Appendix E
Nuclear and Non-Nuclear Testing

E.1 Overview

From 1945 to 1992, the United States conducted both nuclear and non-nuclear testing. After 1992, the United States developed a robust program with which to certify the continued effectiveness of nuclear weapons without the use of nuclear testing.

E.2 U.S. Nuclear Testing Program

The U.S. nuclear testing program began with the *Trinity* test on July 16, 1945 at a location approximately 55 miles northwest of Alamogordo, NM, now called the *Trinity Site*. That test confirmed that the *Fat Man* implosion design weapon would function to produce a nuclear detonation. It also gave the Manhattan Project scientists their first look at the effects of a nuclear detonation.

The United States conducted five additional nuclear tests between 1946 and 1948. By 1951, the United States had increased its ability to produce nuclear devices for testing and conducted 16 nuclear tests that year. Between 1951 and 1958, the United States conducted 188 nuclear tests. Most of these tests had a primary purpose of increasing

the knowledge and data associated with nuclear physics and weapon design. Some of the tests were designed to develop nuclear weapons effects data, and a few were safety experiments. These tests were a mixture of underground, above-ground, high-altitude, underwater, and above-water detonations.

In 1958, the United States instituted a self-imposed moratorium on nuclear tests. In 1961, nuclear testing resumed, and the United States conducted an average of approximately 27 tests per year over the next three decades. These included 24 joint tests with the United Kingdom,[1] 35 tests for peaceful purposes under the *Plowshare* program,[2] seven tests to increase the capability to detect, identify, and locate nuclear tests under the *Vela Uniform* program, four tests to study nuclear material dispersal in possible accident scenarios, and post-fielding tests of specific weapons. By 1992, the United States had conducted a total of 1,054 nuclear tests.

In 1992, Congress passed the legislation that ended the U.S. nuclear testing program, and led to the current policy restriction on nuclear testing.

E.2.1 Early Years of the U.S. Nuclear Testing Program

The first six nuclear tests represented the infancy stage of the U.S. nuclear testing program. The first test at the Trinity Site in New Mexico provided the confidence required for an identical weapon to be employed at Nagasaki. The second and third tests, both in 1946, used identical *Fat Man* design devices to evaluate the effects of airdrop and underwater detonations in the vicinity of Bikini Island in the Pacific. The next three tests were conducted in 1948 on towers on the Enewetak Atoll in the Pacific, testing three different weapon designs. These first six tests began with no previous data, and by today's standards, very crude test measurement equipment and computational capabilities. Because of this, only limited amounts of scientific data were gained in each of these events.

---

[1] The United States and the United Kingdom were preparing to conduct a 25th test when President Clinton announced a moratorium on underground nuclear testing in 1992. Until that point, the nuclear relationship between the United States and United Kingdom as defined by the 1958 Mutual Defense Agreement allowed for the conduct of joint tests between the two nations. This was a great benefit to the United Kingdom—especially following the atmospheric testing moratorium of 1958—because the nation did not have the same access to land that could be used for underground nuclear testing as the United States and the Soviet Union. Following the 1992 testing moratorium, the United Kingdom formally undertook to end nuclear testing in 1995, and the nation ratified the Comprehensive Nuclear-Test-Ban Treaty in April 1998. See Chapter 8: *International Nuclear Cooperation,* for a more detailed discussion of the nuclear relationship between the United States and the United Kingdom.

[2] The Plowshare program was primarily intended to evaluate the use of nuclear detonations for constructive purposes, e.g., to produce craters for the rapid and effective creation of canals.

