, however, the size of the population on the planet that has worked with robots is so small it is buried in the noise. Let's face it, there are just not that many people who have had up close and personal encounters with robots. Should we not have a test for measuring a person's willingness to let a vehicle autonomously drive them from point A to point B? This would be good information to know before fielding a system so the training support package can take this into account. Experience to date indicates that willingness to trust a robotic vehicle differs widely between people. Some personnel become very comfortable quickly with being a passenger in an autonomous vehicle whereas others with weeks of experience in an unmanned autonomous vehicle still hover with their hands inches from the steering wheel, just waiting to take over should the perceived need arise.

Lastly, what about the basic test infrastructure for unmanned ground systems? The DoD has, over the years, developed extensive test facilities, but unmanned systems bring a new set of considerations. For smaller systems, for example, a "road course" may require obstacles such as curbs, winding stairwells, tunnels, puddles, etc. Instrumentation may pose additional challenges. How will test data be collected when the instrumentation suite normally mounted on a vehicle is now larger and heavier than the system being tested (think micro-robots)? With the current emphasis on tunnels, how will the data collected be relayed to the data collection location? Will global positioning systems provide location data on a system in a tunnel, as is currently assumed?

Evaluation considerations

What about metrics when it comes to perception? Much of our testing with regard to perception relies on human feedback. So how do you get an unmanned system to tell you why it did or did not perceive a negative obstacle (examples of a negative obstacle include a hole in the ground or puddle of water)? How good does perception have to be in order to be operationally effective? As good as a human, better than a human, or can we afford to accept less perception for the trade of human standoff and less exposure to risk?

Now that we are talking evaluation, that is a significant challenge for the T&E community. When it comes to a platform without a human operator on board, just what exactly constitutes operationally effective, suitable, and survivable? In many cases the same measures of effectiveness will suffice, but some will not, and some new ones will likely have to be invented. This question really comes down to redefining the trade space—how much performance are we willing to give up for the sake of greater safety and less risk of loss of limb and life?

Let's talk survivability—how does one define survivability when it comes to an unmanned system? The easy part is we do not have a human operator to worry about, but where is the knee of the curve when it comes to how much survivability to design into the system? Where is that limit that says do not make the system so valuable that humans are required to protect it? One thing is certain, the requirements community does not really know, so maybe testers and evaluators can collaborate with users to sort through this new frontier of metrics. Sounds like a perfect collaboration, and the sooner the better.

When evaluating suitability, inevitably the notion of reliability comes into play. This has always been an important aspect but when it comes to unmanned systems, suddenly the implications of reliability take on a new meaning. After all, there will not be an operator around to get down from the vehicle to change the flat tire. Clearly the reliability of a system must be no less than that of its expected mission duration. When it comes to unmanned systems, however, the old limits do not apply. Robots do not have to stop to rest and eat, so mission durations of days, weeks, months, years are being envisioned during the next 25 years. This is not as far fetched as you might think. The Defense Advanced Research Projects Agency is developing a high altitude, long endurance unmanned aerial system expected to provide persistent surveillance during a period of 5 years. When it comes to the air domain we pretty much know how to test for this; after all, our satellites have been doing just fine, thank you.

What about UGVs? Just what exactly is the test methodology to statistically prove the reliability of UGVs expected to conduct missions measured in months and years? After all, the old joke about the trouble with UGVs is that they operate on the ground. Alot of unexpected things can happen on the ground that are not likely scenarios in the air. Do we need to test for every contingency? That is alot of combinations and permutations—so just how do we crack this nut without bankrupting the program budget?

Another thing about reliability—there is a different dimension when it comes to unmanned systems, that of autonomy. How many hours of operation will be required to statistically prove that the system will not go "rogue"? Today we do not test the propensity of a vehicle to suddenly drive off course and into a building when a driver is behind the wheel, but we will have to do just that with UGVs. Not convinced this is an issue? It happened at the Defense Advanced Research Projects Agency Urban Challenge. With no warning one of the vehicles that was doing splendidly suddenly left the course and nearly plowed into a building. It missed impacting the building by mere inches because the manned chase vehicle following it activated the e-stop. When it comes to rogue vehicles, what are all the protocols that will need to be invoked to prove the unmanned system is safe from being "hijacked"? In other words, how will we test the systems to ensure the enemy cannot retask the system to conduct a mission against us? Do we know how to characterize this risk?

Since we are talking safety, do we have all the test infrastructure needed to ensure safe testing? So much of our safety on the test range today depends on humans. What will we need in terms of infrastructure to ensure safe testing when it comes to 9 ton or larger UGVs. How will we ensure safety when testing a large vehicle at high speed over rough terrain? Do we have sufficient e-stop performance? After all, when it comes to large mass at high speed, it takes more time and distance to stop.

So now what?

The discussion up to this point has been more in the nature of posing questions than in recommending answers. That is deliberate. The T&E community will come up with the answers in due time. What the DoD needs is for that "due time" to well precede the point at which unmanned systems will be ready for testing. In order to ensure timely delivery of needed systems to the warfighter, program managers should not be confronted with surprises because of uncertainty regarding how to test the system. What is so encouraging is that the DoD T&E community is not resting on its laurels and assuming that present methodology is sufficient. They are leaning forward in the foxhole to tackle these issues, and they are doing it at a particularly challenging time when there just are not that many systems to learn from. With continued engagement among the T&E community, the robotics experts in the laboratories, industry developers, and the user community, the answers to the questions posed above will be answered in a timely and effective manner, and eventually the population of personnel working with robots will pretty much number everyone in uniform. ❑

Mrs. Ellen M. Purdy is the enterprise director, Joint Ground Robotics within the Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. She is responsible for oversight of the $1.7B unmanned ground systems portfolio and strategic planning for ground robotics. She prioritizes and manages $50M annual Budget Activity 4 and 5 funding, and ensures ground robotic technology development and acquisition complies with statutory and Congressional mandates.

Previously Mrs. Purdy served in the Program Management Office, Future Combat Systems Brigade Combat Team (PMO FCS (BCT)). During 5 years with FCS, Mrs. Purdy served in a variety of positions including manager, Test, Analysis, Modeling and Simulation; deputy director, Combined Test Organization; director, Combined Test Organization; and integrated product team lead for Integrated Simulation and Test. Mrs. Purdy was responsible for a 60 person organization and a $2.2B test program that included testing at the system of systems level and each of the 18 individual platforms plus the network that comprise FCS.

Mrs. Purdy has a bachelor of science degree in chemical engineering a master of science degree in engineering management. She is a member of the Board of Directors for the Association of Unmanned Vehicle Systems International and a member of the Advisory Board for the Virginia Center for Autonomous Systems. Her awards include two Superior Civilian Service Award medals, two Commander's Award for Civilian Service medals, and the Achievement Medal for Civilian Service. Mrs. Purdy has published over twenty technical reports and journal articles.

References

Delude, C. M. 2008. "MIT Neuroscientists See Design Flaws in Computer Vision Tests." McGovern Institute, January 24, 2008.

ITEA Journal 2008; 29: 333–335

Making Improved Reliability a Reality

Ernest Seglie, Ph.D.
Science Advisor,
Operational Test and Evaluation,
Office of the Secretary of Defense, Washington, DC

The September 2008 issue of Information Systems Test & Evaluation focused on improving suitability. It coincided with, and in some cases led, reliability improvement policy and guidance. By early September 2008, all the military departments responded positively to the direction of Under Secretary Young in July 2008 to "establish a reliability improvement acquisition policy to address the problem of inadequate system RAM [reliability-availability-maintainability]." For example, on September 4, 2008, Dr. Donald Winter, Secretary of the Navy, stated, "Having performance is important, but not as important in most cases, as having reliability." I offer several specific examples of progress made toward improving suitability and reliability and a guide on where to find the products.

On June 6, 2008, the Defense Science Board Task Force on Developmental Test and Evaluation (which focused on improving reliability) released its final report. It is available at http://www.acq.osd.mil/sse/dte/docs/DSB-Rpt-DTE-May2008.pdf. Charged to implement key Defense Science Board recommendations, the U.S. Department of Defense (DoD) Reliability Improvement Work Group (RIWG) worked from March to August 2008 to:

- ensure programs are formulated to execute a viable systems engineering strategy from the beginning, including a reliability-availability-maintainability (RAM) growth program, as an integral part of design and development (that is, *Start Programs Right*);
- ensure government organizations reconstitute a cadre of experienced Test and Evaluation (T&E) and RAM personnel (*Re-enforce the Work Force*);
- implement mandated integrated developmental test and operational test, including the sharing and access to all appropriate contractor and government data and the use of operationally representative environments in early testing (*Implement Integrated Testing*).

Ernest Seglie

On September 4 the RIWG published its final report containing implementing actions and products, along with military departments' implementing steps for those actions and products. The report is available at http://www.acq.osd.mil/sse/dte/spec-studies.html. RIWG products were developed by representatives from all military departments and the Defense Information Systems Agency (DISA). It is widely agreed that these products are comprehensive and, if implemented, will be effective in achieving proper defense system reliability and thereby containing system sustainment costs. A short summary of RIWG products includes the following.

Start programs right

Department policy for reliability improvement

Issued on July 21, 2008, it states:

> *"It is Department policy for programs to be formulated to execute a viable RAM strategy that includes a reliability growth program as an integral part of design and development. Additionally, RAM shall be integrated within the systems engineering processes, documented in the program's Systems Engineering Plan and Life Cycle Sustainment Plan, and assessed during technical reviews, test and evaluation, and program support reviews. This policy will be included in the DoD Instruction 5000.2."*

The secretaries of the military departments were directed to establish their own reliability improvement acquisition policy. The policy memo is at http://www.acq.osd.mil/sse/dte/docs/USD-ATLMemo-RAM-Policy-21Jul08.pdf.

Sample reliability language for Sections C, L, and M of acquisition contracts

It also includes a checklist for evaluating reliability program plans and a sample performance incentive fee. For example, from the sample language for Section C (Statement of Work):

> *"The contractor shall develop a reliability model for the system. At a minimum, the system reliability model shall be used to: (1) generate and update the reliability allocations from the system level down to lower indenture levels; (2) aggregate system-level reliability based on reliability estimates from lower indenture levels; (3) identify single points of failure; and (4) identify reliability-critical items and areas where additional design or testing activities are required in order to achieve the reliability requirements."*

The contract language is at https://acc.dau.mil/CommunityBrowser.aspx?id=219127&lang=en-US.

The language and practices are consistent with a new industry standard for reliability: GEIA-STD-0009: Reliability Program Standard for Systems Design, Development, and Manufacturing, August 2008.

Program reliability and maintainability review template

This tool defines relevant reliability activities and evidence of them associated with 16 technical reviews across system acquisition. For example, in the initial technical review during concept refinement, look for evidence of documented reliability assumptions and supporting rationale accompanying reliability requirements for the preferred solution. The template is available at http://www.acq.osd.mil/sse/docs/RAM-Planning-Template.xls.

Standard evaluation criteria

To determine if system contractors employ practices needed to achieve reliability requirements, a scorecard offers criteria and a scoring means. For example, one criterion is "sufficiently-sized reliability engineering staff directly tied to design team." The contractor score is determined by the size of the reliability engineering staff, its workload, and its communication with the system design team. The scorecard is at https://acc.dau.mil/CommunityBrowser.aspx?id=210483&lang=en-US.

RAM champions in each Service

The RIWG recommended designated Service "champions" to ensure reliability initiatives become institutionalized. For example, the U.S. Army has named a headquarters executive to serve as the Department of the Army Reliability Chief.

Re-enforce the work force

Strengthen Defense Acquisition University curriculum

RIWG representatives recommended Defense Acquisition University curricula and work force certification changes to the Overarching–Functional Integrated Product Team. There was broad concurrence that the Defense Acquisition University can further the intent for RAM effectiveness by coordinating the education of the functional work forces that play pivotal roles at various life cycle stages of systems development. Most recently, the RIWG proposed reliability contracting subject matter for contracting work force education.

Implement integrated testing

Integrated testing defined

A memorandum formally defines integrated T&E as a basis for developing further guidance. The memorandum is at http://www.acq.osd.mil/sse/dte/docs/SecDefMemo-Definition-of-Integrated-Testing-25Apr08.pdf

Early T&E involvement in RFPs [Request for Proposals] and contracts. The RIWG included data sharing concepts in a guidebook titled "Incorporating Test and Evaluation into Department of Defense Acquisition Contracts." The guidebook is located at http://www.acq.osd.mil/sse/dte/guidance.html.

New Defense Acquisition Guidebook Guidance and T&E Master Plan format. New guidance in the Defense Acquisition Guidebook Guidance and a new T&E Master Plan format emphasize integrated T&E. It is available at https://akss.dau.mil/DAG/Guidebook/IG_c9.0.asp.

Service implementation responses

Across the DoD, implementation of the actions in the RIWG report is in progress. The Army has been aggressive in emphasizing reliability accountability, contracting for reliability, and establishing an early warning mechanism for potential reliability problems. The U.S. Navy is reinvigorating their RAM processes and implementing reforms such as use of the new GEIA standard in developing program requirements (that standard, GEIA-STD-0009 Reliability Program Standard for System Design, Development and Manufacturing, is available through the ITAA Government Electronics and Information Technology Association On-Line Web Store at http://www.techstreet.com/cgi-bin/detail?product_id=1568406).

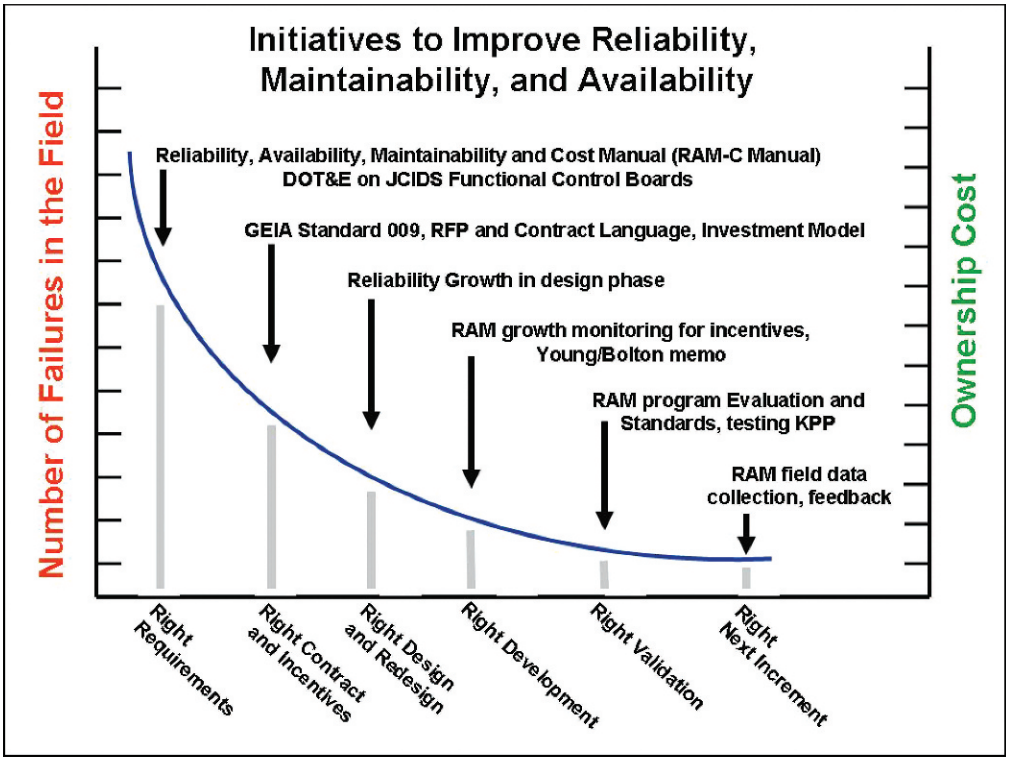


Figure 1. Result of reliability improvement—fewer failures, reduced ownership cost

The U.S. Air Force is examining their policies and processes to improve RAM.

Department leadership reviewed RIWG actions and products and Services implementation in late September and directed a follow-up review in December 2008 to assess the extent of implementation.

With this emphasis across the department, it is reasonable to expect to see improved reliability in programs. We will not see results overnight, as it must begin with requirements and contracts or those opportunities that occur in any program restructuring. The end result of these initiatives will be reduced system ownership costs, fewer failure modes, improved reliability, and improved system value to our warfighters, as *Figure 1* illustrates.

Of course, testers have a vital role. RAM expertise is in short supply across the DoD (and industry). T&E can step forward, join requirements efforts, influence program office System Engineering Plans and T&E Master Plans to properly plan effective reliability programs, and measure and report results. Only with T&E data will programs know where they stand with respect to failure modes, demonstrated system reliability, and the reliability growth potential. ❑

Dr. Ernest Seglie is science advisor, Office of the Director, Operational Test and Evaluation (DOT&E), the Pentagon, Washington, D.C. He provides scientific and technical guidance on the overall approach to Department of Defense (DoD) evaluation of the operational effectiveness and suitability of major DoD weapon systems, provides technical review of test reports, and serves as chief technical advisor to the Director, DOT&E.

ITEA Journal 2008; 29: 337–340

Improving Product Evaluation by Reliability Testing and the DMAIC Method

Guangyu Wu
Universal Lighting Technologies, Madison, Alabama

The first and most important step in the Six-Sigma process is to accurately address customer wants and needs. The customer's words about their wants and needs with regard to reliability requirements are not always clear so they are captured and transferred into quantifiable reliability metrics. The Define-Measure-Analyze-Improve-Control method is used to improve product design. In this article, we will present a sample application illustrating use of the Define-Measure-Analyze-Improve-Control method and reliability analysis to decrease the sound level of the electronic ballast. After that, Accelerated Life Testing is used to verify the improvement and confirm the design.

Key words: Accelerated life testing; DMAIC method; reliability analysis; Six-Sigma process.

The first step in the Six-Sigma process is to accurately address customer wants and needs (Wu 2005a). Unfortunately, the customer's voice is not entirely clear, as is often the case for issues related to electronic product reliability and noise. Customers often express their needs in vague terms like "very reliable" and "low noise level." In addition, there is generally no detailed information on the design requirement or system reliability metrics, e.g., product mean time between failure and stable state availability. We can capture the customer's voice about electronic products in a matrix similar to that used in quality function deployment. *Table 1* illustrates the transfer of the voice of the customer on reliability into quantifiable reliability metrics. *Table 2* is used to identify the Six-Sigma and reliability engineering techniques that could be used to evaluate the product requirements listed in *Table 1* and provide solutions if improvements are necessary.

Sample application

This section presents the analysis of the sound level for electronic ballasts as a case study to illustrate the role and use of accelerated life testing (ALT) within the Define-Measure-Analyze-Improve-Control (DMAIC) methodology.

Table 1. Customer expectation versus product requirements

Customer Expectation	Product Requirements				
	High product MTTF	Low sound level	Product meets UL of safety standard	High MTBF of critical components	Low DPMO
Product has high reliability	X	X		X	
Keep environment quiet when product working		X	X		
Product provides high safety			X		
High initial quality					X
Minimum maintenance	X	X		X	
Convenient to install			X		

DPMO = defect per million opportunities; MTBF = mean time between failure; MTTF = mean time to failure; UL = Underwriters Laboratories, Inc.

Table 2. Product requirements versus technical approaches

Product Requirement	Technical Approaches		
	ALT	FMEA	DMAC
High product MTTF	X		
Low sound level	X	X	
Product meets UL safety standard	X	X	X
High MTBF of critical componenets	X		
Low DPMO			X

ALT = accelerated life testing; FMEA = failure modes and effects analysis; DMAIC = design, measure, analyze, improve, and control; DPMO = defect per million opportunities; MTBF = mean time between failure; MTTF = mean time to failure.

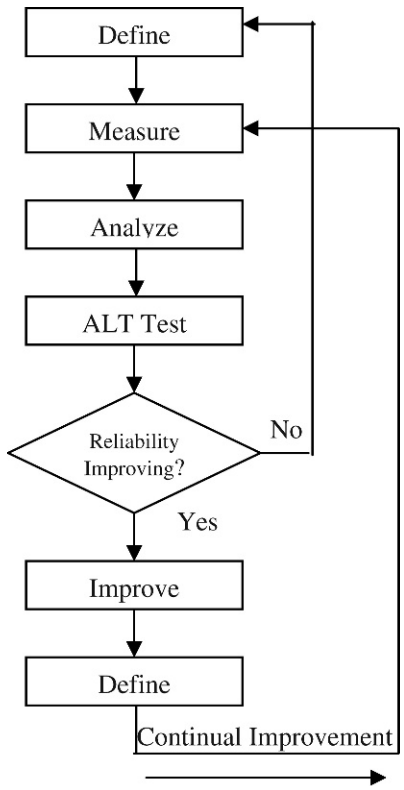


Figure 1. Process map for product improvement

The sound level of ballasts is a key customer requirement. Different environments have different sound level requirements. For example, for applications in libraries and school study halls the sound level requirement must be less than 24 dB (Sound rated A < 24 dB). Applications in residences and quiet offices require sound levels less than 30 dB (Sound rated B < 30 dB). Ballasts used in general office areas and commercial buildings have requirements of less than 36 dB (Sound rated C < 36 dB). In this article, we will present a case study illustrating the use of the DMAIC method and reliability analysis to decrease the sound level of the electronic ballast.

Methodology

In the DMAIC methodology, there is no emphasis on testing to verify improvements. If ALT was incorporated into the DMAIC process, the product improvement could be verified in a short time. *Figure 1* illustrates such a modification to the general DMAIC methodology for this project. Using this approach the customer's needs and wants are achieved *before* the product is released to the manufacturing phase.

Customer's requirement

Universal Lighting Technologies (ULT) plans to improve an electronic product that currently has problems with unacceptable noise levels. The first step was to understand the customer requirements. *Table 3* shows the customer's requirements: average noise level of ballast and mean time between failure (MTBF).

Table 3. Customer's requirements: sound specification and MTBF

Installation In	Average Ambient Noise Level (dB)	Sound Level Rating	MTBF (Year)
TV or Radio Station, Library, Reading Room	20-24	A	>=5
Residence, Quiet Office	25-30	B	>=5
Commercial Building General Office Area	31-36	C	>=5

Results of Six-Sigma project

In order to improve ballast quality and reduce the sound level to the desired level, ULT used the DMAIC methodology to determine the root cause of any sound level problems and eliminate them (Wu 2005b). For example, for Elec1, the vibration of the old board caused the high noise level. We needed to redesign it. One method was a U-shaped part cut in the middle of the board to reduce the vibration. For Elec2, the approach was to use a new magnetic component redesigned to reduce the noise. One method was to use glue to hold the pin in place. *Table 4* shows the results of the analysis.

The ballast Elec1 is used in the reading room, and Elec2 is used in the residence room. *Table 4* shows that all the improved methods are effective, and the results meet the customer requirement of sound level.

Considering another customer requirement for a 5-year life cycle, we ask: which method should we select? For example, for Elec1, the sound level of the new board design (A model) is 17 dB, the sound level of the new board design (B model) is 16 dB, and the sound

Table 4. Method improving sound level

Product Model	Improved Method	Average Sound Level (dB) (Sound Level Rating: A)	
		Improved	Before Improving
Elec1	New design (A model)	17	24
	New design (B model)	16	24
	New design (C model)	18	26
Elec2	New magnetic component (D)	19	30
	New magnetic component (E)	22	31
	New magnetic component (F)	21	30

Table 5. Comparing improvement

Product Model	Improved Method	Average Sound Level (dB) (Sound Level Rating: A)		Estimating MTBF (year)	Select
		Improved	Before Improving		
Elec1	New design (A model)	17	24	4.1	
	New design (B model)	16	24	4.6	
	New design (C model)	18	26	6.5	X
Elec2	New magnetic component (D)	19	30	7.2	X
	New magnetic component (E)	22	31	3.8	
	New magnetic component (F)	21	30	4.7	

level of the new board design (C model) is 18 dB. These three methods have improved the noise level. Which one should be chosen as the final design? In the next step, ALT was used to estimate the MTBF for each method.

Reliability test

For the above sample we chose the following settings for the ALT: There were two stresses in the test: (a) accelerated thermal cycling and (b) accelerated power cycling. The stress strength is generally selected based on the product characteristics, such as *from* ($-30°C$) *to* ($+105°C$) or *from* ($-45°C$) *to* ($+135°C$), the frequency of power cycling 24/d or 48/d. In this experiment, considering the electronic characteristic of ballast models Elec1 and Elec2, the accelerated thermal cycling was selected *from* ($-30°C$) *to* ($+105°C$) and accelerated power cycling for 24/d. Using ALT to do reliability test can save the time and budget.

A Weibull distribution was used to analyze the failure data and estimate the MTBF (Wu 2005a). The Weibull distribution is frequently used in reliability analysis for electronic products. There are three parameters in the Weibull distribution: the shape parameter β, the scale parameter η, and the location parameter γ. In this form both the exponential and Rayleigh distributions are special cases of the Weibull distribution: $\beta = 1$ produces the two-parameter exponential distribution, and $\beta = 2$ produces the Rayleigh distribution.

In the experiment, random samples were chosen for the ALT test. The time-to-failure data were collected, and the Weibull distribution was used to estimate MTBF for different models. The results are shown in *Table 5*. For product Elec1, the new board design (C model) was chosen. Its MTBF is 6.5 years. For product Elec2, we chose the new magnetic component (D), and its MTBF is 7.2 years. These methods meet the life cycle requirement of 5 years. The MTBFs of the other improved methods were below the customer's requirement of a 5-year life cycle, although they did reduce the sound level.

Concluding remarks

This article has shown how improving product quality can be achieved effectively using a combination of the Six-Sigma process and reliability analysis. This approach was illustrated with a case study that focused on improving the sound level of an electronic product designed and produced by ULT. This project was the company's first attempt at decreasing the noise level of electronic products by employing the Six-Sigma methodology and ALT (McLean 2000) together. In the future more parametric studies can be done to provide further guidelines for product design, quality, and reliability improvement and system design. ❑

GUANGYU WU is a Ph.D. candidate at the University of Alabama at Huntsville. He has a master of science degree in reliability and quality engineering from the University of Arizona, a master of science degree and bachelor of science degree in mechanical engineering from Guangxi University, China. He is a reliability engineer at Universal Lighting Technologies. His research interests include: Accelerated Life/Degradation Test (ALT/ADT), Highly Accelerated Life Test (HALT), Reliability Modeling and Data Analysis, Reliability Prediction, Failure Mode and Effective Analysis, System Reliability and Analysis, and Applied Statistics. He is a member of the

Institute of Electrical & Electronics Engineers (IEEE), American Society for Quality (ASQ), Society for Reliability Engineers (SRE), American Society for Engineering Management (ASEM). E-mail: gwu@unvlt.com

References

McLean, H. W. 2000. *HALT, HASS & HASA Explained: Accelerated Reliability Techniques*. Milwaukee, Wisconsin: American Society for Quality.

Wu, G. 2005a. *Six-Sigma Black Belt Project*. Huntsville, Alabama: ISEEM Department, University of Alabama at Huntsville.

Wu, G. 2005b. *Acceleration Factor for Electronic Product in Accelerated Life Test*. Madison, Alabama: Universal Lighting Technologies.

ITEA Journal 2008; 29: 341–345

Revolutionary Imaging: Air Force Contributions to Laser Guide Star Adaptive Optics

Robert W. Duffner
Historical Information Office,
Air Force Research Laboratory, Kirtland Air Force Base, New Mexico

The 178th meeting of the American Astronomical Society (AAS) in Seattle, Washington, in May 1991 provided the ideal setting for Robert Q. Fugate, technical director of the Air Force's Starfire Optical Range at Kirtland Air Force Base (Albuquerque, New Mexico), to make a dramatic announcement. His groundbreaking, classified laser guide star* work during the 1980s had been an unqualified success. Now for the first time, that information could be released to the public.

**A laser guide star is an artificial beacon created by a laser focused in the lower atmosphere. A wavefront sensor on the ground measures distortions in the return light (backscatter) caused by atmospheric turbulence.*

When Bob Fugate walked up to the microphone at one o'clock to deliver his paper, he was a little nervous as he gazed out across the audience to find the room jam-packed. Nearly 400 people had showed up, with some standing two to three deep in the back and along the sidewalls.

Fugate did not disappoint his audience as he got to the heart of the matter right away. The first words out of his mouth were delivered in a confident and deliberate manner, announcing: "Ladies and gentlemen, I am here to tell you that laser guide star adaptive optics works!" To provide historical substance and scientific credibility to his opening statement, Fugate projected two images of the binary star *53 Ursa Major* on a large screen behind him. The uncompensated image shown on the left of the screen appeared as a blank in the heavens. But the image on the right side of the screen, compensated with laser guide star adaptive optics, dramatically revealed a clear image of *53 Ursa Major*, an improvement greater than a factor of 25 over conventional astronomical imaging *(Figure 1)*. Fugate explained that this photo "…was taken while the deformable mirror was continuously correcting atmospheric wavefront distortions." In scientific jargon, this was known as a "closed-loop" system. It consisted of three key components—a wavefront sensor, a high-speed processor, and a deformable mirror—that could keep up with the constant changes in atmospheric turbulence (occurring hundreds of times per second) and produce a high-resolution image.[1]

For a brief moment there was utter silence. The speechless scientists in the audience tried to grasp the significance of Fugate's startling announcement. Within seconds, a steady flow of noisy chatter broke out as they turned to one another and began muttering about the amazing image that they had just seen.[2]

Fugate's presentation that day created a big stir not only in the conference room, but very quickly in the astronomy community as a whole. Bill Thompson, a technical advisor at Phillips Laboratory at Kirtland Air Force Base who led the declassification effort, recalled it was "quite a day" as the astronomers were simply dumbstruck by the impact of the classified information, which was released all at once. "A lot of people in the audience," Thompson observed, "were stunned by the amount of work that had already been done by the Department of Defense…that was presented at the meeting." Wayne Van Citters, from the National Science Foundation, remembered the people listening to Fugate's presentation slowly leaning back in their chairs, mentally regrouping, and reacting with one telling word—Wow![3]

Fugate explained that the *53 Ursa Major* image was made on March 16, 1990, more than a year after his team had closed a laser guide star loop for the first time as part of the Generation I series of experiments conducted at the Air Force Weapons Laboratory at Kirtland. He also told the AAS group that the Air Force had sponsored laser guide star adaptive optics research since the summer of 1982—a shocking revelation to the academic astronomers in the audience—and described his first laser guide star experi-

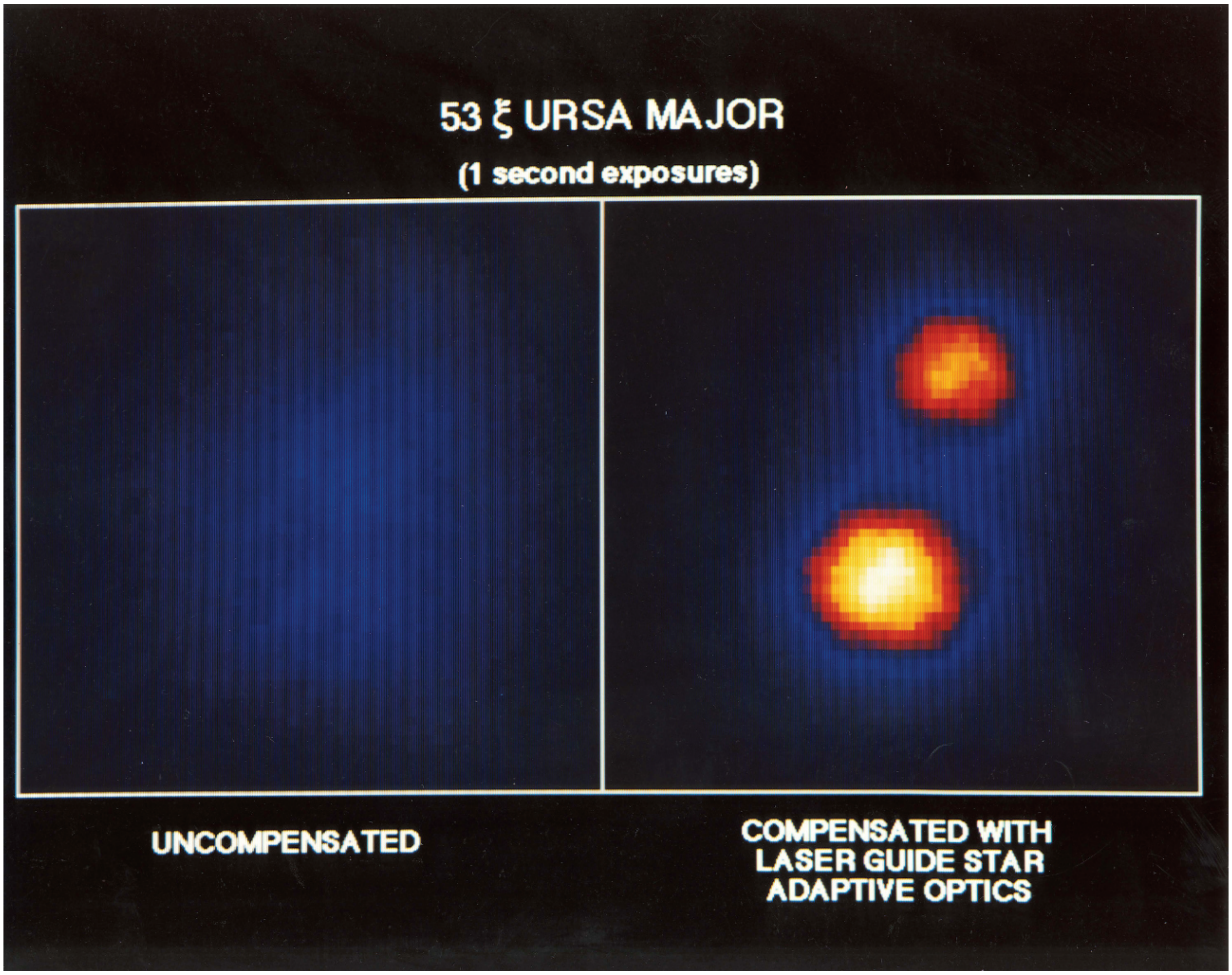


Figure 1. Bob Fugate's first image shown during his presentation at the American Astronomical Society Meeting in Seattle on May 27, 1991.

ment conducted in the fall of 1983. Fugate told the gathering, "Just to convince you that we didn't get one lucky picture," he showed a series of additional compensated images to reinforce the reliability of the laser guide star technique[4] *(Figure 2)*.

Why was there so much commotion over the release of Fugate's guide star work? He and his Air Force colleagues had done something revolutionary—they had begun to conquer the age-old problem of atmospheric turbulence causing distortion in light waves. Distorted light waves produced blurred rather than razor-sharp images of objects in space *(Figure 3)*. Fugate's laser guide star technique was a critical first step in the adaptive optics process that would eventually "compensate" distorted light by removing the effects of atmospheric turbulence, thus enabling high-resolution images. That was important to the military, which wanted to be able to take clear images of satellites, missiles, reentry vehicles, and space debris as part of its space situational awareness mission, and

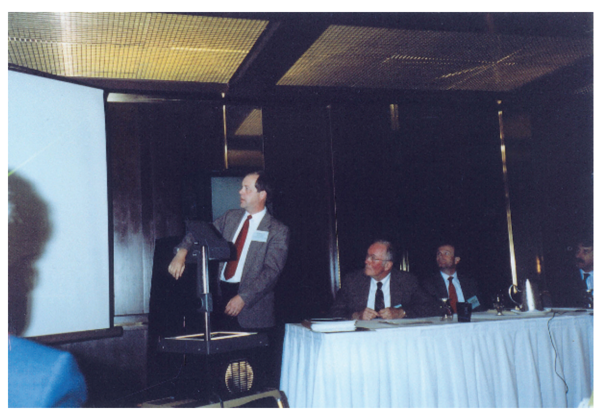


Figure 2. The day after his formal presentation to the American Astronomical Society, Bob Fugate briefed the press on his revolutionary laser guide star findings, while Charles H. Townes (center) and MIT/Lincoln Lab's Charles A. Primmerman (next to Townes) look on.

equally important to astronomers, who wanted ways to improve the image quality of planets, stars, galaxies, and other celestial bodies.[5]

Above all, adaptive optics needed to address atmospheric turbulence caused by temperature fluctuations in the atmosphere. Gases that make up the atmosphere are constantly moving at different speeds, much like the surface water in the oceans. Some sections of the ocean can be perfectly calm with a mirror-like surface, while other regions of the same ocean experience violent, churning surf, and tidal wave conditions. In the atmosphere, similar conditions exist. Temperature changes at various altitudes in the atmosphere result in changes to the air density refractive index, which causes one section of a light wavefront to bend differently and move ahead or lag behind other sections of the same wavefront. This produces the undesirable condition of an uneven wavefront.[6]

In other words, these random temperature fluctuations in different regions of the atmosphere produce a non-uniform and constant swirling mixture of air, which degrades the quality and intensity of a light beam as it moves unpredictably through each sector of the atmosphere. Instead of all parts of the light wavefront traveling in a straight flat line in the same direction, atmospheric turbulence causes the light to follow an erratic path. It is this phenomenon that causes stars to "twinkle." The goal of adaptive optics is to align all sections of the wavefront to move in the same direction and replace the twinkle with a sharp image.[7]

Adaptive optics offered one potential solution by restoring light almost to its original, undisturbed condition outside the atmosphere. Overall, the term *adaptive optics* refers to an optical system that can adapt by compensating for atmospheric distortions induced in light waves. As one expert put it, "It's a method of automatically keeping the light focused when it gets out of focus."[8]

Fugate and his team first attacked the atmospheric turbulence problem by demonstrating a Rayleigh laser guide star in 1983 at a remote optical site at Kirtland *(Figure 4)*. The laser guide star concept relied on a principle of physics called Rayleigh scattering—named after Lord Rayleigh, winner of the Nobel Prize in Physics in 1903—whereby focused laser light is reflected in all directions by molecules (nitrogen, oxygen, and aerosols) in the atmosphere. (Shining a searchlight in the sky at night, with the reflecting light dispersing in all directions, is similar to Rayleigh scattering.) Researchers speculated that if a telescope and an outgoing laser were both pointed towards a prominent object in the sky—such as a star—the Rayleigh-reflected laser light and the starlight would travel downward along a near-identical return path in the atmosphere to the telescope on the ground and encounter nearly identical turbulence.[9]

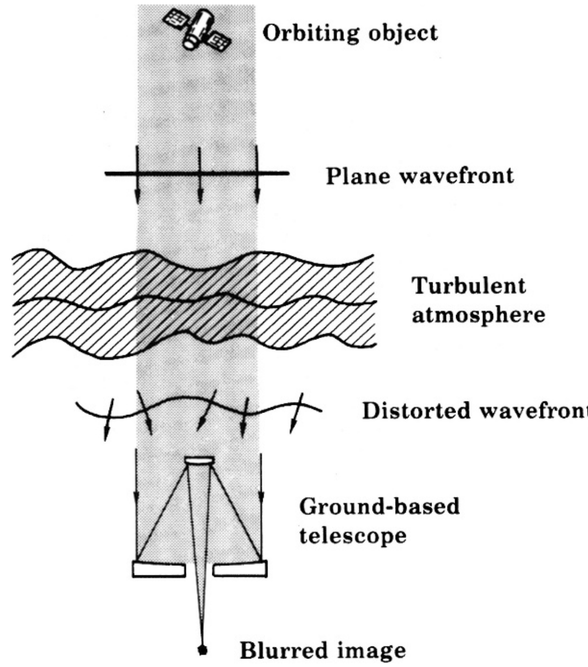


Figure 3. Anatomy of a blurred image

But why did scientists need light from both a star and a Rayleigh laser guide star? Only a tiny percentage of stars are bright enough to deliver enough light to a telescope to determine the amount of distortion across the light's wavefront. That was the main reason scientists began investigating artificial Rayleigh guide stars. The wavefront sensor in an adaptive optics system "consumes" most of the light from a dim star, leaving insufficient starlight to be sent to a camera to image the star. A Rayleigh laser guide star provides additional light to send to the wavefront sensor, enabling the starlight to bypass the measuring device and travel directly to a deformable mirror, and from there to enter a camera and produce a clear image. So, the main advantage of a laser beacon is that it is an artificial bright light source that is independent of the light from the observed object and, therefore, allows all the light from the viewed object to be used by a camera doing the imaging.[10]

Important as it was, the Rayleigh guide star experiment was strictly an attempt at "measurement." As one scientist described it, the guide star experiment was like taking a picture or x-ray of the atmosphere. The question was whether the backscatter from a Rayleigh guide star could be used to measure the extent of distortion (phase errors) induced on the laser wavefront.[11]

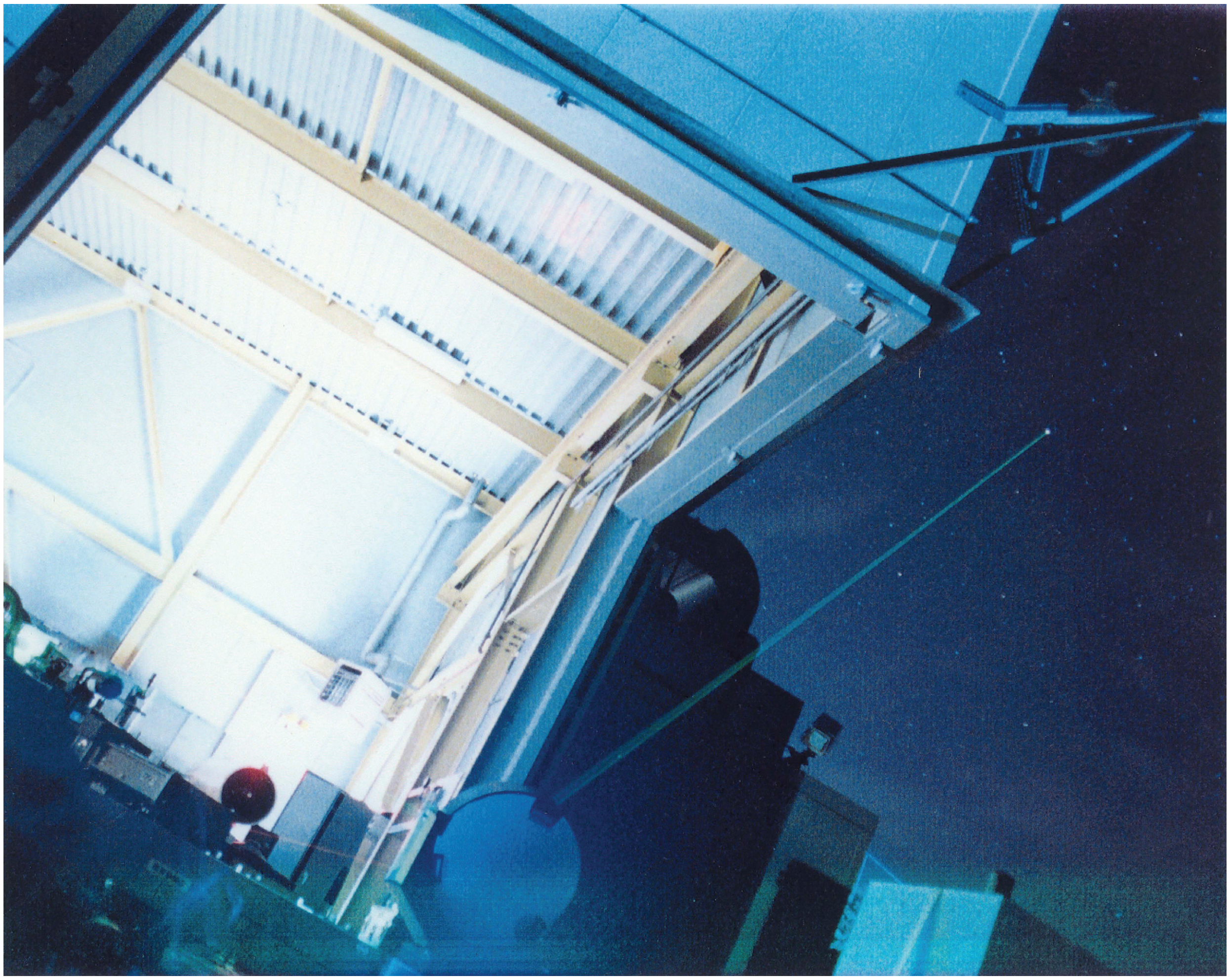


Figure 4. Rayleigh laser guide star experiment (1983) with laser pointing to the star Polaris.