The 188 nuclear tests conducted between 1951 and 1958 included 20 detonations above one megaton (MT), one detonation between 500 kilotons (kt) and one MT, 13 detonations between 150 and 500 kt, and 17 tests that produced zero or near-zero yields, primarily as safety experiments. Many of these tests produced above-ground detonations, which were routine at that time. The locations for these tests included the Nevada Test Site (NTS), the Enewetak Atoll, Bikini Island, the Pacific Ocean, and the Nellis Air Force Range in Nevada. Some of the highest yield detonations were produced by test devices that were far too large to be used as deliverable weapons. For example, the *Mike* device, which produced a 10.4 MT detonation on November 1, 1952 at Enewetak, was almost seven feet in diameter, 20 feet long, and weighed 82 tons. On February 28, 1954, the *Bravo* test on Bikini Island produced a surface burst detonation of 14.8 MT, the highest yield ever produced by the United States. The *Bravo* device was a two-stage design in a weapon-size device, using enriched lithium as fusion fuel in the secondary stage. Figure E.1 is a photo of the *Bravo* fireball shortly after detonation.

Figure E.1 "Bravo" Nuclear Test

During this period, as the base of scientific data grew, and as sensor technology, test measurement, and diagnostic equipment became more sophisticated and more capable, the amount of data and scientific information gained from each test increased. The initial computer "codes" used to model fissile material compression, fission events, etc., were based on two-dimensional models. These computer models became more capable as the scientific data base expanded and computing technology evolved.

E.2.2 Transition to Underground Nuclear Testing

Between October 31, 1958 and September 14, 1961, the United States conducted no nuclear tests because of a self-imposed testing moratorium. The United States resumed nuclear testing on September 15, 1961 and conducted 100 tests over the next 14 months, including underground, underwater, and above-ground detonations. These tests included nine detonations above one MT, eight detonations between 500 kt and one MT, and four detonations between 150 and 500 kt. The locations for these tests included: the NTS; Carlsbad, NM; the vicinity of Christmas Island in the East Indian Ocean; the Pacific Ocean; and Johnston Island in the Pacific. The last four tests of this group were conducted during

the nine day period between October 27 and November 4, 1962. These were the last U.S. nuclear tests that produced above-ground or surface burst detonations.

In compliance with the Limited Test Ban Treaty (LTBT) of 1963, all subsequent U.S. nuclear test detonations were conducted deep underground. (For more information on the LTBT, see Appendix B: *International Nuclear Treaties and Agreements.*) Initially, there was some thought that this restriction would have a negative impact on the program to develop accurate data on the effects of nuclear weapons. The Atomic Energy Commission (AEC) and the Defense Atomic Support Agency (DASA)[3] responded with innovative ways to minimize the impact of this restriction. Through the use of long and deep horizontal tunnels, and with the development of specialized sensors and diagnostic equipment to meet the need, the effects testing program continued effectively.

In the 30 years between November 9, 1962 and September 23, 1992, the United States conducted 760 deep underground nuclear tests.[4] During this period, tests were performed for all of the previously discussed reasons. The locations for these tests included: the NTS; the Nellis Air Force Range in Nevada; the vicinity of Fallon, Nevada; the vicinity of Hattiesburg, Mississippi; the vicinity of Amchitka, Alaska; the vicinity of Farmington, New Mexico; the vicinity of Grand Valley, Colorado; and the vicinity of Rifle, Colorado.[5] The tests during the period between November 1962 and April 1976 included four detonations above one MT, 14 detonations between 500 kt and one MT, and 88 detonations between 150 and 500 kt.[6] Of the 1,054 total U.S. nuclear tests, 63 had simultaneous detonations of two or more devices, and 23 others had zero or near-zero yield.

Generally, a device for a weapons-related underground nuclear test (UGT) (for physics research, to refine a warhead design in engineering development, or for a post-fielding test) was positioned down a deep vertical shaft in one of the NTS test areas. Informally, this type of test was called a *vertical test*. Typically, a large instrumentation package would be lowered into the shaft and positioned relatively close to the device with electrical wires that ran back to above-ground recording instruments. The vertical shaft was covered with earth, and structural support was added to prevent the weight of the earth from crushing the instrumentation package or the device. This closed the direct opening to the

[3] The AEC was a forerunner organization to the current National Nuclear Security Administration (NNSA), and DASA was a forerunner organization to the current Defense Threat Reduction Agency (DTRA).