To confirm that the air turbulence measurements were of the highest quality, Fugate's team compared the Rayleigh experimental data to an independent standard of measurement—the "truth" reference—to show the Rayleigh numbers were correct. Light from a star was the controlled variable or standard of comparison. So experimenters pointed their narrow laser beam to within a few microradians of the star Polaris, the famed North Star *(Figure 4)*. Polaris was chosen because it was one of the rare "bright" stars that could supply an adequate amount of light. The starlight and the laser backscatter light would travel through nearly identical paths from the atmosphere to a telescope, in this case the 1.5-meter telescope at Starfire Optical Range at Kirtland.[12]

Findings showed unequivocally that measurements of distortions in the starlight closely matched those of the guide star scattered light. Fugate's team was pleased by the outcome of the Rayleigh experiment because the data proved the theory of a laser guide star. "These results demonstrate qualitatively," Fugate proudly pronounced, "that laser guide star beacons are effective in measuring atmospheric-turbulence-induced wavefront distortion."[13]

Although Fugate's 1983 experiment was not conducted using an operational adaptive optics system, its success laid the groundwork for development of a closed-loop system. Such a system required a wavefront sensor and a high-speed processor that sent electrical signals to actuators (small pistons) attached to the backside of a deformable mirror. Depending on the strength of the electrical signals, each actuator pushed or pulled to change the shape of the mirror surface. As distorted light struck the irregular mirror surface, the beam was "straightened out" or compensated so a clear image could be formed. It was this kind of closed-loop adaptive optics system that Fugate used in his Generation I experiments to capture the revolutionary compensated images he showed to the astonished crowd of astronomers in Seattle in 1991.[14]

The Rayleigh guide star work marked a milestone in the history of technology that had important consequences. A team of Air Force scientists demonstrated the application of the laser guide star, a revolutionary breakthrough in the annals of optical research that would be pivotal to the development of future adaptive optics systems. Not only did these experiments bolster the Air Force's situational awareness mission, they resulted in a classic case of technology transfer from the military to the civilian sector. Laser guide stars and the subsequent development of sophisticated adaptive optics systems on ground-based telescopes produced hitherto impossible, high-resolution images that were incredibly beneficial to the world's astronomers. Indeed, many considered adaptive optics to be the most important optical advancement in astronomy since the discovery of the telescope.[15] ❑

ROBERT W. DUFFNER is the historian at the Air Force Research Laboratory, Kirtland AFB, Albuquerque, New Mexico. His forthcoming book, The Adaptive Optics Revolution: A History, *will be published by the University of New Mexico Press in the spring of 2009. Email: robert.duffner@kirtland.af.mil*

Endnotes

[1]Paper, Robert Q. Fugate, "Experimental Demonstration of Real Time Atmospheric Compensation," May 27, 1991; Phillips Laboratory News Release (91-36), "Astronomy Breakthrough," May 27, 1991; Interviews with Robert Q. Fugate, April 21 and May 14, 2003.

[2]Interview with William E. Thompson, October 9, 2002.

[3]Ibid; Interview with Wayne G. Van Citters, April 14, 2004.

[4]Paper, Fugate, "Experimental Demonstration of Real Time Atmospheric Compensation," May 27, 1991; Interviews with Fugate, April 21 and May 14, 2003.

[5]Interview with Thompson, October 9, 2002; Interview with Van Citters, April 14, 2004.

[6]Tech report (AFRL-DE-PS-TR-1998-1054, Pt. 1), Glenn A. Tyler et al., the Optical Sciences Company, "Adaptive Optics: Theory and Applications," December 1999, pp. 2–8; Paper, Raymond P. Urtz Jr. (RADC), and James W. Justice (DARPA), "Compensated Imaging" June 12, 1975; Robert Q. Fugate, "Laser Beacon Adaptive Optics," *Optics and Photonics News*, June 1993, pp. 14–19.

[7]See note number 6.

[8]Robert K. Tyson, *Principles of Adaptive Optics* (Boston: Academic Press, 1997), p. 3.

[9]R. Q. Fugate, et al., "Measurements of atmospheric wavefront distortion using scattered light from a laser guide-star," *Nature*, September 12, 1991, pp. 144–146; Interview with Fugate December 16, 2002.

[10]See note number 9.

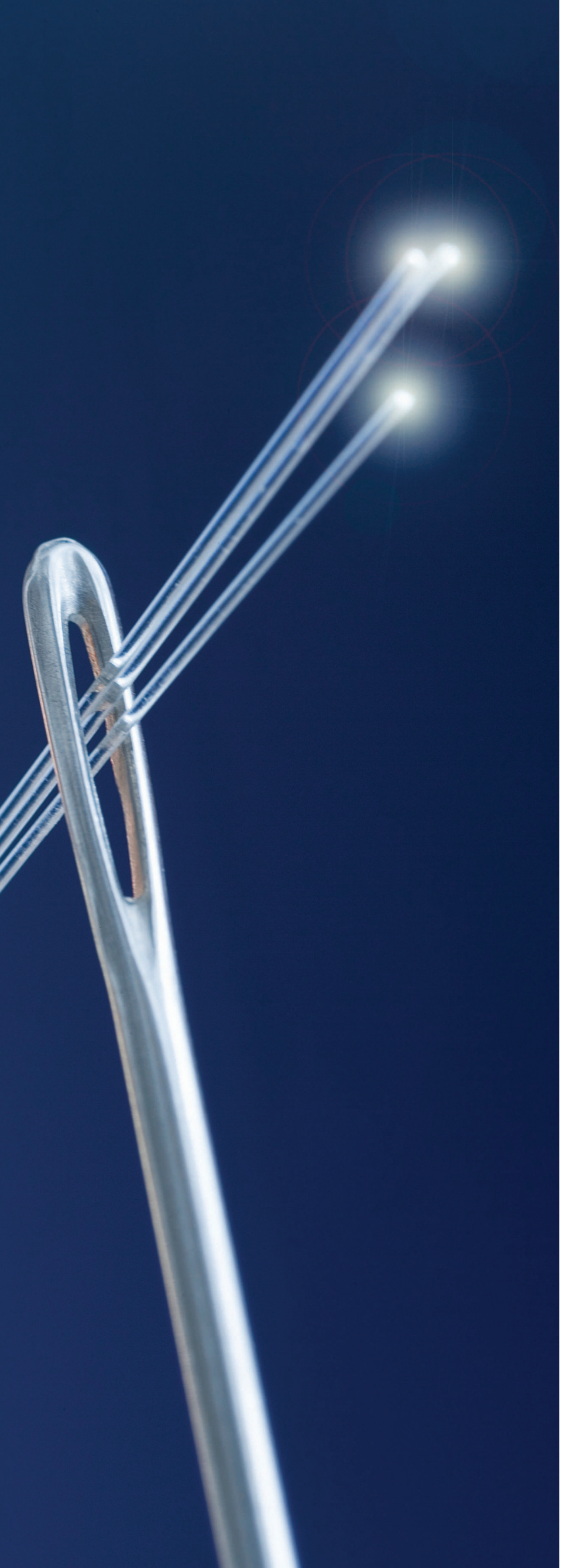
[11]Paper, Fugate, "Experimental Demonstration of Real Time Atmospheric Compensation," May 27, 1991; Interview with Fugate, April 21, 2002; Interview with Thompson, February 11, 2003.

[12]See note number 11; Fugate et al., "Measurement of atmospheric wavefront," September 12, 1991, pp. 144–146; USAF News Release (91-36), "Astronomy Breakthrough," May 27, 1991.

[13]Fugate et al., "Measurement of atmospheric wavefront," September 12, 1991, pp. 144–146; Paper, Fugate, "Experimental Demonstration of Real Time Atmospheric Compensation," May 27, 1991; Interview with Thompson, February 11, 2003; Notes, Fugate to Duffner, subj: Laser guide stars, June 10, 2004.

[14]Interview with Fugate, April 21, 2003; Interview with Thompson, February 11, 2003.

[15]Mount Wilson Observatory, "Adaptive Optics," http//www.mtwilson.edu/ao.



ITEA Journal 2008; 29: 347–350

U.S. Army Aberdeen Test Center Light Armor Range Complex

Col John P. Rooney
Aberdeen Proving Ground, Maryland

The U.S. Army Aberdeen Test Center (ATC) at Aberdeen Proving Ground, Maryland, is the U.S. Department of Defense's lead agency for land-combat, direct-fire, and live-fire vulnerability testing. Adjacent to the Chesapeake Bay and 35 miles northeast of Baltimore, ATC is a multipurpose test center with diverse capabilities and has become a world-class testing and training range that ensures American warfighters receive superior materiel and technology *(Figure 1)*.

One of the many testing capabilities at ATC is the Light Armor Range Complex (LARC). The LARC was built in the 1940s to support testing for the war effort and has provided support to every U.S. war and military conflict since its construction.

History of ATC and the LARC

ATC dates its beginning back to World War I, when artillery testing was moved to the new proving ground from Sandy Hook, New Jersey, to accommodate the increased volume of work and the wartime congestion of New York Harbor. Initially known as the Proof Department, the organization tested only artillery until 1923 when it created two main divisions of ordnance and automotive testing. This structure remained in effect until 1942. Expansion during World War II brought the Proving Center, which became the Ordnance Research and Development Center in 1943.

In 1945, the Ordnance Research and Development Center reorganized into the Development and Proof Services. In 1962, the U.S. Army Test and Evaluation Command was established and its headquarters located at Aberdeen Proving Ground. The test mission and facilities, along with the expertise of the workforce, continued to expand, and in 1968, the Development and Proof Services was renamed the Materiel Testing Directorate. In July 1984, Combat Systems Test Activity was activated as an independent organization within the Army Test and Evaluation Command and was renamed Aberdeen Test Center in 1995.

The LARC was constructed circa 1940s to support the developmental and acceptance testing of armor, weapons, and combat vehicles in support of World War II *(Figures 2 and 3)*. The original complex included three enclosed ranges, seven outdoor ranges, two loading rooms, and three storage rooms to support firing operations.

Traditionally, the workload at the LARC consisted primarily of testing armors for relatively few heavy armored vehicles. Approximately10 years ago, however, the Army and U.S. Marine Corps moved to field lighter and more deployable combat vehicles, which increased the workload at the LARC and strained the capability of the three indoor ranges. During the past 4 years, combat experiences in Operations Iraqi and Enduring Freedom have led the Army and Marine Corps to place an emphasis on armoring nearly all of its tactical and combat vehicle fleet. As a result, the vehicular armor testing workload, as well as personal armor protection testing workload, at the LARC has grown exponentially. Recognizing this growth, ATC set out to bring additional capabilities to the LARC.

Range upgrades

In fiscal year 2007, the LARC comprised three indoor ranges. In fiscal year 2008, ATC more than doubled the range capacity by adding four more indoor

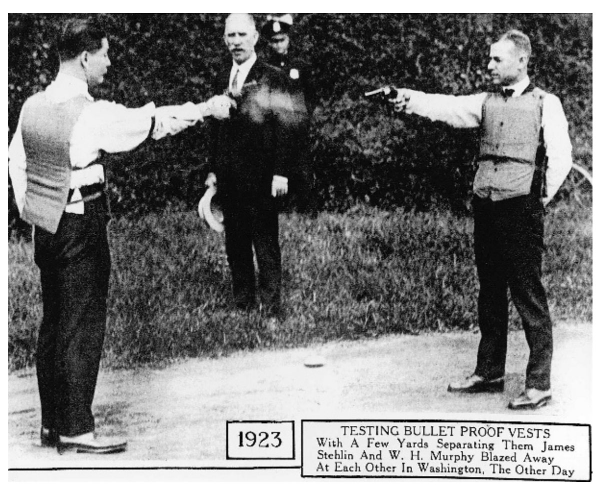


Figure 1. In 1923, two men showcase the survivability of their bullet proof vests in a publicity stunt near Washington, D.C.

Figure 2. A plate being cooled with carbon dioxide before an armor acceptance test in 1962

firing lanes and refurbishing one of the original legacy ranges.

The cornerstone of the LARC upgrades is the four new state-of-the-art indoor ranges, D1–D4. These ranges are designed to provide maximum versatility for indoor armor testing. Traditionally armor targets were tested indoors, whereas vehicles were tested on outdoor ranges. However, ATC recognized that greater test efficiencies could be achieved by moving small arms vulnerability testing of vehicles indoors. As such, the ranges measure 12 feet high, 20 feet wide, and 100 feet long and can accommodate small tactical vehicles such as HMMWVs.

The weapon and target mounts in the new ranges also are designed for test efficiency. Weapons are placed on a mount that enables vertical elevation to be changed with the push of a button. Target fixtures are designed to allow quick and easy adjustments to obliquity, vertical displacement, and horizontal displacement. For larger targets, such as vehicles, the weapon can be mounted on a rail system that enables traverse and elevation changes to be made quickly and effortlessly. The target itself can be placed on a portable turntable that enables attack azimuths to be changed easily.

To provide protection from the noise, blast, and fragments generated during ballistic testing, the interior of the new ranges are layered with AR500 steel capped by 2 inches of ballistic rubber in the form of easily replaceable 2-feet by 2-feet panels. For targets and/or projectiles for which extensive fragmentation is expected, additional shielding can be put in place. At the end of the range is an enclosed 20-feet-long reinforced concrete and sand bullet trap. The net effect of this construction is a range that is equipped to handle up to 30-mm armor piercing projectiles.

Figure 3. In 1967, two workers set up a plate for an armor acceptance test

Of particular concern at the LARC is the ability of the tester to control the temperature and humidity of the test environment. Control of the test environment is now a requirement for many standardized armor test procedures used by the U.S. Department of Defense (DOD), National Institute of Justice, and NATO. These procedures are applicable to both personnel and vehicle armor testing.

The heating, ventilation, and air conditioning system at ranges D1–D4 is a fully automated controlled system. The ranges can be adjusted to any environmental standard set by the customer for their testing needs. The control system is able to maintain indoor climate conditions of 50 percent plus or minus five percent humidity and 68 degrees plus or minus one degree. The control system uses multiple sensors throughout the range, with the primary sensing location down range at the target area. The ability to so accurately control and monitor the temperature and humidity is vital during testing to ensure a valid test.

The heating, ventilation, and air conditioning system also has an elite, closed-loop filtration system. There are two units per range, and there are a total of 24 filters for each range. The filters collect particulate matter such as lead from the air. The pre-filters are changed every 30 days during firing operations, and the HEPA filters are changed every 90 days. Fresh air is continuously circulated to rid the range of fumes.

Test operations for the new ranges are controlled from the test control buildings that are attached to the ranges. The control building houses the ammunition storage and loading room, test control rooms, and a workshop for target preparation and fixture fabrication. Although test efficiency and capability were prime considerations in the design of this state-of-the-art facility, the concept and design of ranges D1–D4 also placed an emphasis on the needs of the customer. As such, real-time viewing of the test event is provided via multiple video monitors in the control room. In addition, the control building houses a conference room for immediate review and discussion of test events and an office space *(Figure 4)*.

Figure 4. One of the new state-of-the-art indoor firing ranges at the Light Armor Range Complex

Instrumentation

ATC test directors at the LARC use various testing instrumentation based on customer requirements. Striking velocity, residual velocity, threat pitch and yaw, threat-target interaction, and behind-armor effects are some of the data elements that may be needed during armor testing. Each test that comes to ATC has its own unique requirements, and ATC test directors work with customers to ensure their testing needs are fulfilled by using a spectrum of instrumentation, technology, and testing knowledge and experience.

Velocity can be measured through a variety of ways. Typically, optical screens and electronic counters are used to determine an instrumental velocity from which a striking velocity is calculated. The system used at the LARC is a double redundant system that uses two sets of optical screens and counters that independently calculate the projectile velocity. A computer program compares the two velocities, ensures they are within 10 feet per second of each other, averages the velocities, and calculates a striking velocity.

If additional data are desired, such as pitch and yaw for example, then ATC has a variety of digital, high-speed video, and flash X-ray systems that can provide the required data. The digital, high-speed cameras used can capture 6,688 frames per second (fps) full resolution and up to 100,000 fps at lower resolutions. X-rays, along with digital image correlation, can provide a three-dimensional reconstruction of the event, enabling critical test data to be gathered.

The laser scanner and associated software is a new technology at ATC. The laser scanning device is used to measure the surface of a material before and after an event. This provides the ability to measure post-shot data on any target and creates a three-dimensional profile of any deformations.

Temperature conditioning chambers are used to prepare the targets to meet customer specifications. Targets can be heated or cooled for any length of time to simulate different conditions.

ATC has the capability to turn around the electronic data gathered from a test that has been verified and validated to authorized customers within a 24-hour period, or 1 business day, so that the customer may make critical testing decisions.

LARC operations

The entire purpose of testing at the LARC is to save lives. How is it possible to ensure that the items we test meet and exceed the standards while setting a wartime testing pace? The LARC in action requires a staff of 20 employees working in two shifts, 6 days a week and working through holidays to keep up with the demand and pace of our customers and the warfighter's needs. The LARC staff currently tests body armor, helmets, armor plate, armor protection kits for vehicles, composites, glass, and ceramics.

Body armor, helmet research, and vehicle armor testing, all top DOD priorities, are being conducted year-round and are expected to continue. In addition, armor production acceptance testing has increased significantly because of accelerated production schedules to meet armor plate demands for the Global War on Terrorism.

On a daily basis, armor mills and manufacturers send samples of their armor to ATC for ballistic acceptance testing. These sample plates are representative of armor lots produced by the mill. Testing is conducted by ATC to assure that the product quality conforms to the specified military technical characteristics and to detect any deteriorations of quality.

During testing, small arms projectiles and fragment simulating projectiles are fired against the armor to determine the V50; i.e., the statistical velocity at which the threat projectile will defeat the armor 50 percent of the time. The threats that are fired and the required V50 that the armor must pass are based on the type and

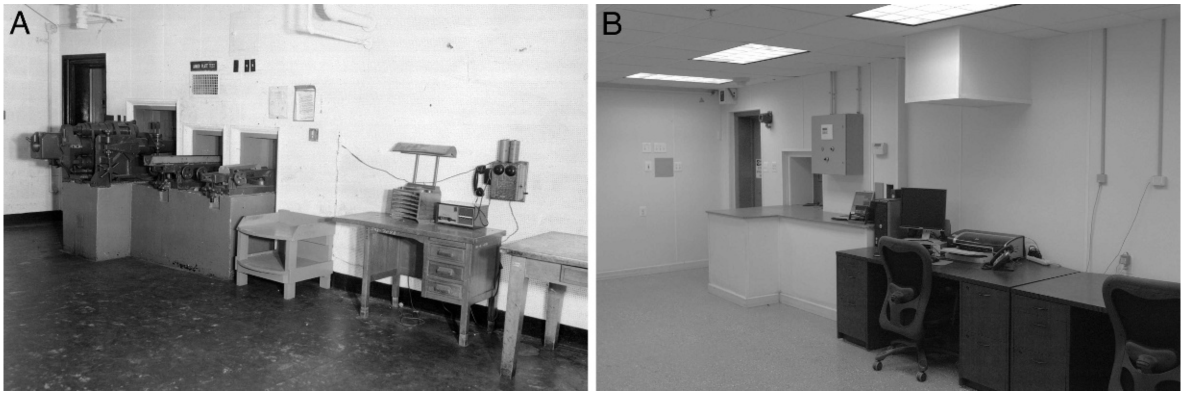


Figure 5A. The control room for legacy range D6 in 1958. 5B. The control room for legacy range D6 in 2008

thickness of the armor and are specified in applicable military standards. Results of the testing are sent to the mill. If the plate(s) successfully completes testing, then the lot of armor represented by the plate(s) is considered ballistically certified by the Army and can then be shipped to manufacturers for use.

The armor acceptance workload has increased dramatically at ATC. More than 5,000 tests have been completed so far this year, which is roughly a 500 percent increase over historical numbers. Because the supply of certified armor is so critical, ATC sets a goal of having plate tested within 2 days of its arrival at the LARC.

Other testing programs at the ranges are always ongoing. At any point, there could be a vehicle undergoing small arms exploitation testing on one range, helmet testing on another, and an armor research and development occurring on a third range.

Future direction

Nearly every future vehicle system will require armor testing at the LARC, and current vehicle systems also will continue to have armor tested at the LARC to arrive at lighter, more efficient solutions. Whereas testing of personnel protective armors was traditionally a relatively small part of the LARC workload, that is no longer true. Like vehicle armor testing, testing of personnel armor systems at the LARC also has increased dramatically. This is because of increased testing of current armor systems, in addition to increased testing of newer, improved armor technologies for future systems. The workload is expected to remain high as quality assurance testing of the new armor systems will be required once production begins.

Additional workload also is expected as the DOD looks to develop improved personnel armors as well as new methodologies and instrumentation to conduct the testing. As a result of this increased workload, ATC is constructing four additional test ranges designed to facilitate the testing of personnel armors.

In fiscal year 2005, the LARC consisted of three World War II era indoor firing ranges. Today it consists of four new state-of-the-art indoor firing ranges and one upgraded legacy range. In the near future, the LARC will consist of four more new state-of-the-art indoor firing ranges and an upgrade of the final two legacy ranges. This 11-indoor-firing-range complex will enable the LARC to meet all of the DOD's needs and requirements *(Figures 5A and B)*.

With the four new ranges, the LARC will be able to provide the full spectrum of indoor personnel and vehicle armor testing. Although the legacy and newer ranges are designed for specific applications, they also are designed to be versatile enough to give ATC and its customers a number of options and a high degree of flexibility in matching ranges to test requirements and workload.

As new materials and threats evolve, so will testing capabilities at ATC, especially in wartime. There will be new testing methodologies developed that will transform testing, just as has occurred in the past, and ATC will continue to evolve and grow to meet the ever-changing world of armor testing. ❑

COL JOHN P. ROONEY is the commander of the U.S. Army Aberdeen Test Center (ATC) located at Aberdeen Proving Ground, Maryland. In 2005, he took command of the Defense Department's lead agency for land-combat, direct-fire, and live-fire vulnerability testing. Since his graduation from the United States Military Academy in 1978, he has had a long, decorated career as an Army officer. His awards and decorations include the Defense Meritorious Service Medal, Legion of Merit Award, Meritorious Service Medal (with five Oak Leaf Clusters), Global War on Terrorism Service Medal, Overseas Service Ribbon, Joint Meritorious Unit Award, and the Parachutist Badge. E-mail: crystal.maynard@us.army.mil

ITEA Journal 2008; 29: 351–359

AFOTEC's Space Test Initiative: Transforming Operational Testing and Evaluation of Space System Capabilities

Major General Stephen T. Sargeant
U.S. Air Force Operational Test and Evaluation Center,
Kirtland Air Force Base, New Mexico
Colonel Suzanne M. Beers
U.S. Air Force Operational Test and Evaluation Center Detachment 4,
Peterson Air Force Base, Colorado

Historically, the value of the operational test and evaluation (OT&E) data has been limited during the acquisition and deployment of space systems because OT&E occurs late in the process, after the satellite is orbiting in space and the ground stations are fielded, well after key acquisition decisions, investments, and critical launch decisions have already been made. This article presents the U.S. Air Force Operational Test and Evaluation Center's Space Test Initiative. The Space Test Initiative delivers an OT&E model that better fits the National Security Space system's acquisition model outlined in NSS 03-01 and delivers better value to both the acquisition and operational decision makers by moving OT&E well before launch.

Key words: Acquisition strategy; integrated testing; investment; OT&E test anatomy; space acquisition; system of systems evaluation.

The U.S. Air Force Operational Test and Evaluation Center (AFOTEC) is responsible for the operational testing and evaluation (OT&E) of all Acquisition Category I and II weapon system programs as well as those on Director of Operational Test and Evaluation oversight, acquired by the Air Force and often our Joint partners, to determine operational effectiveness, suitability, and degree of mission capability in the system's intended operational environment. Since AFOTEC's inception in 1974 and the creation of Air Force Space Command in 1982, OT&E of space systems has occurred after satellites are on orbit and ground stations are fielded. Therefore, AFOTEC could not fully meet its responsibility to provide independent OT&E data to key decision makers in a timely manner with regard to the acquisition and deployment decisions of space systems because the tests occurred after the decisions were already made.

The need for fully informed decisions regarding these increasingly expensive, yet indispensible capabilities is crucial in today's environment of constrained resources. For more than 20 years, AFOTEC and the other service operational test agencies (OTAs) conducted OT&E of space and other high-tech, limited-quantity systems using a model more appropriate for military systems with large-scale production decisions. Using an OT&E model that does not match the system's acquisition strategy renders the results of OT&E largely irrelevant. AFOTEC's "Space Test Initiative" delivers an OT&E model that better fits the National Security Space (NSS) system's acquisition model outlined in NSS 03-01 (DoD 2004) and provides fact-based decision quality data to decision makers in time to support their key space system acquisition decisions.

Figure 1 further illustrates the issue. In a traditional acquisition program governed by Department of Defense Directive (DoDD) 5000.1 (DoD 2003), expenditures are relatively small in the research and development and investment phases compared to the cost of production and system operation. For these traditional acquisitions, operational testing (OT) occurs just before the major investment or production decision and provides data to inform those decisions adequately.

However, most of the investment for space systems occurs early in the program, most often without a major production decision. In the current space OT&E

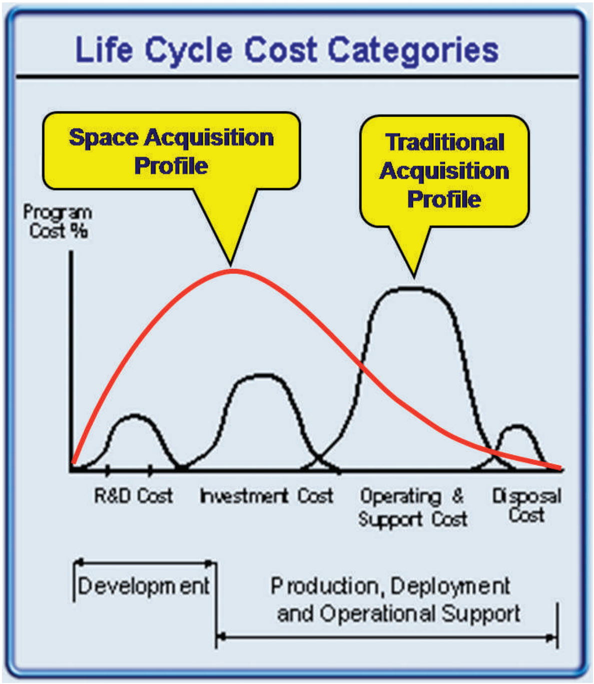


Figure 1. Department of Defense Directive 5000.1 versus National Security Space 03-01 life cycle costs

model, OT&E takes place at the same point in the acquisition cycle as with the DoDD 5000.1 (DoD 2003) programs. However, by this point in NSS 03-01 (DoD 2004) programs, most of the investment has been made, most of the key acquisition decisions have been made, and the critical operational decision to launch the satellite has been made and executed. The ground station and associated software often lag in deployment, making timely post-launch OT&E difficult, if not impossible. Making these key decisions before the execution of OT&E severely limits the value of OT&E.

AFOTEC's Space Test Initiative provides an OT&E model that better fits the space systems acquisition model, delivering better value to both the acquisition and operational decision makers by moving OT&E activity well before launch. The three key tenets of the Space Test Initiative are:

- early and continuous integrated testing involvement throughout the life cycle of the system,
- agile analysis and reporting,
- focus on system-of-system evaluations.

Space test anatomy

AFOTEC's OT&E guide provides an "Anatomy of an OT&E" that describes OT&E activities associated with each phase of a typical acquisition program. The anatomy is built on the DoDD 5000.1 acquisition model, which did not fit well for space system acquisition. In order to guide the OT activities of space systems, a NSS 03-01-focused OT&E anatomy needed development. In July 2008, AFOTEC hosted an Air Force Space Summit at Kirtland Air Force Base, New Mexico, where space acquisition, operations, and testing experts from across the Air Force gathered to build a new test anatomy. After the summit, event organizers socialized the ideas to the broader space acquisition and testing community both inside and outside the Air Force. This action included the other Service OTAs, the Joint Staff, Undersecretary of Defense (Acquisition, Technology, and Logistics), the national intelligence community, and the Director of OT&E. Comments received during that socialization resulted in slight modifications to the summit's model. In this article, we will walk through the resulting anatomy in a phased approach.

The activities shown in orange at the top of the anatomy *(Figure 2)* are conducted by the acquisition community. Those shown in light blue are conducted by the developmental test (DT) community. The grey region with the activities highlighted in yellow are integrated test activities, conducted by both the DT and OT communities. Finally, the blue boxes near the bottom of the anatomy are activities led by the OT community.

Beginning at the left of the anatomy, early in the acquisition process, the acquisition community receives strategic guidance or a description of the operational mission need. The acquisition community begins development of the initial Functional Solution Analysis or system concepts to address the operational mission need.

During the pre-Key Decision Point (KDP)-A period, the integrated test (IT) community begins development of an early involvement strategy. The early involvement strategy tailors this generic model to the specifics of the program, taking into consideration the required decisions, development, testing activities, etc. In addition, during this early phase the group responsible for building operational requirements forms the Integrated Concept Team. Members of the DT and OT communities also form the Integrated Test Team (ITT) and develop the ITT charter.

As the Integrated Concept Team develops the Functional Solution Analysis and the draft Initial Capabilities Document, the IT community is involved in the early reviews of the proposed concepts to generate a Concept Assessment Report. The report provides input to the concept decision, focused on the degree to which the system concept meets the mission needs stated in the strategic guidance.

While the acquisition community moves into the solution definition phase, the IT community partici-

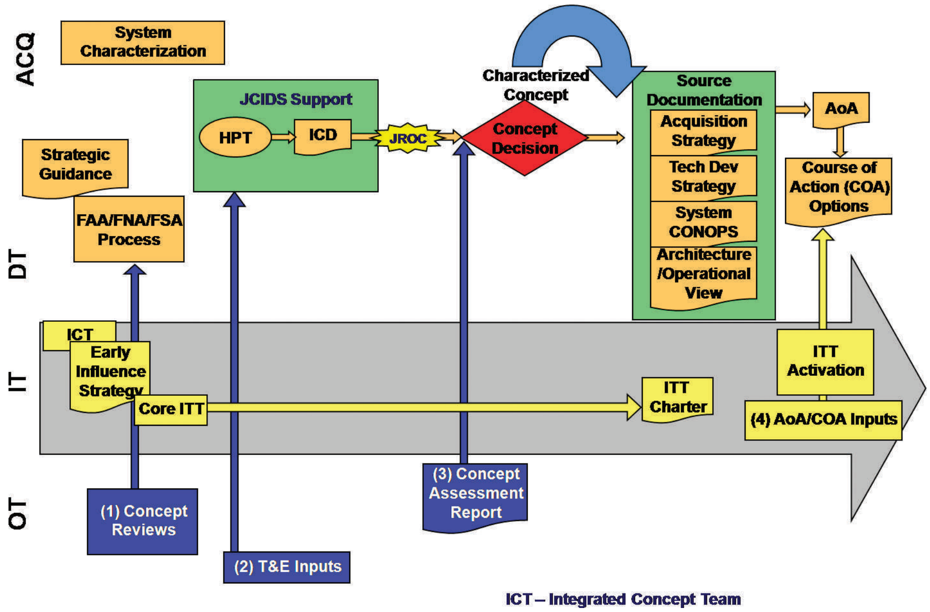


Figure 2. Pre-Key Decision Point-A activities

pates in the analysis of alternatives (AoA) and course of action (COA) development processes. The participation of the ITT in the AoA provides candidate evaluation criteria, potential measures of effectiveness and suitability, and operational scenarios for each alternative being considered. As the acquisition community develops the AoA and COA, the ITT develops the first test and evaluation strategy by melding DT and OT objectives.

The ITT's participation in the AoA/COA culminates in an operational assessment (OA). The resulting OA report informs the KDP-A, Concept Approval, decision. The OA report provides information on the degree of potential operational effectiveness and suitability, highlights any disconnects between the alternatives and the operational mission need, and identifies any potential testing issues of the AoA's alternatives and the COA's acquisition strategies. The OA report does not advocate or recommend an alternative.

Post-KDP-A to KDP-B, concept development phase

Throughout the KDP-A to KDP-B concept development phase *(Figure 3)*, the acquisition community refines the acquisition concept and matures both the technology and functional capabilities of the system. Meanwhile, the ITT continues to refine the test and evaluation strategy and builds the integrated test plan.

During the concept development phase, as the acquisition community translates the operational requirements into a set of technical requirements to serve as the basis of the Request for Proposals, the ITT evaluates the Capability Development Document/ Technical Requirements Document traceability (see *Figure 4*). The look by the ITT at traceability focuses on the translation of operational requirements into the technical requirements that will ultimately serve as the basis for the system design. Throughout the system requirements review and system design review process, the technical maturation and functional development process generates concepts and prototypes. The ITT conducts OAs on these prototypes to evaluate their potential operational effectiveness, suitability, and degree to which they will meet the operational mission need, and to highlight any other operational issues noted during early testing.

The IT planning process culminates in the publication of the initial version of the Test and Evaluation Master Plan describing the integrated test approach. Finally, the IT community conducts an OA to assess the system's concept just before KDP-B to inform the KDP-B decision with an operationally focused evaluation of the system concept (see *Figure 5*).

Post-KDP-B to KDP-C, preliminary design phase

In the KDP-B to KDP-C preliminary design phase *(Figure 6)*, the acquisition community refines the system design through a series of design reviews and technology demonstrations. The IT community further refines their IT planning documents, wrapping up the preliminary design phase with a Test and Evaluation Master Plan update and an initial OT&E

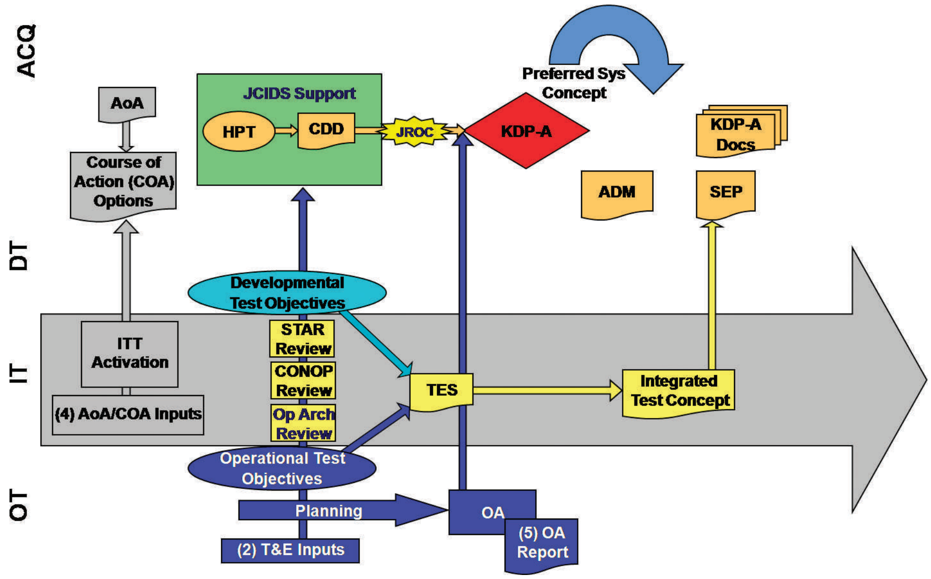


Figure 3. Key Decision Point-A activities

plan that fleshes out the details of how OT objectives will be addressed by traditional dedicated DT testing activities, such as laboratory and chamber testing.

During the preliminary design phase, developers conduct technical demonstrations to evaluate increments or components of the proposed system. The ITT is involved to provide status reports to the system program office on the potential operational effectiveness, suitability, the degree to which they will meet the operational mission need, and any other noted operational issues. In addition, these status reports begin to form an assessment of the system-of-system interfaces required for the system to operate successfully within its operational architecture.

In conjunction with the preliminary design review, the OTA conducts an OA to aggregate the information gathered through the preliminary design review stage to inform the KDP-C, Final Design Entry, decision on the potential operational effectiveness, suitability, and degree to which they will meet the operational mission need. Additionally, if the acquisition authority decides during this timeframe to allow the contractor to procure long lead items, part of the OA evaluates the operational aspects of those system components.

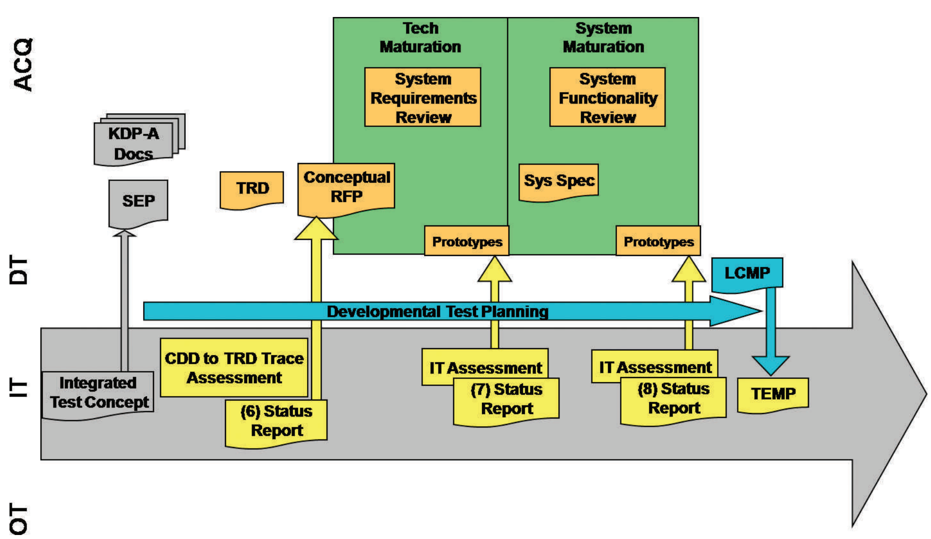


Figure 4. Key Decision Point (KDP)-A to KDP-B activities

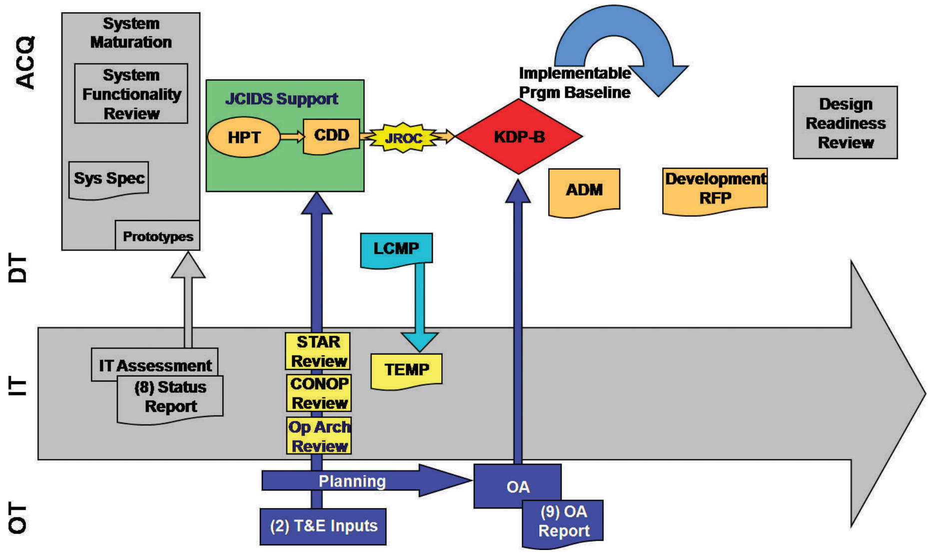


Figure 5. Key Decision Point-B activities

KDP-C to build approval, final design phase

In the final design phase *(Figure 7)*, the acquisition community refines the system design and conducts a series of risk-reduction tests, building up from component tests to subsystems to operational system tests. The IT community is involved with all testing activities. ITT participation is collaborative, and the generated status reports foster open communication between testers and developers as the system design is finalized.

At the conclusion of the critical design review, the OT&E community produces an Operational Assessment Report providing information on the potential operational effectiveness, suitability, and degree to which the proposed design will meet the operational mission need. The critical design review and Design Assessment Report inform the Build Approval decision.

System production to OT&E phase I

After Build Approval, the acquisition community produces the system and conducts a series of test activities, building up from the component to subsystem to full operational system testing. During the system production to OT&E phase I period *(Figure 8)*,

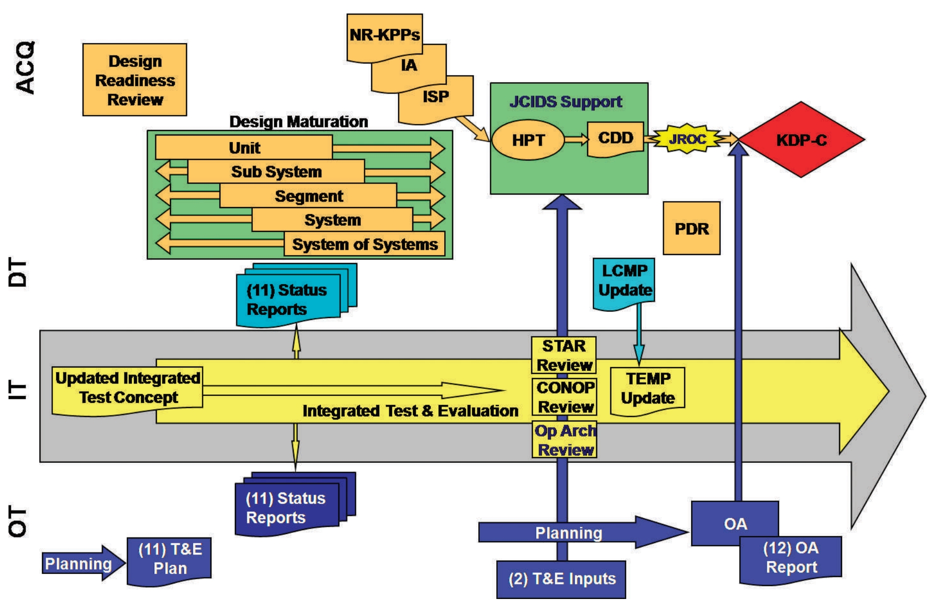


Figure 6. Key Decision Point (KDP)-B to KDP-C activities

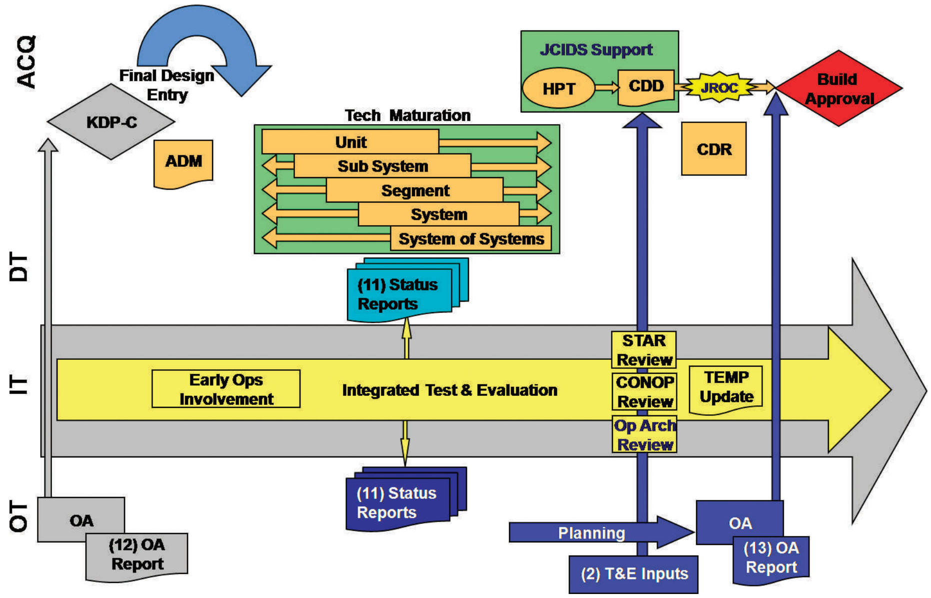


Figure 7. Key Decision Point-C to build approval activities

the ITT participates in the testing, taking full advantage of planned DT events to inject OT test measures and scenarios and gather information to fulfill OT&E test objectives. Status reports informing developers on how the system production is progressing, from both the adherence of the development to specification and the operational community's assessment of meeting operational requirements, keep the lines of communication open between the operational and developmental communities.

The system production period culminates in an OT&E Phase I, with its associated Program Element Officer certification and Test Readiness Review processes. The OT&E Phase I puts the system in as near an operational environment as can be replicated on the ground to support OT&E to inform the Consent to Ship decision. The Phase I OT&E takes into consideration the results of integrated testing, as well as the status of the system-of-systems required to provide mission capability to the warfighter. For example, this report may highlight that the satellite is ready for launch, but the ground segment will not be completed for another 2 years, enabling a conscious decision to delay satellite preparation for launch until the right time to optimize value to the warfighter.

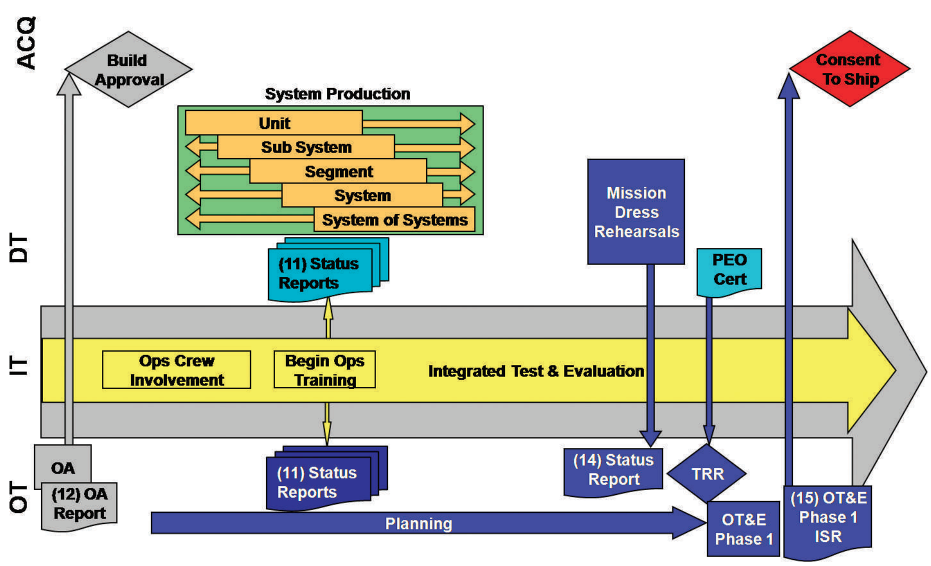


Figure 8. System production to operational test and evaluation Phase 1 activities

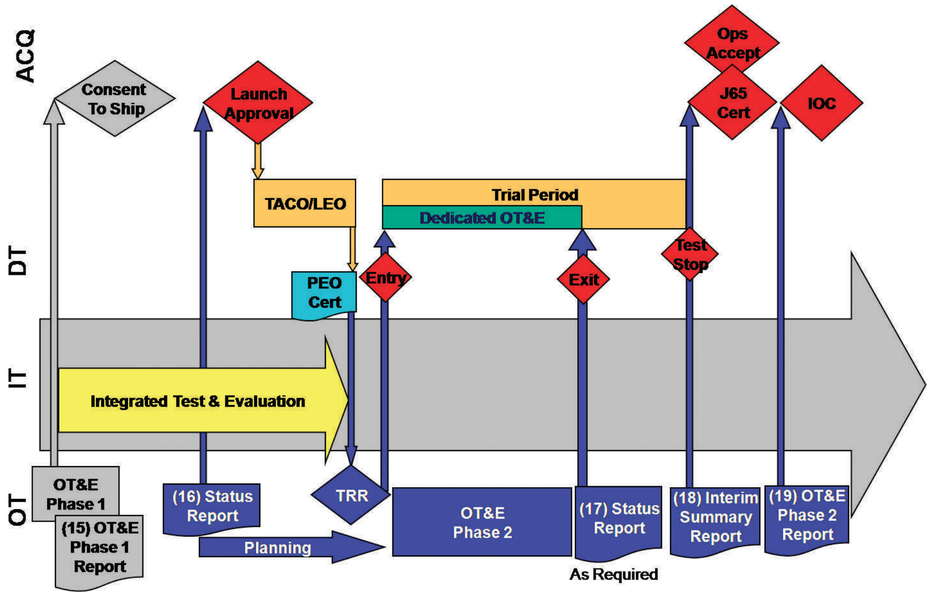


Figure 9. Launch and early orbit operations

Launch range compatibility testing

After deciding to ship the satellite from the manufacturing facility, the system is moved to the launch range, mated with the booster, and final integration and communication testing occurs. Again, integrated testing will inject OT test measures and scenarios into the DT-centric checkout events to provide an operational impact to any technical issues identified during compatibility testing. Integrated testing, documented in a Status Report, informs decision-making at the launch go/no-go decision point.

Launch and early orbit operations, OT&E phase II

After launch and during test and checkout, early orbit operations, and sensor checkout, the operational testing community participates to the greatest extent possible to inject operationally realistic scenarios, backgrounds, and procedures *(Figure 9)*. At the conclusion of the test and checkout period, the Program Element Officer certifies the system is ready to enter OT&E Phase II, the final 10 percent checkout of the operational capability of the system. OT&E Phase II takes a final look at whether the system made its ride to orbit successfully, if the performance reported throughout early integrated testing bears out in the operational environment of space, and that the system-of-system environments represent the true operational architecture and operate as expected.

AFOTEC conducts OT&E Phase II in conjunction with the users' operational trial period to facilitate delivering mission capability to the warfighter. At the conclusion of the OT&E Phase II and exit from the trial period, AFOTEC generates a status report to identify the hard-hitting, show-stopping issues found during this final stage of operational testing. The status report informs the *Operational Acceptance Decision*.

Depending on the program, the interim summary report, an approximately 20-page document that begins to draw conclusions and ratings, informs decisions such as the USSTRATCOM/J65 certification decision. Finally, AFOTEC publishes the OT&E report to provide full details on the results of the analyses. This report informs the Director of OT&E's Report to Congress, Initial Operational Capability decisions, future system upgrade decisions, etc.

Wayahead

To develop the next level of detail and implement the Space Test Initiative, a number of actions are required and in most cases are already in works. These actions include:

Understand/include detailed DT activities. The developmental test activities associated with the design development and maturation phases and system production cycles need further definition and inclusion in this model.

Define necessary policy. Current DoD, Air Force, Air Force Space Command, and AFOTEC policy does not speak to conducting space operational testing in the manner described in the Space Test Initiative. Therefore, AFOTEC initiated a policy crosswalk to determine what is in existing policy and what must be written to allow and direct the Space T&E Anatomy.

AFOTEC, in conjunction with the Air Staff, will draft the necessary policy documentation for incorporation into the current regulations.

Identify and define underlying test and evaluation processes. AFOTEC will define the processes required to execute this Space T&E Anatomy, include details on organizational roles and responsibilities, and entrance/exit criteria for each phase.

Identify and define test personnel resources. The number of personnel required to execute the Space Test Initiative, along with the required skill sets, will be defined. It is likely that AFOTEC will not have, or be able to increase, their personnel pool to provide the technical expertise necessary to execute the Space Test Initiative, particularly the early engineering-focused activities. Therefore, we must build agreements among the members of the integrated test and development community to share personnel resources.

Define capabilities and gaps in test infrastructure. Execution of the initiative's OT&E Phase I test infrastructure requires improvement in order to emulate an operationally realistic test space environment on the ground. For example, OT&E Phase I will have to use vacuum chambers that provide the capability to connect operational communication and command and control links.

Select a long-term candidate program to define cost/benefit. While AFOTEC Detachment 4 intends to apply this concept to all future space OT&E programs, they will select a pilot program to demonstrate and define the cost and benefits of this new approach. In addition, AFOTEC will use the pilot program to refine the concept, adding lessons learned as we execute these ideas from beginning to end on a space program.

Identify and define required contract changes. Most current space acquisition programs, particularly those initiated during the acquisition reform era, provide limited opportunity for government participation or insight into most development activities, or provide for test community access to developmental testing data. We require future contracts be written to allow the integrated test activities, as the ability to implement the Space Test Initiative depends on access to developmental data for analysis.

Space Test Initiative benefits

AFOTEC's Space Test Initiative provides the basis for knowledge-based acquisition and operational decisions throughout the life cycle of our national security space systems. It provides early operational involvement that will deliver a number of benefits, including: (a) ensuring the warfighter receives needed mission capabilities, (b) providing early clarity and continued update of operational requirements, (c) influencing early and continual development and refinement of the Concept of Operations, (d) ensuring frequent reviews of threat documents to ensure the system design addresses current threats, (e) highlighting program shortfalls and benefits throughout the development process when they can be addressed most efficiently and inexpensively, (f) enabling the user to understand and accept acquisition risks and adjust their mission requirements and plans accordingly, and (g) addressing and correcting systemic suitability issues early in the program development.

Other applications

Although AFOTEC's initiative focuses on space systems with its satellite-specific activities of Consent to Ship, Launch, and Early Orbit Operations, the model can be applied to other high-tech, small-quantity programs, such as one-of-a-kind command and control and information systems. Information systems can also benefit from the model of early testing since these programs are similarly front-loaded on investment with relatively little expense on production, operations, and maintenance once fielded. Like most space programs, no two information system programs are the same and few follow the DoDD 5000.1 template exactly. Unlike space programs, however, the DoD does not field information systems at one time (launch). Instead, DoD fields information systems in increments of capability. The fielding difference drives a requirement to test sooner and more often than space programs. However, the Space Test Initiative offers a model for information systems because the fundamental principles apply: (a) early and continuous integrated test involvement throughout the system's life cycle, (b) agile analysis and reporting, and (c) focus on system-of-systems evaluations. If a flexible, agile test approach is not used, the warfighter faces the dilemma of fielding capabilities before testing.

Summary

AFOTEC's proposed Space T&E Anatomy provides a model for testing systems governed by NSS 03-01. It identifies early test, evaluation, and reporting activities to inform acquisition and operational decisions, providing a roadmap for early program influence. The anatomy also provides an overarching model for each individual program's tailored implementation, as no two NSS programs (or DoDD 5000.1 programs for that matter) follow the standard NSS 03-01 model.

The benefit of the AFOTEC Space Test Initiative will be better space warfighting systems acquired through early, continuous integrated testing involvement, providing inputs to the requirements processes to ensure the system addresses the mission capability gap and informing early program decisions when changes are less costly. The initiative focuses the majority of the OT&E effort, conceptually 90 percent of the OT&E community's time, on pre-launch to inform the key Consent to Ship decision. With early and continuous involvement, we will ensure that leaders make conscious, fact-based decisions to send satellites into orbit and field new ground stations when the complete system-of-systems required to deliver warfighting capability is in place. ❑

MAJOR GENERAL STEPHEN T. SARGEANT is the Commander of Air Force Operational Test and Evaluation Center (AFOTEC) at Kirtland Air Force Base, New Mexico. Major General Sargeant reports to the Air Force Chief of Staff regarding the operational test and evaluation of more than 200 major programs being assessed at 22 different locations. He directs the activities of more than 950 civilian and military personnel. As a member of the testing and evaluation community, he works directly with the Office of the Secretary of Defense and Headquarters U.S. Air Force, Washington D.C., to ensure realistic, objective, and impartial operational testing is conducted on Air Force and Joint use systems. Major General Sargeant has served as the commandant of the Air Force Weapons School at Nellis AFB, Nevada, commanded the 8th Fighter Wing at Kunsan Air Base, South Korea, and the 56th Fighter Wing at Luke AFB, Arizona. He has also served in numerous Air Force, Joint, and Coalition staff assignments, including 18 months in Baghdad, Iraq, as the C-5 for CJTF-7 and MNF-I. He is a command pilot with more than 3,000 flying hours in the A-10/A and F-16 A/B/C/D. E-mail: steve.sargeant@afotec.af.mil

COLONEL SUZANNE M. BEERS is the Commander of Air Force Operational Test and Evaluation Center Detachment 4 at Peterson Air Force Base, Colorado. She is responsible for leading, planning, executing, and reporting operational tests and evaluations of 21 space, intercontinental ballistic missile and national missile defense test programs valued at more than $140 billion. Colonel Beers advises and interfaces with Air Force Space Command, U.S. Strategic Command, U.S. Northern Command, North American Aerospace Defense Command, Air Force Materiel Command, and the Missile Defense Agency to ensure test partnerships provide the warfighter with the best possible systems and the operational tests conducted by AFOTEC describe the impact the systems will have in the operational battlespace. Colonel Beers has held assignments in Air Force research and development, acquisition, testing, and sustainment for cruise missiles, satellites, and space command and control centers. She has operational experience with ground launched cruise missiles and the Air Force Satellite Control Network and leadership experience at the acquisition division and operational group level. E-mail: suzanne.beers@peterson.af.mil or suzanne.beers@us.af.mil

References

DoD. 2003. Defense Acquisition Guidebook, DOD Directive 5000.1. The Defense Acquisition System. May 12, 2003. Washington, DC: Department of Defense. Available at https://akss.dau.mil/dag/DoD5000.

DoD. 2004. National Security Space (NSS) Team Acquisition Policy Number 03-01. December 27, 2004. Washington, DC: Department of Defense. Available at https://akss.dau.mil/Documents/Policy/Space%20Acquisition%20NSSAcqPol0301_signed_%2027Dec04%20(GN).pdf.

ITEA Journal 2008; 29: 360–366

Challenges of Embodied Artificial Intelligence and Robotics

Bruce J. MacLennan, Ph.D.
Department of Electrical Engineering & Computer Science,
University of Tennessee, Knoxville, Tennessee

Recent developments in cognitive science, artificial intelligence, and robotics promise a new generation of intelligent agents exhibiting more of the capabilities of naturally intelligent agents. These new approaches are based on neuroscience research and improved understanding of the role of the body in efficient cognition. Although these approaches present many advantages and opportunities, they also raise issues in the testing and evaluation of future artificial intelligence systems and robots. We discuss the problems and possible solutions.

Key words: Adaptation; artificially intelligent agents; connectionist AI systems; embodied cognition; robots; expert systems; learning; training.

Recent research into human and animal cognition has improved our understanding of natural intelligence and has opened a path forward toward artificially intelligent agents, including robots, with much greater capabilities than those implemented to date. Improved understanding of the neural mechanisms underlying natural intelligence is providing a basis for implementing an efficient, robust, and adaptable artificial intelligence (AI). However, the nature of these mechanisms and the inherent characteristics of AI based on them raise significant issues in the test and evaluation of this new generation of artificially intelligent agents. This article briefly discusses the limitations of the "old AI," the means by which the "new AI" aims to transcend them to achieve an AI comparable to natural intelligence, the test and evaluation issues raised by the new AI, and possible means for dealing with these issues to deploy robust and reliable systems capable of achieving mission objectives.

The nature of expertise

It used to be supposed that human expertise consists of internalized rules representing both knowledge and inference. Knowledge was considered a collection of (general or specific) facts that could, in principle, be expressed as sentences in a natural language. Similarly, the process of thought was supposed to be represented by rules of inference, such as the laws of logic, also expressible in natural language. It was granted that natural language was vague, ambiguous, and imprecise, and so artificial languages, such as symbolic logic, were proposed as more adequate vehicles for knowledge representation in the brain. *Knowledge representation languages*, which were often used in AI, were effectively programming languages for operating on knowledge represented by language-like data structures.

A full critique of this model of knowledge and cognition is beyond the scope of this article, so I will just mention a few key points (for more, see, e.g., Dreyfus 1979, Dreyfus and Dreyfus 1986). One objection was that neuroscience provides no evidence that the brain is structured like a stored-program computer. In answer, it was argued that the abstract idea of a general-purpose computer (i.e., of a universal Turing machine) could be implemented in many radically different ways, and so the brain could be such a machine even though it is very different from ordinary computers. It was argued that in any case there was no reason for AI to slavishly imitate the brain; we could use our technologically superior digital computers. Another objection was that whereas we are sometimes conscious of following verbalizable rules, much of our intelligent behavior takes place without conscious rule following. In answer it was argued that well-learned behaviors were "compiled" into unconscious neural operations, much as programs written in high-level languages are compiled into machine code. A third objection was that, although it might be plausible that human knowledge and inference were represented in language-like rules, this was implausible as a model for nonhuman animal cognition, especially in simpler animals with no language-using ability. One answer was that nonhuman animals do not have conceptual knowledge, which is "true knowledge," as opposed to concrete memory and instinctive stimulus-response behaviors; only humans exhibit "true" cognition. An overarching defense of rule-based models of knowledge and inference was that they are "the only game in town"; that is, that there were no defensible alternative models. Nevertheless, there are additional objections to rule-based approaches.

Even for humans, who have complex and expressive linguistic abilities, research shows that rules do not account well for expert behavior. As an example, I will use the book of Hubert L. Dreyfus and Stuart E. Dreyfus (1986), who summarize much of the research. Based on characteristic cognitive processes, they identify five levels of expertise: (1) novice, (2) advanced beginner, (3) competence, (4) proficiency, and (5) expertise (Dreyfus and Dreyfus 1986, pp. 16–51). They apply this classification to "expert systems," which are rule-based AI systems incorporating appropriate inference rules and a large knowledge base oriented toward some domain of knowledge. They argue that these systems operate at best at the "competence" level, which is characterized by goal-directed selection and application of rules (Dreyfus and Dreyfus 1986, pp. 23–27, 101–121). However, expert systems cannot perform at the "proficient" level, which is characterized by unconscious, similarity-based apprehension of the situational context in which cognition should occur rather than by conscious, rational "calculation" (rule-based determination) of the context (Dreyfus and Dreyfus 1986, pp. 27–30). This apprehension of context is critical to proficient behavior, as it enables cognition to focus on stimuli that are relevant to the situation, without wasting time considering and rejecting those that are not. Experts apply rules, if at all, in a flexible, non-rigid, context-sensitive way, which is why it is difficult to capture expertise in rules (Dreyfus and Dreyfus 1986, pp. 30–36, 105–109). How, then, can we design artificially intelligent agents that exhibit true expertise?

Connectionism

The rule-based approach to knowledge representation and inference continued to dominate AI so long because there did not seem to be any viable alternative. However, H.L. Dreyfus (1979) and others pointed the way to a different approach. First, since human and animal intelligence is realized in the physical brain, it seemed apparent that an AI would be possible, although the AI system might have to be more like a brain than a conventional computer. Second, in the 1960s and 1970s, Pribram, Dreyfus, and others observed that human pattern recognition and memory seemed to have properties similar to optical holograms, as did simple models of neural networks (e.g., Anderson, Pellionisz and Rosenfeld 1990, chapter 7; Dreyfus 1979, pp. 20, 25, 51; Dreyfus and Dreyfus 1986, pp. 58–63, 90–92, 109; Haugeland 1978; Hinton and Anderson 1989; Pribram et al. 1974). These considerations helped to revitalize, in the early 1980s, the study of neural network computation, which had been languishing for about a decade (for more on the history of neural networks and connectionism, see MacLennan 2001, seminal papers are collected in Anderson and Rosenfeld 1988, Anderson et al. 1990, Haugeland 1997).

Connectionism is used to refer to approaches to knowledge representation and inference that are based on simple neural network models. In rule-based approaches, knowledge is represented in language-like discrete structures, the smallest units of which are *features*: predicates for which many languages have words (e.g., "feathered," "winged," "two-legged," "egg-laying" are some features of birds). Connectionist representations, in contrast, are based on large, unstructured arrays of *microfeatures*. A microfeature is a property localized to one of a large number of parts of a sensory or memory image (e.g., a green pixel at a particular location in an image), which generally does not have any meaning in isolation. They are *not* the sorts of things for which natural languages have words, because they are not normally significant, or even perceptible, in isolation. In a typical neural net representation, the activity level of a neuron (usually a continuous quantity) represents the relative degree of presence of a corresponding microfeature in the representation (e.g., the amount of green at that location in the image). As a consequence of the foregoing, connectionist representations are typically holistic in that individual elements have meaning only in the context of the whole representation.

Connectionism derives its name from the fact that knowledge is encoded in the connections between neurons. Because these are connections among (typically large) numbers of neurons representing microfeatures, connectionist knowledge representation is characteristically distributed and nonlocal. It is *distributed* in that the representation of what we normally think of as one fact or behavior is distributed across many connections, which affords connectionist knowledge representations a high degree of useful redundancy. It is *nonlocal* in that each connection participates in the representation of a large number of facts and behaviors. Therefore, a large number of connections in a neural network can represent a large number of facts and behaviors, but not in a one-to-one manner. Rather, the entirety of the connections represents the entirety of the facts and behaviors.

Biological neurons are notoriously slow compared to contemporary electronics; their maximum impulse rate is less than 1 KHz. And yet brains, even of comparatively simple animals, solve problems and coordinate activities that are beyond the capabilities of state-of-the-art computers, such as reliable face recognition and locomotion through rough and complex natural environments. How is this possible? Part of the answer is revealed by the "100-Step Rule" (Feldman and Ballard 1982). This is based on the simple observation that if we take the time for a simple cognitive action, such as recognizing a face ($\ll$1 second) and divide it by the time it takes a neuron to fire ($\gg$1 millisecond), we find that there can be at most about 100 sequential processing steps between sensation and action. This reveals that brains process information very differently from contemporary computers. Information processing on traditional computers is *narrow-but-deep*, that is, it depends on the sequential execution of very large numbers of very rapid operations; even if execution is not completely sequential, the degree of parallelism is very small compared to that of a brain. In contrast, information processing in brains is *shallow-but-wide*: there are relatively few sequential layers of information processing, as reflected in the 100-Step Rule, but each layer is massively parallel on a scale that is qualitatively different from contemporary parallel computers. For example, in the retina approximately 100 million retinal cells preprocess visual data to be transmitted by approximately one million optic nerve fibers, which indicates the degree of parallel processing in visual information processing. Because neural density is at least 146,000 neurons/mm^2 throughout the human cortex (Changeux 1985, pp. 51), most neural modules operate with degrees of parallelism on the order of hundreds of thousands or millions.

Another difference between most contemporary computers and biological neural networks is that neurons are fundamentally *analog* computing devices. Continuous quantities are represented by the frequency, and in some cases the phase, of neural impulses propagating down the axon of a neuron (output fiber). Knowledge is stored in the "strength" of the connections between neurons, which depends on the diffusion of chemical signals from a variable number of sources to a variable number of receptors and is best treated as a real-valued "weight." The resulting electrical signals propagate continuously down the dendrites (input fibers) of a neuron, obeying electrical "cable equations" (Anderson 1995, pp. 25–31), and are integrated in the cell body into a continuous membrane potential, which governs the frequency and phase of the spiking behavior of the neuron (Gerstner and Kistler 2002).

It should be noted that analog signal processing in the brain is low-precision: generally continuous quantities are estimated to be represented with a precision of less one digit. Paradoxically, humans and other animals can perform perceptual discriminations and coordinate sensorimotor behaviors with great precision, but brains use statistical representations, such as "coarse coding" and other population codes (Rumelhart et al. 1986, pp. 91–96; Sanger 1996), to achieve high-precision representations with low-precision components. These techniques, which exploit large numbers of neurons, have additional benefits in terms of reliability, robustness, and redundancy.

Similarly, artificial neural networks are usually based on analog computing elements (artificial neurons or *units*) interconnected by real-valued weights. Of course these continuous computational systems, like other continuous physical systems, can be simulated on ordinary digital computers, and that is the way many artificial neural networks are implemented. However, many advantages can be obtained by implementing artificial neural networks directly in massively parallel, low-precision analog computing devices (Mead 1989), a topic outside the scope of this article (MacLennan in press).

The ability to adapt to changing circumstances and to learn from experience are hallmarks of intelligence. Further, learning and adaptation are critical to many important applications of robotics and AI. Autonomous robots, by their very autonomy, may find themselves confronting situations for which they were not prepared, and they will be more effective if they can adapt appropriately to them. Autonomous robots should also be able to adapt as the parameters and circumstances of their missions evolve. It is also valuable if AI systems and robots can be trained in the field to perform new tasks and if they can generalize previous training to new, unanticipated situations. How can learning, training, and adaptation be accomplished?

An important capability of connectionist AI systems is that they can learn how to do things that we do not know how to do. This is the reason that connectionist systems are said to be *trained*, but not *programmed*. In order to program a process, you need to understand it so well that it can be reduced to explicit rules (an algorithm). Unfortunately, there are many important problems that are not sufficiently well understood to be programmed, and in these cases connectionist learning may offer an alternative solution. Many connectionist (or neural network) learning algorithms have been developed and studied during the last several decades. In *supervised learning*, a network is presented with desired input-output pairs (e.g., digital images and their correct classifications), and the learning algorithm adjusts the network's interconnection weights so it will produce the correct outputs. If the training is done properly the network will be able to generalize from the training inputs to novel inputs. In *reinforcement learning*, the network is told only whether it has performed correctly; it is not told the correct behavior. There is a very large literature on neural network learning, which is beyond the scope of this article (see, e.g., Haykin 1999).

One characteristic of connectionist learning is that, although connectionist systems can sometimes adapt very quickly, they can also adapt gradually, by subtle tuning of the interconnection weights. Rule-based systems can also adapt, but the fundamental process is the addition or deletion of a complete rule, a more brittle procedure. Thus, connectionist systems are better able to modulate their behavior as they adapt and to avoid instability.

Embodied cognition

An important recent development is the theory of *embodied cognition* and the related theories of *embodied AI* and *embodied robotics*. The theory of embodied cognition addresses the important—indeed essential—role that the body and its physical environment plays in efficient cognition. As Dreyfus (1979, pp. 248–250, 253) observed long ago (1972), there are many things that humans know simply by virtue of having a body. That is, there is much knowledge that is implicit in the body's state, processes, and relation to its physical environment, and therefore this knowledge does not need to be represented explicitly in the brain. The theory of embodied intelligence has its roots in phenomenological philosophy (e.g., Dreyfus 1979, pp. 235–255) and the pragmatism of William James and John Dewey (Johnson and Rohrer 2007).

For example, we swing our arms while we walk, which helps maintain balance for bipedal locomotion, but our brains do not have to calculate the detailed kinematics of our limbs. Rather, our limbs, joints, etc. have their characteristic frequencies etc., and all our brain must do is generate relatively low-dimensional signals to modulate these physical processes to maintain balance, as monitored by bodily sensors (inner ear, skin pressure, joint extension, etc.). The brain's goal is not to *simulate* the physical body in motion (a computationally intensive task), but to *control* the physical body in interaction with its physical environment in real-time by means of neurally efficient computation. As opposed to a computer simulation of a robot, the brain's computations constitute a complete description of the body's motion only in the context of a specific physical body in its environment.

Because, in effect, an animal's brain can depend on the fact that it is controlling a body of a specific form, it can offload some information processing tasks to its physical body. For example, rather than calculating from first principles the muscle forces that will move its limb to a particular location, it can leave this "calculation" to the physical limb itself by learning correlations between effector signals and corresponding sensory responses (for which neural networks are ideally suited). Therefore also, if a weight (such as a cast) is put on a limb, or its motion is restricted by pain or an injury, an animal can adapt quickly to the change (an important goal for our robots too).

The power and efficiency of embodied cognition is exemplified by insects and other simple animals that behave very competently in their environments but that have very small brains. Understanding how they exploit embodiment for information processing—or, more precisely, how they obviate the need for information processing—will help us to design more competent autonomous robots, especially insect-size or smaller robots.

Studies of natural intelligence, and in particular of how the brain exploits the physical characteristics of the body and of its environment to control the body in its environment, has contributed to and will continue to contribute to the design of future robots (Brooks 1991, Pfeifer and Bongard 2007, Pfeifer et al. 2007). We are inclined to think of these problems in terms of equations and calculations (i.e., rule-like information representation and processing), but natural intelligence teaches how to use neural networks for efficient and competent behavior in real-world environments. This is a critical goal for future autonomous robots and indeed for AI embedded in other physical systems.

Challenges

We have argued that connectionist AI, based on neural network models and embodied cognition, provides a sounder, more effective basis for future AI and robotic systems than does rule-based knowledge representation and processing. Indeed there is widespread (though not universal) agreement on this, and many projects are pursuing these approaches. Therefore it is important to acknowledge that connectionism and embodiment present challenges for the test and evaluation of the systems in which they are used.

One problem is the *opacity* of neural networks. In a rule-based system the rules are expressed in an artificial language with some similarity to natural languages or to symbolic logic. The basic terms and predicates, in terms of which the rules are expressed, are generally those of the problem domain. Therefore the knowledge and rules of inference used by the system are *transparent*, that is, potentially intelligible to human beings. In a neural network, in contrast, the knowledge and inferential processes are implicit in real-valued connection weights among myriads of microfeatures. Further, representations are nonlocal and distributed. Therefore, individual microfeatures and connections do not usually have meanings that can be expressed in the terms of the problem domain.

Many people are troubled by the opacity of neural networks compared to the (potential) transparency of rule-based systems. With a rule-based system, they argue, you can look at the knowledge base and inferential rules, understand them, and see if they make sense. A human can, in effect, see if the system is making its decisions for the right reasons, or at least that it is not making them for the wrong reasons (e.g., on the basis of irrelevant factors). In contrast, a trained neural network might perform some task very well, but we will be unable to understand—in human-comprehensible terms—the bases on which it is doing its job. Perhaps it has found some totally irrelevant cues in the training and test data that enable it to perform well on them, but it will fail dismally when deployed.

These are legitimate concerns, but unavoidable. As we have seen, rule-following is characteristic of merely "competent" behavior, and therefore behavior that *can* be expressed in human-comprehensible rules will not surpass the merely competent level. Conversely, expert behavior—which is our goal for AI and autonomous robotics—will entail subtle discriminations, integrative perceptions, and context sensitivities that cannot be expressed in human-comprehensible terms. How then can we come to trust a connectionist AI system? In the same way we come to trust a human expert: by observing their reliably expert behavior in a variety of contexts and situations. The situation is similar to that with the use of unsupervised trained animals to perform some mission. We cannot look into *their* heads either, but we can test their behavior in a variety of mission-relevant situations until we have sufficient confidence.

Much of the inflexibility and brittleness of rule-based systems—and indeed of many digital computer programs—is a consequence of their behaving the same in all contexts, whereas natural intelligence is sensitive to context and can modulate its behavior appropriately. Because of the ability of artificial neural networks to integrate a very large number of microfeatures, which may be individually insignificant, they can exhibit valuable context sensitivity. However, this presents a test and evaluation challenge for connectionist systems, as we cannot test such a system in a single or simple context (e.g., in a laboratory) and assume that it will work in all contexts. Rather, it is important to identify the contexts in which the system may find itself and ensure that it operates acceptably in all of them.

Context sensitivity and embodied cognition both necessitate use of the *implemented* robotic or AI system in almost all phases of test and evaluation. As previously mentioned, one of the advantages of connectionist AI is that it can be sensitive to the context of its behavior, but this implies an early transition of the test and evaluation activity into realistic physical contexts (i.e., field testing). Because we want and expect the system to make subtle contextual discriminations, it cannot be adequately tested or evaluated in artificially constructed situations that do not demand this subtlety. The same applies to the system's (hopefully robust) response to novelty. Further, embodied intelligence depends crucially on the physical characteristics of the system in which it is embedded and on its physical relationships to its environment. Although preliminary testing and evaluation can make use of simulations of the physical system and its environment, such simulations are always incomplete and are more computationally expensive the more complete they are. Whereas to some extent conventional AI systems can be tested and evaluated offline, embodied AI systems cannot. Therefore physical prototypes must be integrated earlier into the development cycle.

In effect, test and evaluation of embodied connectionist AI and robotic systems is no different from that of vehicles, weapons systems, and other physical devices and equipment. The difference is in our expectations, for we are used to being able to test and evaluate software systems offline, except in the later stages in the case of embedded software.

Finally, as discussed above, embodied connectionist systems are to some degree opaque, that is, their *cognitive* processes are not fully transparent (intelligible) to humans. Of course, neural networks and their embodiments obey the laws of physics and are intelligible in *physical* terms, but that level of explanation is of limited use in understanding the intelligent behavior of a system. This seems like a distinct disadvantage compared to abstract, rule-based systems but, as we have argued, it is a necessary consequence of expert behavior. In this regard, the test and evaluation of embodied connectionist systems is not much different from that of other physical systems, for which abstract models and simulations are insufficient in the absence of field-testing.

Further, the deployment of embodied connectionist systems is not qualitatively different from the deployment of trained animals or humans. Being able to recite memorized rules of procedure or to perform well in laboratory environments does not substitute for performance testing and evaluation in real, or at least realistic, situations.

Conclusions

We have argued that embodied connectionist AI and robotics promises a new generation of intelligent agents able to behave with fluent expertise in natural operational environments. Such systems will be able to modulate their perception and behavior according to context and to respond flexibly and appropriately to novelty, unpredictability, and uncertainty in their environments. These capabilities will be achieved by understanding natural intelligence, its realization in neural networks, and its exploitation of embodiment, and by applying this understanding to the design of autonomous robots and other intelligent agents.

However, a more natural intelligence is also an intelligence that responds more subtly to its environment and guides its body in a fluent dance with its physical environment. As a consequence, such systems cannot be adequately tested or evaluated independently of their physical embodiment and the physical environment in which they act. Naturally intelligent systems typically lack both transparency of behavior and independence of information processing from physical realization, which we have come to expect in AI.

Nevertheless, such systems may be tested and evaluated by similar approaches to those applied to other inherently physical systems; it is really only a shift of emphasis from abstract rules and programs to concrete physical interaction with the operational environment. ❏

BRUCE MACLENNAN received a bachelor of science in mathematics from Florida State University in 1972, a master of science in computer science from Purdue University in 1974, and a doctor of philosophy in computer science from Purdue in 1975. From 1976 to 1979 he was a Senior Software Engineer at Intel and contributed to the architecture of the 8086 and iAPX-432 microprocessors. In 1979 he joined the computer science faculty of the Naval Postgraduate School in Monterey, California, where he was Assistant Professor, Associate Professor, and Acting Chairman. Since 1987 he has been a member of the computer science and electrical engineering faculty at the University of Tennessee, Knoxville, where he investigates novel computing technologies, nanotechnology, and biologically inspired computation. E-mail: maclennan@eecs.utk.edu

References

Anderson, J. A. 1995. *An Introduction to Neural Networks*. Cambridge: MIT Press.

Anderson, J. A., Pellionisz, A., and Rosenfeld, E. (eds.) 1990. *Neurocomputing 2: Directions for Research*. Cambridge: MIT Press.

Anderson, J. A., and Rosenfeld, E. (eds.) 1988. *Neurocomputing: Foundations of Research*. Cambridge: MIT Press.

Anderson, M. L. 2003. "Embodied Cognition: A Field Guide." *Artificial Intelligence*. 149, pp. 91–130.

Brooks, R. 1991. "Intelligence Without Representation." *Artificial Intelligence*. 47, pp. 139–159.

Changeux, J.-P. 1985. *Neuronal Man: The Biology of Mind*. translated by L. Garey. Oxford: Oxford University Press.

Clark, A. 1997. *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.

Dreyfus, H. L. 1979. *What Computer's Can't Do: The Limits of Artificial Intelligence.* rev. ed. New York: Harper & Row.

Dreyfus, H. L. and Dreyfus, S. E. 1986. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.

Feldman, J. A. and Ballard, D. H. 1982. "Connectionist Models and Their Properties." *Cognitive Science*. 6(3), pp. 205–254.

Gerstner, W. and Kistler, W. M. 2002. *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge: Cambridge University Press.

Haugeland, J. 1978. "The Nature and Plausibility of Cognitivism." *Behavioral and Brain Sciences*. 1, pp. 215–226.

Haugeland, J. (eds.) 1997. *Mind Design II: Philosophy, Psychology, Artificial Intelligence.* rev. & enlarged ed. Cambridge: MIT Press.

Haykin, S. 1999. *Neural Networks: A Comprehensive Foundation.* 2nd ed. Upper Saddle River: Prentice-Hall.

Hinton, G. E., and Anderson, J. A. (eds.) 1997. *Parallel Models of Associative Memory.* updated ed. Hillsdale: Lawrence Erlbaum.

Johnson, M. and Rohrer, T. 2007. "We are Live Creatures: Embodiment, American Pragmatism, and the Cognitive Organism." In *Body, Language, and Mind,* vol. 1. Edited by J. Zlatev, T. Ziemke, R. Frank, and R. Dirven, pp. 17–54, Berlin: Mouton de Gruyter.

MacLennan, B. J. 2001. "Connectionist Approaches." In *International Encyclopedia of the Social and Behavioral Sciences,* edited by N. J. Smelser and P. B. Baltes, 2568–2573. Oxford: Elsevier.

MacLennan, B. J. "Analog Computation." In *Encyclopedia of Complexity and System Science,* edited by R. A. Meyers. New York: Springer. In Press.

Mead, C. 1989. *Analog VLSI and Neural Systems*. Reading: Addison-Wesley.

Pfeifer, R. and Bongard, J. C. 2007. *How the Body Shapes the Way We Think—A New View of Intelligence*. Cambridge: MIT.

Pfeifer, R., Lungarella, M. and Iida, F. 2007. "Self-Organization, Embodiment, and Biologically Inspired Robotics." *Science*. 318, pp. 1088–1093.

Pribram, K. H., Nuwer, M. and Baron, R. J. 1974. "The Holographic Hypothesis of Memory Structure in Brain Function and Perception." In *Contemporary Developments in Mathematical Psychology*, Vol. 2. edited by D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes, 416–457. New York: Freeman.

Rumelhart, D. E., McClelland, J. L. and the PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol 1: Foundations*. Cambridge: MIT Press.

Sanger, T. D. 1996. "Probability density estimation for the interpretation of neural population codes." *Journal of Neurophysiology*. 76, pp. 2790–2793.

ITEA Journal 2008; 29: 367–374

Test and Evaluation of Fully Autonomous Micro Air Vehicles

Robert C. Michelson
Millennial Vision, LLC, Atlanta, Georgia

The test and evaluation of unmanned systems present special challenges, but these challenges are amplified when one moves into the realm of micro air vehicles. Further complications arise when these micro air vehicles are fully autonomous. The following discussion explores the difficulties in testing not only the physical flight characteristics of small flying things, but their behavioral characteristics as automatons. A fully autonomous flapping-wing micro air vehicle known as the Entomopter will be used as a robust example that embodies a greater range of operation than typical fixed wing micro air vehicles.

> *"Micro Air Vehicle is a most unfortunate name given to this class of air vehicles, because none are truly 'micro' and the original official Defense Advanced Research Projects Agency (DARPA) vehicle definition requiring a maximum 15 cm dimension confirmed the name to be a total misnomer."*
>
> (Michelson 2004)

Key words: Behavioral testing; cognitive systems; emergent behaviors; Entomopter; mobility; perception; physical flight testing; planetary flight testing; problem solving; reasoning; wind tunnels.

Originally touted by the military as an asset that could be carried by every warfighter (thus enabling the soldier to "see over the next hill" on a moment's notice and packaged like meals-ready-to-eat), the micro air vehicle (MAV) has proven to be a more elusive panacea primarily due to its small, albeit not "micro," size. Because MAVs are on the size scale of a small bird, they are not rugged. They must be light enough to fly while exhibiting a useful endurance. Being small and light, they are thus at the mercy of weather effects, most notably wind and, to a lesser extent, rain. Their small size also dictates antenna aperture size, which normally cannot greatly exceed the maximum dimension of the MAV. Therefore, communication frequencies must be higher, and hence, more directional to achieve reasonable link margins.

Another serious drawback is command and control of an air vehicle that cannot be seen by an operator at distances greater than about 50 meters. Command and control is also limited by the available onboard energy that can be devoted to in-flight video transmission for teleoperation.

A solution to these problems is autonomy. A fully autonomous MAV containing sufficient onboard intelligence to carry out useful missions has various advantages including:

- extended range because high frequency line-of-sight links are obviated,
- quicker reaction time to atmospheric perturbations and obstacle avoidance than can be afforded by a teleoperator,
- potentially greater stealth due to lower bandwidth emissions,
- the ability to operate indoors or in urban canyons where communication is not possible,
- jam resistance,
- the potential for beneficial emergent behaviors leading to higher probability of mission success.

Difficulties in testing these tiny MAVs fall into two categories when they suddenly are given the power of autonomy: physical flight testing and behavioral testing. To exemplify some of the issues involved in each, consider the flapping wing MAV known as the Entomopter.

The Entomopter was designed from inception to be a fully autonomous MAV for use in indoor reconnaissance. Initial development began at the Georgia Tech Research Institute by the author under an independent research and development program and was later funded by the Defense Advanced Research Projects

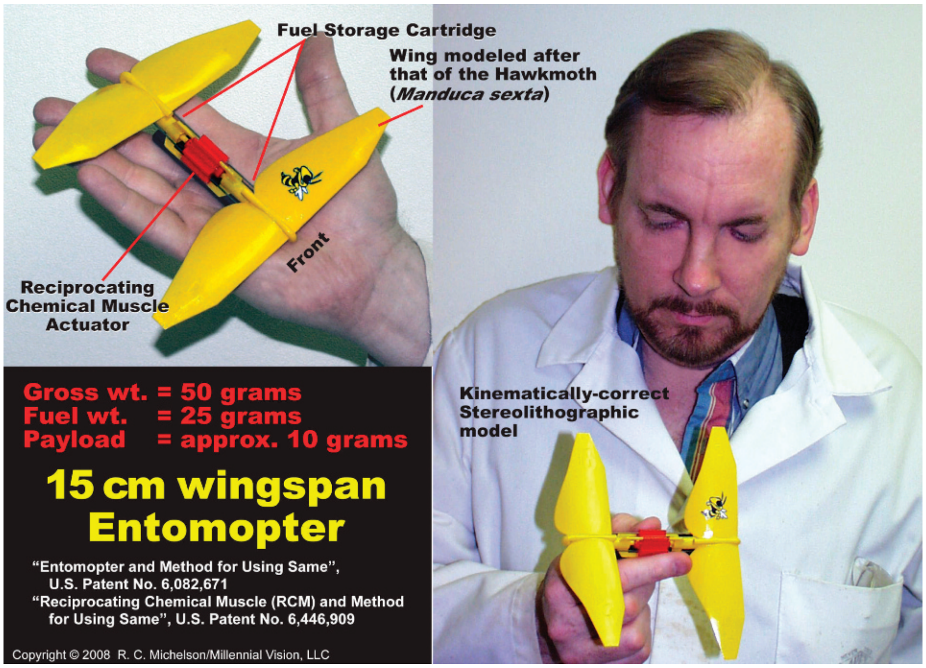


Figure 1. A 15-centimeter terrestrial Entomopter

Agency's Mesomachines program to demonstrate feasibility of such a device for indoor flight. The U.S. Air Force Research Laboratory then provided funding to extend the Entomopter's chemically-fueled propulsion system into its fourth generation. Subsequently, the National Aeronautics and Space Administration (NASA) Institute for Advanced Concepts became interested in the Entomopter's unique flight capabilities, which make slow flight in the lower Mars atmosphere possible. A feasibility study was then funded to show how an Entomopter-based Mars Surveyor could enhance the science missions envisioned for Mars.

All of these programs involved analytical substantiation of the Entomopter in various environments ranging from low Reynolds number flight in the lower atmosphere of Earth to low Reynolds number flight in the lower atmosphere of Mars. Also common to these programs was the fact that the Entomopter was to be fully autonomous and never teleoperated. Full autonomy was essential for indoor operation where communication and global positioning system signals were not available, and it was likewise essential for Mars operations where the latency of control (10–15 minutes depending upon the distance of Mars from Earth) necessitated a vehicle that could carry out missions unassisted.

The Entomopter

The terrestrial Entomopter *(Figure 1)* is a multimode autonomous robot capable of flight and limited ambulatory behaviors. Autonomous navigation is based on a combination of attraction and avoidance behaviors deriving input from both an integrated optic-olfactory sensor for detection of chemical species (or, alternatively, a sensor for a specific type of radiation), and an ultrasonic swept beam ranging device.

The terrestrial Entomopter eventually found potential applications on Mars by virtue of its unique "blown" flapping wing (U.S. Patent No. 6,082,671 and U.S. Patent No. 6,446,909). Present planetary surface rovers have shortcomings that NASA could address with a slow flying aerial platform, however flight on Mars is complicated by the fact that the atmosphere is rarefied, thereby making it difficult to generate lift with conventional wings. In fact, fixed wing vehicles must have enormous wings and travel at speeds in excess of 300 kilometers per hour to stay aloft in the Mars atmosphere (Colozza et al. 2000). Turn radii are on the order of kilometers, making it inefficient to return to points of interest, and high-speed traverse across the surface at lower altitudes causes smearing of sensor data, thereby negating any beneficial increase in resolution that may have otherwise been gained (Colozza et al. 2002).

NASA recognized that the ability of the Entomopter to fly in low Reynolds number conditions without the need for air-breathing propulsion made it a natural candidate for flight in the rarefied atmosphere of Mars, albeit in a larger incarnation. Unlike fixed wing flyers, an Entomopter-based Mars surveyor would be able to cover a wide area while still being able to fly slowly and return to a refueling rover.

The Entomopter began as a biologically inspired design, but rather than attempting to replicate biological kinematics and aerodynamics, improved systems have been devised to leverage what is observed in biological systems to produce a machine that is manufacturable, controllable, and able to generate the power necessary to fly from onboard energy sources (Michelson 2004).

The Hawk Moth (Manduca *sexta*) was chosen as a baseline model for the wing aerodynamics. The University of Cambridge in England was part of the initial Entomopter design team because it had studied Hawk Moth wing aerodynamics for more than a quarter of a century and had produced seminal works describing the leading edge vortex and its effects on the flapping wing (Ellington et al. 1996, Liu et al. 1998, van den Berg and Ellington 1997, Willmott and Ellington 1997, Wilmott et al. 1997). The flapping mechanism for the Entomopter has been extended beyond that of the Hawk Moth to provide a resonant single-piece construction that takes advantage of torsional resonance in the Entomopter fuselage to recover flapping energy common to flying insects that temporarily store potential energy in either muscles or exoskeletal parts (resilin).

In the terrestrial version, the same structure that provides wing flapping also scans a frequency modulated continuous wave ultrasonic beam to provide front, side, and down-looking range measurements for obstacle avoidance and altimetry. It also has the potential to track and follow free-moving agents in a fashion similar to that employed by bats.

Stability and control in flight as well as navigation are achieved by actively modifying the lift of each wing on a beat-to-beat basis using pneumatic control of the air circulating over the beating wing. Also, as demonstrated in the Georgia Tech Research Institute's wind tunnels where pneumatically controlled wings were shown to develop positive lift at negative angles of attack (α) as great as -70 degrees (Englar et al. 1994), Entomopter wings (unlike those of the Hawk Moth) should be able to generate positive lift not only on the downbeat but the upbeat as well. These wind tunnel tests have shown that coefficients of lift exceeding the theoretical maximum by 500 percent for the given wing shape can be achieved without the complexity of active angle-of-attack modulating mechanisms (Michelson and Naqvi 2003).

A chemically fueled reciprocating chemical muscle has been designed and is in its fourth generation of development. This actuator system has demonstrated 70-Hz reciprocation rates with throws and evolved power levels necessary to support flight of a fully autonomous Entomopter system (Michelson and Amarena 2001). The reciprocating chemical muscle uses the energy locked in various monopropellants to produce reciprocating motion for propulsion as well as waste gas products for the operation of gas bearings, an ultrasonic obstacle avoidance ranging system, and full flight control of the vehicle.