[4] Four of these were surface experiments, without a nuclear detonation, to study plutonium scattering.

[5] After May 17, 1973, all U.S. nuclear tests were conducted at the NTS.

[6] 81 of the 90 are listed in the unclassified record with a yield between 20 and 200 kt.

surface and precluded the fireball from pushing hot radioactive gases up the shaft into the atmosphere. When the detonation occurred, the hundreds or thousands of down-hole instruments momentarily transmitted data, but they were almost immediately consumed in the fireball. The preparation for a vertical UGT took months and included drilling the vertical shaft and preparation of the instrumentation package, which was constructed vertically, usually within 100 meters of the shaft. The instrumentation package was typically 40 to 80 feet high, several feet in diameter, and surrounded by a temporary wooden structure. The structure would have floors approximately seven to eight feet apart and a temporary elevator to take technicians to the various levels to place and prepare the instruments. The test device would be lowered into the shaft, followed by the cylindrical instrument package. After the test, the earth above the detonation would often collapse into the cavity left by the cooling fireball, forming a subsidence crater on the surface directly over the test location.[7] See Figure E.2 for a photograph of a preparation site for an underground nuclear test.

Figure E.2
Underground Nuclear Test Preparation

Generally, a UGT device for an effects test was positioned in a long horizontal tunnel deep into the side of one of the mountains in the Yucca Mountain range at the north end of the NTS. Informally, this type of test was called a *horizontal test*. The tunnels were relatively large, usually more than 30 to 40 feet across, and ran several miles into the side of the mountain. Typically, the tunnel had a small-scale railroad track running from the entrance to the deepest part of the main tunnel, which included a train to support the logistics movement of workers and equipment. The main tunnel would have many long branches, called side-drifts, each of which could support a UGT. Instruments were positioned at various distances from the device, and a huge "blast door" was constructed to permit the instantaneous effects (nuclear and thermal radiation, X-rays, and electromagnetic pulse) to travel to instruments at greater distances but to close prior to the arrival of the blast wave. After the detonation, instruments outside the blast door would be recovered, and the side-drift would be closed and sealed with a large volume of earth.

For both vertical and horizontal UGTs, the device would be prepared in a laboratory environment and transported to the test site, usually only a few days prior to the test date.

[7] The collapse that caused the subsidence crater could occur at any time from minutes to months after the detonation; the time of the collapse was unpredictable.

On the test date, the NTS operations center would continuously monitor wind direction and wind speed to determine where any airborne radioactive particles would travel in the unlikely event of a "venting" incident.[8] If the wind conditions could blow venting gases to a populated area, the test was delayed until the wind conditions changed. Frequently, UGTs were delayed hours or days.

The Threshold Test Ban Treaty (TTBT) was signed by the United States in 1974; the treaty was not ratified until 1990, but in 1976 the United States announced that it would observe the treaty pending ratification. The treaty limited all future tests to a maximum yield of 150 kt; this presented a unique problem because, at the time, each of the three legs of the nuclear triad required new warheads with yields exceeding 150 kt. This compelled the weapons development community to make two major changes to nuclear weapons development. First, new warhead designs were limited to using tested and proven secondary stage components, which provide most of the yield in high-yield weapons. The rationale for this change was that if previous testing had already determined the X-ray output required from the primary stage to ignite or "drive" the secondary, and if testing had also determined the output of the secondary, then all that would be needed was a test to determine if the new primary would produce a yield large enough to drive the secondary. Of the 1,054 U.S. nuclear tests, at least 82 had yields that exceeded 150 kt. Another 79 may have had yields exceeding 150 kt, but they are listed in unclassified source documents only as being between 20 to 200 kt. Many of these tests provided the data for scientists to determine the required information (ignition threshold, yield output, etc.) to certify several different secondary stage designs, which would produce yields greater than 150 kt. See Figure E.3 for a summary of U.S. nuclear tests by yield.

The second change was that to test any new warhead with a yield greater than 150 kt, the warhead would have to be reconfigured to ensure that it would not produce a yield in excess of 150 kt. Thus, the newest strategic warheads would not have a nuclear test (in their new configuration) for any yields above 150 kt.

By the 1980s, the U.S. nuclear test program had evolved into a structure that categorized tests as physics research tests, effects tests, warhead development engineering tests,

[8] Venting incidents occurred very few times during the history of U.S. underground nuclear testing. Venting can occur when a vertical UGT shaft is close enough to an unknown deep underground cave system that leads to the surface and permits the expanding fireball to push hot radioactive gases through the underground cave system to the surface and into the air. Instruments to determine geology thousands of feet underground were not precise enough to detect all possible underground caves/cavities. Venting can also occur if the blast door for a horizontal UGT is not strong enough to contain the blast wave.

Time Period	Yield					
	Zero or Near-Zero	> 0 to 150 kt	Possible > 150 kt	> 150 to 500 kt	> 500 kt to 1 MT	> 1 MT
1945 - 1948	0	6	0	0	0	0
1951 - 1958	17	137	0	13	1	20
1961 - 11/04/62 *	0	79	0	4	8	9
11/9/62 - 03/17/76 **	5	391	79	9	14	4
May 76 - 1992	1	257	0	0	0	0
Total:	23	870	79	26	23	33

* Last U.S. above-ground or surface detonation.
** Last U.S. detonation above 150 kt.

Grand Total: 1,054 Nuclear Tests

Figure E.3 U.S. Nuclear Tests by Yield

and post-fielding tests. Physics research tests contributed to the scientific knowledge and technical data associated with general weapons design principles. The effects tests contributed to the base of nuclear effects data and to testing the vulnerability of key weapons and systems to the effects of nuclear detonations. Development tests were used to test key aspects of specific designs or to refine specific designs to increase yield output or to improve certain nuclear detonation safety features. Post-fielding tests were conducted to provide stockpile confidence and ensure safety. For each warhead-type, a stockpile confidence test (SCT) was conducted between six and 12 months after fielding. This was intended to check the yield to ensure that any final refinements in the design that were added after the last development test and any imperfections that may have resulted from the mass-production process did not corrupt the designed yield. Post-fielding tests were also used to confirm or repair safety or yield problems when non-nuclear testing, other surveillance, or computer simulation detected possible problems, especially unique abnormalities with the fissile component. If a problem was confirmed and a significant modification applied, a series of nuclear tests could be used to validate the modification to ensure that fixing one problem did not create a new issue.

E.2.3 The Transition to 3-D Codes

By the early 1980s, the United States had conducted more than 970 nuclear tests, most of which had the basic purpose of increasing the scientific data associated with weapon design or refining specific designs. The physics laboratories had acquired the most capable

computers of the time and were expanding the computer codes to analyze fissile material compression, fission events, etc., in a three-dimensional (3-D) model. By the mid-1980s, use of 3-D codes had become routine. The 3-D codes provided more accurate estimates of what would be achieved with new designs or what might happen (for nuclear detonation safety considerations) in an abnormal environment.

With the 3-D codes, the labs evaluated a broader range of abnormal environments for fielded warhead-types, e.g., the simultaneous impact of two high-velocity fragmentation pieces. This led to safety experiments and safety improvements that might not have otherwise occurred.[9] The increased computational modelling capability with the 3-D codes also helped scientists to refine the near-term nuclear testing program to include tests that would provide maximum value-added to the base of scientific knowledge and data. Each year, the results of the nuclear testing program increased the labs' computational modeling capabilities.