Testing MAVs (physical realm)

Rigorous testing of MAVs usually begins in a wind tunnel with airfoil sections and eventually with the entire air vehicle. A problem in wind tunnel testing is that many tunnels are not configured to handle such small test objects and the balances may not have the desired resolution to see what is happening under various test conditions. Further, the low Reynolds numbers at which MAVs operate pose scaling problems. Most MAVs in the Defense Advanced Research Projects Agency 15-centimeter size region are at the transition between classical aerodynamics and low Reynolds number aerodynamics. As has been observed in certain thin airfoil low Reynolds number experiments at the University of Bristol, external acoustic energy can actually affect the measurements (Grundy et al. 2001).

One advantage that MAVs offer during wind tunnel testing is the ability to easily power the model to get true readings concerning the forces involved and the behavior of vortices over the entire body of the craft. Powered wind tunnel testing is facilitated when considering fixed wing and rotary wing MAVs, however the same tests on a powered flapping wing MAV become problematic. Conventional flapping wing MAVs induce vertical loads into the wind tunnel balance that are larger than the aerodynamic wing forces being measured. This requires the development of a special balance that can withstand the inertia of the wing/body interaction while still being able to sense the minute aerodynamic forces created by different flow regimes over the wings and fuselage. The Entomopter has flapping wings, but there are two sets of wings flapping equally and oppositely across the fuselage as a fulcrum (like two seesaws 180 degrees out of phase). As a result, it is one of the few flapping wing designs that can be tested with a conventional wind tunnel balance designed for fixed wing MAVs. Control forces on the Entomopter wing are a result of differences in lift produced by circulation-controlled air foils, which employ the Coanda Effect to increase wing lift by keeping flow attached or decrease it by allowing the flow to detach into shed vortices. The effect as seen by the balance is similar to that of a conventional fixed wing with elevons.

Because of the low risk involved in MAV flight, air vehicle tests proceed from the wind tunnel to free flight outdoors much sooner than with larger aircraft that can ill afford to crash during development. Free flight testing of MAVs suffers from the inability to adequately instrument the vehicle because of its very limited payload capacity. Most MAVs operate with 50 percent of their gross takeoff weight being energy (battery, fuel) and propulsion; the remaining 50 percent comprises the airframe and the payload. A 50-gram vehicle such as the terrestrial Entomopter devotes 25 grams to fuel and 15 grams to the airframe/propulsion system (which are highly integrated). This leaves a mere 10 grams for payload. The addition of test instrumentation is at the expense of fuel and endurance. Unlike a fixed wing battery-operated MAV, the Entomopter has the luxury of consuming its fuel and getting ever lighter as the mission progresses, so a cutback in fuel to accommodate test instrumentation actually has less of an impact on endurance than its electrically-driven cousins.

Once in free flight, monitoring of the MAV is totally reliant on stored or telemetered data because the vehicle behavior to control inputs (whether teleoperated or autonomously generated) is largely unobservable from the ground. Because MAVs have short wing spans (or disk areas for rotary wing MAVs), they have low wing loading and are especially sensitive to roll induced by wind gusts. The mass moment of inertia scales as the fifth power of a given dimension, so the smaller the vehicle, the higher the bandwidth of the control system necessary to stabilize rotational instabilities. Analysis of in-flight oscillatory behaviors usually becomes a trial-and-error process where modifications to control surfaces and airfoils are made on the ground after analysis of the flight test data (as opposed to real-time adaptive control modifications). Again, this is a function of the inability to carry significant test payloads but is also encouraged by the relatively low cost and low risk of MAV free-flight testing.

The challenges increase by orders of magnitude when trying to test a planetary flight vehicle. In the case of the Mars Entomopter, Earth conditions actually preclude free-flight testing because the Mars Entomopter would be optimized for the reduced gravity on Mars, which enables a heavier vehicle to be used than that which would be possible on Earth.

The atmosphere on Mars has a mean surface level pressure of 600 Pascals (0.087 pounds per square inch) compared to Earth's 101,300 Pascals (14.7 pounds per square inch at sea level). This makes flight on Mars very difficult. A conventional fixed wing flight vehicle would have to travel excessively fast (perhaps greater than 300 kilometers per hour) simply to stay aloft without stalling and plummeting to the surface; however, at these speeds detailed survey of the planet's surface is impractical, as would be landing for refueling to extend the mission. The flapping Entomopter wings can in essence flap at more than 300 kilometers per hour, and coupled with its circulation-controlled airfoils that can generate up to seven times greater lift than theoretically possible for its wing shape, the Entomopter can fly slowly in the rarefied Mars atmosphere. A 15-centimeter terrestrial Entomopter flying in Earth's atmosphere is therefore equivalent to a 1-meter Entomopter flying in Mar's thin atmosphere. Since the gravity of Mars is about 37 percent that of Earth, the larger Mars Entomopter can weigh three times as much as the terrestrial Entomopter and still fly like its miniature earthbound cousin. An Entomopter with an approximate 1-meter wingspan flies in the same Reynolds number regime (and therefore generates lift in the same manner) as small insects do in the denser atmosphere of Earth.

So how does one conduct flight tests of a planetary flight vehicle here on Earth? In the wind tunnel, the air speed can be scaled to correspond to low Reynolds number flight for a given sized vehicle. Alternately, there are a few wind tunnels in the world that can be pumped down to a low atmospheric pressure to make unscaled measurements. Free flight is even possible by taking air vehicles to high altitudes and releasing them in the thin atmosphere at altitude.

Mars flight testing is still problematic because of the reduced gravity. A larger, heavier vehicle that might be used in the lower atmosphere of Mars, even if dropped from altitude, will still be under the influence of Earth's higher gravitational field. One suggestion that has been considered is to conduct wind tunnel tests in a pumped-down tunnel while using magnetic levitation to off-load the appropriate amount of vehicle weight. Still, this technique would suffer from inertial effects as some of the weight to be off-loaded might be in the moving wing structures, the weight of which would be hard to control while in motion. The best solution to date is to design analytical models for testing, then build surrogate testbeds that approximate the actual vehicle with full knowledge of the underlying assumptions.

Testing autonomous MAVs (behavioral realm)

In addition to the testing of the physical performance of MAVs under controlled (wind tunnel) or real-world (free-flight) conditions, autonomous MAVs require another level of testing to assess intelligent behavior. Autonomous MAVs must ultimately perform a mission function effectively. MAV autonomy will be based on intelligent cognitive behaviors that direct logically expected maneuvers, leading to mission success. Cognitive systems may evolve "emergent behaviors" that may still lead to mission success, but in ways that are unexpected by the designer. Emergent (unpredictable) behavior is a consequence of a decentralized system with reactive properties, indicative of swarms of independent MAVs. The individual MAV's own high level cognitive system will gravitate toward centralized forms of behavior, and therefore emergent behavior is less likely but still possible. This makes testing of autonomous MAVs an even greater challenge.

Autonomous MAVs will possess various behavioral traits that, when combined, will elicit useful high level behavior leading to mission success. These traits are discussed below:

Perception and mobility

Autonomous MAVs must be able to perceive their environment and control their motion through it. To find their targets, for example, MAVs must be able to identify their goal through some means of object recognition and move toward their goal based on a path planning algorithm that avoids threats (which may be physical obstacles or ephemeral things such as adverse weather or hostile agents). The more information that MAVs can perceive to assist them in navigating to their goal, the greater the odds of achieving mission success.

Temporal reasoning

In order to succeed at their assigned mission, autonomous MAVs must perceive and reason about events that occur during various temporal intervals. The relations of these intervals to one another is important to the outcome of future events. For example, were a MAV to be tracking a moving target and the target moves behind an obstacle (becoming occluded) only to emerge from the other side of the obstacle moments later, it would be reasonable to assume that the newly acquired target is in fact the same target that was being tracked. The correct autonomous MAV behavioral response might therefore be to continue tracking the reacquired target rather than beginning a search pattern on the far side of the obstacle.

Logical deduction, falsification, default reasoning, and explanation

MAV intelligence presupposes the fact that the MAV can draw conclusions from its perception. These conclusions may require deductions to be made based solely on inferences drawn from its world model and its finite body of stored time-weighted (most recent = most accurate) perceived data. Consider the terrestrial Entomopter tracking a moving object in a building. The object turns out to be a rat running across the floor. Suddenly the Entomopter loses track because the target disappears—the rat ran into a hole in the wall. The Entomopter must deduce from its observations that the target was lost at the point it intersected the wall. It must conclude as false the assumption that the wall is a solid object. The Entomopter must then reason that the wall can be penetrated, so its explanation is that the target escaped its view by exiting through a hole in the wall. The behavioral response of the Entomopter might then be to give up the track or instead it might begin a search on the opposite side of the wall if it has a path to the other side stored in its track history or perhaps it may even begin a search for a path to the other side in order to reacquire the target.

Testing for these kinds of behaviors is very difficult in an autonomous system because there is no deterministic solution unless the observer has the same database as the MAV, and even then there may be several correct responses that might appear as emergent behavior on the part of the MAV. Questions may include "what can you assume and why?"; "what does it take to falsify an assumption?"; "when is there more than one explanation for an event?"; and "which explanation will the MAV choose?"

Belief revision and reason maintenance

An autonomous MAV may find that it has to revise its beliefs based on contradictory facts. Because the MAV could have inferred more facts based on the originally assumed fact, revising its belief about the original fact is far more complicated than simply retracting it. The MAV must retract *all* beliefs it inferred using the original fact if they are not substantiated by independent facts that are still believed to be true.

Again consider the Entomopter tracking the rat that escaped through a hole in the wall. The explanation arrived at by the Entomopter based on its world model is that the rat has escaped from the current room into an adjoining room. In the process of searching for a path to the adjacent room, were the Entomopter to discover that the wall is not a single plane but a hollow wall with two faces, it can no longer assume that the rat must be in the adjacent room. The onboard cognitive system of the Entomopter must now revise its beliefs about the nature of walls and, in the process, discard assumptions that rooms are separated by a single solid plane with holes. In fact, the rat could be hiding

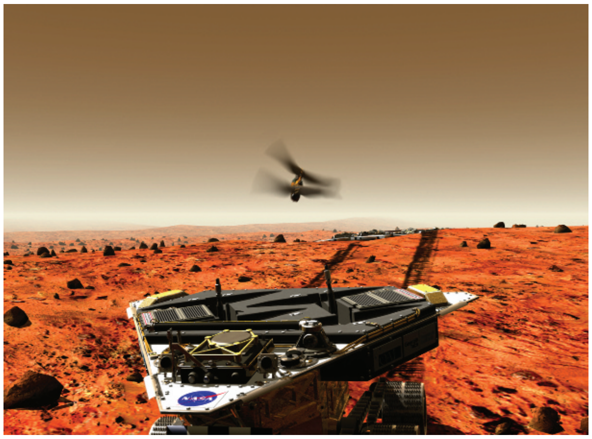


Figure 2. Mars Entomopter operating in the vicinity of its refueling rover

"inside" the wall. Based on this new conclusion, the Entomopter may now decide that the cost in energy to continue the search for a path to the adjacent room may not be worth the expense because the likelihood of finding the rat in the adjacent room has been significantly diminished.

Planning, searching, problem solving

Events often have more than one possible outcome, and MAVs can execute more than one action at any given time. The sequence of possible inferences and actions about event outcomes creates an enormous space of possible world states. The MAV's onboard cognitive system must choose a sequence of inferences and actions to reach a desired state. In addition, because the MAV is a flying automaton, it cannot hesitate in the making of these decisions. It does not have the luxury of stopping to think. Therefore, the cognitive systems employed by MAVs must be highly efficient in their ability to process data and draw conclusions.

Consider the Mars Entomopter returning to its refueling rover when it detects something moving on or near the surface of the planet *(Figure 2)*. Moving objects other than the refueling rover are not expected. Is the object in fact the refueling rover? Is it a dust devil? Is it alive? The Entomopter must solve this problem quickly because the object may be of paramount scientific importance, but on the other hand, the Entomopter is low on fuel and must ensure that whatever action it takes will not jeopardize its ability to return to the refueling rover. The solution to the problem may lie in the relative location of the object, the appearance of the object, the speed of the object, and the apparent direction of its travel. This must be coupled with the amount of fuel remaining and the distance of the refueling rover from the Entomopter and from the detected moving object. Any information about headwinds will play a factor.

As the problem becomes more complex, the ability to test and evaluate the response correctness becomes very difficult. The problem presents various optional responses, some of which would be better than others, whereas some would lead to disaster. Due to the immense size of the problem space, such situations are ripe for the occurrence of unforeseen emergent behaviors and are practically impossible to test. Even in simulation, the selection of the given set of parameters defining the problem space will likely not match what is actually encountered during a mission. The best that can be hoped for is to correctly evaluate the behavior after the fact based on stored/recovered data in the hope that unforeseen bizarre behaviors can be eliminated during future missions.

Probabilistic inference

In the real world where mission success or failure may result from more than one possible outcome, some events are more likely than others. "Uncertain reasoning" or "probabilistic inference" must be used by the MAV to decide which of the many outcomes is most likely when presented with scenarios involving several possible outcomes. Considering the previous example, were the Entomopter to have a priori knowledge that a storm is approaching and that its refueling rover is definitely located in the opposite direction from the presently unidentified object moving across the Mars landscape, the probability of that object being a weather-related phenomenon (and therefore of lower interest) would bias the decision to continue toward the refueling rover.

Social reasoning, communication, and human-machine interaction

Under some circumstances, fully autonomous MAVs might be members of a cooperative "swarm" of other MAVs or might be a member of a network involving other autonomous systems (unmanned ground vehicles, ground stations, or even humans). When interacting with these other intelligent or even autonomous entities, the MAV must be able to reason about that entity's mental state. Depending on the entity type, this implies that the behavior of the MAV will require it to consider the emotions, beliefs, desires, personality traits, etc. of the other entities with which it is interacting. The MAV's actions will depend to a degree on the perceived responses of its counterparts. If the Mars Entomopter knows that its refueling rover is apt to stop upon sensing that the Entomopter is returning to refuel, then the Entomopter will adjust its flight approach accordingly. Alternately, were the

Entomopter to attempt to land on unknown terrain, it would assume a degree of unpredictability about its touchdown location.

Conclusions

MAVs will incorporate greater degrees of autonomy as they increase in capability and popularity among the mainly military user base. All of the challenges associated with aircraft testing and evaluation are present with MAVs, but are compounded by the different physical environment in which they operate. The environment differs not only in terms of the mission space which, unlike manned aircraft and larger unmanned aerial vehicles, includes confined spaces such as urban canyons, building interiors, and natural formations such as caves, but there is also a difference in the very medium through which the MAVs operate: a more viscous, low Reynolds number environment.

Flapping wing implementations are more likely MAV implementations than for larger aircraft. The testing of flapping wing MAVs in the wind tunnel and in free flight presents special challenges.

The physical realm in which MAVs will be tested is therefore more challenging in many respects, but the ability to test the entire vehicle with its propulsion system intact and operating in the wind tunnel enables a level of convenience and accuracy that may not be present in larger test objects. However, these advantages evaporate when extraplanetary MAVs are considered. Simulation of different atmospheres and particularly different gravitational fields is problematic.

When full autonomy is applied to MAV designs, the ability to test them in the behavioral realm becomes an even greater challenge because of the size of the MAV and the difficulty in monitoring and communicating with them in free flight. The payload capacity of MAVs is necessarily small, but the non-scaling items such as antenna apertures present the evaluator with serious real-time monitoring issues. As the MAV's level of cognition increases, predicting its behavioral response in a free-flight real world environment becomes statistically impossible in real time as the size of the problem space expands.

A recommended testing regime would decouple the physical realm testing from that of the behavioral realm. Controlled wind tunnel testing to corroborate simulations such as those derived from computational fluid dynamics analyses will yield baseline performance results that can predict free-flight behaviors given that actual conditions can be recreated in the wind tunnel environment. Free-flight testing will then validate the controlled wind tunnel measurements.

Once the MAV is endowed with a sentient nature and cognition to interpret its perception of the environment with its assumed physical laws, our ability to "test" MAV behavior may be limited and instead will devolve into more of an "evaluation" of observed behavior. ❑

Robert C. Michelson is a graduate of Virginia Polytechnic Institute & State University and the Georgia Institute of Technology with degrees in electrical engineering. He worked on aerospace radar systems at the U.S. Naval Research Laboratory in Washington D.C. prior to his 30-year career at the Georgia Tech Research Institute, where he taught unmanned aerial vehicle/micro air vehicle avionics design as adjunct associate professor to the School of Aerospace Engineering. He is currently principal research engineer emeritus with the Institute. In 2004 he created Millennial Vision, LLC to continue research into biologically inspired aerial robots and remote sensing (see http://angel-strike.com/rcmbio/RCM/MICHELSON.bio.html). E-mail: millennialvision.llc@gmail.com

References

Colozza, A., et al. (2002). "Planetary Exploration Using Biomimetics." NASA Institute for Advanced Concepts Phase I Final Report, November 30, 2000. Atlanta, Georgia: NASA Institute for Advanced Concepts.

Colozza, A., et al. (2002). "Planetary Exploration Using Biomimetics—An Entomopter for Flight on Mars." Phase II Final Report, NASA Institute for Advanced Concepts Project NAS5-98051, October 2002. Atlanta, Georgia: NASA Institute for Advanced Concepts.

Ellington, C. P., van den Berg, C., Willmott, A. P. and Thomas, A. L. R. 1996. "Leading-edge Vortices in Insect Flight." *Nature*. 38(D), pp. 626–630.

Englar, R. J., Smith, M. J., Kelley, S. M. and Rover III, R. C. 1994. "Development of Circulation Control Technology for Application to Advanced Subsonic Transport Aircraft, Part I: Airfoil Development." AIAA Paper No. 93-0644, Log No. C-8057. *AIAA Journal of Aircraft*. 31(5), pp. 1160–1168.

Englar, R. J., Smith, M. J., Kelley, S. M. and Rover III, R. C. 1994. "Development of Circulation Control Technology for Application to Advanced Subsonic Transport Aircraft, Part II: Transport Application." AIAA Paper No. 93-0644, Log No. C-8058. *AIAA Journal of Aircraft*. 31(5), pp. 1169–1177.

Grundy, T. M., Keefe, G. P. and Lowson, M. V. 2001. "Effects of Acoustic Disturbances on Low Re Aerofoil Flows." In *Fixed and Flapping Wing Aerodynamics for Micro Air Vehicle Applications*. Vol. 195, pp. 91–112. Reston, Virginia: American Institute of Aeronautics & Astronautics.

Liu, H., Ellington, C. P., Kawachi, K., van den Berg, C. and Willmott, A. P. 1998. "A Computational Fluid Dynamic Study of Hawk Moth Hovering." *Journal of Experimental Biology*. 201, pp. 461–477.

Michelson, R. C. 2004. "Novel Approaches to Miniature Flight Platforms." Proceedings of the Institute of Mechanical Engineers. *Journal of Aerospace Engineering*. 218(G) Special Issue Paper, pp. 363–373.

Michelson, R. C. and Amarena, C. S. 2001. "4th Generation Reciprocating Chemical Muscle: Reciprocating Chemical Muscle (RCM) for Specialized Micro UAVs and Other Nonelectric Anaerobic Aerospace Actuation Applications." Prepared under Grant No. F086300010007 to the U.S. Air Force Research Laboratories (AFRL/MNGN). Eglin Air Force Base, Florida: U.S. Air Force Research Laboratory.

Michelson, R. C. and Naqvi, M. "Extraterrestrial Flight (Entomopter-based Mars Surveyor)." von Karman Institute for Fluid Dynamics RTO/AVT Lecture Series on Low Reynolds Number Aerodynamics on Aircraft Including Applications in Emerging UAV Technology, Brussels Belgium. November 24–28, 2003.

van den Berg, C. and Ellington, C. P. 1997. "The Vortex Wake of a 'Hovering' Model Hawk Moth." *Philosophical Transactions of the Royal Society B*. 352, pp. 317–328.

Wilmott, A. P., Ellington, C. P. and Thomas, A. L. R. 1997. "Flow Visualization and Unsteady Aerodynamics in the Flight of the Hawk Moth Manduca *Sexta*." *Philosophical Transactions of the Royal Society B*. 352, pp. 303–316.

Willmott, A. P. and Ellington, C. P. 1997. "The Mechanics of Flight in the Hawk Moth Manduca *Sexta*. I. Kinematics of Hovering and Forward Flight." *Journal of Experimental Biology*. 200, pp. 2705–2722.

ITEA Journal 2008; 29: 375–379

Biomimetic Applications of Arthropod Visual Systems

Jeffrey S. Sanders, Ph.D.
Trideum Corporation, Huntsville, Alabama

Phylum Arthropoda, which includes the classes of insects and crustaceans, contains more than a million representatives that comprise the majority of all known living species. Across this phylum there is a huge diversity in morphology, including the visual systems of these animals. The primary external component of the insect and crustacean visual system is the compound eye, which has evolved with time into a range of forms that optically sample the space around an animal. This optical sensing system enables these animals to hunt for food, find mates, navigate in complex environments, and perform whatever other tasks that are required for the species' survival. Researchers around the world are studying the visual systems of arthropods to find solutions to modern engineering problems. With so many evolutionarily successful, natural examples it is clear that valuable lessons can be learned from these highly specialized designs. This article provides an overview of natural compound eyes and covers examples of their applicability to modern problems in optical sensing and guidance and navigation.

Key words: Apposition compound eyes; Arthropoda; autonomous agents; biomimetic systems; compound optical systems; insect; superposition optics; unmanned aerial vehicles; visual systems.

Throughout the history of engineering, scientists have often looked at the single aperture human eye to determine the basic design of optical systems. This was a fundamentally sound decision because most optical systems needed the final image to be presented in a format that was compatible with the human visual system. This concept was applied from the ancient "burning glass" of the Greeks to Galileo's telescope and to the advanced cameras and telescopes of today (Hecht and Zajac 1979). Because of the long history and success of single aperture optical systems it is quite easy to see how other types of biological optical systems have been ignored as paradigms for human-made sensors.

With the great advances in solid-state detectors and digital computing, single aperture optical sensors have become even more entrenched in the fields of robotics, image recognition, and military target detection. These new technologies enabled the transfer of an image created by an optical system directly to a massive data processing system, again mimicking the human visual system. However, the computing power and algorithms necessary to recreate the human visual function do not yet exist and may never be fully implemented. These single aperture paradigms have led to vision systems that emphasize single or few fields of view, fine segmentation of the image, and massive digital processing for feature extraction. Alternative paradigms modeled after the multi-aperture optical systems of arthropods offer new ways to segment the object space of a sensor, increase the field of view (FOV), and perform low-level visual functions relatively easily, inexpensively, and quickly.

The multi-aperture optical systems of the arthropods are more commonly known as compound eyes. The compound eye has evolved into many forms, each of which is specifically adapted to the optical environment of the insect or crustacean upon which it resides. The compound eye performs visual functions from the darkness of the ocean depths (Land and Nilsson 1990) to the brilliance of the sunlit desert (Horridge 1977). It provides the host creature with a wide FOV that is extremely beneficial for survival, as well as sufficient sensitivity to light as nature dictates (Kirschfeld 1974). The compound eye also offers a way for an animal to evolve an optical system that is highly customized (and optimized) for its environment and the visual tasks it needs to accomplish (Schwind 1980).

Biomimetic is a term that refers to human-made processes, substances, devices, or systems that imitate nature. The purpose of arthropod visual system biomimetic research is not to develop sensors and

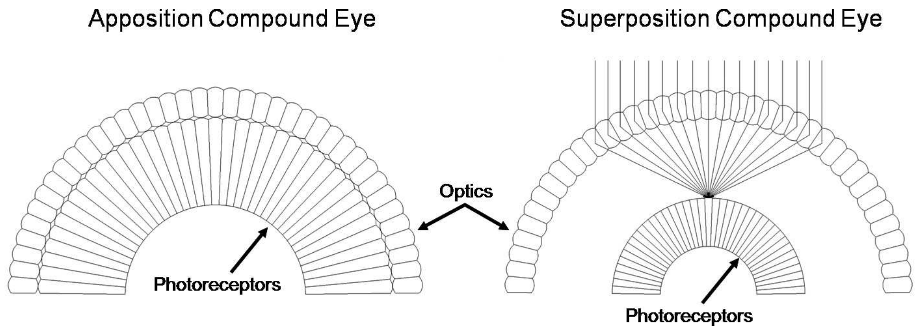


Figure 1. Apposition and superposition compound eyes

processing systems that outperform traditional systems for *all* applications. However, by investigating arthropod visual systems it can be determined if they can be mimicked in a human-made sensor system to provide advantages over traditional sensors for certain tasks. The justification for research in this area is best provided by a quote by Michael F. Land, who states that:

> *"Compared with the relatively standard pattern of vertebrate eyes, those of invertebrates present an enormous diversity. They vary greatly in size, the method of formation of the image, resolving power and sensitivity, and this variety is a reflection partly on the different evolutionary histories of the different kinds of eye and partly on the different visual needs of the animals that bear them. It would be wrong, however, to regard even the simple eyecups of flatworms as less successful, evolutionarily, than the large and sophisticated compound eye of a dragonfly or the fish-like eyes of an octopus. It is more profitable to regard each as being a perfected compromise suited to the lifestyle of the animal in question than it is to classify some as 'primitive' and others 'advanced'. It is true that some eyes are capable of supplying more information than others, but at a higher cost in terms of the space they take up and the metabolic energy they require."* (Land, 1981)

Interest within the U.S. Department of Defense in small autonomous agents is increasing with each passing year. Applications include autonomous weapons seekers, micro air vehicles (MAVs), and small terrestrial robotic applications. In addition to the compound eye, arthropods have evolved many other types of useful sensors that could be considered innovative and even revolutionary if they were engineered for use in autonomous robots, but these will not be covered here. Advances in arthropod biomimetics could lead to new technologies in microsensors, wide FOV situational awareness, and large numbers of mobile sensors populating the battlefield. These technologies will provide for huge advances in battle space surveillance but with an associated increase in bandwidth and connectivity requirements. Battlefield sensor networks of hundreds of sensors today may be tens of thousands of sensors at some point in the future. The test and evaluation (T&E) challenges associated with effectively evaluating systems of this complexity could be potentially enormous and require new methodologies and mindsets.

The two main types of compound eyes

The vast diversity of phylum Arthropoda makes it difficult to infer morphological generalizations over its different classes and orders. There are at least seven different arthropod compound eye types with distinctive optical mechanisms (Warrant and McIntyre 1990). However, the optical characteristics of arthropod eyes clearly lend themselves to division into two main classes: apposition compound eyes and superposition compound eyes. *Figure 1* shows the imaging characteristics of these two main types.

The apposition eye consists of an array of lenses and photoreceptors with each of the lenses focusing light from a small solid angle of object space onto a single set of photoreceptors. Each lens-photoreceptor system is referred to as an ommatidium, or "little eye" (Horridge 1977). A typical apposition eye has thousands of these ommatidia packed in non-uniform hexagonal arrays. The number of ommatidia can range from as little as 130 in the single eye of the water flea to more than 25,000 in the compound eyes of dragonflies.

The optical superposition compound eye (hereafter referred to simply as a superposition eye) has primarily evolved on nocturnal insects and deep water crustaceans, and it combines light from many ommatidia to form a single image on a photodetector layer within the eye. The superposition eye optical system is, in essence, an omnidirectional camera system imaging on a spherical image surface. One potentially huge advantage of this type of imaging system is the lack of off-axis optical aberrations that plague traditional optical systems. These

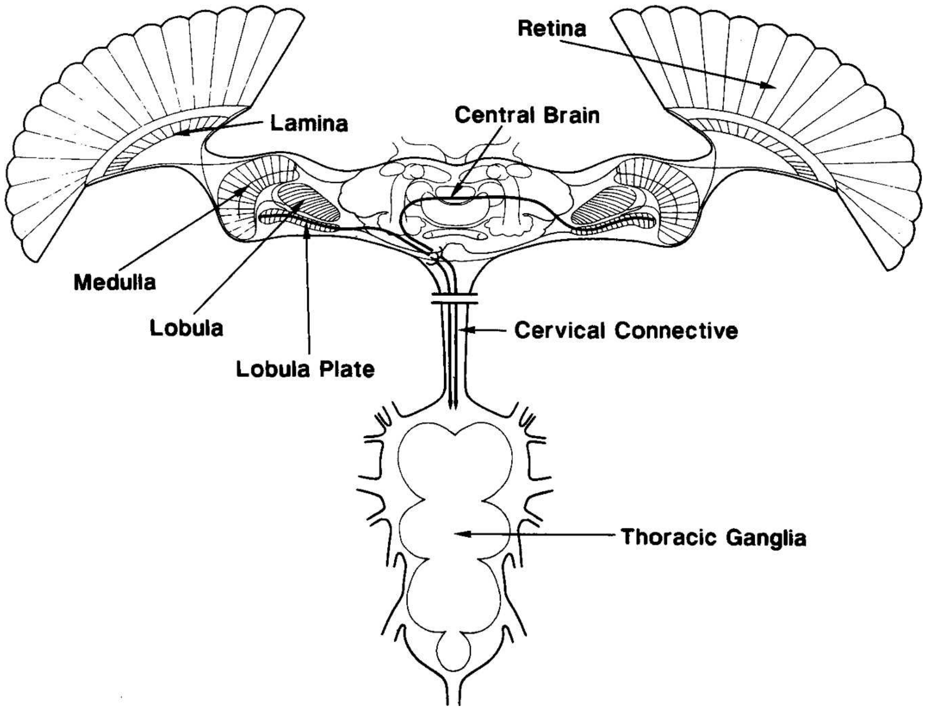


Figure 2. Visual pathway of a fly

aberrations are missing because this eye geometry has no optical axis, it is simply imaging in all directions at once.

Apposition compound eyes are highly adaptive optical systems as they can conform to surfaces, have variable sampling rates over their FOV (even more than one fovea), and can vary spectral and polarization sensitivity across the eye surface. In an environment with bright ambient light sources this type of eye excels at providing a highly complex conformal optical sensor for a relatively small mass. The superposition compound eye, on the other hand, is more like a traditional single aperture optical sensor in that it images on a single image surface, albeit in a very complex manner and over a very wide FOV.

When studying arthropod visual systems to learn ways to enhance human-made sensors it is best to focus on the behavior and the environment of an animal that matches the environment of the human-made autonomous system of interest. For autonomous robots that will operate in signal rich environments (daytime on the surface), apposition eyes are the target compound eye of choice for study. However, if the need is for wide FOV situational awareness in a signal poor environment (nighttime or underground), superposition compound eyes provide potential insights to advantageous sensor designs. The most important thing to note about compound eye design is that the design and layout of the eye is intimately tied to the neural processing layers beneath the eye. Apposition eyes cannot be studied solely as optical systems without including the underlying neural pathways in the analysis.

Arthropod neural processing and the visual pathway

Arthropod brains (and the term is used loosely) are hard-wired for a variety of functions, and it is possible to investigate these relatively simple functions using visual stimuli and electrophysiological techniques. There are multiple advantages of arthropods over higher life forms when studying basic neural function. One is that the subjects themselves are very inexpensive and relatively simple systems when compared to animals such as primates. Specific neural pathways can be isolated and their functions studied without the need for keeping the subjects alive over extended periods, in essence you have "disposable subjects." Numerous ethical issues are avoided because the odds are low that anyone would ever care if a pest insect such as a blowfly or fruit fly loses its life in the pursuit of science.

A typical insect visual pathway (a fly) is shown in *Figure 2*. Visual processing of retinal input occurs in several stages starting with motion detection in the lamina and a variety of higher functions in the medulla, lobula, and lobula plate. Covering the functions of the arthropod visual pathway is a subject far beyond the scope of this article, but the important data for the reader to absorb is that the visual processing in a fly brain is done with highly specialized neurons that represent a relatively small mass when compared to the optical system that feeds it information from the outside world. Higher life forms have eyes that are much smaller than their brains. Arthropods with apposition eyes already have a conformal optical system that minimizes its mass, yet it still represents a relatively large mass when compared to the neurons that it feeds. Even with these tiny processing systems, insects such as the fly can perform amazing feats of target detection, visually-based guidance and control, and collision avoidance that are difficult to achieve in human-made systems with modern technology.

All of this highly specialized optical sampling and neural connectivity in an apposition compound eye/brain combination means that insects such as the fly do not see the world as we see it. In fact, the majority of the eye/brain design is geared toward interpretation of optic flow as the fly traverses the environments in which it lives. Optic flow can be simply described as the motion of contrasts across the eye's visual field, and insects use this for flight and gaze stabilization, navigation (including distance measurement), speed control, landing, passive ranging, and collision avoidance (Horridge 2005). All of these functions happen in real time using a biological processor with a volume measured in cubic millimeters and weighing tens of milligrams or less.

An example of optic flow that the reader can relate to is how a human perceives the world going by while riding in a car. Images of closer objects flow across our visual fields faster, and we intuitively know that the faster moving objects are closer that the slower moving ones. This is a classic example of passive ranging with optic flow, and insects are hard-wired to take advantage of this rudimentary information about the world around them to do what it takes to survive. Although most of the energy and space budget in apposition eye/brain systems is expended on motion detection, insects are also capable of pattern recognition, and use both color and polarization information to navigate, find mates, and search for food.

Applications of compound eye optical designs

Descriptions of arthropod sensing capabilities described so far may inspire the reader to abandon traditional optical designs for many applications and exploit natural compound eye designs. However, compound eye–based optical designs are only truly applicable to meet the following optical sensing requirements:

- fields of view exceeding 2π steradians (greater than 180 degrees),
- non-traditional shapes (conformal optics),
- extremely small size/weight budgets (micro-optics),
- non-foveated wide FOV optical systems (greater than 90 degrees).

Unless you have one of these very challenging requirements, single aperture sensors will almost always outperform a compound eye design. Some robotic sensing applications have exploited circular arrays of simple sensors for robotic navigation algorithm development and testing (Franceschini et al 1991), and this is a rudimentary form of an apposition eye. There have been numerous examples of two-dimensional compound eye optical systems developed but almost all have been for specific imaging applications and only extend across a relatively small range of angles (Duparre 2004). The most complex and realistic artificial apposition compound eye to date was fabricated using self-forming polymer waveguides (Jeong et al 2004). This eye has the classic hexagonal packing of long tapering optical systems with a spherical outer surface and is on a scale with natural apposition optical systems (sub-millimeter diameter). Although apposition compound eyes present many fabrication challenges due to their inherently large parts count, advances are being made in the technologies and techniques required for their fabrication.

Superposition compound eye optical systems have many applications, ranging from grazing incidence x-ray optics to missile approach warning sensors and missile seeker design. Human-made superposition optical system designs have been ray-traced and shown to meet or exceed the resolution of natural superposition optics (Sanders 1994), but practical optical superposition hardware has yet to be described in the open literature. In addition to the complexity of the superposition optics, this type of imaging system requires detectors on a spherical surface, and this technology does not yet exist. However, recent advances in flexible polymer electronics offer hope that this problem may be solvable in the near future and make superposition optical systems a practical alternative to challenging sensing problems.

Applications of arthropod visual processing

One of the most stunning real-life examples of arthropod visual processing in action ever witnessed by this author was a high-speed target intercept video. This particular interceptor system had a gimbaled, multi-foveated, multi-spectral optical sensor that could acquire targets in an FOV exceeding 180 degrees in all directions. The system could lock onto a moving target, provide launch timing and initial thrust vector data, and then provide proportional navigation inputs until target intercept occurred within a few hundred milliseconds after initial detection of the target. The truly amazing thing is that this was not a military system at all, but a biological interceptor known as a dragonfly.

Examples like this abound in nature, and in the last decade biologists have made significant advances in understanding how insects perceive the world around them and process visual information to perform basic flight control tasks. Engineers and scientists have taken this biological data and successfully applied it to problems ranging from autonomous platform collision avoidance and terrain following to automated landing of helicopters and missile guidance. Of particular interest is the applicability of this processing paradigm to the control of unmanned aerial vehicles (UAVs) and the emerging generation of micro-UAVs (Massie et al 2003). Researchers have demonstrated micro-helicopter flight stabilization using an optic flow processor weighing only 200 milligrams (Franceschini et al 2007). Accomplishments such as this, along with advances in micro-UAV platform development, represent enabling technologies for the proliferation of autonomous flying agents on the battlefield of the future.

Conclusions

There are innumerable lessons that can be learned from nature and applied to the tough problems faced by today's engineers. Seemingly simple yet quite complex creatures such as insects and crustaceans are very evolutionarily successful and provide a huge resource for ideas and techniques for enhancing the capabilities of autonomous agents. The concepts presented in this article focus only on the visual capabilities of these creatures but there are numerous other lessons to be learned ranging from mechanical flight control to chemical sensing. Autonomous agents are getting smarter and more capable thanks to advances in computer processing, but as traditional guidance and control algorithms have become more complex there have been corresponding increases in the mass and energy required to execute these algorithms. Thanks to lessons learned from nature it is becoming possible to perform challenging flight control tasks with minimal processing power, energy, and mass. This new paradigm for vision-based self-control could enable a proliferation of autonomous flying agents on the battlefield of the future with the associated complexities of integrating them into legacy systems and testing and evaluating their capabilities. ❑

JEFFREY S. SANDERS is a principal engineer at Trideum Corporation in Huntsville, Alabama. After receiving a Ph.D. in engineering with a dissertation involving arthropod biomimetics, he has spent the last 15 years supporting various Army agencies in the area of target signatures and ground target surrogate design and development. He is currently a principal investigator on an Air Force program that is developing open-source arthropod vision modeling software. Dr. Sanders received a bachelor of science degree in electrical engineering, a masters of science degree in electrical engineering, and a Ph.D. in engineering from Memphis State University in 1987, 1990, and 1994, respectively.

References

Duparre, J. 2004. "Artificial Apposition Compound Eye Fabricated by Micro-Optics Technology." *Applied Optics*. 43(2), pp. 4303–4310.

Franceschini, N., Pinchon, J. M. and Blanes, C. 1991. "Real-Time Visuomotor Control: From Flies To Robots." Fifth International Conference on Advanced Robotics. *Robots in Unstructured Environments*. 2, pp. 931–935.

Franceschini, N., Ruffier, F. and Serres, J. 2007. "A Bio-Inspired Flying Robot Sheds Light on Insect Piloting Abilities." *Current Biology*. 17, pp. 329–335.

Hecht, E. and Zajac, A. 1979. *Optics*. Reading, Massachusetts: Addison-Wesley.

Horridge, G. A. 1977. "The Compound Eye of Insects." *Scientific American*. 237, pp. 108–122.

Horridge, A. 2005. "The Spatial Resolutions of the Compound Eye and Its Neuro-Sensory Feature Detectors: Observation vs. Theory." *Journal of Insect Physiology*. 51, pp. 243–266.

Jeong, K., Kim, J. and Lee, L. 2004. "Biologically Inspired Artificial Compound Eyes." *Science*. 312(773), pp. 557–561.

Kirschfeld, K. 1974. "The Absolute Sensitivity of Lens and Compound Eyes." *Zeitschrift für Naturforschung*. 29, pp. 592–596.

Land, M. F. 1981. *Comparative Physiology and Evolution of Vision in Invertebrates Vol. VI/B, Invertebrate Visual Centers and Behavior I, Optics and Vision in Invertebrates*. New York: Springer-Verlag.

Land, M. F. and Nilsson, D. E. 1990. "Observations on the Compound Eyes of the Deep-Sea Ostracod Macrocypridina Castanea." *Journal of Experimental Biology*. 148, pp. 221–233.

Massie, M., Baxter, C., Curzan, J. P., McCarley, P. and Etienne-Cummings, R. 2003. "Vision Chip for Navigating and Controlling Micro Unmanned Aerial Vehicles." *Proceedings of the 2003 International Symposium on Circuits and Systems, Vol. 3*. May 25–28, 2003. pp. 786–789.

Sanders, J. S. 1994. "Design and Analysis of Electro-Optic Sensors Based on the Compound Eyes of Arthropods." Memphis State University Ph.D. dissertation.

Schwind, R. 1980. "Geometrical Optics of the Notonecta Eye: Adaptations to Optical Environment and Way of Life." *Journal of Comparative Physiology*. 140, pp. 59–68.

Warrant, E. J. and McIntyre, P. D. 1990. "Limits to Resolution in Superposition Eyes." *Journal of Comparative Physiology A*. 167, pp. 785–803.

ITEA Journal 2008; 29: 380–387

Testing the Intelligence of Unmanned Autonomous Systems

Miles Thompson
Trideum Corporation,
Aberdeen Proving Ground, Maryland

There is a common misconception in the testing industry that all unmanned autonomous systems can be tested using methodologies developed to test manned systems. A tester can incorrectly state that driving a manned vehicle is conceptually identical to driving an unmanned vehicle; the only difference lies in the position of the operator. In actuality, the main difference lies in the unmanned autonomous system's role in the decision process. Because testers cannot make assumptions about the decision process within an unmanned autonomous system, there is a need for a methodology that completely tests this decision process without biasing the system into a default "human" solution. This article presents a comprehensive methodology for testing the intelligence of an unmanned autonomous system within a given set of requirements.

Key words: Autonomous systems; decision process; intelligence; unmanned autonomous system.

There is a common misconception in the testing industry that all unmanned systems can be tested using methodologies developed to test manned systems. The operation of some unmanned systems is conceptually identical to operating manned systems except for the position of operator. However, other unmanned systems differ from manned systems not only in the position of the operator, but also in their role in the decision process. In the robotics industry, these unmanned systems are referred to as intelligent or autonomous.

The terms unmanned ground vehicle (UGV) or unmanned air vehicle (UAV) are much more common within the military, but these terms are not descriptive enough for this methodology. It is possible to have a UGV or UAV that has no choice in the decision process; radio controlled cars and planes are good examples. Instead, an intelligent unmanned system is a system that takes data, perceives information, possibly compares the information against knowledge, and makes a decision based upon this information. In this way, a teleoperated, unmanned system could be considered intelligent only if the system being tested included the operator or the system makes some decisions without the operator. The term autonomous is also used interchangeably with intelligent, giving rise to the name unmanned autonomous system (UAS). For the purposes of this article, UAS describes an unmanned system that makes decisions based on gathered information.

Because testers should not make assumptions about the decision process within a UAS, there is a need for a methodology that completely tests this decision process without biasing the system into a default "human" solution. In this article, the reader will discover a new, comprehensive methodology for testing the intelligence of a UAS while testing explicitly to a given set of requirements.

Unmanned autonomous system testing (UAST)

To test any system effectively, a tester must be given three things: the system to be tested, the documentation associated with operation and maintenance of the system to be tested, and the requirements against which the system will be tested. An intelligent unmanned system is no different. For testing the intelligence of a UAS, this does not necessarily mean an entire vehicle. The system under test could instead be the removed computer from a UAS as a preliminary subsystem test in simulation. In fact, testing the decision making capability of the computer before its being placed within a vehicle is not only cheaper, it is much safer. The point to remember here is that not all UAS testing requires the entire system and that these selected tests can be run in preparation for tests of the entire system. At the same time, no testing process is complete without at some point testing the entire system.

The documentation for the operation and maintenance of the UAS is critical to developing representative tests. This documentation gives the testers insight into how the developers envision the end-users will operate and maintain the UAS. Additionally, the documentation associated with a UAS makes testers aware of safety concerns through extensive hazard analysis. This analysis should show that the UAS would fail gracefully and without incident.

Finally is a list of requirements against which the system can be tested. Requirements are specifications that a test article must meet or exceed. Metrics, which are quantifiable data that pertain to a specific requirement, are used to assess satisfaction of requirements. Given just these three things, testers need to create an effective and efficient test schedule. The process to design this test schedule is the methodology all UAS testers should use for consistent evaluation.

Test to requirements

The requirements are the best place to start designing a methodology because they are the basis of all testing. Requirements are specifications that a test article must meet or exceed. Many requirements dictate the missions and environments the test article will encounter in the field and are required for creating the critical test matrix. The critical test matrix is an array of test scenarios with environments along one dimension and missions along another. The difference between the critical test matrix and any other test matrix is that these are the bare minimum test scenarios necessary to observe the behavior of a UAS and to record enough data for later virtual environment verification. Other requirements can fit into classes of requirements that are important for parallel testing. Parallel testing is a strategy that tests more than one requirement per test. Testers then use classes of requirements to identify specific measurement and evaluation tools necessary for the test. Each class represents a type of sub-test within each test scenario in the critical test matrix, and the set of all the sub-tests creates a complete test. In order to show the process in action, here is an example based on typical UAS requirements.

The UGV must be able to create a suitable path plan of movement to a goal waypoint and execute the plan under the following conditions:

1. Dynamic environment with moving obstacles possibly up to 65 MPH;
2. Must recognize all impassable obstacles at time 0.75 seconds before predicted impact at current speed along linear path to obstacle and avoid collision with any obstacle;
3. Must recognize all pedestrian human obstacles at time 1.5 seconds before predicted impact at current speed along linear path to obstacle and avoid collision while adjusting speed appropriately;
4. Maintain these top speeds: Off-Road Flat—40 MPH, Highway—55 MPH, Off-Road Complex/Forested—25 MPH, Human Within Possible Collision Trajectory—15 MPH;
5. Maintain mobility for 95 percent of mission duration up to 4 hours;
6. Allow an operator to take control of the vehicle for teleoperation at any point in the mission and later relinquish control back to the UGV;
7. Allow an operator to send goal waypoints to the UGV as a queue or interrupt a current waypoint with a new one.