E.2.4 End of Underground Nuclear Testing

In 1992, in anticipation of a potential comprehensive test ban treaty, the United States voluntarily suspended its underground nuclear testing program. Public Law 102-377, the legislation that ended U.S. nuclear testing, had several key elements, including a provision for 15 additional nuclear tests to be conducted by the end of September 1996 for the primary purpose of applying three modern safety features—enhanced nuclear detonation safety (ENDS), insensitive high explosive (IHE), and fire-resistant pit (FRP)—to those warheads planned for retention in the reduced stockpile under the proposed Strategic Arms Reduction Treaty (START) II. (For more information on START II, see Appendix B: *International Nuclear Treaties and Agreements*.) With a limit of 15 tests within less than four years, there was no technically credible way (at the time) to certify design modifications that would incorporate any of the desired safety features into existing warhead-types. Therefore, the legislation was deemed too restrictive to achieve the objective of improving the safety of those warhead-types lacking all of the available safety enhancement elements, and it was decided that the United States would not conduct any further tests. The last U.S. underground nuclear test, *Divider,* was conducted on September 23, 1992 (Figure E.4).

The Fiscal Year 1994 National Defense Authorization Act (Public Law 103-160) called on the secretary of energy to "establish a stewardship program to ensure the preservation

[9] For example, an interim fix for one of the Army warheads was fielding a "horse-blanket" to be draped over the container to provide fragmentation/projectile shielding for transportation and storage; the ultimate fix put the shielding inside the container.

of the core intellectual and technical competencies of the United States in nuclear weapons." The Stockpile Stewardship Program, a science-based approach to ensure the preservation of competencies as mandated by Public Law 103-160, has served as a substitute for nuclear testing since 1992. (For more information on the Stockpile Stewardship Program, see Chapter 7: *U.S. Nuclear Infrastructure*.)

Figure E.4 Preparation for Divider, the Last U.S. Underground Nuclear Test

E.3 Quality Assurance and Non-Nuclear Testing

The goals of the U.S. nuclear weapons quality assurance (QA) programs are to validate safety, ensure required reliability, and to detect, or if possible, prevent, problems from developing for each warhead-type in the stockpile. Without nuclear testing, the current stockpile of nuclear weapons must be evaluated for quality assurance through only the use of non-nuclear testing, surveillance, and—to the extent applicable—modelling. The Department of Energy (DOE) Stockpile Evaluation Program (SEP) has evolved over decades, and currently provides the information to support stockpile decisions and assessments of the safety, reliability, and performance of the stockpile. This program is designed to detect stockpile defects, understand margins at a component level, understand and evaluate changes (e.g., aging), and (over time) predictably assess the stockpile. The overall QA program includes: laboratory tests, flight tests, component and material evaluations, other surveillance evaluations and experiments, the reported observations from the Department of Defense (DoD) and the DOE technicians who maintain the warheads, continuous evaluation for safety validation and reliability estimates, and the replacement of defective or degrading components as required.

No new replacement warheads have been fielded by the United States for over two decades. During that time, sustaining the nuclear deterrent has required the United States to retain warheads well beyond their originally programmed life. As the warheads in the stockpile age, the SEP has detected an increasing number of problems, primarily associated with

non-nuclear components. This led to an expanded program of refurbishments, as required for each warhead-type, and a formal process to manage it. The SEP program has been very effective for quality assurance. Even though it has been almost 20 years since the last U.S. nuclear test, approximately one dozen different warhead-types serve as the backbone of the nation's nuclear deterrent, each with annual safety validations and very high reliability estimates.

Because the warheads of the stockpile continue to age beyond any previous experience, it is anticipated that the stockpile will reveal age-related problems unlike any other time in the past. As a part of proactive quality assurance management, the DOE has recently established a Surveillance Transformation Project (STP). Its focus is on the creation and maintenance of a knowledge-based, predictive, adaptable, and cost-effective evaluation program. This section of the appendix describes the many activities associated with the quality assurance of the U.S. nuclear weapons stockpile. These activities take place in multiple DOE locations, and many of them occur at the Pantex Plant in Amarillo, Texas (Figure E.5).