Keep in mind, these requirements for the UGV are only pertaining to the autonomous or intelligent behavior and control and therefore qualify as UAS. There are additional requirements for mobility, sensing, and reliability that would use traditional manned vehicle test methodologies. In order to create a test matrix, testers need to extract a list of missions and environments from the requirements documentation. A paired mission and environment constitutes a scenario, and the test matrix will consist of all combinations of missions and environments, also known as all the scenarios. In this example, there is only one mission and that is to get to the end waypoint. From the requirements, we can also determine there are only four environments specified: highway, off-road flatlands, forest, and off-road complex. The testers also know that all the environments are dynamic with moving obstacles up to 65 MPH.

With the goals and environments clearly defined, the scenarios for critical test are set. Now the testers need to know exactly which metrics to run during each of these scenarios. The classes for sub-tests include, but are not limited to, object recognition, collision avoidance, operator control, path planning, time limit, and mobility control. Within object recognition, there are two sub-tests: human recognition and other impassable obstacle recognition. It can be inferred from the requirements that all environments have the possibility of human pedestrians as obstacles and therefore should be tested in every scenario. Collision avoidance only has one requirement; that is to avoid all impassable obstacles. Operator control has two different sub-tests: teleoperation and waypoint directive. The path planning class has an inherent sub-test that

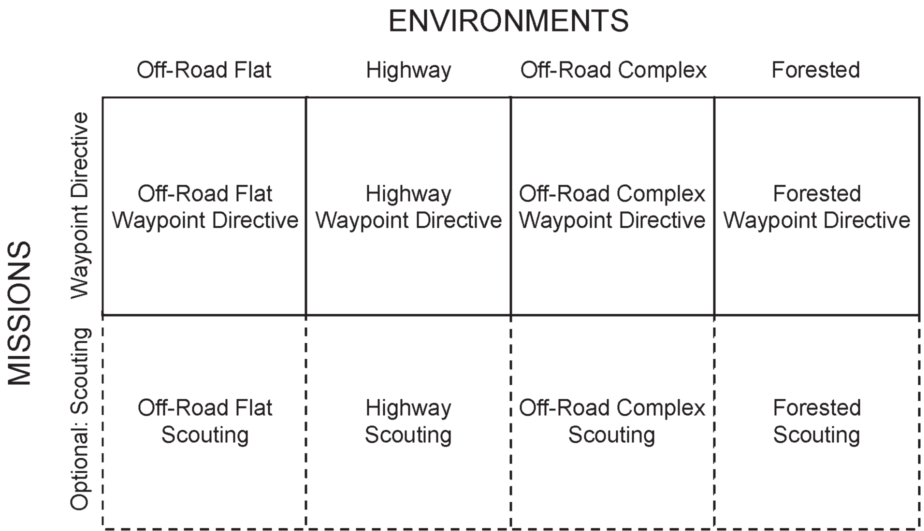


Figure 1. An example of this methodology creating a Critical Test Matrix

can be considered a pass if the UAS meets the collision avoidance and waypoint mission requirements. In the time limit class, the only time requirement is for the UAS to maintain mobility for 4 hours during mission. Mobility is an interesting class because there are two distinct sub-tests. The first is the obvious, maintaining mobility 95 percent of the time during the first 4 hours. The second is making sure the UAS does not exceed its maximum speed requirements in its given environments.

For this simple example, we end up with four mandatory tests and four optional tests *(Figure 1)*. The test matrix that is formed from this process is the critical test matrix because it contains all the necessary sub-tests to ensure that every requirement and every test scenario is tested at least once. During each of these tests, every identified sub-test will occur and the performance of the UAS will be measured with related metrics defined by each sub-test's requirement class *(Figure 2)*. Additionally, it is possible to have requirements be mission or environment dependent or even exclusive. In these cases, those sub-tests are only run when the scenario has the necessary mission or environment. Unfortunately, this amount of testing is not necessarily complete testing, because while the process generated the necessary scenarios, there was no iteration through the various parameters of each scenario. For example, if the sun had been 20 degrees lower in the sky from the zenith position, the UAS may have performed differently.

Complete testing—the critical and the unique test matrix

The critical test matrix is a good start to complete testing because it can give testers quick analysis of a UAS function in typical missions and environments. However, because it does not incorporate a scientific step through the scenario and mission parameter values, it can hardly be considered complete. Traditional Design of Experiments would suggest a full factorial test design or possibly a Plackett-Burman test design for creating a complete multivariate test matrix (NIST 2008). The full factorial test design just ends up being impractical. For example, a UAS that is placed in 10 scenarios with an average of 30 parameters each and 20 discreet values for each parameter has a test matrix with an order of magnitude of 10^{40} tests. The Plackett-Burman design would be better, even if only to reveal the most influential parameters, but this test design grows exponentially for each value over two for each parameter, bringing it to an unmanageable size very quickly. A new method of Design of Experiments is required to test a UAS comprehensively while at the same time limiting the size of the test matrix or developing ways to increase the speed and decrease the cost of testing.

Parameter effect propagation

The Robotic Intelligence Evaluation Program (RIEP) at Aberdeen Test Center (ATC) has designed a methodology to both comprehensively test the intelligence of a UAS and limit the size and complexity of the tests. Through parameter effect propagation, testers can limit the size of the test matrix by running only those tests that provide a unique situation to a UAS. Parameter effect propagation is the process of recognizing all the individual sets of parameter values (i.e., all the possible scenarios) and estimating the changes to the perception and interaction factors of the UAS. Factors are test article based variables that alter

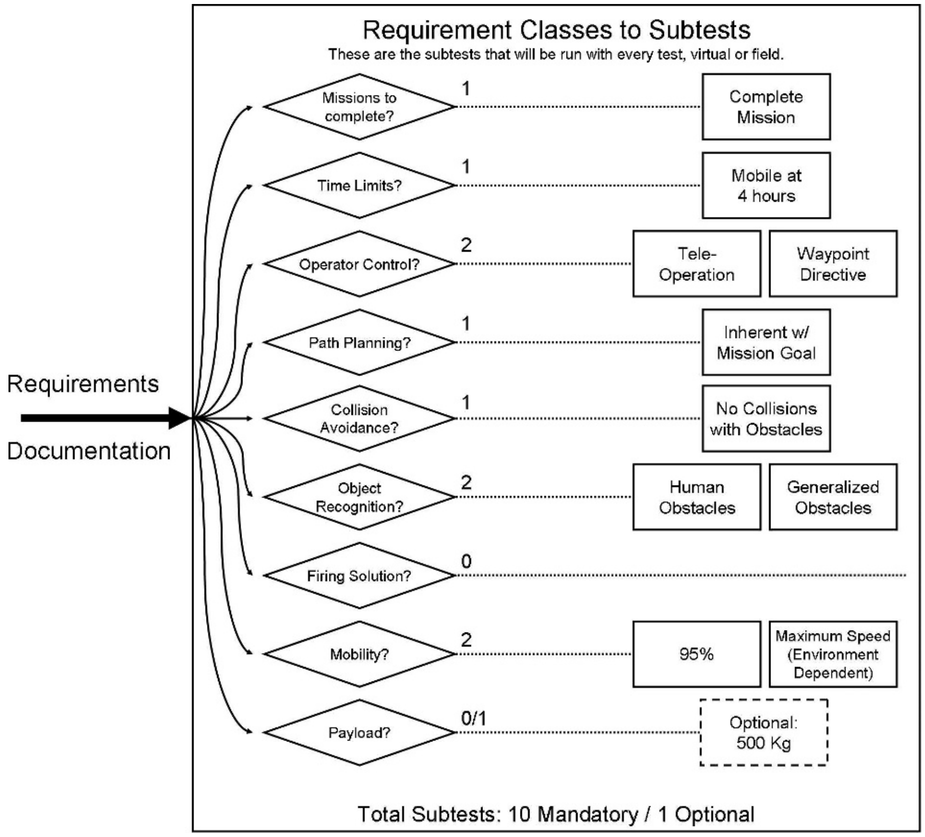


Figure 2. The complete set of sub-tests for every test scenario

the function of the system (e.g., coefficient of friction, camera visibility, battery power), and many are determined by the scenario parameters *(Figure 3)*. For example, an unlit video camera would return much less information during a night scenario compared to a day scenario. At the same time, a night scenario might return the same information as a broken camera scenario, and the UAS will treat them as the same.

The Parameter effect propagation process requires an intimate understanding of the subsystems of a UAS. Each subsystem of the UAS, including redundant subsystems, contributes to the overall collection of

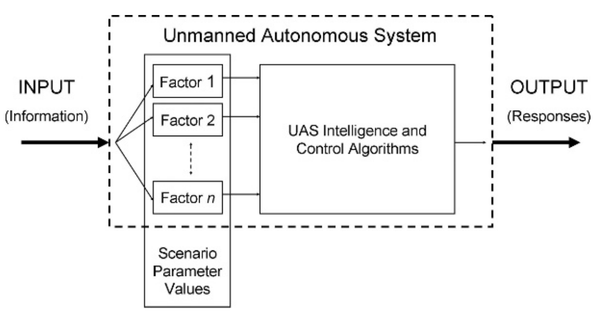


Figure 3. The influence a parameter value has on the system can be better represented by the effect the value has on a system factor

factors. For example, a UAS with eight similar sensors will have at least eight factors that are sensor specific. The data from each of these sensors are a response that could be altered by the sensor factors. A change in the response of any subsystem could alter the capability of the system as a whole.

Before testers can determine how the parameters of a scenario will uniquely affect a UAS, they need to establish the parameters. Using the list of scenarios developed from the requirements documentation, testers can produce a list of scenario parameters for each. The list of scenario parameters is determined by examining the variables in a scenario's mission along with all the ways the environment can interact with the system. Initially, this method might appear to yield an enormous amount of parameters, but the environment can only interact with the system in a few ways. In addition to the physical interaction and electromagnetic interference between the environment and the UAS, the only other inputs into the system are determined by the sensor subsystems. In other words, just like a human, a UAS can only base decisions on what it detects.

With a complete list of parameters, the testers next need to know which values to use for each. Discreet

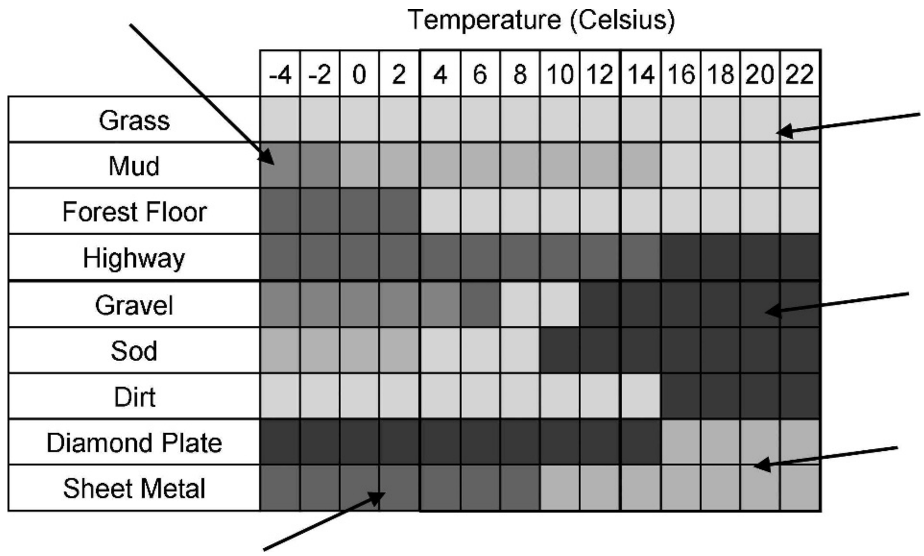


Figure 4. An example of how parameter effect propagation leads to unique testing regions. These unique testing regions give testers the necessary tests to construct a unique test matrix

parameters are mostly straightforward but if the parameter is continuous or there is a multitude of discreet values, there is an approach for selecting the proper values. When traditionally testing a system, testers tend to choose values at the 95 and 99 percent extremes on the probability distribution function of a parameter. With a system containing so many parameters, testing only the extremes would leave possible destructive daily parameter interactions untested. However, selecting parameter values along a probability distribution function at a given percentage step size not only allows the tester to select the extremes but also more of the most common values. The percentage step size is determined by running a sensitivity analysis on the UAS factors by this particular parameter.

Finding the effects on the UAS factors is the next step. Effects on subsystem factors can be predicted through analytical calculations and then verified through field testing or through empirical experimentation. Testers should not be concerned with the multitude of parameter value combinations, but rather concerned with the sets of factor effects derived from the parameter permutations. Running 100 different tests that induce the same factor effect set in the system will not tell the tester nearly as much as running 100 different tests where each of which induces a unique factor effect set. A test team should step through each parameter value for each scenario and calculate a corresponding factor effect set. It is important to take the set of parameter values that define a scenario together so that the factor effect set can accurately represent interaction between multiple parameters. For example, smoke and lighting can each affect optical sensors, but the combination of different values of these two parameters can produce drastic effects on a system. Further, without lighting at all, it makes no difference to the optical sensor on a UAS how much smoke is in its path because it may not be able to "see" at all. When one parameter's effect makes another's effect on a factor inconsequential, it is called parameter dominance. If all the effects of all the values of one parameter dominate all the effects of all the values of another parameter, there is no reason to have the second parameter and it can be removed from the scenario.

Unique test matrix

In order to produce the unique test matrix, the testers need to have produced a factor effect set for each parameter value set. Many of these factor effect sets will be indistinguishable from one another given the precision used by the decision making process of the UAS. In order to limit the size of the test matrix while still representing a significant portion of the scenarios a UAS would encounter in the field, the testers remove extra parameter value sets that correspond to indistinguishable factor effect sets. In *Figure 4*, there is a simple example of this process where two parameters, terrain material and temperature, both affect a factor on the UAS—traction with the ground. The different shaded regions each represent a different value for traction and are also known as unique testing regions. To the system being tested, each similarly shaded case would be identical because the system is unable to realize a difference between them. If testers were to run every case, there would be 126 tests just for traction alone. Instead, choosing one case from each of the unique testing regions leaves testers with five tests and with a significant portion of the information they would have had running all 126 tests.

The resulting test matrix might be many orders of magnitude less than the original full factorial test design. This test matrix is called the unique test matrix because each test to be run should propose a unique problem to the UAS resulting in the greatest probability of inducing emergent and possibly unwanted behavior. A UAS with 12 distinct subsystems may only end up with a few hundred unique factor effect sets for each scenario. For the same 10 scenarios from the example before, the unique test matrix could have an order of magnitude of 10^3 test elements. Although 1,000 tests still sounds like a lot, testers can use simulation software to run virtual tests and identify possibly problematic field tests beforehand. In addition to the unique test matrix, the critical test matrix should be run in a virtual environment as well for verification of the virtual environment. Although executing the virtual tests depend on many conditions, such as the length of the scenario, the devoted processing power, and the discreet time interval in the test, 1,000 virtual tests could take as little as 1 day.

Necessities for virtually testing an UAS

Running an entire unique test matrix in the field is both temporally and financially impractical. Realistically, a unique test matrix will have anywhere from 1,000 to 1,000,000 tests for a complicated UAS, and these are only for testing the intelligent aspects of the system. To run 1,000,000 tests in the field would take 547 years at five tests a day 365 days a year. Even 1,000 tests would take the better half of a year at this same rate. In addition to the temporal and financial issues of running field tests, safety is becoming a much larger issue. Virtual testing allows a system to be preliminarily tested without the risk of bodily harm or damage to the test article. Virtual testing must become a standard complement to field-testing UASs if the testing community is ever going to be able to test an intelligent UAS safely and comprehensively. That being said, what are the necessary pieces to virtual testing?

Virtual environment resolution

To conduct a virtual test, a tester must have a virtual environment in which to run the test. Depending on the UAS, these environments could include ground, air, underwater, surface, space, or any combination of the above. The important thing is that the virtual environment is identical to the real environment in which the UAS would be operating. Identical sounds difficult, but it is really a relative term. If a UAS has no way to detect its environment other than touch, then what does it matter if the environment is all one color? If all the physical obstacles and surfaces in the real environment match the physical obstacles and surfaces in the virtual environment, this UAS would not know the difference and the environments are identical.

This leads to the point that the virtual environments must be interaction based. The term interaction is used because it is not enough to have excellent sensor models if the system cannot take action in the virtual environment. Unmanned ground systems can be mobile but very few actually sense terrain properties such as a friction, roughness, and material, which are all important from an environment perspective. Using lean software development is also very important when designing virtual environments for specific systems. Creating complex acoustic environments is a waste of time if the UAS does not possess any acoustic equipment. To create an effective virtual environment, testers need to study the UAS and understand not only what information it gathers, but also what environmental information is necessary for all interactions. For example, if a UAS has a range detecting LADAR that gathers ranges to millimeter precision, that data may be converted to a lower resolution for path planning. The virtual environment needs to be able to return the more precise data that the UAS will gather in the field to make sure that the UAS processes the information correctly.

Communications and control

From the earlier discussion, a UAS gets its perceived information from two sources: sensors and communications. While communications could be considered a form of sensor, it has its own special place in testing because the commands sent over communications are not determined by the environment the UAS is in. Commands may be sent at any time during a mission and may be as brief as a requesting status update to as verbose as an operator taking over the controls in teleoperation mode. Any commands that the UAS could receive in the field need to be accurately and completely represented in virtual testing. Some UASs even act as communication repeaters and need to be able to send communications on to other agents. This capability must also be modeled accurately within a virtual environment.

Expanding communications to operator control is another important aspect. A virtual environment must be able to interface with the command and control devices that will be used by the operator in the field and be able to update them with the information the operator needs for making decisions. Even autonomous vehicles that do not have a teleoperation mode require an operator to issue high order commands. A virtual environment that encapsulates these operator devices can also be used for operator workload tests without the risks of damaging expensive equipment.

Networking issues

If the virtual environment already needs to support operator equipment, then it might as well be distributed and scalable. This way, multiple UASs can take part in the same test as needed and operators can be stationed at different locations. The virtual test can be scaled up and interconnected with multiple test centers for a distributed test across many assets with many entities.

Aside from these preferences for complex virtual tests, latency is a major issue with networked virtual tests. Tests run in real time across a network introduce outside latency that must be minimized or accounted for. Tests run in virtual time with a higher rate of time passage would enable faster testing tests but follow the same latency issues. Unfortunately, using a faster (or even slower) virtual time is no longer testing the UAS in a realistic environment. Most UASs are already maxing out the computational load in real time and would gain a computational advantage in a slower-than-real-time test and have a disadvantage in a faster-than-real-time test.

Hardware-in-the-loop and vehicle-in-the-loop

Hardware-in-the-loop is a necessity of any virtual test environment. To make sure the UAS is tested on its own hardware, it needs to be able to directly connect to the virtual environment. There are a few ideas on how to do this. The least invasive way is to have a connection from the virtual environment to all the UAS's inputs from sensors and communications. In other words, take the UAS's computer out of the remainder of the system and connect the virtual sensors and communications in the exact same ports that the real ones use. This requires very intimate knowledge of the raw data each sensor produces and knowledge of the connection types. Another approach is to connect directly to the system bus. In order for this to work, messages coming off the sensors need to be transformed to represent what the virtual sensors are seeing and not what the real sensors see.

Vehicle-in-the-loop is a different and interesting new field. Instead of connecting the computer to the environment and seeing how it would perform in the virtual environment, testers are now able to leave the entire vehicle connected. This means a command to drive forward would not only move the virtual vehicle, but also the real vehicle. Already used for ground vehicles, the technology requires the UAS to be dynamically supported on what is known as a Roadway Simulator such as one at the ATC. The advantage of vehicle-in-the-loop testing is that it can be used to determine the effectiveness of the UAS control algorithms in addition to the decision process in a perfectly safe and contained environment where no one can get hurt and the UAS cannot be damaged.

A general procedure for testing a UAS is progressive. It is recommended that the platform is testing for structural integrity and controllability first. Along with these tests, it is important to test the internal communications and the ability of components to interact within the system. Next is hardware-in-the-loop intelligence testing followed by the vehicle-in-the-loop testing. Finally, actual field tests are necessary to verify and validate the results from the previous tests.

Verification and validation

"*Validation* is the process of determining the extent to which the models and simulations accurately represent the real world from the perspective of the intended use of the models and simulations" (ATEC 2007).

During a pre-test meeting, the testers, UAS developers, and end-users should have come to a decision regarding the validity of the testers' approach, execution, and analysis of the UAS and prepared a written statement stating this fact. If so, the validation aspect is complete.

"*Verification* is the process of determining if a model and/or simulation (M&S) accurately represents the developer's conceptual description and specifications and meets the needs stated in the requirements document …" (ATEC 2007).

To verify the credibility of the virtual tests as a representation of the real world, testers must compare the system's performance in the virtual test with the same system's performance in the related field test. However, before testers can compare performance between field and virtual tests, the environmental parameters must be compared to ensure that the virtual environment is representative of the field environment in every significant way. As testers of robotic intelligences, the most important metric for verification is replication of the decision process in the virtual and field tests. If, in a virtual test, a UAS successfully completes a mission but does not complete it in the field test, there is some environmental stimulus missing from the virtual test that is present in the field test or vice versa. Software systems are deterministic and will generate the same output for the same input so testers know there is an incorrect input in the virtual test. This

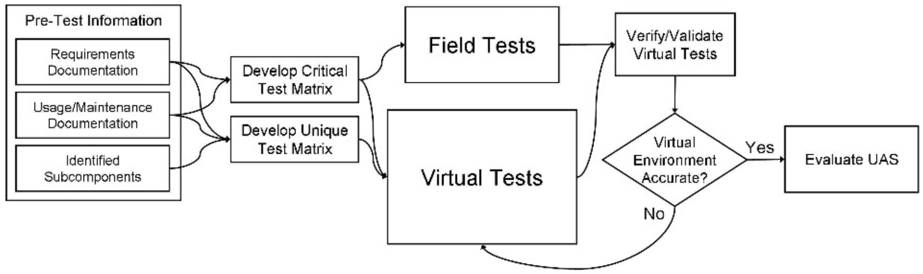


Figure 5. The test design process for testing an unmanned autonomous system's decision process or intelligence

is even true for probabilistic algorithms, as the output should be measured as the probability distribution for actions and not necessarily the action taken.

Testers should examine the data from all replicated tests to make sure the UAS decisions are identical for each pair of tests. Recall that the replicated tests are those from the critical test matrix because they were run in both the field and virtual environments. Comparing the UAS logs from the virtual and field tests is vital to this process. There is no definitive way to understand a UAS decision process without acquiring the logs from that system. If there are any discrepancies, an iterative process is used to modify the simulation software so that it more accurately represents the stimuli the UAS acquires with its sensors. The ATC RIEP recommends that all simulation software be tested in advance for fidelity with sensor models and compared with a base test UAS or sensor capture system in the field.

Conclusion

Ultimately testers are left with a linear flowchart for testing a UAS. As can be seen in *Figure 5*, given the right starting information a test team can develop a very sophisticated test regimen for any UAS. One last reminder: the methodology discussed in this article is for testing the intelligent aspects of a UAS and is not a complete reference. For a complete testing of any system, testers need to incorporate traditional system testing to ensure that the physical and psychological aspects are tested as well. ❑

MILES THOMPSON is a Robotics Test Engineer for the Trideum Corporation based in Huntsville, Alabama. A 2006 graduate of Carnegie Mellon University, his work at Aberdeen Test Center is largely centered on the fields of robotics and mathematical models of systems. He is currently on the Robotic Intelligence Evaluation Program (RIEP) at Aberdeen Test Center. The goal of the program is to provide the U.S. Army, and the rest of the testing industry, methodologies for testing the intelligent systems. Because these methodologies rely heavily on simulation testing, the program intends to release tools for managing these simulations and collecting and analyzing data. E-mail: miles.thompson@atc.army.mil

References

ATEC. 2007. "Test and Evaluation Modeling and Simulation Verification, Validation, and Accreditation Methodology." ATEC Pamphlet 73-2113. April 2007. Alexandria, VA: US Army test and Evaluation Command.

NIST (National Institute of Standards and Technology). 2008. NIST/SEMATECH e-Handbook of Statistical Methods, Gaithersburg, MD: NIST. Available at http://www.itl.nist.gov/div898/handbook/. Accessed January 9, 2008.

ITEA Journal 2008; 29: 388–395

The Test and Evaluation of Unmanned and Autonomous Systems

Fil Macias
U.S. Army White Sands Missile Range,
White Sands Missile Range, New Mexico

Current Department of Defense test and evaluation capabilities and methodologies are insufficient to address testing of weapon systems operating in non-deterministic and unscripted modes characteristic of the unmanned and autonomous system. Task complexity and adaptability to the environment are critical for evaluation of unmanned and autonomous performance. This represents a new challenge for the test and evaluation community. Verification of system performance and interactions will require the tester to understand the nuances of multiple technical domains such as physical/Battlespace, information/knowledge, and cognitive/social.

Key words: Autonomy; behavior adaptation; cognitive agents; complex systems; test arena.

Those of us who can recall portions of our misspent youth watching *Lost in Space*, *Star Wars*, and *Star Trek* may find it somewhat ironic that we would be involved in testing robotic/intelligent machines for the U.S. Department of Defense (DoD). Although the technology is still nascent and advancing, we are faced with the innovator's dilemma and the need to improve unmanned autonomous systems (UASs) through advancing test and evaluation (T&E). We certainly do not test B9s or R2D2s at our test ranges and can agree that it is much too early to invoke Asimov's Laws of Robotics. Yet, intelligent machines in the form of UASs are now rapidly finding their way into the hands of the warfighter and are envisioned to provide amazingly new tactical capabilities in the near future in a variety of areas including mission assurance; command and control; and intelligence, surveillance, and reconnaissance. UASs will enable the warfighter to expand capabilities in the large arena of dull, dirty, and dangerous jobs typically performed by the soldier.

Introduction

The evolutionary nature of UAS acquisition must be met with evolutionary test capabilities yet to be discovered and developed. Test capabilities must be deployed at a faster pace than UAS deployment to satisfy the demand for warfighter improvements. The DoD is stimulating this new area of innovation with ongoing plans to invest more than $11 billion during the FY 2007–2013 timeframe out of a projected $24 billion-plus total budget for unmanned systems as referenced by the Unmanned Systems Roadmap 2007–2032 (DoD 2005).

The Pentagon's investment in UAS technology will drive innovation toward technology with increasing levels of intelligence and autonomy. UASs will enhance the tactical capabilities of warfighters by advancing decision making while performing dangerous, difficult, and monotonous tasks. Consider the example of a Predator unmanned autonomous vehicle designed for surveillance yet now also configured for weapons delivery. The specific needs of UAS employment are mission driven. As UASs become more intelligent, the potential for cognitive agents to enable and enhance these areas is expected to accelerate. UAS missions are expected to expand beyond simple remote sensing to include target illumination and weapons delivery. Physics-based testing is fundamental for knowledge rendering, expanded decision making, and collaborative action. The issue for T&E is that it must become more prominent during the acquisition process and frontloaded to create an advantage. What is that advantage? The advantage is to create a more viable test infrastructure that transfers UAS technology to the warfighter in a fast, efficient, and cost effective manner. Hence, testing is an important step in the evolutionary progress of any innovation.

Current T&E techniques are suitable for systems with tightly coupled tethered operations. As we approach infinitesimally close to fully autonomous systems over a 30-year horizon, testing becomes enormously more difficult. Therefore, to address these and other testing limitations, the Test Resource Management Center established the Unmanned and Autonomous System Test (UAST) focus group, whose mission is:

1. Develop the technologies required to test and evaluate our transforming military capabilities. This includes any system that makes our warfighter more survivable and effective in combat:

a. lethal and non-lethal weapons;

b. manned and unmanned ground, sea, air, and space systems;

c. intelligence, surveillance, and reconnaissance systems;

d. information systems.

2. Provide the required T&E technologies in time to support developmental and operational tests to verify performance before production or deployment.

UASs as complex systems

Recognition that UASs represent a new type of technology with a new engineering genus is key to development of a UAST strategy. In fact, these systems may be characterized as complex systems (Braha et al 2006). Why? Consider the fundamental differences between a traditional system (what we test) and a complex system (what we must test). Systems designed under traditional means are expected to perform predictable tasks in bounded environments. A complex system, on the other hand, functions and operates in open, non-deterministic environments. These systems are composed of interconnected parts having one or more properties (behavior among the possible properties). The complex system is always greater than the sum of its parts. For example, humans are complex systems and so are robots. Complex system testing involves more than optimized testing procedures because many interactions happen among systems and subsystems. Testing is expected to be more fluid coupled with a component of uncertainty, commonly referred to as chaos.

Testing a UAS cannot be limited to the physical domain aspects of the individual system but must consider the systemics of the entire collaborating unit: humans, systems, and mission. Complex systems are studied by many areas of natural science, mathematics, and social science. Fields that specialize in the interdisciplinary study of complex systems include systems theory, complexity theory, systems ecology, and most importantly cybernetics.

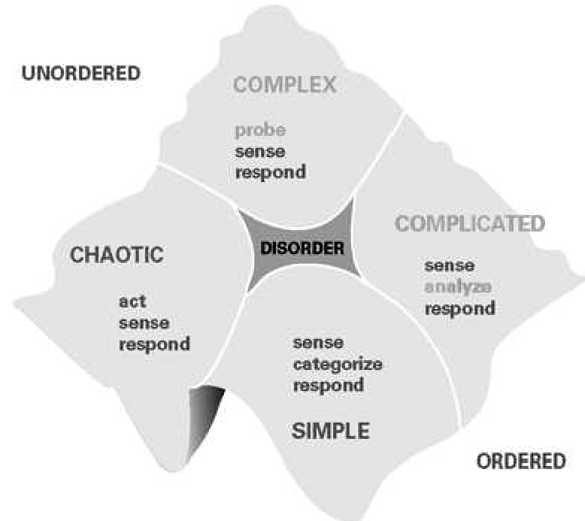


Figure 1. The Cynefin framework (Kurtz and Snowden 2003)

Traditional T&E is limited to single system focus with life cycle development numbered in years. System Interoperability is managed through interface requirements and integration of components. The focus is on reliability, maintainability, and availability within a centralized acquisition and management framework. Needs and requirements are primarily considered fixed at concept phase. The Cynefin framework is a sense-making device developed to help people make sense of complexities, see *Figure 1* (Kurtz and Snowden 2003).

In the ordered world of T&E, we have well developed guidance for simple and complicated systems but not for complex systems. As UASs trend to more complex and chaotic dynamics we must relax our strategies to enable us to establish emergent practices for complex systems and novel practices for chaotic systems deployment. During the next 30 years, both industry and academia are expected to provide UASs that are emergent and unanticipated, and T&E must be prepared to handle both ordered and unordered test requirements as capabilities mature. Testing UASs demands new tools and methods to address complex systems and action-based environments for chaotic system scenarios.

Autonomous, intelligent systems such as a UAS and its operators, will execute outside of predictable, stable behavior within carefully optimized situations. In order to be useful to the warfighter, a UAS must have the capacity for adaptation and change and be able to perform the unexpected. Consider the acquisition, deployment, and current mission/utilization of the MQ-1 Predator, a medium-altitude, long-endurance, remotely piloted air vehicle. The U.S. Air Force considers the Predator not just an aircraft, but a system consisting of four aircraft, a ground control station, and other pertinent equipment. The Predator failed Operational Test and Evaluation (OT&E) (Coyle 2003). The problem for T&E is that the Predator went on to huge success on the battlefield. Even though the Predator failed to meet measures of performance, nothing is in place to give equal significance to measures of effectiveness. The assembly line process of acquisition and T&E failed to ensure that a viable solution could be passed. In this situation, T&E recommendations were overridden and the system was deployed. Although it would be easy to conclude that we no longer need T&E, this would be a hasty ill-advised conclusion with significant long-term consequences. Test, as it is practiced today, has huge overhead and is highly optimized for yesterday's problems. This is the reason why we need to think about new ways of testing and focus more on mission, capabilities, and effectiveness instead of the measures of performance that failed Predator.

In the future that we are facing today, UASs will be deployed on a timeline of months instead of years. Systems that are being developed in industry and academia have utility today to a warfighter who is facing enormous challenges. The question of how to start testing these systems in parallel with development may require us to move beyond the traditional test focus and towards a test strategy that covers the entire acquisition cycle from cradle to grave. The challenges of testing UASs are moving from simple system test toward the world of complex systems engineering.

Terminology is important in understanding the concepts of complex systems. Consider the following definition of a complex system (Kirshbaum 2008):

> *A* complex system *is any system that involves a number of elements, arranged in structure(s) that can exist on many scales. These go-through processes of change are not describable by a single rule nor are reducible to only one level of explanation; these levels often include features whose emergence cannot be predicted from their current specifications. Scientists are finding that complexity itself is often characterized by:*
>
> - *self-organization,*
> - *non-linearity,*
> - *order/chaos,*
> - *emergent properties.*

Classical test approaches emphasize real-time support, communications, networking, and command and control that continue to be optimized to satisfy problems with declining returns. Current methodologies are no longer relevant due to inherent need, expanding system interdependencies, and increasing complexity. A UAS has multiple interactions between many onboard components and an increasing set of external agents. The properties of a complex system cannot be completely explained by understanding only its component parts. In the case of a UAS, the complexities of human interaction, multi-system operation, cloud computing knowledge frameworks, sophisticated behavior models, collaboration, and expanding mobility all combine to create emergence that leads to even more complex adaptive behavior. Only by instituting positive feedback and negative feedback test frameworks, can these systems be sustained through intergenerational development. UAST provides the potential for expanding unmanned-based warfighter capabilities in requisite ways for better addressing mission and sustainability.

The latest Defense Acquisition Guidebook (DAG), Section 4.4.11.8, specifically devotes a section to Unmanned Systems (UMS) stating "…UMS and unmanned variants of manned systems are being rapidly developed and fielded to meet critical warfighter capability needs…" (DoD 2007). Testing for UAS is currently being fast tracked but within a traditional T&E infrastructure that is limited in its ability to address the challenges of UAST. These challenges, however, are increasingly a function of human/UAS awareness/interaction (cognition) and autonomous control levels (autonomy) *(Figure 2)*. Complexity at the systems level is also matched by even more complexity when these systems are aggregated with other UAS and manned systems in system-of-systems and complex system scenarios. In addition, the requisite variety of products being fielded in the UAS technology sector is a function of both emergent needs and accelerating technology. Yet, T&E remains very systems-centric while the scope of UAS development has expanded to include human dynamics, cognition, knowledge representation, and autonomy (Braha et al 2006).

Missiles, aircraft, and ships are all complicated with well-defined and well-understood boundaries for test. They are expected to perform predictable tasks under the watchful eye of well-trained operators. Traditional engineering practices are ordered and linear. It is a goal oriented process that seeks to achieve known specific ends through well-defined means (Cook 2008). This process can be described through the following milestones:

- Functional Specification (what the system is expected to do),
- Design (how the system and components may look and function),
- Testing and Validation (procedure that sets procedure intended to establish the quality,

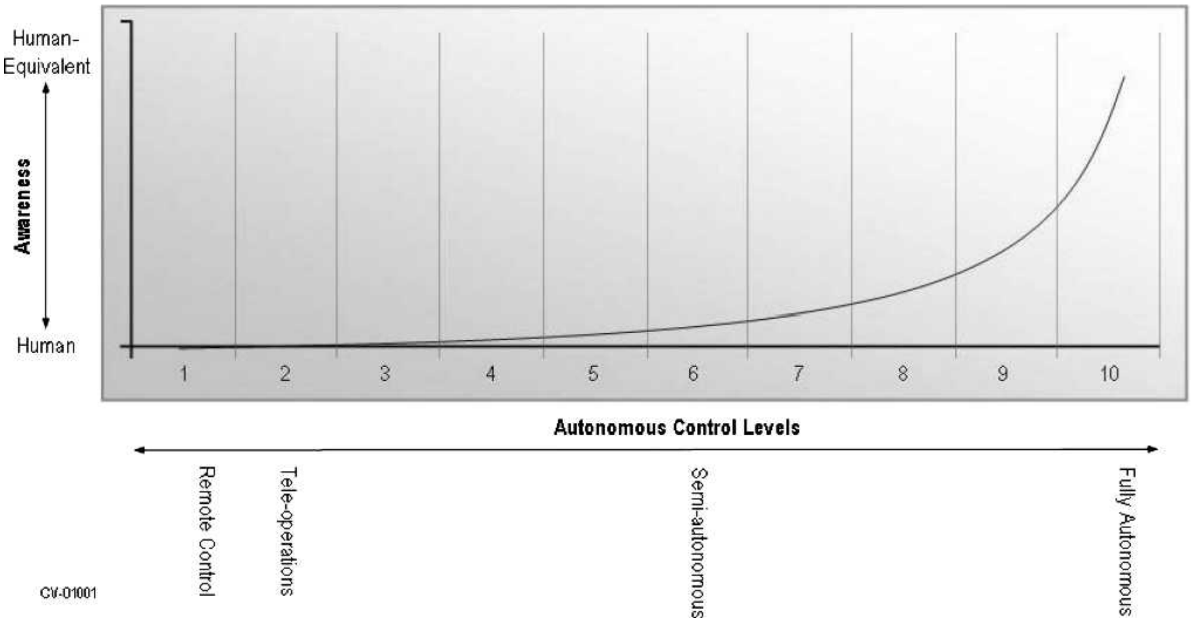


Figure 2. Unmanned systems levels of awareness versus levels of control (Braha et al 2006)

performance, or reliability of something; conditions designed to recreate reality to ensure it performs as needed to discover flaws and correct them),

- Production and Manufacturing (once designed and tested, copies are made).

Current T&E techniques are suitable for systems with manned operations and limited autonomous behavior. The nature of the UAS makes testing considerably more difficult. Whereas the mechanism for creating affordable systems with operational effectiveness is well understood, there are aspects of UAS that create new challenges for UAST. *Figure 3* from the latest Defense Acquisition Guidebook, Section 4.4, (DoD 2007) describes affordable system operational effectiveness with emphasis on measures of performance and measures of effectiveness, including emphasis on the traditional "-ilities" (reliability, maintainability, and supportability). However, for UAST, new "-ilities" have come to the forefront of discussion, namely flexibility, adaptability, and composability, especially as we target system-of-system and complex system scenarios.

UAS are inherently "complex systems" and traditional engineering and subsequently testing practices are not sufficient to address a complex support paradigm characterized by:

- systems-of-systems and complex systems;
- collaboration, extended enterprises, and federated operations;
- net-centric and mobility communications;
- capability driven acquisition and testing;
- sustainability of systems;
- design for flexibility, adaptability, and composability;
- uncertainty in the environment and predictability of human/systems collective action;
- model-based engineering and mission driven development;
- spiral processes, evolutionary acquisition, and open-source dynamics.

A framework for UAST

The UAST Framework consists of four categories: models and architecture, aspects and protocols, testbeds and test environments, and analytics. Each of these categories have been selected to enable a framework of dialog for research and development of a UAST infrastructure and supports an unfolding technology sector with expanding diversity and requisite variety. *Figure 4* illustrates the key areas of the UAST Framework, which will hereafter be referenced as the four technical topic domains.

Models and architecture

UAST models and architecture will become increasingly important based on trends toward model based systems engineering, the guidance of the DoD Architecture Framework, and the increasing relevance of Enterprise Architecture Framework (or Architecture Framework, for short). An architecture framework defines how to organize the structure and views associated within enterprise architecture. In this

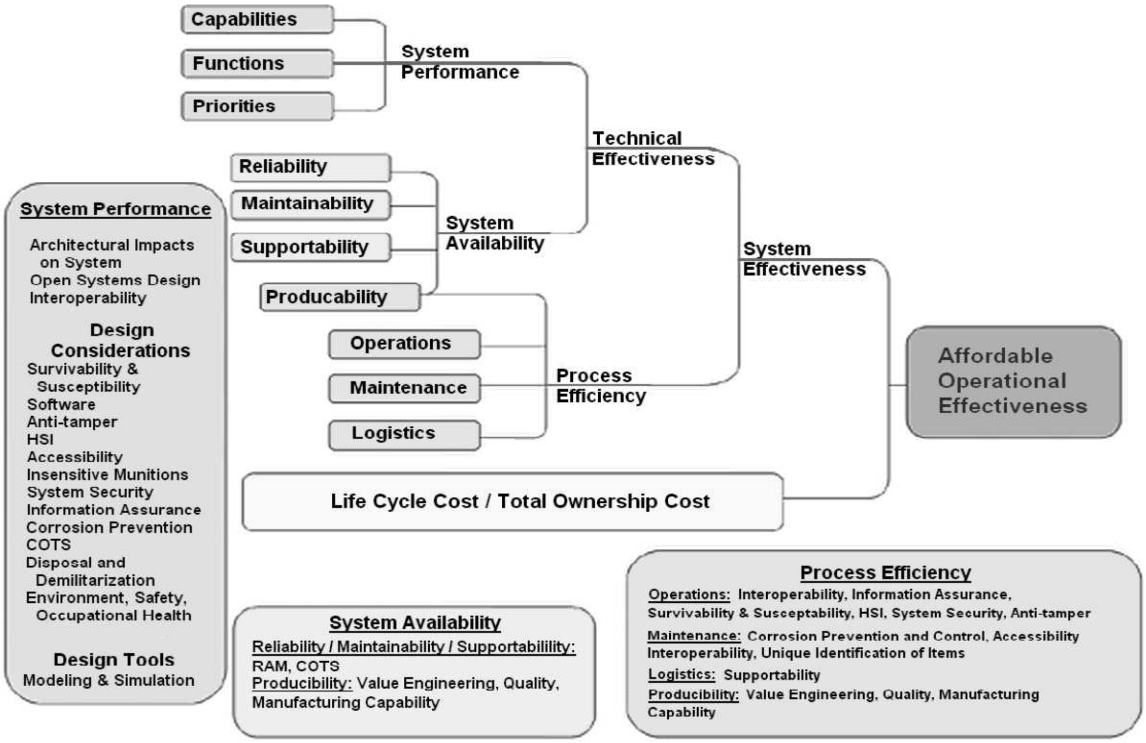


Figure 3. The Defense Acquisition Guidebook operation effectiveness diagram (DoD 2007)

specific case, the enterprise is the mission of establishing a UAS test framework that complements and augments existing initiatives in UAS testing while concurrently working toward a flexible and efficient approach for UAST. Because the discipline of enterprise architecture is broad and because the enterprises it describes tend to be large and complex, the models associated with a particular technology sector can also be large and complex. To manage this scale and complexity, an architecture framework defines

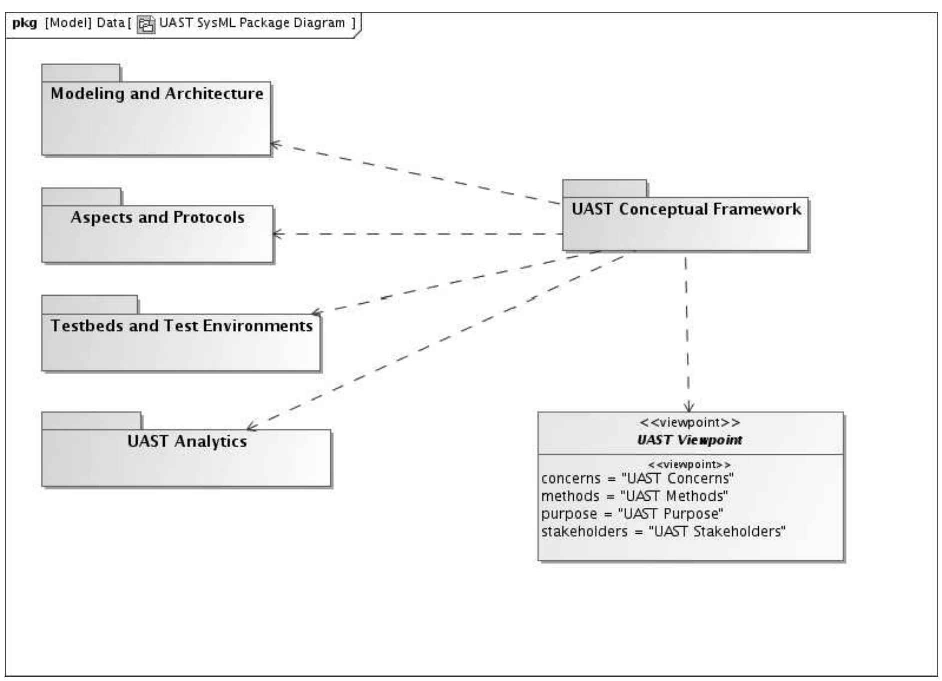


Figure 4. Unmanned and Autonomous System Test framework

complementary projections of the enterprise model, where each projection is meaningful to different system stakeholders. This category of the UAST framework seeks to expand and augment the development of architecture frameworks for UAST in areas such as adaptive architecture frameworks, collaborative tools, human-systems interactions, and decision making.

Analytics

The UAS analytics domain has been established to capture information from systems that inherently are based on awareness and control requirements for varying degrees of onboard intelligence and data gathering capacity. A UAS might be configured to have some level of on-board processing, analysis, and data mining that transforms data to knowledge. A key capability is to identify what information to share (vs data) to enable decision making in support of capability driven development. By understanding what the systems deem to be important information to make decisions, we can support test driven development. This enables identification of what information is important to the system while it is being developed. The information; however, must be assembled with other non-system information to produce knowledge. This knowledge can then be used to facilitate advanced decision making (for development). UAST decision making includes both verification and validation to answer the issues of "system right" and "right system." As we shift from a systems focus to a capabilities focus with the increased emphasis on system-of-systems and complex systems, then the core question becomes what platform can we establish to measure both performance and effectiveness, and how can this information be gathered in data centers to facilitate knowledge creation and system design decision making. The result is that information and knowledge from the testing supports UAS development. This category therefore seeks to identify and mine the information enabling a UAS to focus on mission effectiveness and task completion.

Testbeds and test arenas

Capabilities for UAST are based on sensing, knowledge acquisition and representation, decision making, and autonomous behaviors. In addition, these technologies will be structured to support testing or experimentation in an environment that provides for rigorous, transparent, and repeatable testing of UAS, sensors, simulators, computational tools, and other new technologies. The term for this environment varies across many disciplines; however, it will be referred to as a UAST test arena or testbed. This will support T&E of concurrent real sensor data and simulated entities (Hybertson and Sheard 2008) and it will:

- enable repeatable, controlled, and reproducible measurements and procedures to support federated testing involving multiple command groups and participants;
- provide a robust infrastructure that supports measurement of the UAS and subsystem's performance against *truth-data*;
- provide instrumentation that tracks UAS positions and orientations;
- support advanced sensors to characterize the UAS test environment and provide insight into the UAS functioning and health/status.

A test arena is seen as a set of assets (e.g., instrumentation, targets, etc.), each of which has specifiable relevance to a specific test request, but only if related assets are also available. In addition to test requests, maintenance and calibration requests for each asset must be included. The resulting map of test requests, asset conjoint relevance, and calendar time becomes a complex data structure that must be processed to inform test arena managers in their critical decisions regarding resource allocation and scheduling. We need to compute the test arena commitments for scenarios that will satisfy the greatest throughput of testing per unit of time. To do this, we need (a) a parametric (tailorable) model of a test arena, (b) a language for expressing the test requirements and implied assets, (c) resource allocation and scheduling algorithms, and (d) a dashboard that provides quantitative information about the asset utilization and user satisfaction, all within the context of flexible and efficient testing.

Aspects and protocols

The UAST aspects and protocols category is based on the proposition that UAS can better be understood by using multi-aspect analysis and protocols that offer procedural methodology in the design and implementation of test. In the 30-year timeline over which UASs are expected to fully evolve, a shift from systems assessment to capabilities assessment will be necessary per guidance and by necessity. The T&E of disruptive technology and capability with existing approaches has not kept up with the pace of deployment driven by accelerating need. A shift from an emphasis on system measures of performance toward system measures of effectiveness and mission assessment will continue. A simple component model equipped with protocol framework could be extended to provide aspect analysis. For example, a sensor (component) may have a goal (aspect) to reduce false alarms. Given this goal, it knows to request additional system information (via a protocol). The protocol of an aspect observes the service requests and replies from components and reacts. A nice feature of this model is that an assembly of aspect and components can be transformed back into an assembly of components. All this is done without breaking the black-box nature of the component. A future model of this approach would be protocols frameworks for evaluating knowledge representation, decision making, and collaboration across the cooperating elements of the scenarios involving both systems and humans.

A test driven approach

The goal of UAST is to refine theory and practice for UASs. This is the scientific method of hypothesis and test toward verified conclusions. Without sophisticated UAST that is faster and prepositioned to meet future test challenges, we will never reach requisite variety in UASs without UAST that can match the pace and tempo of UAST capability deployment. An objective of UAST is to develop the basis of a framework that will enable a new testing strategy that establishes confidence and is constantly evolving and adapting to new challenges. It must offer a value proposition that supports safe test conduct and delivers the information and knowledge to support the decision process that expands opportunity and enables risk reduction. A test driven approach is a front-end process approach to test that considers chaotic and complex test scenarios in a rapid pace environment as real possibilities. This approach will require pre-concept test frameworks that enable a T&E strategy for integrated testing in a test continuum that executes in cycles for periods of months and not years (Braha et al 2006). The software world continues to successfully implement these ideas for the high pace rapid deployment world of internet cloud computing and enterprise software.

If an unmanned autonomous complex system is required to achieve a certain objective, the challenge to the tester is to create a learning environment that can deal with chaotic, complex, complicated, and simple (well known) scenarios. Providing reference problems with measures of performance enables researchers to compare implementations, communicate results, and leverage toward specification. It will always be important to develop test artifacts and measurement methodologies to capture performance data in order to focus research efforts. However, reductionist approaches toward testing often result in problem space explosion that translates into testing strategies that require years when the need is in months. Understanding UAST involves recognizing that traditional approaches to T&E are not sufficient to test the UAS over the 30-year period in which UASs are expected to become more intelligent and collaborative. Collaboration across systems and with personnel leveraging internet cloud computing environments provides unforeseeable challenges. Traditional test/engineering methodologies continue to lag behind. We need a new strategy that takes into account the theories proposed by a complex systems community researcher like Yaneer Bar-Yam (Braha 2006). He lays out an intuitive framework for several concepts that can be applied to UAST as it continues to evolve:

- focus on creating an environment and process rather than a product,
- continually build on what already exists,
- individual components must be modified in-situ,
- operational systems include multiple versions of functional components,
- utilize multiple parallel development processes,
- evaluate experimentally in-situ,
- increase utilization of more effective parts gradually,
- effective solutions to specific problems cannot be anticipated,
- conventional systems engineering should be used for non-complex components.

Conclusions—a community of interest

UAS will be an unfolding challenge that must be met with an equally adaptive UAST. This is not an area where we can come with a single solution and believe that it will apply to the next 30 years. There is no universal model for a UAS and for that matter UAST. In the recently published biography on John Adams, he is quoted as stating the following: “Our different views of the same subject are the result of a difference in our organization and experience” (Coates 2008). This seems appropriate to describe the current state of UAST. UAST will be unable to succeed in delivering a value proposition to the testing community unless the testing community is involved. We need tester views, opinions, and experience to create a new trajectory for T&E that is UAST. UAST is currently supported by a focus group staffed by exceptional individuals from all over DoD. This group forms the basis of a community of interest to support UAST. This community of interest needs to consider the complexity of testing cognitive agents (UAS complex systems) with increasing autonomy involved in collaboration to achieve mission goals (Hybertson and Sheard 2008):

- Early tester participation,
- Multi-level assessment: Monitoring, assessment, and response occur at multiple levels.
- Plan-based assessment: Monitoring is triggered by an assessment of dependencies and constraints on plan execution.

- Capability-based assessment: Ongoing assessment of vehicle mission-related capabilities is based on subsystem and environment status.
- Predictive assessment: Monitoring and assessment anticipate future events or conditions.
- Team-based assessment: Assessment occurs not just of individual vehicles, but at the team level as well.

This is the challenge of UASs. UASs are needed for increased variety in warfighter capabilities but only with UAST can we test and evaluate for the requisite variety critical to the warfighter. For the T&E community, this should be seen as an opportunity to significantly improve the suitability and sustainability of this emerging technology sector. By enabling the requisite flow of UAS capabilities coming off the assembly line of industry we can better meet the ever-expanding array of problems related to traditional, irregular, catastrophic, and disruptive operations. ❑

FILIBERTO 'FIL' MACIAS is an electronics engineer in the Systems Engineering Directorate at White Sands Missile Range where he serves as supervisor of the Networks and Control Division. Fil has more than 24 years of T&E experience to include missile telemetry, test sensor/instrumentation development, and real-time modeling and simulation. He also serves on numerous advisory panels investigating the application of advanced technology to test and evaluation. In addition, he serves as the executing agent for the Unmanned and Autonomous Systems Project. He has bachelor's in electrical engineering and a masters of business administration from the University of Texas at El Paso. E-mail: Filiberto.a.Macias@us.army.mil

References

Braha, D., Minai, A. and Bar-Yam, Y. 2006. *Complex Engineered Systems*. Cambridge, Massachusetts: Springer-NECSI.

Coates, E. R. Sr. 2008. Life of Thomas Jefferson Reconciliation With John Adams. Available from http://etext.virginia.edu/jefferson/biog/lj37.htm. Accessed August 10, 2008.

Cook, S. 2008. "The Science of Systems Engineering and the Engineering of Complex Systems." *INCOSE Insight Magazine*. 11(1), pp. 8–11.

Coyle, P. E. 2003. Is Missile Defense on Target? *Arms Control Today*. Washington, DC. Arms Control Association. Online journal. Article available at http://www.armscontrol.org/act/2003_10/Coyle_10. Accessed August 10, 2008.

DoD. 2007. "Unmanned Systems Safety Guide For DoD Acquisition." First Edition (Version .96), January 2007. Washington, DC: OUSD (AT&L) Systems and Software Engineering/Developmental Test & Evaluation.

DoD, Office of the Secretary of Defense. 2007. "Unmanned Systems Roadmap 2005–2032." Washington, D.C.: Government Printing Office.

Hybertson, D. and Sheard, S. 2008. "Integrating and Unifying Old and New Systems Engineering Elements." *INCOSE Insight Magazine*. 11(1), pp. 5–7.

Kirshbaum, D. 2008. Complex Adaptive Systems. Available from http://www.calresco.org/lucas/cas.htm. Accessed August 10, 2008.

Kurtz, C. F. and Snowden, D. J. 2003. "The New Dynamics of Strategy: Sense-Making in a Complex and Complicated World." *IBM Systems Journal*. 42(3), PP. Available at http://en.wikipedia.org/wiki/Cynefin. Accessed August 11, 2008. pp. 467–470.

ITEA Journal 2008; 29: 396–401

Test and Evaluation of Cognitive and Social Capabilities of Collaborative Unmanned Autonomous Systems

Nikita Visnevki
General Electric Global Research Center, Niskayuna, New York
Michelle Bezdecny
Rensselaer Polytechnic Institute, Troy, New York

Testing and evaluation (T&E) of the effectiveness of unmanned autonomous systems (UAS) is a very serious challenge. UAS are ideal target platforms for implementation of intelligent and cognitive capabilities. UAS are often used in hostile and harsh environments where human presence is not desirable, including air, space, surface, or undersea operations. This requires a high degree of UAS autonomy and human independence. To increase effectiveness, UAS may exhibit some social behavior, interacting with each other in a dynamic and flexible fashion. Finally, UAS are capable of carrying lethal weapons. Design and implementation of such intelligent systems is a very difficult task indeed, but even more difficult is the task of T&E of capabilities and operational readiness of these systems. In this article we attempt to lay the system-theoretic foundation for analysis and testing of cognitive and social capabilities of UAS. We propose a cognitive analysis framework and then show how this leads into analysis of the social interaction capabilities of UAS. This cognitive framework is a key to development of comprehensive UAS T&E methodology.

Key words: Cognition; cooperation; emergent behavior; system autonomy; social systems.

During the last decade, various branches of science and engineering have seen tremendous growth of interest in cognition. Cognition is a concept that spans a very large number of scientific disciplines and affects many areas of engineering. Numerous scientific conferences and industrial workshops have been held in the recent past that were entirely dedicated to the goal of understanding cognition and its potential impact on society.[1,2] The fruits of these activities have so far been twofold.

On one hand, new disciplines have emerged and are gaining serious momentum. Specifically, cognitive wireless communications systems (Mitola and Maguire 1999, Haykin 2005a), cognitive radar and sensor technologies (Haykin 2005b), and cognitive human-machine interfaces (Neerinex et al. 2006) are being proposed, built, and evaluated.

At the same time, there is a growing sense of misunderstanding and concern associated with the concept of cognition. The very term "cognition" means different things to different communities of interest. All attempts to come up with a unifying definition seem to fail. Questions are being raised more and more often on why we even care about cognition and what benefit it really offers.

Cognition per se is very important to the researchers in the fields of psychology, brain physiology, medicine, philosophy, and other non-engineering disciplines as well as to the enthusiasts who care about cognition for the sake of cognition itself (IEEE 2008). Engineers and engineering scientists, on the other hand, are preoccupied with solving practical problems in two fundamental application domains:

- systems focused on assisting humans in performing certain difficult/dangerous tasks,
- systems that eliminate humans from performing certain difficult/dangerous tasks.

The key element of the first application domain is human-machine interaction, and the key element of the second domain is system autonomy—the key domain of interest for this article.

One can argue that anything we do can be traced to either of these fundamental domains. In fact, for most practical applications, these domains are typically jointly coupled in the sense that we try to assist humans in performing certain higher-level tasks by eliminating the need for humans in performing lower-level tasks and vice versa.

For a significant portion of the last century, computing has been the tool of choice in assisting scientists and engineers in developing the systems from the above application domains. While using the computing systems available today, we inevitably face two large classes of problems:

- problems that computers solve in a much more efficient manner that humans do
- problems that human can solve much better than any computer currently available.

The first class includes applications that require a high degree of systematic rigor, predictability, and handling massive amounts of data (databases, automatic safety-critical control systems, etc.). The second class is characterized by the need of handling inference, learning, and building associations (face recognition, fast suboptimal routing, decision making with incomplete or conflicting information, etc.). The second class of problems forces us to look at human-style cognition and attempt to implement it in an artificial fashion. Engineers and engineering scientists are being driven toward studying human-like cognition by the need for computing systems that can solve problems from the second class more efficiently.

So, how do we define cognition? The only acceptable definition we find is the one inspired by the history of American politics:

> *"We do not know how to define cognition, but we recognize it every time we see it."*[2]

In other words, we recognize that defining cognition is a futile activity. Instead, what we would like to propose in this article is a unifying system-theoretic capability-focused analysis framework that includes cognitive systems at the very top and shows relationships between cognitive, adaptive, and simple reactive systems. At the heart of this framework is the Systems Capability Technical Reference Model (SC-TRM) that ties all the familiar relevant concepts together and serves three fundamental purposes:

- define the meaning of the concepts commonly attributed to cognitive systems,
- clearly illustrate the relationships between these concepts and their association with specific analytical techniques and tools currently available for addressing aspects of cognitive systems,
- serve as a reference model for understanding how current and future systems fit into the framework and what capabilities they enable.

The premise of our research investigation and the objective of this article is very simple—no matter how smart/intelligent/cognitive unmanned autonomous systems (UAS) are, the infrastructure involved in testing and evaluation (T&E) of such UAS must be smarter, more intelligent, and exhibit higher levels of cognition than the system under test. Therefore, understanding the fundamentals of cognition is an essential part of laying the foundation of an effective UAS T&E roadmap.

The rest of this article is organized as follows. Section 1 introduces the SC-TRM and describes some essential points that this TRM highlights. Section 2 discusses social and emergent behavior of groups of UAS and ties them to various levels of SC-TRM from Section 1. Finally, Section 3 offers some final concluding remarks and identifies future directions of research.

1. SC-TRM

A keynote speaker from the Center of Neural Computations of the Hebrew University in Jerusalem delivered a talk on the information-theoretic aspects of the "Perception-Action" cycle at the latest Cognitive Information Processing Workshop.[2] The talk was followed by a roundtable panel discussion focused on cognition, and very familiar themes quickly surfaced—what is cognition, why do we need it, and what systems should be viewed as cognitive. A question came from the audience:

> *"So, if the perception-action cycle is an essential identifying feature of cognitive systems, does this mean that a thermostat is cognitive?"*

This section and the rest of this article will attempt, among other things, to answer this question.

System theory is a branch of engineering science that is focused on analysis and synthesis of large-scale complex systems, families of systems, and systems of systems (Kossiakoff and Sweet 2002, Blanchard and Fabrycky 2005). Therefore, it seems appropriate to apply the fundamental principles of system theory to cognitive system analysis. The two critical and very powerful concepts of systems engineering are a framework and a technical reference model (TRM), both abstractions.

The concept of a framework is based on the fundamental engineering principle of "divide and conquer" and offers a set of semi-independent views into complex systems that isolate and clearly represent certain aspects of a system's structure, behavior, and relationships with the outside world. An example of such a framework is a Department of Defense Architecture Framework (DoDAF) (DoD 2004a,b). It offers three views into any system—an operational view, a systems view, and a technical standards view. Each view is further broken down into 26+ artifacts called "products" that focus on descriptions of specific aspects of system composition and behavior and offer an overview of relevant industry standards and practices that affect the development and evolution of the system. It is not our intention to develop a DoDAF version of the cognitive system framework. Rather, we refer to DoDAF as a useful example of what a framework may offer—a means of systematic analysis of complex large-scale systems.

The concept of a TRM is a consequence of the same engineering principle of divide and conquer. In contrast to the concept of a framework, TRMs are aimed at defining hierarchical levels or layers of capabilities or functionality in the system as well as their relationships to each other. The most well known TRM is the Open System Interconnection (OSI) reference model for communication systems (Day and Zimmermann 1983). This model defines any communication system as a seven-layer system. Each of these layers addresses a specific aspect of the communication system from physical, to data, to network, to transport, to session, to presentation, to the final application. This model revolutionized the field of networked communication and enabled tremendous growth and variety in the field. TRMs are examples that illustrate another important system theoretic principle—enabling design enhancements by introducing systematic design constraints. Although somewhat counterintuitive, this principle states that system designers achieve better results by eliminating certain degrees of freedom in their approaches. Nowhere is this principle so productive as in software engineering—application of advanced design patterns and usage of well thought-through coding styles and design guidelines yield higher quality software and greater degrees of interoperability, flexibility, and maintainability of the code.

This article offers the beginnings of what we refer to as the SC-TRM and the systems capability analysis framework that extends into the domain of cognitive and social systems. Similar attempts have already been made in the past (Wang et al. 2006). However, in our opinion, the approach described in Wang et al. (2006), although very novel, has two fundamental disadvantages: it is focused on defining the TRM of the brain, and it is, like the OSI model, one-dimensional.

Study of the brain and its organization is of course very important. However, the brain is only one example of a cognitive system. As described earlier, we are interested in practical capabilities of cognitive systems. Had the Wright brothers focused on replicating the flight dynamics of the birds in heavier-than-air aviation, who knows, we might still be traveling the world in steamboats. Bird flight dynamic studies provided useful clues to the design of the aircraft, but a departure from the biologically inspired flight principles led to the creation of better, faster, and more productive aviation. Perhaps similarly, a more pragmatic approach might be fruitful in the study of cognition.

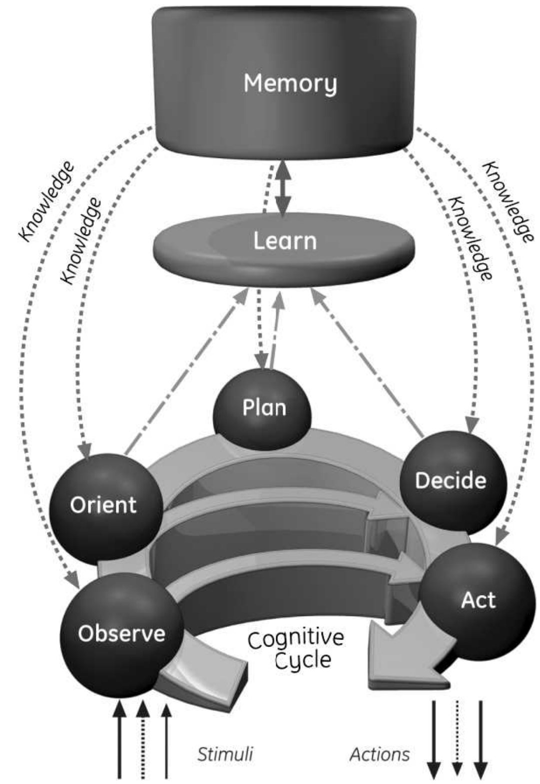


Figure 1. Systems Capabilities Technical Reference Model (SC-TRM)

The dimensionality of the TRM is where the bio-inspired argument is actually appropriate. The brain is not a linear, single dimensional structure; neither is it a two-dimensional plane. It is a complex three-dimensional (3D) hierarchy, which tells us that perhaps some of the cognitive capabilities cannot be easily unfolded and represented by a stack or a two-dimensional plane. The SC-TRM that we are offering here is a 3D representation of the systems' capabilities.

The SC-TRM is shown in *Figure 1*. At the center of the TRM is a cognitive cycle that consists of stages of cognition such as "Observe," "Orient," "Plan," "Decide," and "Act." A system interacts with the environment through external stimuli received from the environment and the actions the system generates in response to these stimuli. Every stage of the cognitive cycle is supported by the explicit or implicit knowledge from the system's memory (structural, logical, procedural, and all other forms of memory one can imagine). Knowledge can be either built in or acquired through a learning process that supports higher-level activities in the cognitive cycle, such as "Orient," "Plan," and "Decide."

The 3D nature of the model illustrates the levels of sophistication, complexity, or significance of the elements of the diagram in *Figure 1*. This implies that "Plan" is a higher-level activity than "Orient," and "Decide" is a higher-level activity than "Act." As is often the case, independent of the level of intelligence capabilities of any system, there might be needs to react without deciding or decide without planning. This possibility is included in the SC-TRM in *Figure 1*.

Obviously, not all systems will include every element of the SC-TRM, as not all communication systems implement every layer of the OSI TRM (Day and Zimmermann 1983). Some elements or capabilities shown in this TRM will be degenerate in some systems. This leads us to the introduction of the systems capabilities framework (SCF) that includes three views:

- *L0*—reactive capabilities,
- *L1*—adaptive capabilities,
- *L2*—cognitive capabilities.

The *L0* view includes representation of the system's reactive capabilities. It includes an "Observation-Action" loop with all the system knowledge and memory necessary to perform reactive operations. The *L1* view focuses on the adaptive capabilities of the system, including the learning, knowledge, orientation, and decision making necessary to support these capabilities (Haykins 2001). The *L2* view focuses on cognitive capabilities and includes the learning, knowledge, and planning necessary to support these capabilities.

Using these three views, we can now define three capability levels of systems based on the SC-TRM:

- *L0*-systems or *reactive systems*: When their capabilities are described through the views of SCF, views *L1* and *L2* are empty or degenerate. This is the simplest class of systems from the capability standpoint.
- *L1*-systems or *adaptive systems*: When their capabilities are described through the views of SCF, view *L2* is empty or degenerate.
- *L2*-systems or *cognitive systems*: All three views of the SCF can be populated with meaningful content.

The fundamental difference between *L1* adaptive and *L2* cognitive systems can be summarized as follows. Adaptive systems are built around a single cost/goal function, and they continuously optimize their response behavior to counteract non-stationarities of their operational environment. This cost function was known to the system designers at the time the *L1* system was conceived and engineered. Cognitive or *L2* systems are capable of auto-synthesizing arbitrary cost/performance functions that they can then use as a basis for their behavior. Such a system can come up with its cost function and use it while appropriate conditions exist. Once conditions change to the point that the current cost function is no longer adequate, a cognitive system can "invent" a new function, and organize its behavior around it, etc. The "Planning" stage of the cognitive cycle at the center of the SC-TRM is an enabling force behind it.

This distinction of the capabilities of the *L1* and *L2* systems requires serious thought. For example, let us say we are sending a deep space probe to the new planet to take pictures of its surface. As the probe approaches the planet, it discovers that the sunsets on that planet are very beautiful. The probe decides to deviate from its original trajectory to take a better picture of the sunsets to send back to Earth. The question is, can we enable such a capability in an autonomous probe today? Many of us would say definitely yes.

The more interesting question is not whether we can enable this capability, but rather, how we can do it. We can probably define the concept of a "beautiful sunset" for the probe in some anthological terms that are machine-interpretable in an adaptive system and alter the cost function of the control system of the probe as to check occasionally for the "beautiful sunset" possibility. Should this possibility arise, we will execute the probe maneuver to optimize the chance of photographically capturing a "beautiful sunset." By doing this, we essentially are turning a control cost function of the probe into a compound cost function with multiple independent or semi-independent terms. We can add all sorts of interesting maneuvers to the probe's behavior (discovering new planets and asteroids on its way, etc.). However, we have to anticipate all these conditions and build them into the design of the probe, creating a very complex adaptive *L1* system.

In order for the probe to be cognitive, or an *L2* system, it must be capable of dealing with a "beautiful sunset" concept on its own. It must generate a new cost function for its behavior, without having had it built into its design by humans. A cognitive probe will detect a beautiful sunset, change its trajectory, take photos, send them to Earth, and perhaps even ask: "do you like this?" At which point a human operator may say: "hey, good job, send us some more if you get a chance," or perhaps: "hey, do not do this again, you are wasting fuel."

At this time, we are finally ready to address the issue of the cognitive capabilities of a thermostat that we mentioned in the beginning of this section. When viewed in the context of the SC-TRM introduced in this section, this question is painfully and obviously absurd (as it was intended to be). The correct question to ask would be:

"Does a thermostat fit into the Systems Capability TRM that extends into the domain of cognitive systems, and if it does, at what level?"

The short answer to this question is yes. It does, and it is an *L0* reactive system. Therefore, the cognitive capabilities of a thermostat are highly degenerate.

2. Social interactions of UAS

The social aspects of cognitive systems deserve separate attention. Socialization, when applied to system capabilities, is a loose term implying various forms of interaction among autonomous systems and between cognitive systems and human subjects. This aspect is not directly addressed by the SC-TRM and SCF introduced in Section 1, yet it is intimately related to them.

When social interactions among cognitive systems appear, they inevitably take either a *centralized* or a *distributed* form. Most human societies are organized according to a centralized scheme with governments, executive and legislative branches, etc. Wolf packs always have a leader, and every member of the pack has a predefined role. This enables us to conjecture that systems with higher degrees of intelligence always seek centralized forms of social organization by default, as centralized forms of social organization are amenable to analysis. In other words, *L2* cognitive systems will always attempt to build centralized social networks.

Ants and other insects have been shown to form completely decentralized, distributed societies, and their complex interaction patterns have been shown to be a result of *emergent behavior* in the massively distributed collection of individuals (Resnick 1997). In fact, emergent behavior is entirely the result of decentralized, massively distributed interactions of *L0* reactive systems (Resnick 1997).

These two extreme examples illustrate how the SC-TRM introduced in Section 1 explains the interaction of systems with various capability levels. The real model of social interaction of practical systems will be somewhere between the two extremes, with *L2*-systems gravitating to centralized schemes, and *L0*-systems having no real choice but interacting through emergent behavior.

The question is, can emergent behavior and decentralized interaction appear among *L1*- and *L2*-systems? The answer is a definite yes. If the complexity of the environment in which the systems interact exceeds the cognitive and analytic capabilities of these systems, they will have no choice but to interact with each other in a distributed fashion, engaging in various emergent and unpredictable behavior as a group. The same is also true if the most sophisticated *L2*-system is unable to obtain adequate global information from or about the social network to support its planning or reasoning capabilities. It will have to rely on local information to make either adaptive or reactive decisions, and an emergent behavior in the social group of these systems becomes possible (see the highway traffic jam analysis problem in Resnick 1997). This is why the transitions from "Observe" to "Act" and from "Orient" and "Decide" in *Figure 1* are crucial even in the *L2*-systems. Understanding these aspects of human-machine or machine-machine interactions is essential to develop adequate T&E solutions that support cognition and system autonomy.

3. Conclusion

Many techniques and tools to enable systems at all three levels of the SC-TRM introduced in Section 1 are currently available. It is just a question of how to put them together in such a way as to enable these capabilities. Cognition in artificial systems is possible now, perhaps not to the extent of a cognitive space probe described in Section 1, but at least to some degree. Understanding and analyzing systems capabilities in a rigorous and systematic fashion will enable better UAS and better UAS T&E of systems to be developed.

One of the weaknesses of the SC-TRM presented in Section 1 is that it does not define the concepts of knowledge, knowledge representation, and knowledge utilization. Knowledge is a critical element in cognition, and many aspects of knowledge are still very poorly understood and misinterpreted. Yet, knowledge is central to both cognition and the SC-TRM presented in this article. More research is required to flesh out knowledge-related aspects of this TRM. We also see at least three additional directions of science and technology investigations that should be taken:

- A comprehensive analysis of various relevant artificial intelligence, learning, optimization, and adaptation techniques within the context of the introduced SC-TRM is necessary. The objective of this activity should be to identify where all these techniques fit and how they can enable desired levels of capabilities in the systems.
- The two fundamental application domains share many similarities, but there are also some fundamental differences between approaches supporting these domains. The SC-TRM is application-domain agnostic in a sense that it equally applies to systems that are assisting humans and to systems that are striving for human independence. The particular techniques for both implementation and testing of systems in these two domains are, however, very different. A good strategy is to divide and conquer—identify two independent focus areas of cognitive system T&E/science and technology, each of which is

focused on its application domain, and establish effective collaboration strategies between the two.

- A T&E roadmap for effective test methodology development should be established. This roadmap should be coordinated with the roadmap of UAS development and acquisition, but it should also look ahead and address the fact we mentioned in Section 1 that any T&E methods for intelligent and cognitive systems should be more intelligent and exhibit higher levels of cognitive capabilities. ❑

NIKITA A. VISNEVSKI completed his undergraduate studies in computer and software engineering at the St. Petersburg Academy of Aerospace Engineering, St. Petersburg, Russia, in 1995. He received his master of sciences degree in electrical engineering from Texas Tech University, Lubbock, Texas, in 1997, and his Ph.D. in electrical engineering from McMaster University, Hamilton, Ontario, Canada, in 2005. His industrial experience includes working for the Boeing Company, Ford Motor Company, and The MathWorks, Inc. Since August 2005 he has been a staff scientist and system architect at the Global Research Center of the General Electric Company. He is a member of the Signal Electronics & Embedded Systems Laboratory of the Electronics and Energy Conversion Organization. His primary research focus areas include intelligent and cognitive embedded systems as well as large-scale systems architectures. E-mail: Nikita.Visnevski@ge.com

MICHELLE BEZDECNY received her bachelor of sciences in mechanical engineering in 2002 and a master of sciences in manufacturing engineering in 2004, both from the University of Wisconsin-Madison. Her industrial experience includes four years at General Electric's Global Research Center in Niskayuna, NY as a mechanical engineer. She is currently pursuing her doctoral degree at Rensselaer Polytechnic Institute in Troy, NY. Michelle's main research interests are cognitive engineering and adaptive systems.

Endnotes

[1]Discussions with participants at the Special Workshop on Cognitive Dynamic Systems, Niagara-on-the-Lake, Ontario, Canada, May 26–28, 2008, http://soma.mcmaster.ca/cds2008.php

[2]Discussions with participants at the IAPR/IEEE/EURASIP Workshop on Cognitive Information Processing, June 9–10, 2008, Santorini, Greece, http://cip2008.di.uoa.gr/index.html

References

Blanchard, B. S., Wolter, J. and Fabrycky, W. J. 2005. *Systems Engineering and Analysis, 4th ed*. Upper Saddle River, New Jersey: Prentice Hall.

Day, J. D. and Zimmermann, H. 1983. "The OSI Reference Model." *Proc. of the Institute of Electrical and Electronics Engineers (IEEE)*. 71, pp. 1334–1340.

DoD. 2004a. "DoD Architecture Framework Version 1.0: Deskbook." Washington, DC: DoD Architecture Framework Working Group.

DoD. 2004b. "DoD Architecture Framework Version 1.0: Volume I Definitions and Guidelines." LOCATION: DoD Architecture Framework Working Group.

DoD. 2004c. "DOD Architecture Framework Version 1.0: Volume II Product Descriptions." LOCATION: DoD Architecture Framework Working Group.

Haykin, S. 2001. *Adaptive Filter Theory, 4th ed*. Upper Saddle River, New Jersey: Prentice Hall.

Haykin, S. 2005a. "Cognitive Radio: Brain-Empowered Wireless Communications." *IEEE Journal on Selected Areas in Communications*. 23(2), pp. 201–220.

Haykin, S. 2005b. "Cognitive Radio: Brain-Empowered Wireless Communications." *IEEE Journal on Selected Areas in Communications*. 53(3), pp. 472–480.

IEEE Spectrum. 2008. *The Singularity–A Special Report*. New York: IEEE. Available at http://www.spectrum.ieee.org/singularity.

Kossiakoff, A., William, N. and Sweet, W. N. 2002. *Systems Engineering Principles and Practice*. New York, New York: Wiley-Interscience.

Mitola, J., III and Maguire, G. Q., Jr. 1999. "Cognitive Radio: Making Software Radios More Personal." *Personal Communications, IEEE*. 6(4), pp. 13–18.

Neerincx, M. A., Lindenberg, J., Smets, N., Grant, T., Bos, A., Olmedo-Soler, A., Brauer, U. and Wolff, M. 2006. "Cognitive Engineering for Long Duration Missions: Human-Machine Collaboration on the Moon and Mars." *Second IEEE International Conference on Space Mission Challenges for Information Technology, 2006. SMC-IT 2006*. Pasadena, California. July 17–20 2006. p. 7.

Resnick, M. 1997. *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*. Cambridge, Massachusetts: The MIT Press.

Wang, Y., Wang, Y., Patel, S. and Patel, D. 2006. "A Layered Reference Model of the Brain." *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. 36(2), pp. 124–133.

Acknowledgments

The authors would like to thank the Test Resource Management Center (TRMC) Test and Evaluation/Science and Technology (T&E/S&T) Program for their support. This work is funded by the T&E/S&T Program through the White Sands Missile Range, New Mexico, contract W9124Q-08-C-0562.

ITEA Journal 2008; 29: 402–410

A Sensor Network Architecture: Information, Control, and Behavior Definitions for Large-Scale or Systems-of-Systems Testing

Howard E. Michel, Ph.D. and Hemant Joshi
Electrical & Computer Engineering Department,
University of Massachusetts Dartmouth, North Dartmouth, Massachusetts

This article envisions a plug-and-play architecture for test and evaluation that will allow engineers to rapidly and robustly define and configure test environments and scenarios. The architecture described here is based on a layered functional decomposition of the three aspects of test: information flow, control flow, and behavior. These individual layered decompositions are presented first, then as an integrated technical reference model. Having an integrated technical reference model is crucial to developing affordable and robust systems that are self-aware, self-healing, and adaptable within a resource-constrained environment. Having these capabilities will become increasingly important as test scenarios become increasingly complex, such as distributed system-of-systems testing or as information volume becomes increasingly more demanding, yet unpredictable, as in the case of continuous test during operational missions. This work is presented as a natural extension of the military's Network Centric Warfare model.

Key words: Adaptable systems; autonomous sensor networks; control distribution; information processing; integrated technical reference models; intelligence; plug-and-play sensors; sensor network architecture; information processing.

Consider a test scenario with 500,000 individual sensors, where commanders on the ground, not test directors, control test execution and where the test and evaluation (T&E) tasking is to report on the effectiveness of each weapon system, including the soldier in the loop. Further consider that you have only months to plan this test, and weeks to report the results. Such might be the situation when testing the Army's future combat system in a force-on-force exercise. And simultaneously, with this tasking, you have the requirement to identify and report on any system, subsystem, or component that has produced an anomalous response, either as a result of previously unobserved combinations of environmental conditions or random situations that had not been observed in previous testing.

Or consider the task of testing swarms of autonomous air vehicles. These vehicles will exhibit adaptive, collaborative behaviors that respond to changes in the environment and their individual and cooperative capabilities. How do you define a test scenario for a continuously adapting and changing system?

Clearly, the state-of-the-art in T&E for military systems today could not handle this tasking. What is required is a catalog of nonintrusive instrumentation—sensors, processors, storage devices, and software—to continuously monitor key parameters, accompanied by user-friendly interfaces that can be used to rapidly configure these components into a system for each test. Furthermore, if this system is to be cost-effective, it should be designed around modular components that are built with open interface standards for hardware, software, and data and metadata formats. If the system is to be robust, it needs to be self-aware, self-healing, and adaptable within a resource-constrained environment.

This article proposes a methodology for describing the functionality of such a system in a manner enabling industry and the military to develop a plug-and-play catalog of sensors, processors, and software modules, allowing test engineers to easily and cheaply configure test systems to meet the requirements of future T&E.

The discussion begins with describing a model represented as a three-sided layered pyramid. The first face of this abstraction describes data and information flow; the second face describes the control flow; the third face represents behavior abstractions. These models are then combined into an integrated technical reference model (I-TRM) and a simple architecture based on the I-TRM is proposed. These developments are illustrated with a simple T&E example.

Models and architectures

A model is used for representing a set of components of a process, system, or subject area, or for developing, understanding, analyzing, improving, and/or replacing a process (ICH 2008). It is also used to define the meaning of common concepts, illustrate the relationships between those concepts, and assist in understanding how current and future systems fit into that model (Visnevski and Bezdecny 2008). A technical reference model (TRM) is used to formulate definitions and provide a formal structure for describing implicit and explicit concepts and operations.

One popular TRM, the International Organization for Standardization's (ISO) Open System Interconnection (OSI) TRM focused on internetwork communication by putting forward a seven-layered abstraction of the functions required in computer communications (Day and Zimmermann 1983, ISO 1983, Tannenbaum 1996). This TRM was very successful in establishing a framework for describing and developing operating concepts, but it was less successful in establishing implementable standards and products. The Internet today is based on the transmission control protocol/internet protocol (TCP/IP) architecture and associated standards (Tannenbaum 1996).

A TRM is different from an architecture. An architecture is used to describe the arrangement of system, function, and design components and interfaces that comprise a solution satisfying a set of requirements (IEEE 1999). A well designed TRM can be used in developing such an architecture.

An example of this model-to-architecture process can be found in the Embedded Instruments System Architecture program, now the Non-Intrusive Instrumentation (NII) program run by the Test and Resources Management Center's Test and Evaluation/Science and Technology Program NII office through the Naval Underwater Warfare Center.

Under contract to this program, General Electric (GE) developed the Open Modular Embedded Architecture (OMEA). The OMEA architecture is a working incarnation of the Information Centric TRM proposed by Michel and Fortier (2005) and Fortier and Michel (2005). The OMEA test suite was fully implemented on Agilent, Yokogawa, and Video instruments. OMEA also implemented synthetic instrumentation and a user interface, and was demonstrated in 2007 (Visnevski 2008).

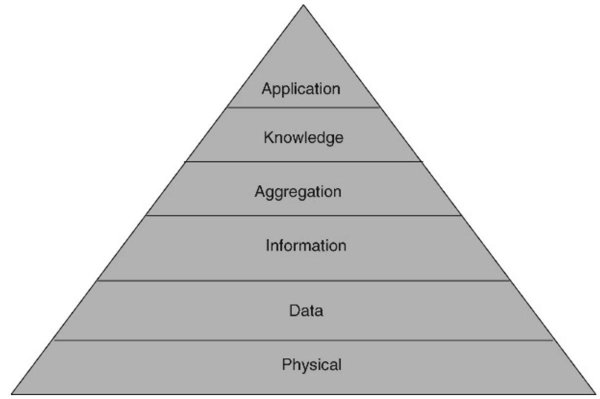


Figure 1. The information-centric face

Proposed integrated technical reference model, the I-TRM

The proposed I-TRM is composed of three faces of a pyramid that describe data, information, and knowledge flow; control flow; and behavior, respectively. The faces are composed of six layers that span their respective domain, but each has a common physical layer at the bottom and a common application at the peak of the pyramid. The data–information–knowledge face and the control face will be described in terms of transformations or functionality, but it is envisioned that the formal specification of the interfaces between these layers would be data and metadata definitions. The specification of the behavior layers is and would be in the form of algorithms. Additionally, the layers are not meant to define distinct hardware layers or software layers because engineering efficiency may dictate integrated products that span multiple layers.

The information-centric face (IC-face)

The IC-face provides a description of the functions associated with data collection, information aggregation, and knowledge generation. It does not specify details related to the control mechanism that manages how and where the data are collected. It is based on the Information-Centric Technical Reference Model (IC-TRM) proposed by Michel and Fortier (2005) and Fortier and Michel (2005) and refined by Joshi and Michel (2007, 2008).

This six-layered view of the IC-face is shown in *Figure 1* and described further on. Lower layers deal with an enormous amount of data that have very low information value. As we move up the layers, the data volume decreases but the information value of that data increases. All data are transformed through these six layers, although some of these layers may be minimal in certain situations. The individual layers, described from the lowest layer up to the top layer, are as follows.

The physical layer. The physical layer gathers and manipulates raw data in unformatted, unverified, and transitory format and deals with the electric and mechanical characteristics of the system. It is composed of electromechanical sensors. An example of data at the physical layer might be the voltage output from a thermocouple. These data are clearly transitory and volatile. Metadata associated with the physical layer would be the sensor type, serial number, location, and calibration status. These metadata would generally exist in a stable form as part of the physical sensor.

The data layer. The data layer performs extraction and transformation of data into digital form and checks the authenticity of the measurements. In our given example, the voltage from the physical layer is transformed into a byte or a word using a prescribed (although possibly variable) process involving amplifiers, filters, and analog-to-digital converters. Variable parameters could include sampling rate, digitization accuracy, filter cutoff frequency, amplifier gain, etc. Metadata generated at this level could include these parameters, plus a time tag, a verification bit to indicate that the sensor is calibrated and operating properly, etc. Metadata from the physical layer and data layer would be bundled with the data to form an informative data packet.

The information layer. The information layer correlates data with scaling, location, type of measurement, etc., to produce information about the system or environment. Continuing with our simple scenario, the data and metadata from the data layer would be combined to produce information that reports, for example, that the temperature at the 12-o'clock position in the combustion chamber of the number one engine was 1,000°F at $T + 1.0$ seconds from test start, and that this measurement should be believed with a high degree of confidence. Notice, that the sensor serial number, location, calibration status, etc., are still available in the data, but hidden. At this level information is made available to the system and users of the system. Metadata created at this level would involve defining the state of these data-converting processes, including for example, the criteria used to define "a high degree of confidence."

The aggregation layer. The aggregation layer performs knowledge aggregation by goal-directed information merging from various sources, as per the requirements of the system or subsystem under test. Continuing now with our jet-engine example, there may be temperature sensors located at the 3-, 6-, and 9-o'clock positions in the engine combustion chamber. There probably are also temperature sensors located at the inlet and exhaust of the engine, as well as pressure sensors, fuel flow sensors, etc. Virtual instruments can be created at this level. For example, readings from multiple temperature sensors, with synchronized time tags, could be combined to give an instantaneous view of the temperature gradients within the combustion chamber. Additionally, a moving window of a time-sequenced series of readings could be combined to provide the dynamic response to changes in the system. Temperatures, pressures, and fuel flows could be combined to create a measure of engine efficiency. The aggregation layer produces information at a higher level of value and lower rate than the information layer below it. Metadata created at this layer would include information about the processes used to aggregate the information from the lower layers.

The knowledge layer. The knowledge layer transforms aggregated information into knowledge by processing it against intrinsic and extrinsic information, and knowledge available. For example, external information such as engine-temperature redline limits could be brought into the model at this level. If the engine temperature approached or exceeded this value, warnings could be issued, or commands could be issued to lower layers in the T&E system to increase sampling rate or accuracy of the engine temperature sensors so a more accurate post-test analysis could be conducted.

The application layer. The application layer concentrates on user–system interaction. It provides a means of accessing and using information for the user in a consistent format, from the system.

The control face (C-face)

The C-face of the I-TRM defines a layered abstraction for describing tasks in a hierarchy from high-level goal definition, through task validation, translation, distribution, and execution. The C-face is derived from the Control Technical Reference Model proposed by Dipple and Michel (2006) and refined by Joshi and Michel (2007, 2008). The C-face concentrates on hierarchical control and task distribution, and primarily builds on the initial work done in the field of control architecture by Albus at the National Institute of Standards and Technology (Albus et al. 1981, 1989; Albus and Ripley 1994; Barbera et al. 1982). Other

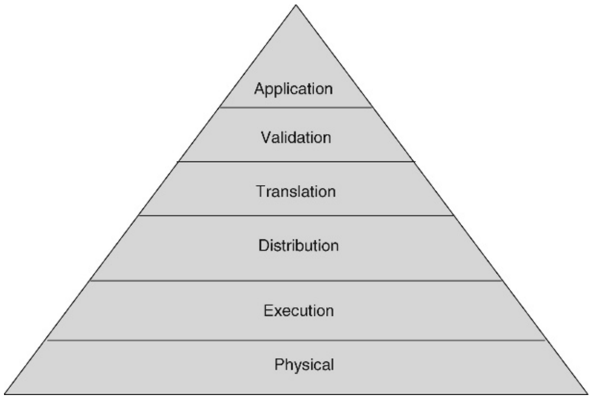


Figure 2. The control face

examples of control-centric definitions are the Mission Oriented Operating Suite (MOOS) (MOOS 2008) and the Joint Architecture for Unmanned Systems (JAUS) (JAUS 2008).

The C-face thus describes the transformation of high-level user-friendly goals into executable tasks. In following our engine performance example, it would relieve the test engineer from having to keep track of individual sensor serial numbers, exact locations, calibration history, etc., and the proper configurations and operating parameters to create synthetic instruments, etc. This information would be created automatically using the flow up of metadata and applying rules specified in defining the T&E case. It would also allow the system to retask individual sensors or groups of sensors in response to predetermined conditions, and thus possibly provide vital information in anomalous situations at the cost of not having multiple copies of slow-changing redundant data. The six layers of the C-face are shown in *Figure 2* and described in the following paragraphs, from the application layer down to the physical layer.

The application layer. The application layer focuses on user–system interaction. It provides an interface for the user to interact with the system to define mission goals. Our simple test scenario illustrating data transformation from an engine temperature sensor (described in the IC-face narrative) may be part of a larger T&E goal, producing a family of engine performance curves.