Figure E.5 Pantex Plant

E.3.1 The Evolution of Quality Assurance and Sampling

The Manhattan Project, which produced one test device and two war-reserve (WR) weapons that were employed to end World War II, had no formal, structured QA program. There were no safety standards or reliability requirements to be met. QA was the sum of all precautions thought of by weapons scientists and engineers and the directives of Dr. Oppenheimer and his subordinate managers. History proves that the Manhattan Project version of QA was successful in that it accomplished an extremely difficult task without a catastrophic disaster.

The first nuclear weapons required in-flight insertion (IFI) of essential nuclear components, until which time the weapons were unusable. Once assembled in flight, the weapons had none of the modern safety features to preclude an accidental detonation. The QA focus was on ensuring the reliability of the weapons because they would not be assembled until they were near the target. In the early 1950s, as the U.S. nuclear weapons capability

expanded into a wider variety of delivery systems, and because of an emphasis on more rapid response times for employment, IFI became impractical.

The development of sealed-pit weapons to replace IFI weapons led to requirements for nuclear detonation safety features to be built into the warheads.[10] (See Chapter 5: *Nuclear Safety and Security*, for a detailed discussion of nuclear detonation safety and safety standards.) During this time, the concern for safety and reliability caused the expansion of QA activities into a program that included random sampling of approximately 100 warheads of each type, each year. Initially, this was called the New Material and Stockpile Evaluation Program (NMSEP). "New material" referred to weapons and components evaluated during a warhead's development or production phase. (See Appendix D: *U.S. Nuclear Weapons Life-Cycle*, for a description of nuclear weapon life-cycle phases.) New material tests were conducted to detect and repair problems related to design and/or production processes. The random sample warheads were used for both laboratory and flight testing, and they provided an excellent sample size to calculate reliability and to stress-test the performance of key components in various extreme environments. This was unaffordable for the long term, and within a year or two, the program was reduced to random sampling of 44 warheads of each type. This sample size was adequate to calculate reliability for each warhead-type. Within a few more years, that number was reduced to 22 per year and remained constant for approximately a decade. Over time, the random sample number was reduced to 11 per year to reflect fiscal and logistical realities. With the implementation of the Surveillance Transformation Project, each weapon system was re-evaluated with respect to approach to sampling, accounting for the specific technical needs of each system, and new approaches to evaluation tests being implemented. As a result, some system samples were reduced from 11 per year to lower numbers.

In the mid-1980s, the DOE strengthened the significant finding investigation (SFI) process. Any anomalous finding or suspected defect that might negatively impact weapon safety, reliability, or control is documented as an SFI. The QA community investigates, evaluates, and resolves SFIs.

The NMSEP is a part of today's Stockpile Evaluation Program. At the national level, random sample warheads drawn from the fielded stockpile are considered to be a part of the Quality Assurance & Reliability Testing (QART) program. Under the QART program, additional efficiencies are gained by sampling and evaluating several warhead-types as a warhead "family" if they have enough key components that are identical. Until 2006, each

[10] Sealed-pit warheads are the opposite of IFI—they are stored and transported with the nuclear components assembled into the warhead, and they require no assembly or insertion by the military operational delivery unit.

warhead family had 11 random samples evaluated each year. The sample size of 11 per year enabled the QA program: to provide an annual safety validation; to provide a reliability estimate semi-annually; and to sample any randomly occurring problem was present in ten percent or more of the warheads of that type, with a 90 percent assurance, within two years.