The validation layer. The validation layer provides a mechanism for authenticating the semantic correctness of the goal and for determining whether the goal is accepted or not. These processes are based on intrinsic and extrinsic information and knowledge. This layer also verifies the probability of accomplishing the goal with the resources available. In our continuing example, metadata from the lower layers would be evaluated here to assess the readiness of the T&E system to produce the family of engine performance curves requested. Are all of the required sensors in place, calibrated, and operating correctly? Is the software in place to create the required synthetic instruments? Are there other, higher priority tasks causing conflicting demands on T&E resources?

The translation layer. The translation layer decomposes valid goals into functional tasks based on knowledge about the lower layers. This layer provides a mechanism to register low-level system components and their physical capabilities. This layer would, in our example, create the need for synthetic instruments to produce synchronized temperature and pressure readings in the combustion chamber of engine one, and synchronize them with corresponding readings for fuel flow and aircraft altitude. It would pass these test requirements down to the distribution layer.

The distribution layer. The distribution layer, based on available spatial and temporal information (passed up from lower layers as metadata), organizes system tasks by decomposing the task groups into subtasks and assigning priorities to them in accordance with pre-established or dynamic goals. Continuing our example, one of the tasks received by the distribution layer might be to create an instrument to record a series of temperature readings correlated with pressure readings from engine one. Software at this level might understand, for example, that there are four individual temperature sensors in the combustion chamber, and that if they all are reading within 50°F of each other, then any one of them can be used to report the subject temperature. Furthermore, software at this level would understand that these temperature readings need to be closely synchronized with pressure and fuel flow readings, but that because of physical characteristics, synchronization at the fraction of a second level is sufficient, rather than at a microsecond level of synchronization. These commands are then passed down to the execution layer.

The execution layer. The execution layer receives directives from the distribution layer and transforms them into control signals for the physical layer. For example, the execution layer, based on its detailed understanding of the sensors involved, would issue a command to temperature sensor number XYZ to stream temperature data (assuming it has a stream mode) at 100 samples/s, using gain N starting at time T.

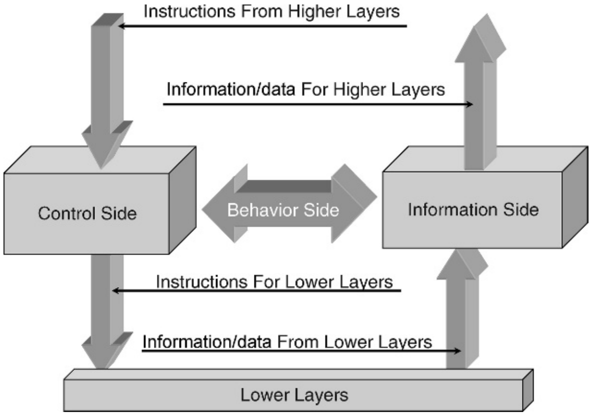


Figure 3. Control loops in the I-TRM

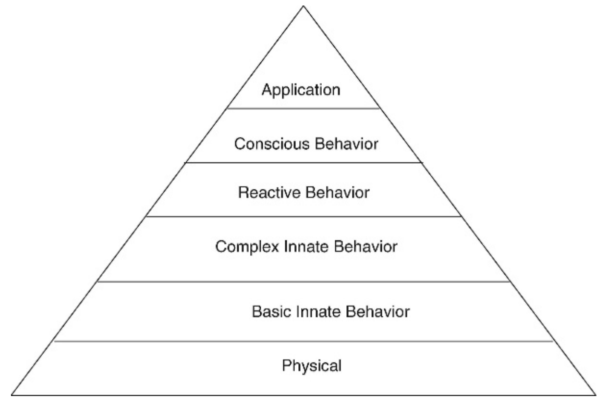


Figure 4. The behavior face

The physical layer. The physical layer constitutes sensors and mechanical units. It executes actions as directed by higher layers.

The behavioral face (B-face)

The B-face describes the intelligence (deliberative and reactive behaviors) of the system and acts as a bridge connecting the IC-face and C-face. This bridge is implemented using control loops based on a classical closed loop control system methodology (Nagrath and Gopal 1981) as adapted in Joshi and Michel (2007, 2008). As shown in *Figure 3*, commands and goals flow down, data and metadata (or information and meta-information) flow up, and the behavior at each layer interprets the execution of the commands or processes the data based on the system status as reported in the metadata or meta-information.

The B-face is a hierarchal arrangement of behaviors into layers based on the scope of control and responsibility of each function. This hierarchical distribution is based on Arkin's work, according to which behaviors can be divided into three major categories: innate behaviors, reactive behaviors, and conscious behaviors (Arkin 1998). The six layers of the B-face are illustrated in *Figure 4* and described in the following paragraphs, from the physical layer up to the application layer.

The physical layer. The physical layer constitutes sensors and mechanical units. It is the same hardware as described in the IC-face and C-face.

The basic innate behavior layer. The basic innate behavior layer implements primitive reflexive behavior and stimulus–response behaviors of the system. It combines the execution layer procedure–execution to produce the relevant data. In terms of smart sensors implied in the on-going example, one basic innate behavior would be to produce a data-word associated with a physical temperature. Another basic innate behavior would be to report metadata when queried.

The complex innate behavior layer. The complex innate behavior layer is the highest reflexive layer. It implements procedures that connect the information extracted from the data and metadata passed up with task execution distribution passed down. Complex innate behaviors are composed of one or more basic innate behaviors structured in a predefined manner to produce a higher-level user-friendly interface. An example of a complex innate behavior might be to "continuously sample and stream data at a particular rate, gain setting, and filter characteristics." Another complex innate behavior might be a self-calibration mode.

The reactive behavior layer. The reactive behavior layer provides a mechanism for dealing with information collaboration from various modules into one structured data unit (local model) (Norvales et al. 2006). It also provides procedures for translating goals into submodules in compliance with the state of the environment. This behavior requires a sophisticated understanding of the state of the various system sensors and rules to interpret the goal-oriented system tasking. This understanding and the rules would be test-specific, but easily configured from more general rule prototypes. In continuing with our aircraft engine example, creating virtual instruments that span either temporal or multisensor data streams would occur at this level. Software at this level would understand that it has various independent sources of temperature and pressure readings from the correct locations that can be correlated to create the appropriate synthetic instrument. The software would also know how to react and retask these sensors if priorities or system status changes, as might be the case if a sensor failed.

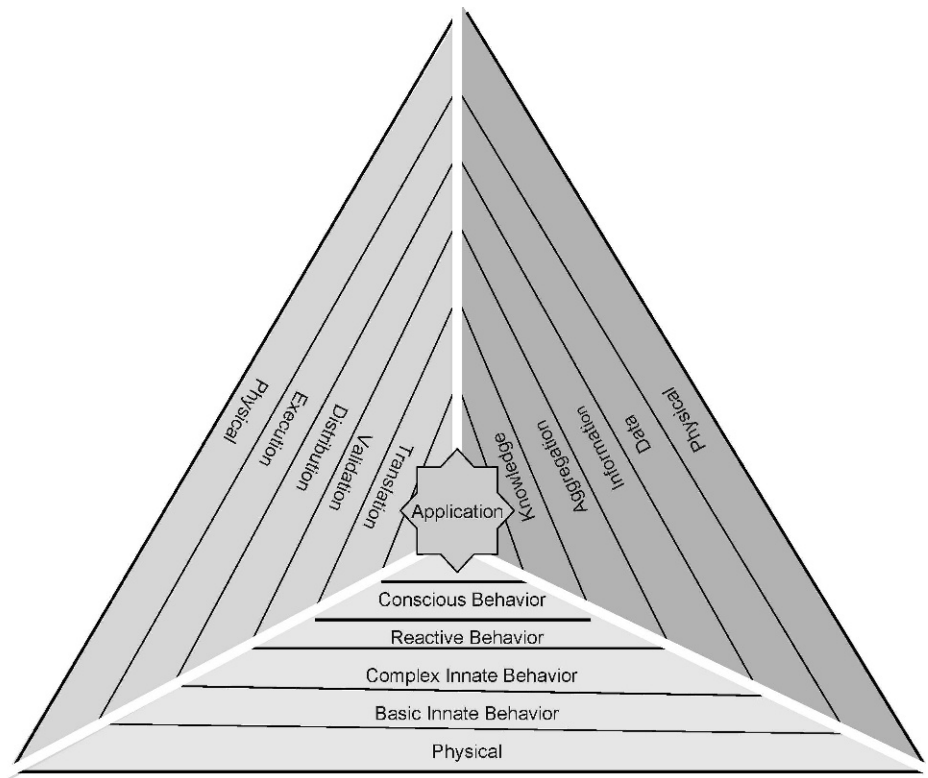


Figure 5. Integrated technical reference model

The conscious behavior layer. The conscious behavior layer provides schemas for checking goal validity and feasibility in the given situation by checking goals against intrinsic and extrinsic knowledge (global model) (Novales et al. 2006). It determines which goals should be accepted. It manages deliberative actions of the systems. In continuing our example, this layer would essentially perform a complete self-test when the system is powered up, and for example, report that all required hardware is operational and configured correctly, or possibly that the suggested synthetic instrument asked for could not be created because of limited bandwidth.

The application layer. The application layer is responsible for user interactions. It decides what information should be furnished to whom and when. It either consumes knowledge passed up to it (in the case of an intelligent program) or passes the knowledge to the user. It receives commands

The integrated technical reference model

The I-TRM illustrated in *Figure 5* is a refinement of architectural principals that have been suggested by Joshi and Michel (2007, 2008). It is an integrated view of the IC-face, C-face, and B-face.

The I-TRM has six layers: (a) physical; (b) data, execution, and basic innate behavior; (c) information, distribution, and complex innate behavior; (d) aggregation, translation, and reactive behavior; (e) knowledge, validation, and conscious behavior; and (f) application.

It is envisioned that each layer is a combination of hardware and/or software that can be specified through its interfaces, specifically its input and output data structures (data/metadata, information/meta-information, knowledge, status, control, and goal structures) and its ability to transform those data structures (behaviors).

Commercial devices that span one or more layers could be built with companies differentiating their products by the implementations behind the open specification of the interfaces. This is already happening with the adoption of the IEEE-1451 family of standards for smart sensors. These smart sensors can implement the lower two or three layers of the I-TRM. An equivalent effort to standardize the upper layers of the pyramid would allow companies to develop and market software products to simplify or automate test setup, operation, and data interfaces.

Sensor network architecture

Consider a sensor network architecture built using the principals of the I-TRM. Sensor networks are used for taking environmental measurements and are distributed systems. The proposed architecture is implemented as a three-tiered hierarchy. The bottom tier consists of sensor nodes. The second tier is composed of cluster heads, and the third tier is composed of a single root node.

Sensor nodes are equipped with various sensors and are capable of performing basic networking, computing, and sensing tasks. A group of sensor nodes are connected through a local one-hop network to a cluster head, the next tier level.

Cluster heads are functionally more powerful units with more computational power, advanced data traffic and networking capabilities, and rich power resources for maintaining its one-hop communication with all subordinating sensor nodes and the root node. Cluster heads process data traffic and handle complex data processing to enrich the informational value associated with the data.

The root node is at the top of this hierarchy. The manner in which sensor nodes, cluster heads, and root node are connected is a tree arrangement. This arrangement provides an easy mechanism for efficient distribution of computational and power resources. This architectural arrangement is inspired by the architecture of a sensor network described in Stojmenovic (2005) and the IEEE 1451 standards. It was proposed in Joshi and Michel (2008) and by Joshi (2008) in his thesis. The reader is referred to these references for a description of the various class diagrams associated with this software architecture.

Sensor node description

A sensor node is the most repetitive physically instantiated unit. As discussed earlier, a sensor is the simplest of the functional units. Its functional blocks span the I-TRM layers 1 to 3.

Layer 1 is a physical layer that consists of sensors and mechanical units. This layer executes actions as directed by higher layers and possesses no intelligence of its own. It gathers raw data in unformatted, unverified, and transitory format. It deals with the electric, mechanical, and procedural characteristics of the system including the working of transducers. Layer 1 implements transducer working commands, transducer data, and basic sensor behaviors (sensor sensing, transmitting, sleeping, and hibernating). The transducer working commands are the set of commands for controlling the Layer 0 transducers, whereas transducer data are the output of the transducer. The basic sensor behaviors are the functionality or behaviors of the sensor that do not require any external information.

Layer 2 contains objects of more informational value and executes more complex commands and behaviors. Sensor data are created by calibrating transducer data. This processing is performed when the sensor is in the sensor-sensing state. Because this design is inspired by IEEE 1451 standards, the implementation of the transducer electronic data sheet (TEDS) is required. It allows for self-calibration of the transducer in the sensor. TEDS would be implemented in the system by introducing calibration data, a data structure. Self-calibration would be initiated with a transducer calibration command.

Cluster head description

A cluster head is a link between sensor nodes and the root node, and spans Layers 3 and 4. It is responsible for data aggregation and data consolidation. There are five different types of parallel running processes in this functional unit. Two types of processes are used to maintain correspondence between a sensor node and the cluster head, that is, one for issuing commands to a sensor node and a second for gathering information from a sensor node. The third process is responsible for performing data aggregation by combining data from various sensor nodes into a single data unit, cluster information. A fourth process controls the communication between the cluster head and root node. It is envisioned that these processes would be configurable through a user interface, the fifth process. A specific example of this would be a graphical or menu-driven user interface that allowed the test conductor to associate specific sensors to control aggregation to create synthetic instruments. Another example would be the ability of this aggregation process to respond to changes in the state of meta-information and adapt the lower-level system tasking to optimize system performance.

Within this architecture, there is a data object, the Sensor Instance Information object, related to each sensor. This object contains data from its associated sensor and the metadata related to that sensor. This metadata makes hot swapping possible. Processes, which maintain the correspondence between a sensor unit and a cluster head, are dedicated processes for each node, so each process has its own copy of every required component and runs independently of the other.

Root node description

The root node is at the top of the sensor network architecture hierarchy and spans I-TRM layers 5 and 6. It is a functional unit that performs knowledge extraction from aggregated information from various cluster heads, accepts and validates the test scenario, and runs the primary user interface software. It is situated at the base station.

The five main processes are:

1. maintaining correspondence between the root node and cluster heads
2. extracting knowledge from the aggregated information from the various cluster heads

3. updating the global knowledge base
4. validating tasking
5. providing a user interface

The processes for maintaining communication between the root nodes and various cluster head can be subcategorized into two groups, one for issuing commands to cluster heads and the second for gathering information from cluster heads. To have the validation feature, proxies for the root controller are generated for each application level process.

Conclusions

This article presents a framework for discussing, designing, and developing sensor network architectures based on modeling system components from three orthogonal abstractions: control distribution, information processing, and intelligence (behavior). This methodology can produce robust systems that are self-aware, self-healing and adaptable within a resource-constrained environment. It provides a potential mechanism for realizing interoperability between various subsystems—both hardware and software—across various manufacturers. ❑

HOWARD E. MICHEL retired from the U. S. Air Force in 1994, having served as a pilot, satellite launch director, engineer, and engineering manager. Dr. Michel earned his doctorate from Wright State University in 1999 and is currently Associate Professor of Electrical and Computer Engineering at the University of Massachusetts Dartmouth. He holds two patents and has published over 40 articles on intelligent systems, artificial neural networks, and optical computing. His research interests include autonomous vehicles, artificial neural networks, and distributed sensor networks. Dr. Michel is a senior member of the IEEE and IEEE Region 1 Director (2008–2009). E-mail: hmichel@umassd.edu

HEMANT JOSHI received his master of science in computer engineering from the University of Massachusetts Dartmouth. His master's thesis was entitled "Autonomous Mobile Sensor Networks Architecture for Hazard Detection and Surveillance." He received a bachelor's degree in engineering (Electronics & Communication Engineering) from Visvesvaraya Technological University, Belgaum, India. Mr. Joshi worked as a lecturer in the ECE Department, Maharishi Arvind Institute of Engineering, Jaipur, India and is currently employed by VMware as a development engineer. E-mail: g-hjoshi@umassd.edu

References

Albus, J. S., Barbera, A. J. and Nagel, R. N. 1981. "Theory and Practice of Hierarchical Control." *Proceedings of the 23rd IEEE Computer Society International Conference*. Washington, D. C. September 15–17.

Albus, J. S., Lumia, R., Fiala, J. and Wavering, A. 1989. "NASREM—The NASA/NBS Standard Reference Model for Telerobot Control System Architecture." *Proceedings of the 20th International Symposium on Industrial Robots*. Tokyo, Japan. October 4–6, 1989.

Albus, J. S. and Rippey, W. G. "RCS: A Reference Model Architecture for Intelligent Control." *Perception to Action Conference, 1994, Proceedings*. September 7–9, 1994, pp. 218–229.

Arkin, R. 1998. *Behavior-Based Robotics*. Cambridge, Massachusetts: MIT Press.

Barbera, A. J., Fitzgerald, M. L., and Albus, J. S. 1982. "Concepts for a Real-Time Sensory-Interactive Control System Architecture." *Proceedings of the Fourteenth Southeastern Symposium on System Theory*. Blacksburg, Virginia, April 15–16, 1982.

Day, J. D. and Zimmermann, H. 1983. "The OSI Reference Model," *Proceedings of the IEEE*. 71(12), pp. 1334–1340.

Dippel, H. A. and Michel, H. E. 2006. "The Control Technical Reference Model." *International Conference on Artificial Intelligence*, Las Vegas, Nevada, June 2006.

Fortier, P. and Michel, H. E. 2005. "Comparison of the EI TRM versus TENA," ITEA Technology Review Workshop, Atlanta, Georgia, July 12–14, 2005.

ICH Architecture Resource Centre. 2008. "Interoperability Clearing House Glossary of Terms" Available at http://www.ichnet.org/glossary.htm. Accessed August 28, 2008.

IEEE. 1999. *IEEE Std 1220-1998*. New York: Institute of Electrical and Electronics Engineers.

ISO. 1983. "Open system interconnection, basic research model, ISO International Standards IS 7498, ISO/TC 97/SCI6. Geneva, Switzerland: International Organization for Standards.

JAUS. 2008. Domain Model Volume I, Joint architecture for unmanned system. Available at http://www.jauswg.org/. Accessed August 28, 2008.

JAUS. 2008. Reference architecture specification volume II, part 1, architecture framework. *Joint Architecture for Unmanned System.* Available at http://www.jauswg.org/. Accessed August 28, 2008.

JAUS. 2008. Reference architecture specification volume II, part 2, message definition. *Joint Architecture for Unmanned Systems.* Available at http://www.jauswg.org. Accessed August 28, 2008.

JAUS. 2008. Reference architecture specification volume III, part 3, message set. *Joint Architecture for Unmanned Systems.* Available at http://www.jauswg.org. Accessed August 28, 2008.

Joshi, H. and Michel, H. E. 2007. "Integrating Information-Centric, Control-Centric and Behavior-Centric Technical Reference Models for Autonomous Sensor Networks." *International Conference on Wireless Networks.* Las Vegas, Nevada. June 2007.

Joshi, H. and Michel, H. E. 2008. "Integrated Technical Reference Model and Sensor Network Architecture." *International Conference on Wireless Networks*, Las Vegas, Nevada. June 2008.

Joshi, H. 2008. "Autonomous Mobile Sensor Networks Architecture for Hazard Detection and Surveillance," University of Massachusetts Dartmouth Master's thesis. January 2008.

Michel, H. E. and Fortier, P. 2005. "Development of an Embedded Instrumentation System Architecture and Its Comparison to the Test and Training Enabling Architecture." *Defense Transformation and Network-Centric Systems*, Proceedings of SPIE Vol. 6249. April 2005.

MOOS. 2008. *Mission-Oriented Operating Suite*. Available at http://www.robots.ox.ac.uk/~pnewman/TheMOOS/. Accessed August 28, 2008.

Nagrath, I. J. and Gopal, M. 1981. *Control System Engineering*, Second Edition. New Delhi: Wiley.

Novales, C., Mourioux, G. and Poisson, G. 2006. "A Multi-Level Architecture Controlling Robots from Autonomy to Teleoperation." *First National Workshop on Control Architectures of Robots*. Montpellier, Vermont. April 6, 2006.

Stojmenovic, I. 2005. *Sensor Networks Algorithms and Architecture*. New York: Wiley Publications.

Tanenbaum, A. S. 1996. *Computer Networks*, Third Edition. Upper Saddle River, New Jersey: Prentice Hall.

Visnevski, N. 2008. "Embedded Instrumentation Systems Architecture." *Instrumentation and Measurement Technology Conference Proceedings, IMTC 2008*. May 12–15, 2008, pp. 1134–1139.

Visnevski, N. and Bezdecny, M. 2008. "Test & Evaluation of Cognitive and Social Capabilities of Collaborative Unmanned Autonmous Systems." *ITEA Tech Review Conference*, Colorado Springs, Colorado. July 2008.

Acknowledgment

This work was partially supported by the Test and Resources Management Center's (TRMC) Test and Evaluation/Science and Technology (T&E/S&T) Program Non-Intrusive Instrumentation office through Naval Underwater Warfare Center (NUWC). The views and opinions expressed or implied in this article are those of the authors and not necessarily those of the Department of Defense or any of its subordinate agencies.

ITEA Journal 2008; 29: 411–419

A Methodology to Assess Lethality and Collateral Damage for Nonfragmenting Precision-Guided Weapons

Amanda Humphrey and Judi See, Ph.D.
Science Applications International Corporation (SAIC), Albuquerque, New Mexico
David Faulkner
Decisive Management Professionals International (DMPI), Woodbury, Tennessee

A methodology was developed to assess lethality and collateral damage for the Focused Lethality Munition (FLM) program. FLM is a new nonfragmenting, precision-guided weapon with damage effects mechanisms that differ from the principal fragmentation damage effects for traditional weapons. To date, guidelines to determine lethality, based on mannequin test data, have not been articulated for nonfragmenting warheads such as FLM. Medical and military documentation was surveyed to derive lethality criteria for four FLM damage effects mechanisms and establish guidelines to address combination effects. The criteria were successfully applied to assess FLM military utility and preliminary validation of the procedures was conducted. Future plans include further augmentation of the model as additional data from program office continuation testing and operational weapon use become available.

Key words: Lethality; collateral damage; mannequin test data; focused lethality weapon; wound severity assessment.

The Focused Lethality Munition (FLM) program was conducted to assess the military utility of a focused-lethality precision-guided weapon. In the modern urban battle space, adversaries routinely place legitimate military targets near civilians or objects such as hospitals and churches protected under the Law of Armed Conflict. The ability to use air power for targets requiring minimal collateral damage is currently limited by the weapon fragmentation effects of available air-to-ground weapons, which can cause significant collateral damage. The FLM weapon was designed specifically to address high-value target prosecution, while minimizing collateral damage outside the focus area.

The FLM weapon combines two technologies to offer a more localized kill mechanism compared with the current steel case warhead, which has a fragmentation effect of 2,000 feet or more.[1] First, the multiphase blast explosive technology uses tungsten fill to increase the explosive weight and enhance near-field blast, as compared with conventional high-explosive fills. Second, the case surrounding the tungsten fill is composed of carbon fiber, which requires less energy to rupture than a comparable steel case. Upon detonation, the composite breaks into small, nonmetal fibers, thereby minimizing warhead fragmentation effects.

The FLM warhead is integrated into the Small Diameter Bomb I in place of the current warhead. The FLM warhead has the same weight, center-of-gravity tolerances, and outer mold line as the Small Diameter Bomb I, and operators use the same mission planning tools. The only modifications include incorporation of FLM weaponeering characteristics and collateral damage estimation calculations.

Science Applications International Corporation (SAIC) provided FLM assessment support to the United States (U.S.) Central Command from February 2007 through May 2008 under the Trusted Agent contract, with weaponeering expertise from the Decisive Management Professionals International (DMPI) subcontractor. The assessment team collected FLM lethality, collateral damage, and accuracy data during five static detonation events at Eglin Air Force Base, Florida, and 11 F-15E live-fly events at White Sands Missile Range, New Mexico. Primary data sources included human surrogates (i.e., full-weight mannequins, gel men, wooden dummies, and blast test devices) arranged in operationally realistic scenarios (*Figure 1*).

Figure 1. FLM test setup

Problem description

Mannequins and blast test devices provide a readily available and cost-effective mechanism to address damage incurred during weapons tests. However, they do not provide a definitive means to determine lethality because mannequins are not living. Some damage can be observed (e.g., severed limbs). Many other types of injuries are not directly visible (e.g., internal injuries) and must be inferred to estimate the impact on lethality. To date, guidelines to determine lethality, based on mannequin data have not been articulated for nonfragmenting warheads such as FLM. Whereas the principal damage effects mechanism for traditional weapons is fragmentation, FLM damage results from a combination of other factors:

- Blast pressure impulse exerts G-forces on the body that can damage the spine, neck, and appendages.
- Blast overpressure can compress and damage air-filled structures, such as the lungs, ears, and gastrointestinal tract.
- Thermal effects can burn the skin and respiratory structures.
- Secondary weapon component debris, though minimal for FLM, can penetrate the organs and soft tissues.

To assess FLM military utility, a set of lethality guidelines was developed and applied for nonfragmenting warheads, based on research of medical and military literature.

Development of the solution

To develop a methodology to determine lethality for nonfragmenting weapons, the definition of a "serious/lethal" wound was first tackled. The Joint Munitions Effectiveness Manual (JMEM) Weaponeering System, the standard U.S. Armed Forces weaponeering tool, does not directly define a serious/lethal wound (JTCG-ME 2006; DIA 2003). However, it is widely accepted throughout the targeting and assessment community that *serious/lethal* is a wound category between the JMEM *serious* (i.e., one that requires hospitalization) and *lethal* (i.e., one that causes immediate death) categories. Thus, a serious/lethal wound is characterized by sufficient injury to cause death within 4 hours of the kinetic event when competent medical attention is unattainable.

Second, attention turned to the Department of Defense (DoD) severity scale to identify the numbers and types of severity categories needed (Chairman of the Joint Chiefs of Staff 2006). The DoD severity scale is used to describe personnel injury during battle damage assessments by the DoD. The DoD scale classifies the severity of human wounds into one of five categories:

- deceased (lethal)
- very serious (life is imminently endangered)
- serious (immediate concern, but no imminent danger to life)
- incapacitated (hospitalization required)
- not seriously injured (no wounds or minor injuries that do not require hospitalization)

A similar five-category scale was developed for FLM purposes: *lethal*, *severe*, *moderate*, *light*, and *no injury* (Rows A and B in *Figure 2*). For FLM assessment purposes, the DoD category for *not seriously injured* was subdivided into *light* and *no injury* to facilitate calculation of the collateral effects radius (distance

A	FLM Scale	Lethal		Severe	Moderate	Light	No Injury
B	DoD Scale	Deceased	Very Serious	Serious	Incapacitated	Not Seriously Injured	
C	Distance Propelled	≥ 10 feet		> 5 feet but < 10 feet	1 to 5 feet (depending on impact geometry)		
D	Blast Overpressure Percent Lung Contusion	Severe Area > 50%			Moderate Area > 10% but < 50%	Slight and Trace Area > 0% but < 10%	No Injury Area = 0%
	Blast Overpressure Qualitative Severity Level	Very serious injury/lethality predominant		Some severe injury	Some moderate injury	Injuries greater than trace unlikely	Injury unlikely
E	Thermal Injury Severity	Major 1st / 2nd degree > 25%, 3rd degree, > 10%		Moderate 1st / 2nd degree 15 to 25%, 3rd degree, 2 to 10%		Minor 1st / 2nd degree < 15%, 3rd degree, < 2%	No thermal injury
F	Required Secondary Debris Wound Treatment	Immediate medical care for survival or penetrates skull		Less urgent surgery for survival	Simple medical care in treatment facility	Self-help treatment	

Figure 2. FLM severity scale components

associated with $P \leq .10$ of serious/lethal wound to a standing human) and the risk estimation distance (distance associated with $P \leq .001$ of human injury, considering posture, warning level, and terminal ballistic condition of warhead). Further, the DoD categories for *deceased* and *very serious* were combined into a single category representing *lethal*, on the basis of the JMEM description of serious/lethal wounds. All of the categories were color coded, with red for *lethal* and blue for *no injury*, for ease of interpretation.

Third, the medical and military literature was researched to clarify the individual damage effects mechanisms for FLM and to establish criteria for each of the five categories in the FLM wound severity scale.

Blast pressure impulse effects. Blast pressure impulse is the primary damage effect mechanism for FLM. Observations of full body translation caused by the pressure impulse of kinetic events in recent combat operations indicate a relationship between the distance the human body is propelled and the severity of the injury incurred (JTCG-ME, 2006).[2] Thus, the FLM wound severity ratings were linked to distance propelled to provide criteria for assessing impacts of blast pressure impulse (Row C in *Figure 2*).

- Humans propelled a distance of 10 or more feet suffer sufficient G-forces to produce very serious or lethal injury, corresponding to the FLM *lethal* category. The cause of death in these instances is cardiac arrest, severe neck or spinal injury, severe brain trauma, or traumatic amputation of a major limb (arm or leg).
- Humans propelled more than 5 but less than 10 feet suffer serious injury, corresponding to the FLM *severe* category. Injuries may include brain concussion, hemorrhaging of the brain and vital organs, severe ligament damage, and bone fractures.
- Humans propelled between 1 and 5 feet either suffer no injuries or are incapacitated for a short period, depending on the geometry of the impact. The injuries sustained correspond to the FLM *moderate*, *light*, or *no injury* categories.

The primary factor differentiating among these three categories is the portion of the body taking the brunt of the impact. A higher severity rating of *moderate* is assigned if the body falls head or face first or impacts another stationary object such as a wall or vehicle. In some rare cases, humans propelled less than 5 feet receive serious injury requiring extended medical attention (often associated with blunt trauma from impacting a hard surface or object).

Blast overpressure effects. Blast overpressure is a secondary human lethality mechanism for FLM. Blast overpressure produces a crushing effect on the human body, potentially causing severe injury to the lungs, ears, other organs, and soft tissue (Pennardt 2007). Blast overpressure lung injury creates contusions that cause hemorrhaging, swelling, and fluid accumulation, leading to labored and progressively less efficient breathing (De Lorenzo and Porter 1991). Additional symptoms include disturbances in consciousness, small strokelike symptoms, and bloody sputum. Use of cardiopulmonary resuscitation or mechanical respira-

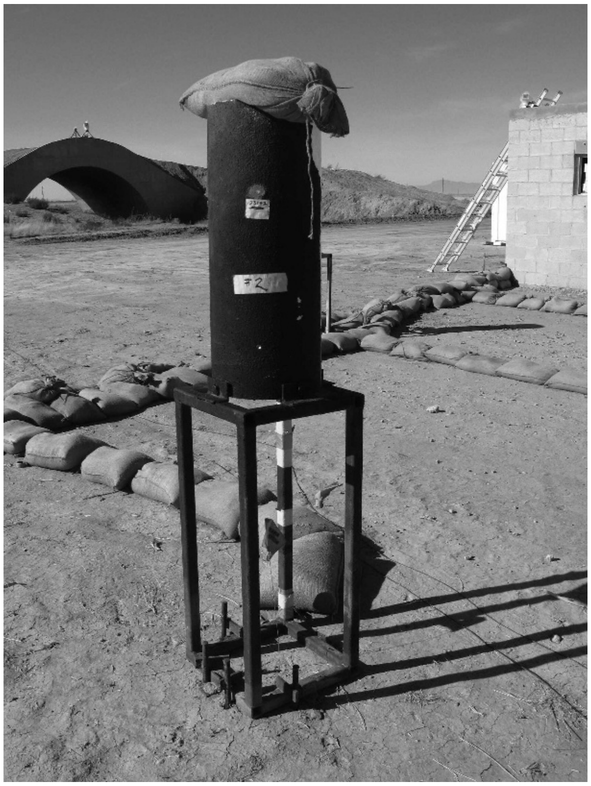


Figure 3. Blast test device

tors on individuals with blast overpressure lung injuries may release air bubbles into the bloodstream, which can cause severe injury or death if they reach the heart or brain.

Blast test devices are frequently used to gather blast overpressure data during weapons effects tests. A blast test device is a rigid cylindrical device about the size of a human torso that measures external pressure loading due to blast overpressure *(Figure 3)*. Measurements obtained from four pressure transducers evenly spaced around the circumference of the cylinder at midheight represent the pressure felt on the chest, right side, left side, and back of a human thorax. Each measurement is saved as a separate data file constituting a pressure-versus-time trace for a given location on the blast test device. Data are then entered into the INJURY 8.2 software program to predict lung injury from blast overpressure.[3]

The FLM wound severity ratings were correlated to two different types of blast overpressure estimates provided by INJURY 8.2 (Row D in *Figure 2*). Qualitative estimates provide labels for an easy-to-read designation of the severity of injury expected. Quantitative estimates provide probabilities associated with the degree of lung contusion expected, as characterized by the percentage of total lung surface area contused (directly related to lung hemorrhage).

Table 1. INJURY 8.2 output for blast overpressure damage

	INJURY 8.2 output (probability)			
Example	**Severe**	**Moderate**	**Slight/trace**	**No injury**
Case 1	0.00	0.01	0.16	0.84
Case 2	0.05	0.42	0.23	0.30
Case 3	0.40	0.40	0.15	0.05

In general, the FLM severity category associated with the highest INJURY 8.2 probability indicates the most likely severity of injury, following the JMEM 50-percent lethality criterion rule:

- Any category associated with an INJURY 8.2 probability greater than .50 was always selected as the final severity category (e.g., *no injury* for Case 1 in *Table 1*).
- In cases where none of the probabilities exceeded .50, the team located the largest probability and inspected the probability in the next most severe category. If the combined values produced a probability greater than .50, then the more severe category was selected. If the combined values failed to produce a probability greater than .50, then the category with the largest value was selected for the final rating (e.g., *moderate* for Case 2 in *Table 1*).
- In the case of a tie, the more severe category (e.g., *severe* for Case 3 in *Table 1*) was always selected.
- If the INJURY 8.2 *severe* category met the criteria described in the preceding paragraph, an FLM rating of *lethal* was always assigned if the probability was .50 or greater; otherwise, the FLM rating remained at *severe*.

Thermal effects. Thermal is a secondary human lethality mechanism for FLM. The severity of thermal injury or burns is characterized by (a) degree, based on the severity of the tissue damage that may extend to the underlying fat, muscle, or bone; and (b) amount of body surface area involved. Burn degree is designated as either *first-degree* (redness and swelling in the outermost layers of skin), *second-degree* (redness, swelling, and blistering, with damage extending beneath epidermis to deeper layers of skin), or *third-degree* (full-thickness burn that destroys the entire depth of skin, causing significant scarring).

Amount of body surface area involved is expressed in terms of the "rule of nines" used by health-care professionals for adult burn patients (Arizona Burn Center 2008). For the rule of nines, each arm with its hand included constitutes 9 percent of the body surface area; the front of each individual leg with its foot is 9 percent; the back of each individual leg with its foot is

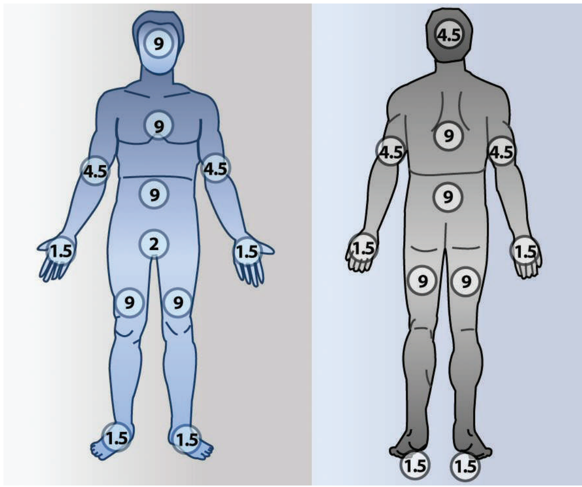


Figure 4. Weighted rule of nines

9 percent; the chest is 9 percent; the abdomen is 9 percent; the back is 9 percent; the buttocks are 9 percent; the face, back of the head, and neck are 9 percent; and the genital area (perineum) is 1 percent.

Burns involving the hands, feet, face, eyes, ears, or genitals are considered especially critical because the skin is thinner. To account for this fact, a weighting scheme was applied to the traditional rule of nines *(Figure 4)*. Whereas hands and feet are normally lumped in with their limbs, ratings were assigned for each limb and additional values for each hand or foot. Thus, maximum values of 9 percent were applied to each arm (4.5 percent for the front and 4.5 percent for the back), 3 percent to each hand (1.5 percent for the front and 1.5 percent for the back), 18 percent to each leg (9 percent for the front and 9 percent for the back), and 3 percent to each foot (1.5 percent for the top and 1.5 percent for the bottom). The genitals and the face were assigned double values—9 percent for the face alone and 2 percent for the genitals. Altogether, the weighting permitted an increase of up to 17.5 percent of the total body surface area burned if critical body parts were affected.

Taken together, the burn degree and percentage of body surface area affected are used to identify burn severity:

- Minor burns are (a) first- or second-degree burns covering less than 15 percent of an adult's body or (b) third-degree burns covering less than 2 percent body surface area. Minor burns, which may be treated at home or in a doctor's office, are linked to an FLM severity rating of *light* (Row E in *Figure 2*).
- Moderate burns are (a) first- or second-degree burns covering 15 to 25 percent of an adult's body or (b) third-degree burns covering 2 to 10 percent body surface area. Moderate burns, which should generally be treated at a hospital, are linked to FLM severity ratings of *severe* or *moderate*.
- Major burns are (a) first- or second-degree burns covering more than 25 percent of an adult's body or (b) third-degree burns covering more than 10 percent body surface area. These burns are the most serious and should be treated in the specialized burn unit of a hospital, correlating to an FLM severity rating of *lethal* (Thivierge 2008).

Secondary debris penetration. A tertiary human lethality mechanism for FLM is secondary debris, either from weapon components or objects in the area (e.g., vehicles, buildings, and other objects). Secondary debris may contribute to lethality within the target area; however, there are no reliable methods or models to predict the effects of secondary debris for all target environments. The criteria in Row F of *Figure 2* were developed for FLM severity ratings of secondary debris penetration.

Criteria for the *lethal* category were based on the following considerations. Wounds that penetrate the skull are usually immediately lethal. Very serious wounds that require immediate treatment include injuries that disturb consciousness, breathing, the airway, or circulation; major injuries to the head or torso; or major hemorrhaging. For example, major hemorrhaging is generally the imminent threat for most wounds involving the abdomen and chest because they house the vital organs (Owen-Smith 1981). Shock, which is considered very serious and requires immediate treatment, can arise from major muscle damage, especially when associated with a major fracture, severe burns, major hemorrhaging, multiple wounds, and pericardial injuries.

The remaining FLM severity categories represent less serious injuries. Injuries that require surgery or intensive medical care, but will not cause death if delayed, received FLM ratings of *severe* (e.g., weapon component debris penetrates the abdomen but does not affect any internal organs). Injuries that require medical care, but can be managed by simple treatment and dressing provided in a medical care facility, are categorized as *moderate* (e.g., abdominal debris that causes moderate bleeding). Injuries that can be treated by self-help are categorized as *light* (e.g., small surface wounds) (Owen-Smith 1981).

Finally, after researching the individual damage effects mechanisms for blast pressure impulse, blast overpressure, thermal, and secondary debris penetration, guidelines were established for estimating combined effects. The postdetonation condition of an individual is not always caused by only a single mechanism. Combinations of effects may cause multiple injuries that can lead to a higher severity rating than would be associated with any one injury by itself. The following factors were considered when addressing combined effects for mannequins with multiple injuries:

- amount of potential blood loss
- location of injuries, with the head, chest, abdomen, and genitals being the most vulnerable in the case of blunt trauma and penetration injuries
- head injuries combined with other injuries
- injuries affecting the airway or respiratory system
- injuries combined with major burns

For the FLM program, the overall severity level was assigned on the basis of the highest severity level of the individual sustained injuries. For example, if blast pressure impulse was *moderate* and thermal injury was *severe*, the overall severity rating was *severe*. If there were multiple injuries at the same severity level, that severity level was assigned. The question of whether multiple injuries at the same severity level translate into a higher overall severity level has not been adequately resolved.

Criteria application

After each FLM test event, photographs of the test site were taken for comparison to pretest setup, displacement of the mannequins from their original positions measured, and extent of any damage (e.g., burns, punctures, missing limbs) thoroughly documented to enable application of the preceding criteria. The criteria for blast pressure impulse, thermal damage, and combination effects resulting from these two damage effects mechanisms were successfully applied. Data for blast overpressure were collected but not used because of data corruption. The collected data did not meet the quality parameters for input into INJURY 8.2. Secondary debris effects were documented for completeness but not included in the scoring because of the unpredictable and unrepeatable nature of secondary debris.

Of the two damage effects mechanisms assessed, thermal presented the most challenges. Application of the criteria for blast pressure impulse was relatively straightforward because it primarily involved simple measurement of the distance each mannequin was propelled. Assessing thermal damage to the skin on mannequins entailed determining the degree of the burn, the percentage of body surface area affected, and the specific body parts burned. FLM thermal injuries may result from fire associated with the explosion or from the tungsten fill of the warhead. Tungsten can potentially cause significant thermal injury when it impacts the skin because of the high temperature and velocity of the tungsten at the time of impact. The chief difficulty was translating guidelines developed for human burns to evidence obtained from mannequins, which cannot exhibit the same thermal effects as humans because they do not have skin.

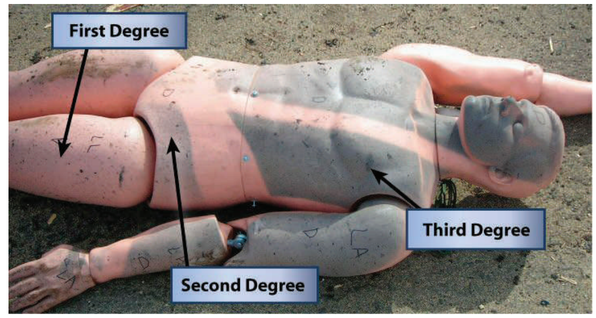


Figure 5. Burns in mannequins

During the initial static tests, primary reliance was on observations of thermal effects in commercially butchered pigs to identify applicable thermal guidelines for mannequins. These guidelines were based on the professional expertise of Reddoch Williams, M.D.[4] By his estimations, if the skin was covered with a significant layer of tungsten (i.e., no skin visible underneath), the burn was classified as third degree. A moderate layer of tungsten (i.e., skin barely visible underneath) led to a classification of second-degree burns. A light dusting of tungsten (i.e., skin clearly visible underneath) was classified as first-degree burns.

When applied to mannequins, these guidelines translated into third-degree burns if the surface of a mannequin was charred with the outer layer of rubber cracking. When the surface of the mannequin presented a gritty gray-to-black appearance to the point where the discoloration and grit could not be removed by wiping, the mannequin was characterized as having second-degree burns. When the surface of the mannequin presented a gritty gray-to-black appearance, but the discoloration and grit could be removed by wiping, the mannequin was characterized as having first-degree burns. *Figure 5* provides examples for each degree of burn.

When assigning thermal severity ratings, the severity level was increased if the injuries were associated with the respiratory system or eyes. Burns to the faces of mannequins were recorded and assessed for airway burns by degree of charring to the mouth and nose and for burns to the eyes by degree of charring or tungsten on the eye area. There was no method to directly determine thermal inhalation and vascular injuries to mannequins.

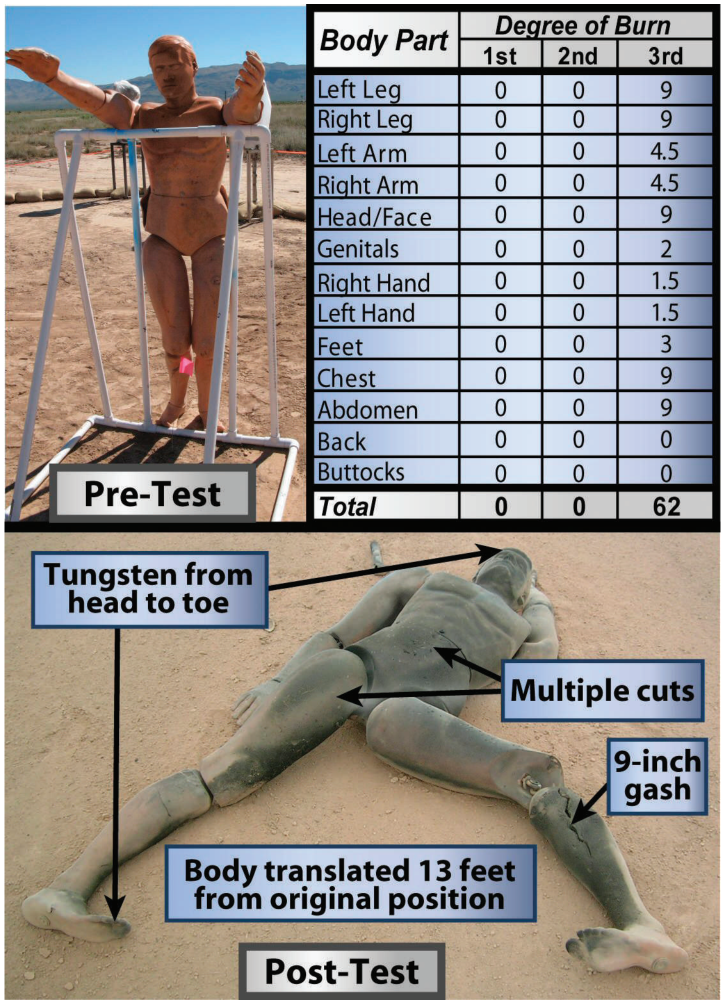


Figure 6. Sample validation slide

Criteria validation

Two methods were used for preliminary validation of the criteria and assessment procedures. Both techniques provided considerable support for the validity of the FLM severity criteria.

Logistic regression analysis. First, the results of a logistic regression model provided initial confirmation for the legitimacy of the lethality evaluation criteria. The logistic regression analysis was conducted to provide supporting evidence for the lethality/nonlethality cutoff distance the team estimated based on visual observation of the outcome for each mannequin (i.e., criteria were applied to derive a lethality rating for each mannequin and then the results were visually inspected to determine the distance representing the cutoff between lethal and nonlethal). For a more objective, mathematical approach, a logistic regression analysis was completed—a statistical technique was used to predict lethality, based on distance from the impact. The analysis included 64 FLM test articles (32 designated as lethal and 32 designated as nonlethal). The mathematical model obtained from the logistic regression correctly categorized 59 of 64 cases (92 percent). Further, the predicted cutoff distance between lethality and nonlethality exactly matched the distance determined on the basis of visual inspection alone. The orderliness of this outcome lends credence to the validity of the underlying procedures used to determine lethality.

Independent application of criteria. Second, an independent verification of criteria application was conducted to address validity. Two people with no previous connection to the FLM program were asked to provide independent lethality ratings for a sample of 20 mannequins (12 *lethal*, 1 *severe*, 2 *moderate*, 2 *light*, and 3 *no injury*). The independent assessors had several days to review the FLM severity criteria and descriptions before individually completing the validation task. A briefing provided background information on the test setup, the targets, and the collateral concerns; depictions of the pre- and post-test layouts; and a separate slide for each mannequin with pre- and post-test photographs, descriptions of injuries, distance the mannequin was propelled, and thermal injury data (degree of burn, percentage of body surface area affected, and body parts burned) *(Figure 6)*. The independent assessors provided ratings for blast pressure impulse, thermal effects, and overall lethality as well as descriptions of their rationale for each rating.

The *kappa* statistic for inter-rater agreement was .76, a value that indicates "excellent" agreement among the three sets of ratings for the sample of 20 mannequins (Fliess 1981). When disagreements occurred, the independent assessors tended to assign more severe ratings. In particular, the independent assessors experienced difficulty deciding whether to assign *moderate*, *light*, or *no injury* ratings when the body was propelled 1 to 5 feet. They had trouble evaluating the critical deciding factor of impact geometry for these cases, in part because it was difficult to discern from photos alone whether the body had been thrown face first or simply fell forward after the blast.

Future plans

The criteria provided a defendable and repeatable approach to determine lethality for nonfragmenting, precision-guided weapons such as FLM. The FLM operational manager at the U.S. Central Command was able to use the final report, delivered in May 2008, to provide a military utility recommendation for the FLM weapon and develop future plans. As additional data from program office continuation testing and operational weapon use become available, it is expected that the methodology will be augmented. The methodology presented in this article represents the preliminary development, application, and validation of procedures and guidelines. With additional research and data, the methodology can easily be expanded to provide robust and repeatable procedures. Several focus areas for the future include:

- development and verification of more concrete guidelines for thermal effects
- instrumentation with blast test devices that meet specifications for use with INJURY 8.2, with attention to proper setup (e.g., anchoring to prevent tip over during blast; coating of pressure sensors and wires for sufficient protection against flames, heat, tungsten, particles, and light)
- use of autopsy reports for definitive determination of lethality
- more robust validation of lethality determinations (including initial decisions regarding damage effects and criteria application)
- overall enhancement of criteria to boost specificity and eliminate as much subjectivity as possible
- verification of the full model with all four damage effects mechanisms and combination effects ❑

Amanda Humphrey is an assessment analyst at SAIC, Inc., in Albuquerque, New Mexico. She has a bachelor of science degree in biological anthropology from the University of New Mexico. Ms. Humphrey led the effort to develop and apply the lethality criteria. E-mail: amanda.r.humphry@saic.com

Dr. Judi See is a human factors analyst at SAIC, Inc., in Albuquerque, New Mexico. She holds doctoral and master's degrees in experimental psychology from the University of Cincinnati and a bachelor of arts degree in psychology from Urbana University. Dr. See led the FLM validation activities and provided technical and quality assurance oversight during project completion.
E-mail: judi.e.see@saic.com

David Faulkner, the founder and president of the DMPI defense consulting firm, has more than 20 years of experience in DoD senior leadership and targeting positions. He has a master of arts degree in Near Eastern studies from the University of Arizona and a bachelor of science degree in physics from Appalachian State University. Mr. Faulkner applied his expertise in munitions effectiveness and employment to assist in development of the FLM lethality criteria.
E-mail: david.i.faulkner@dmpi-llc.com

Endnotes

[1]The Air Force Research Laboratory Munitions Directorate at Eglin Air Force Base, Florida, developed both FLM technologies.

[2]Based on extensive operational observations of lethality induced by full body translation, per U.S. Central Command Director of Joint Targeting and Assessment.

[3]INJURY 8.2 is a medical research product developed by the U.S. Army Medical Research and Materiel Command's Military Operational Medicine Research Program.

[4]Reddoch Williams, M.D. (E-mail: Reddoch@aol.com). Dr. Williams received formal training at Emory University and has experience treating human burn wounds. At the time of the FLM test, he served as a Flight Surgeon, Air Force Reserve, attached to the Command Surgeon's office at Headquarters Air Force Special Operations Command.

References

Arizona Burn Center at Maricopa Medical Center. 2008. Guidelines for Early Care and Transfer of Burn Patients, 2008, Available at http://www.saems.net/Downloads/50152_Burn%20Center2col_f.pdf. Accessed October 8, 2008.

Joint Technical Coordinating Group for Munitions Effectiveness (JTCG-ME). 2006. "JMEM Air-to-Surface Weaponeering System (JAWS)," CD-ROM, V2.4, 61A1-3-11-CD, October 2006 (classified).

Chairman of the Joint Chiefs of Staff. 2006. "Health Service Support," Joint Publication 4-02, October 2006. Washington DC: Joint Chiefs of Staff.

Defense Intelligence Agency (DIA). 2003. "Battle Damage Assessment (BDA) Quick Guide," DI-2820-4-03, February 2003. Washington DC: Department of Defense Intelligence Production Program.

De Lorenzo, R. A. and Porter, R. S. 1991. "Care of Blast and Burn Casualties." In *Tactical Emergency Care: Military and Operational Out-of-Hospital Medicine*. Upper Saddle River, NJ: Prentice Hall.

Fleiss, J. L. 1981. "The Measurement of Interrater Agreement." In *Statistical Methods for Rates and Proportions,* Second Edition. New York: John Wiley & Sons, pp. 212–304.

Owen-Smith, M. S. 1981. "Primary Treatment of the Wounded." In *High Velocity Missile Wounds*. London: Edward Arnold, pp. 43–145.

Pennardt, A. December 2007. "Blast Injuries," *eMedicine*. Available at http://www.emedicine.com/emerg/TOPIC63.HTM. Accessed October 8, 2008.

Thivierge, B. 2008. "Burns," In *Medical Encyclopedia*. Farmington Hills, MI: The Gale Group.

Acknowledgment

The authors wish to acknowledge the support and contributions of the entire FLM team—the 681st Armament Systems Squadron Program Office, U.S. Central Command, Trusted Agent, SAIC, and DMPI—during completion of this paper.

ITEA Journal 2008; 29: 420–428

Methodology for Estimation of Operational Availability as Applied to Military Systems

Gary A. Pryor

U.S. Army Training and Doctrine Command, Fort Leonard Wood, Missouri

Operational availability (Ao) is an important consideration during the evaluation of system effectiveness and sustainability. Ao is sometimes specified as an attribute within military requirements documents, at the discretion of the proponent. Recently, however, the Chairman of the Joint Chiefs of Staff Manual (CJCSM) 3170C mandated the establishment of materiel availability as a sustainment Key Performance Parameter (KPP). KPPs are defined to be those attributes of a system that are considered critical or essential to the development of an effective military capability. However, test and evaluation of availability is problematic because it is highly dependent on the response and delay times associated with the maintenance and logistics support structures, which are not normally in place prior to fielding. This often leads to the evaluation of Ao via analysis or simulation—measuring the systems reliability and maintainability characteristics—and applying an estimate of the effect of the logistic support system. This article provides a brief background of Ao as well as a comparison of several methodologies for measuring and estimating Ao. Although KPPs are required by CJCSM 3170C to be testable, it is clear that it is necessary in most cases to measure the inherent reliability and maintainability of an item and to apply modeling and/or simulation techniques to evaluate the actual Ao. The equations and methodologies in the article describe the most common of those techniques, as well as their limitations and shortcomings.

Key words: administrative and logistics, corrective and preventive, delay time, downtime, failure frequency, maintenance time, material availability, total active time, uptime.

Operational availability (Ao) is widely used as a readiness-related objective in the specification of requirements for military systems. Its definition can be found in a number of military sources and is fairly consistently represented. (See Sidebar 1)

In general terms, Ao is the proportion of time a system is either operating or is capable of operating (called *uptime*) while being used in a specific manner in a typical maintenance and supply environment. In other words, Ao is the ratio of uptime to *total time*, or more correctly, *total active time*. *Active time* refers to a period in the item's life in which it is being utilized in the environment in which it is intended to perform its primary function. The counterpart, *inactive time*, refers to the period of time in which an item is not being utilized, such as time that it is in storage or being refurbished, or even while undergoing long-range transport (e.g., by sea) or being utilized as a spare or float. Ao is only applicable to active portions of an item's life.

The basic mathematical definition of Ao is

$$\text{Ao} = \text{Uptime/Total (Active) Time}$$

$$\text{(note: active time is assumed from here on)}$$

and since an item can be either operable (uptime) or inoperable (*downtime*), mutually exclusively, *Ao* is acceptably defined as shown in Equation 1.

$$Ao = \frac{Uptime}{Uptime + Downtime} \tag{1}$$

Though the definition of *Ao* is fairly straightforward, the calculation of *Ao* can vary somewhat depending on the definitions of uptime and downtime. The inclusion of certain "operational states" can be legitimately expressed as either uptime or downtime; and as long as the definition is clear there can be various "correct" interpretations.

Sidebar 1: Definitions of Operational Availability (Ao)

At its web site, the Defense Acquisition University (DAU) Glossary of Defense Acquisition Acronyms and Terms defines Ao as "The degree (expressed as a decimal between 0 and 1, or the percentage equivalent) to which one can expect a piece of equipment or weapon system to work properly when it is required, that is, the percent of time the equipment or weapon system is available for use. [Ao] represents system "uptime" and considers the effect of reliability, maintainability, and mean logistics delay time. [Ao] may be calculated by …" [calculation will be covered later in this paper].

(DAU, 2005, website accessed September 12, 2008)

DA PAM 73-1 defines Availability as "the probability that a piece of equipment is in an operable and committable state at a given (random) point in time. Repair, maintenance, and administrative and logistics downtime [ALDT] are the most common causes of equipment non-availability for use. A system's availability is a function of its reliability and maintainability."

(DA PAM 73-1 2003, page 213)

DA PAM 70-3 gives a similar definition "A readiness parameter that is a measure of the degree to which a system is either operating or is capable of operating at any time when used in its typical operational and support environment. Normally, it is most sensitive to the responsiveness of the logistics support system and the system's op-tempo. Because the [materiel developer] has such a principal role in the factors that dominate reliability and maintainability (R&M) and such limited control over the factors that dominate availability, this section is focused on R&M."

(DA PAM 70-3 2008, page 87)

Uptime and downtime

Downtime can be defined as the time during which an item is incapable of performing its primary functions. The most common cause of downtime is a reliability failure and the subsequent maintenance and logistics delays associated with restoring the item to an operational state. The next most common cause is the performance of scheduled maintenance or other maintenance *not necessarily associated with a critical reliability failure* during the performance of which the system cannot be operated.

There are other possible system states, such as relocation time, that are sometimes defined as downtime, but the inclusion of these states as downtime tends to confuse the issue. *Relocation time* is defined as the time spent transporting an item from one place to another in an inert or packaged state, i.e., moving a bulldozer on a trailer from one jobsite to another. True, the bulldozer cannot operate while on the back of a trailer, but neither has the bulldozer lost the capability to perform its primary function—the bulldozer could be used if called upon and unloaded. Categorizing relocation time as downtime unfairly penalizes the item and confuses the issue. If the bulldozer has a low Ao, is it because it is constantly failing or is it because it is frequently transported from jobsite to jobsite? There are other similar categories of time that inherently confuse the issue: setup and teardown time, preventive maintenance checks and services, even re-fueling time can fit in this category. It's much cleaner if downtime is defined to include only that time the system is not functional due to either essential preventive or corrective maintenance, or administrative or logistics delays associated with the repair of reliability failures.

Uptime can be defined as the time an item is capable of performing its primary function if called upon to do so. And, since uptime and downtime are mutually exclusive, uptime is the amount of time left after subtracting downtime from the total available time. (Although technically true, using this approach for estimating *Ao* can lead to erroneous estimates of *Ao*, as will be discussed later.)

In order to properly estimate *Ao*, it is also necessary to further refine the definition of uptime—specifically in such a way as to define the duty cycle of the item. How much of the time is the item expected to operate? This is accomplished by dividing uptime into useful subcategories. The most common of these subcategories are (a) standby time (ST), time in

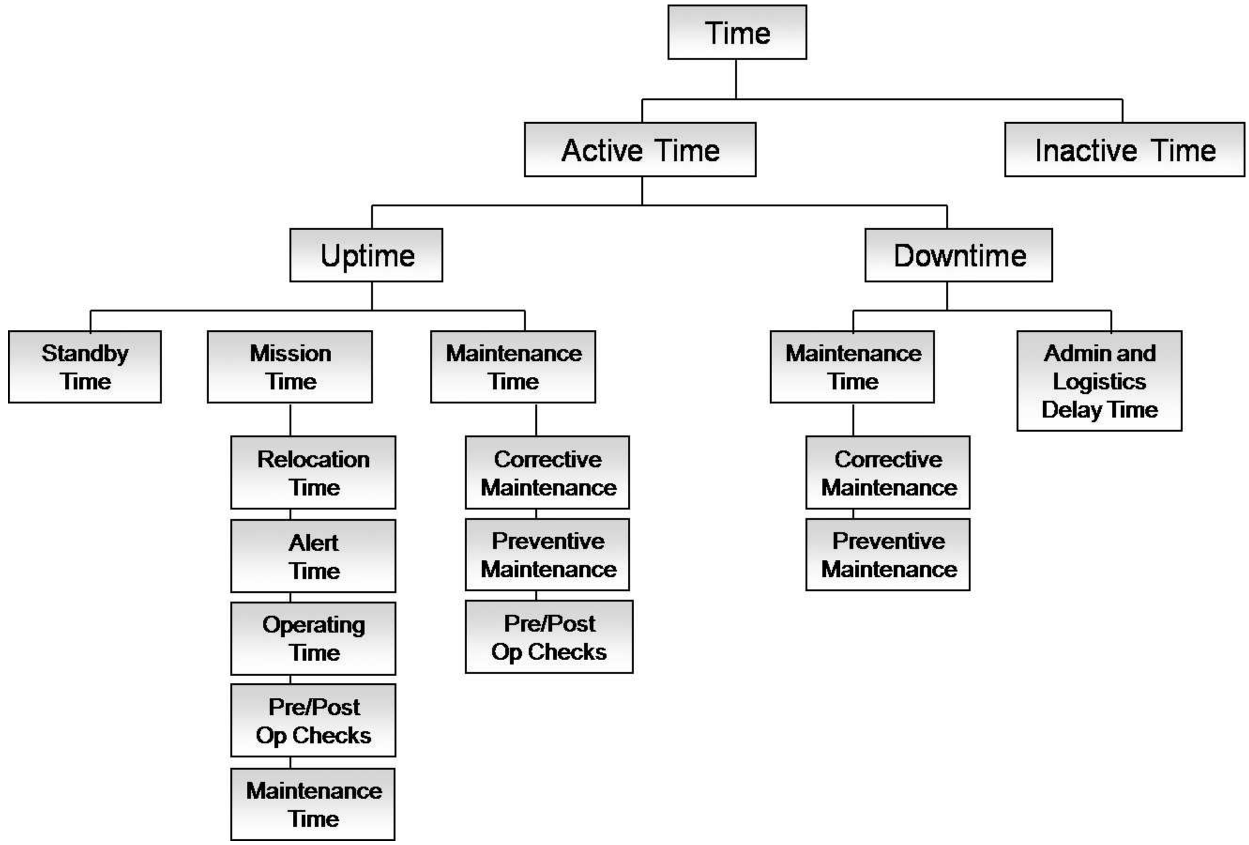


Figure 1. Categories of time for use in defining Ao

which the system is not performing and is not dedicated to performing its primary function; (b) mission time, time in which the system is dedicated to performing its primary function (sometimes incorrectly used interchangeably with operating time); (c) relocation time (defined above); (d) maintenance time, time during which the item is undergoing some scheduled or unscheduled maintenance or pre-/postoperational checks and services; and (e) operating time (OT), time during which an item is actively performing one or more of its primary functions.

There are various ways of defining and organizing these time categories. For example, relocation time can occur either between missions or during a mission, or perhaps both. Maintenance time can occur prior to, after, or during a mission, or any combination thereof.

One common representation of time is shown in *Figure 1* below.

Of all the various time categories, the ones that have the greatest influence on Ao are the OT (since the more the system operates, the more often failures occur), the corrective and preventive maintenance times, and the administrative and logistics delay time (ALDT).

Measurement and evaluation of operational availability

In an operationally realistic test environment, when the actual maintenance and logistics support structures are utilized, the operational availability can be measured by summing the uptime (usually inclusive of OT and ST) and the downtime (usually inclusive of total corrective maintenance [TCM] time, total preventive maintenance [TPM] time, and total ALDT [TALDT]) as shown in Equation 2.

$$Ao = \frac{OT + ST}{OT + ST + TPM + TCM + TALDT} \qquad (2)$$

As indicated by DA PAM 73-1, normally the TALDT will greatly outweigh the other factors of downtime. Unfortunately, that is also the hardest parameter to accurately measure during testing. While an item is under development and until well after fielding, the supply system is not fully stocked with spare parts for the item. For test purposes, the contractor normally stockpiles the necessary spare parts so the delay is fairly short, or the opposite occurs and longer-than-normal delays are incurred because the spare parts do not yet exist in sufficient numbers and testing has exhausted the supply.

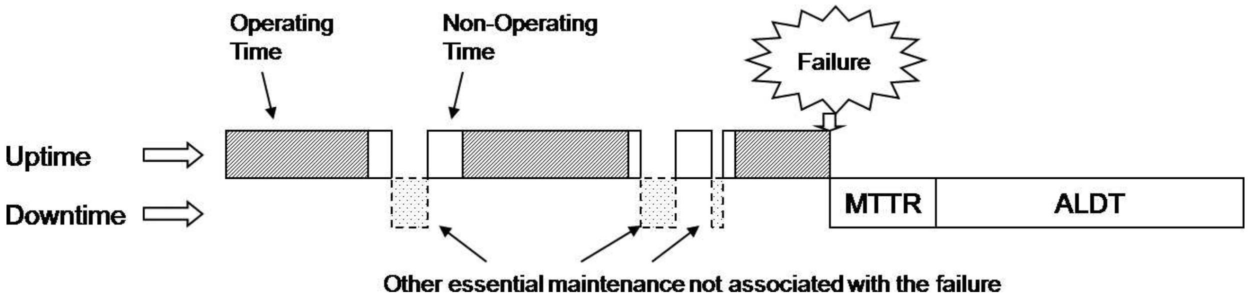


Figure 2. Uptime and downtime distribution

DOD 3235.1-H (1982) states, "One significant problem associated with determining Ao is that it becomes costly and time-consuming to define the various parameters. Defining ALDT and TPM under combat conditions is not feasible in most cases. Nevertheless, the operational availability expression does provide an accepted technique of relating standard reliability and maintainability elements into an effectiveness-oriented parameter."

And, because it is necessary to estimate the *Ao* during the establishment of system requirements, long before any test data is available, an analytical technique is needed to estimate or calculate the expected *Ao*.

Estimation of operational availability

There are three commonly used approaches to the estimation/calculation of *Ao*. The first, using Equation 2, is commonly misused and should be avoided. The results of the other two approaches track fairly well in most cases, but can significantly diverge under certain conditions.

Approach 1: Equation 2

$$Ao = \frac{OT + ST}{OT + ST + TPM + TCM + TALDT}$$

The methodology involved in using Equation 2 to estimate *Ao* for a given calendar time (usually a calendar year) is to

- project the annual *OT* and *ST* (uptime);
- use the projected annual OT and estimated reliability to calculate the annual number of failures expected and thus an annual amount of total corrective maintenance (TCM) time and TALDT (*TCM* = annual *OT*/*MTBF** mean corrective maintenance time (*MCMT*); *TALDT* = annual *OT*/*MTBF* * *ALDT*; where *MTBF* is mean time between failures and *MCMT* is);
- estimate the annual total preventive maintenance (TPM) time;
- sum all uptime (*OT* + *ST*) and downtime (*TCM* + *TPM* + *TALDT*) and plug into Equation 2 to calculate *Ao*.

This approach will almost always result in an incorrect denominator. Because we are calculating *Ao* on an annual basis, the denominator by default has to be one year (8,760 hours). However, in practice one will usually get a denominator either much *larger* or *smaller* than 8,760 hours. Denominators over 8,760 often result because the analyst assumes from the start that *OT* and *ST* together will equal 8,760; the projected downtime is then added to 8,760 to result in an incorrect denominator. When the analyst uses only *OT* to estimate uptime and to project downtime, the resultant denominator can be either larger or smaller than 8,760 hours; only a tiny fraction of the time would the analyst inadvertently obtain an 8,760-hour denominator. Even if that were the case, it's still not the best way to estimate *Ao*.

Approach 2: Subtracting expected downtime from total time

Approach 2 overcomes the denominator issue by initially setting it to the desired value (e.g., 8,760 hours).

Then, because by definition $Uptime = Total\ Time - Downtime$,

$$Ao = (Total\ Time - Downtime)/Total\ Time.$$

In fact, a commonly used equation can be easily derived, where *TT* is the total time:

$$Ao = \frac{TT - Downtime}{TT}$$

$$Downtime = TPM + TCM + TALDT$$

$$(TCM + TALDT) = \#\text{ of failures} * (MCMT + ALDT)$$

$$(TCM + TALDT) = \frac{OT}{MTBF} * (MCMT + ALDT)$$

$$Ao = \frac{TT - TPM - \frac{OT*(MCMT + ALDT)}{MTBF}}{TT \cdot MTBF}.$$

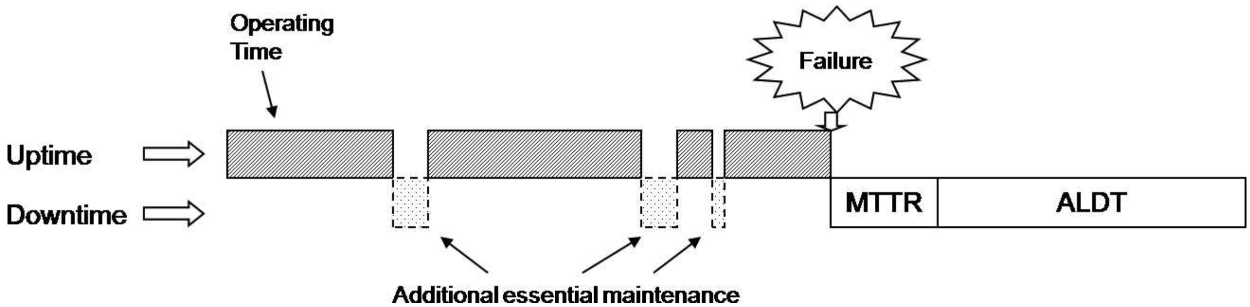


Figure 3. Uptime and downtime for continuously operating system

Dividing numerator and denominator by TT:

$$Ao = 1 - \frac{OT^{*}(MCMT + ALDT)}{TT^{*}MTBF} - \frac{TPM}{TT}.$$

Or, because TPM is normally a very small contributor to downtime, the equation is sometimes truncated and represented as

$$Ao = 1 - \frac{OT}{TT} * \frac{(MCMT + ALDT)}{MTBF}. \qquad (3)$$

The derivation is simple and straightforward, but there are problems with this methodology. In certain cases, Equation 3 can result in a negative number! This occurs when the sum of the estimated *OT* and the projected downtime exceeds *TT*. This will occur if the estimated *OT* is a very high percentage of *TT*, or when combinations of reliability and restore times result in significant amounts of downtime. The use of this equation is widespread but can give erroneous results (comparisons of Equation 3 with the results of the third [and preferred] approach will be shown later in this article).

Approach 3: Availability based on failure frequency

Figure 2 shows a normal operate–failure–repair cycle for an item.

The failure annotated in *Figure 2* indicates the occurrence of a *critical* failure—a failure resulting in an item incapable of performing its primary function. Critical failures result in downtime for corrective maintenance as well as administrative and logistics delays. The repair of other *noncritical* failures as well as preventive maintenance is annotated in *Figure 2* as "Other essential maintenance not associated with the failure." Theoretically, noncritical failures may also result in ALDT; however, by their nature the repair of noncritical failures can be deferred until parts arrive—eliminating most if not all ALDT from consideration.

As can be seen in *Figure 2*, the item goes through a period of OT and non-OT, with occasional downtime for miscellaneous essential maintenance, until a critical failure occurs. At that time, the system must be restored to an operational state and, therefore, experiences downtime due to corrective maintenance and administrative and logistics delays. This cycle is continually repeated. Thus, the *Ao* can be represented by a single failure cycle. Ignoring the other essential maintenance (including preventive) for the moment:

$$Uptime = \text{item's } MTBF + \text{non} - OT = MCTBF,$$
$$Downtime = MCMT + ALDT,$$

where *MCTBF* is the mean calendar time between failures.

$$Ao = \frac{MCTBF}{MCTBF + MCMT + ALDT} \qquad (4)$$

The definition of *Ao* in DAU (2008; Reference 1) includes the following: "Ao may be calculated by dividing Mean Time Between Maintenance by the sum of the Mean Time Between Maintenance, Mean Maintenance Time, and Mean Logistics Delay Time (MLDT), that is, Ao = MTBM/(MTBM + MMT + MLDT)."

DOD 3235.1H (1982) contains the same definition, with an important distinction (the note is underlined in the referenced text). "Ao = MTBM/(MTBM + MDT) <u>Note that the above definition assumes that standby time is zero</u>."

Since the majority of downtime is usually associated with unscheduled maintenance, the DAU (2008) expression is often simplified by considering only the unscheduled portion, and of that, only the portion related to correcting failures. Thus, the Reference 1 expression of *Ao* can be reduced to Equation 4, with an important caution. The numerator of Equation 4 is often expressed using the parameter *MTBF* (without the "calendar" time distinction). This leads to the possibility of mistakenly dismissing the standby time as unimportant. This can be a major mistake—in reality, it is the mean *calendar* time between failures that determines the frequency at which downtime occurs.

The *MCTBF* is dependent on the item's *MTBF* and the duty cycle. The item's inherent reliability (*MTBF*)

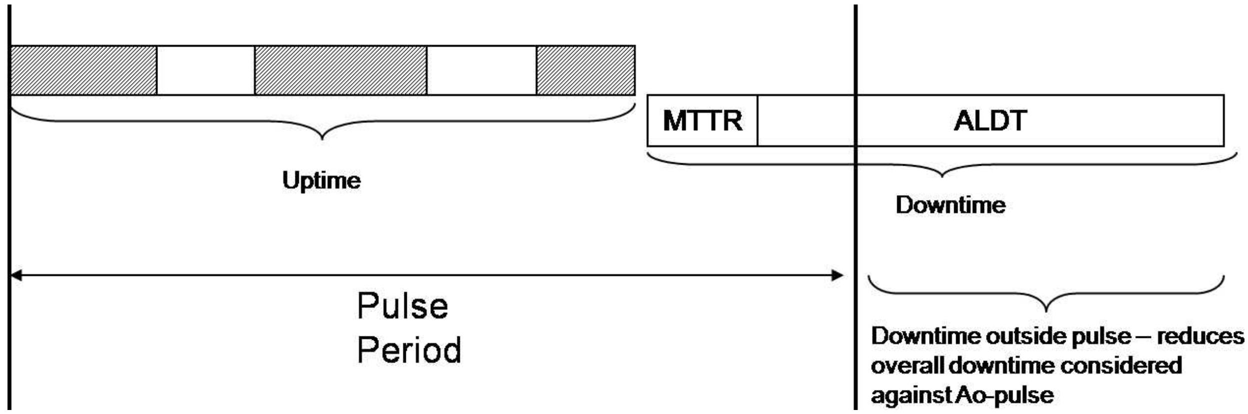


Figure 4. Uptime and downtime for pulse Ao *calculation*

is converted to calendar hours by dividing by the duty cycle, or operating rate (*OPR*). For continually operating systems, *MTBF* = *MCTBF*, but for all others, the duty cycle must be considered. For example, a system with an *MTBF* of 100 operating hours and that operates 12 hours per day (half the time; OPR = 50%) will average 200 calendar hours between failures. In summary, *MCTBF* = *MTBF*/*OPR*.

The additional essential maintenance was excluded from the above discussion because, when included, the estimate of *Ao* will vary somewhat depending on the usage rate and the amount of additional essential maintenance. In *Figure 2*, the additional essential maintenance lowers the uptime and is subtracted from the numerator but does not change the denominator (the calendar time between failures remains the same). This is because in *Figure 2*, there is ample time to perform the additional essential maintenance between usages.

Figure 3 shows another case that represents a continuously operating system.

For a continuously operating system, there is no planned period of nonusage. The *MCTBF* equals the *MTBF* equals the uptime. However, the denominator (which previously was *MCTBF* + *TCM* + *TALDT*) is increased by the amount of additional essential maintenance. It takes longer for the failure to occur, but in the process adds *downtime*, unlike the ST, which increases the *uptime*. Therefore, depending on the situation, the formula for calculation of *Ao* with additional essential maintenance varies—the presentation of other possible alternatives will be discussed in a follow-up to this article.

Choosing an appropriate OPR

The conversion of *MTBF* to *MCTBF* has already been discussed, but it is important to distinguish between and choose the right *OPR* that corresponds to the situation you are trying to represent.

If a system is used intensively part of the year and then used little or not at all the remainder of the year, using the annual *OPR* to estimate *Ao* represents neither the intensive usage period nor the nonintensive period. It might be close to a weighted average, but does not mean much to the casual observer. In a case like this example, the annual *OPR* would not be a good choice. It would be more appropriate to focus on the intensive usage; or calculate parameters for both intensive and nonintensive usage periods. However, if the intensive usage periods are relatively short and spread throughout the year, it is probably acceptable to use the annual *OPR*. The analyst will have to make sure to choose a suitable *OPR*.

Another similar consideration is the calculation of *Ao* for short, intensive usage periods that are followed by low or nonusage periods. This is sometimes referred to as *pulse Ao*. These cases are also very different. *Figure 4* represents a typical pulse *Ao* usage period.

The pulse *Ao* will always be greater than or equal to the steady-state *Ao* because some of the downtime that was induced during the pulse period will extend outside of the pulse period; therefore, it is not counted against the pulse *Ao*. The difference between pulse *Ao* and steady-state *Ao* varies depending on the reliability relative to the pulse period, the relative duration of the downtime, and, of course, the planned usage during the pulse period.

If the *MTBF* is significantly greater than the length of the pulse, there is a high chance of completing the pulse without a failure. Then, a failure will occur only for a small percentage of missions, and only that small percentage of missions will experience any downtime during or extending beyond the pulse. For this case there will not be much difference between the pulse *Ao* and the steady-state *Ao*.

If the *MTBF* is such that there is a good chance of experiencing one or more failures during the pulse and if the average downtime is also high (relative to the pulse), then a significant portion of downtime can be expected to extend beyond the pulse period, and the pulse *Ao* will differ significantly from the long-term *Ao*.

For situations where many failure–repair cycles occur during the pulse period (i.e., very long pulses or low *MTBF* combined with low downtimes), many failures and repairs can occur during the pulse. In that case, since there are multiple failures and repairs during the pulse, most of the associated downtime occurs during the pulse, and only downtime from the last failure can extend beyond the pulse. In this case, there will be some, but perhaps not significant, difference between the pulse *Ao* and steady-state *Ao*.

There is a methodology developed for consideration and estimation of pulse *Ao*. However, the details and discussion are outside the scope of this article and will be discussed in a follow-up article.

Estimation of operational availability using Approach 3

The remainder of this article will develop equations for calculating steady-state *Ao*, including the consideration of additional essential maintenance, and a separate equation for continuously operating systems. Now, please refer back to *Figure 2*.

For now we will assume that there is enough standby time to accomplish the additional essential maintenance. Thus,

$$Uptime = MTBF/OPR - (\text{amount of additional essential maintenance}).$$

The decision was made to represent the amount of additional essential maintenance in terms of a clock-hour maintenance ratio—the amount of additional essential maintenance is dependent on the amount of operating time. Preventive maintenance is also usually prescribed in terms of both calendar time and usage. Therefore, it is convenient to express our amount of additional essential maintenance as a function of operating time. The clock-hour maintenance ratio for additional essential maintenance (CMR_{ESS}) is expressed in terms of maintenance clock-hours per operating hours. For example, if an item normally operates for 500 hours per month and requires 5 clock-hours of additional essential maintenance downtime (in addition to that for repairs of critical failures), the CMR_{ESS} is equal to 5 clock-hours of maintenance/500 hours of operation = 0.01 maintenance clock-hours per operating hour. Our *OT* during a single failure cycle is equal to *MTBF*. Therefore,

$$Uptime = MTBF/OPR - MTBF * CMR_{ESS}$$

and

$$Downtime = MCMT + ALDT + MTBF * CMR_{ESS}.$$

Ao is calculated as the ratio of uptime over total time:

$$\begin{aligned} Ao &= \frac{Uptime}{Uptime + Downtime} \\ &= [MTBF/OPR - (MTBF * CMR_{ESS})] \\ &\quad \div [MTBF/OPR - (MTBF * CMR_{ESS}) \\ &\quad + MCMT + ALDT + (MTBF * CMR_{ESS})] \\ &= \frac{MTBF/OPR - (MTBF * CMR_{ESS})}{MTBF/OPR + MCMT + ALDT}. \end{aligned}$$

Now dividing numerator and denominator by MTBF and OPR:

$$\frac{1 - (OPR * CMR_{ESS})}{1 + OPR*[(MCMT + ALDT)/MTBF]}. \quad (5)$$

Note the uptime was decreased by subtracting the amount of additional essential maintenance. As alluded to earlier, this maintenance time cannot exceed the time remaining after daily operations and has to be completed during the available standby time. We can define an upper limit on the additional essential maintenance, not to exceed the available standby time.

The additional essential maintenance ($MTBF * CMR_{ESS}$) must be less than the available non-OT:

$$\begin{aligned} \text{Available non} - OT &= MCTBF - MTBF \\ &= MTBF/OPR - MTBF \end{aligned}$$

$$\text{Downtime for addt'l essential maint} = MTBF * CMR_{ESS}$$

Additional essential maintenance downtime must be less than non-OT:

$$\text{Addt'l essential maint time} \le \text{non} - OT$$

$$MTBF * CMR_{ESS} \le (MTBF/OPR - MTBF).$$

Dividing both sides by MTBF,

$$CMR_{ESS} \le 1/OPR - 1.$$

So, if $CMR_{ESS} - (1 - OPR)/OPR \le 0$, then Equation 5 can be used.

Now, if the additional maintenance time exceeds the available standby time, we cannot use Equation 5. Therefore, another equation is needed. *Figure 3*, described as the failure cycle description for a continuously operating system, also describes the case where the additional essential maintenance time exceeds the standby time.

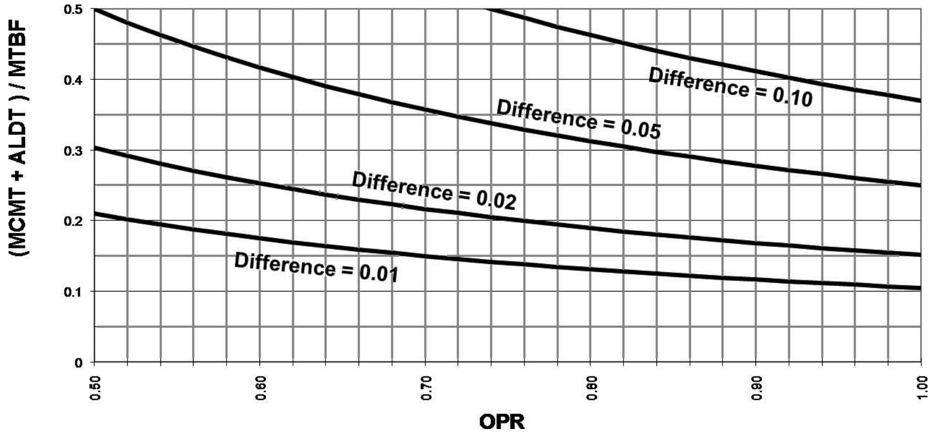


Figure 5. Constant difference between Equations 3 and 6 as a function of reliability, maintainability, logistics delay, and operating rate

This is because, as is true for the continuously operating item, there is no excess non-operating or standby time. What was previously available standby time is now used for additional essential maintenance. Since

$$Uptime = MTBF \text{ and } Downtime = MCMT + ALDT + MTBF * CMR_{ESS},$$

a complementary equation can be easily derived.

Ao is calculated as the ratio of uptime over total time:

$$Ao = \frac{Uptime}{Uptime + Downtime} = \frac{MTBF}{MTBF + MCMT + ALDT + (MTBF * CMR_{ESS})}.$$

Dividing numerator and denominator by *MTBF* gives

$$Ao = \frac{1}{1 + (MCMT + ALDT)/MTBF + CMR_{ESS}} \tag{6}$$

The results of Equation 3 are always less than Equation 6. However, only under certain circumstances are the differences very significant—that is, when the ratio of average downtime ($MCMT + ALDT$) to *MTBF* is greater than 1/5 and increasing with the OPR. *Figure 5* shows the resultant differences between the two equations (with $CMR_{ESS} = 0$ in Equation 6 for consistency). To use *Figure 5*, look up the intersection of the ($MCMT + ALDT$)/$MTBF$ on the ordinate and the *OPR* on the abscissa—your general location will give an approximate difference as indicated by the lines of constant difference shown on the chart. For example, if our ($MCMT + ALDT$)/$MTBF$ ratio is 0.2, and our *OPR* is 0.76, the intersection hits directly on the 0.02-difference contour line, and we know that the difference between Equations 3 and 6 is 0.02 (with Equation 3 giving the lower result).

Conclusion

The equations described herein provide a technique and methodology for measuring and estimating *Ao*. The results of Equations 5 and 6 and their expansions not covered in this document (including the important pulse *Ao*) closely match results of Monte Carlo simulations written to specifically measure *Ao* over typical operating cycles. Although KPPs are required by CJCSM 3170.01C (2007) to be testable, it is clear that it is necessary in most cases to measure the inherent R&M of an item and apply modeling and/or simulation techniques to evaluate the actual *Ao*. The equations and methodologies in this article describe the most common of those techniques, as well as their limitations and shortcomings. ❑

Gary A. Pryor is a civilian employee of the U.S. Army Training and Doctrine Command, serving as a RAM engineer supporting the US Army Maneuver Support Center (MANSCEN), Fort Leonard Wood, MO. He received an undergraduate bachelor of science degree in aeronautical and astronautical engineering from the Ohio State University and a master of science degree in computer science from the University of Missouri, Rolla (now the Missouri University of Science and Technology). He has been working for the Army in the RAM Engineering Field since 1983. He is a current member of the Society of Reliability Engineers (SRE) and was an ASQC Certified Reliability Engineer from 1990–2002. His contributions to the RAM field include the development, support, and free distribution of the Reliability Test Planner software application.

References

Chairman of the Joint Chiefs of Staff Manual 3170.01C, 1 May 2007.

DA PAM 71-3. 2003. "Test and Evaluation in Support of Systems Acquisition." Pamphlet, May 30, 2003. Washington, DC: Department of the Army.

DA PAM 70-3. 2008. "Army Acquisition Procedures." Pamphlet, January 28, 2008. Washington, DC: Department of the Army.

Defense Acquisition University. 2005. *Glossary of Defense Acquisition Acronyms and Terms*. 12th ed. Available at https://akss.dau.mil/pv/glossary.aspx. Accessed September 12, 2008.

DOD 3235.1-H. 1982. *Test and Evaluation of System Reliability, Availability, and Maintainability, A primer,* March 1982. Washington, DC: Office of the Under Secretary of Defense for Research and Engineering.

2009 ITEA Journal *Themes*

The ITEA Publications Committee has established themes for the 2009 issues of *The ITEA Journal* and invites articles in the following areas:

Test and Evaluation of Highly Complex Systems (March issue). Complex systems embody the idea that the whole is greater than the sum of the parts – the large scale behavior cannot be predicted from knowledge of the individual constituents. Many of these systems are network enabled and must be tested in the presence of the network effects, including interactions with other networks, with services, and with applications. Service-oriented architectures are being integrated with legacy systems and sorting out behaviors and performance requires insight beyond the system under test. The human element – the cognitive domain – is an essential piece and is the place where perception, awareness, understanding, beliefs and values reside. This issue casts a broad net that our current test environment is being asked to accommodate: joint and distributed testing, chaos and complexity theory, emergent behaviors, virtual testing, modeling and simulation, cognition and autonomy, and assessing mission effectiveness from individual human, component, and system performance. ***(Manuscript deadline: December 1, 2008)***

The Future of Test Facilities (June issue). Shrinking RDT&E budgets signal a coming reduced workload for test facilities. To remain viable, commercial organizations seek to expand the use of facilities by redefining their markets, for example, aerospace wind tunnels are used for flow around cars, buildings, downhill skiers, and bobsledders. Government programs seek economies by partnering with industry and academia, overlapping developmental and operational testing, and attempting a peaceful co-existence between test and training. Opinions are sought on both sides of the test and training issue – what can and should be done and what can never be – as well as lessons learned from past efforts. Ideas are welcome on the synergy and benefits of other common use of test facilities: non-Department of Defense testing, non-traditional testing, extending the customer base, inclusion of experimentation, expanded use of commercial and academic test capabilities, and collaborating with international partners. ***(Manuscript deadline: March 1, 2009)***

Integrating Test and Evaluation (September issue). In December 2007 the Department of Defense issued new policies that represent a shift in emphasis from test to evaluation and promote an emphasis on integrated test and evaluation throughout the system life cycle in a seamless continuum. This issue addresses implementation and follows up the new policies to examine integrating contractor, developmental, operational, and live fire testing and the renewed role of developmental evaluation. The issue also examines the ethics and obligations of test and evaluation. Test and evaluation exist to serve the customer and must coach the customer to ask the right questions, must report the truth, report in a timely manner, and report assertively to ensure appropriate attention is paid. Questions arise from using data from Iraq as test data: how do we instrument systems to get data? How do commercial entities get customer data? How do non-military government organizations get data in areas for which they are responsible? The issue also asks: what are the impediments to realizing integrated test and evaluation and what are the limitations of doing so? ***(Manuscript deadline: June 1, 2009)***

Air & Space (December issue). The year 2009 marks the 40th anniversary of the first moon walk by Neil Armstrong. The first powered flight at Kitty Hawk, North Carolina occurred 106 years ago in December. Today space is more than exploration and air is more than airplanes. Earth is blanketed by countless satellites viewing, recording and communicating; the international space station is an orbiting laboratory; government and private organizations are pursuing commercial access to space. The national and international airspace is pushed to record densities at a time when unmanned air vehicle use is booming, and the Federal Aviation Administration is being asked to incorporate these unpiloted and remotely piloted flying creatures into the soup. This issue takes a retrospective look at how we arrived here, where technology is taking us, and the demands that will be placed on test and evaluation. Air and space constitute the realm of rockets, missiles, weapons, satellites, aircraft of every pedigree, transportation, intelligence, sensors, communications, hypersonics, and so much more. ***(Manuscript deadline: September 1, 2009)***

In addition: T&E articles of general interest to ITEA members and *ITEA Journal* readers are always welcome. Each Issue includes specialty features, each 2-3 pages long: **"Featured Capability"** describes unique, innovative capabilities and demonstrates how they support T&E; **"Historical Perspectives"** recall how T&E was performed in the past, or a significant test or achievement, often based on personal participation in the "old days" of T&E.; **"TechNotes"** discusses innovative technology that has potential payoff in T&E applications or could have an impact on how T&E is conducted in the future. **Interested authors**: should submit contributions to the **ITEA Publications Committee Chairman (itea@itea.org, attn.: Dr. J. Michael Barton)**. Detailed Manuscript Guidelines can be found at **www.itea.org** under the ITEA Publications tab.

ITEA Journal 2008; 29: 437

ITEA Author Index, Volume 29, 2008

ITEA Journal 2008; 29: 438–444

ITEA Subject Index, Volume 29, 2008

K

L

M

N

S

T