Weapons drawn for surveillance sampling are returned to the NNSA Pantex Facility near Amarillo, Texas, for disassembly. Generally, of the samples selected randomly by serial number, two to three are used for flight testing and the remainder are used for laboratory testing and/or component and material evaluation (CME). Surveillance testing and evaluation may be conducted at Pantex or at other NNSA facilities. Certain components are physically removed from the weapon, assembled into test configurations, and subjected to electrical, explosive, or other types of performance or stress testing. The condition of the weapon and its components is carefully maintained during the evaluation process. The integrity of electrical connections remains undisturbed whenever possible. Typically, one sample per warhead family per year is subjected to non-nuclear, destructive testing of its nuclear components and cannot be rebuilt. This is called a destructive test (D-test), and the specific warhead is called a D-test unit. Depending on the availability of non-nuclear components, and the military requirement to maintain stockpile quantities, the remaining samples may be rebuilt and returned to the stockpile.

E.3.2 Surveillance Transformation Project

Much of the current surveillance methodology is based on the original weapon evaluation programs, relying mainly on random stockpile sampling applied to flight tests, subsystem go/no-go testing, and selected component evaluations to search for design and manufacturing "birth" or aging defects. This approach gives a current snapshot of the condition of that warhead-type but provides little ability to predict future stockpile problems. The ability to predict a problem is becoming more important as the current warheads of the stockpile continue to age beyond the experience of stockpile scientists and engineers.

In June 2006, the NNSA chartered a complex-wide team to integrate efforts to develop a comprehensive plan for achieving surveillance transformation. The STP is a plan to define a road map to begin transformation to a more knowledge-based, predictive, adaptable, and cost-effective evaluation of current and future stockpile health. It sets the nuclear weapons complex on a course to transform surveillance across four major objectives:

- *Rigorous Requirements Basis*: create a strong technical requirements basis for stockpile evaluation;

- *Evaluating for Knowledge*: design and execute an evaluation program that responds to changing evaluation data needs over the weapon system life-cycle;
- *Predictive Assessment*: develop the capabilities to predict the state of health of the enduring stockpile through end-of-life projections, reliability assessments, predictive performance assessments in areas beyond reliability (i.e., safety/survivability/use control/nuclear performance), and risk-based responsiveness for replacement and refurbishment decisions; and
- *Premier Management and Operations*: create a strong program management team to make the best decisions based on defensible cost-benefit criteria.

With the implementation of STP, additional emphasis has been placed on component and material evaluation for the early detection of signs of aging. Identified aging mechanisms would then be used to predict when the changes as a result of aging would begin to negatively affect system performance so that prophylactic measures can be taken.

E.3.3 Stockpile Laboratory Testing

For each warhead family, the NNSA laboratory evaluation program strives to examine each possible operational use of the warhead, potential environmental conditions, safety and use control features, and the end-to-end process required for nuclear detonation. Several critical system-level functional performance aspects are verified and the data to support reliability assessments are obtained with capabilities to address the spectrum of environmental and operational conditions increasing with the current enhanced surveillance investments. The system-level testing program also examines safety components to determine if there is any concern for the overall safety of the weapon.

Laboratory non-destructive testing can include activities such as radiography and gas sampling. Stockpile lab testing includes, for example, fuzing mode tests, environmental sensing unit tests, trajectory sensing device tests, functioning of firing sets tests and neutron generators, performance of stronglinks and other safety devices, and for weapons so equipped, permissive action link (PAL) tests and command disable function tests. The NNSA testing program emphasizes testing at the highest possible system or subsystem levels. Diversification of tests is used as necessary to address certain aspects of weapon performance under specific use conditions and with maximum realism. Figure E.6 depicts a laboratory in which tests are performed.

Joint Integrated Laboratory Tests (JILTs) evaluate interconnected DoD and NNSA weapon components. For example, the DoD arming and fuzing mechanism would be tested in conjunction with a NNSA denuclearized warhead firing system. These system-level tests

Figure E.6 Gloveboxes

are conducted at either NNSA or DoD facilities.

Normally, the nuclear explosive package from the D-test unit is destructively tested to look for any changes in dimensions or material composition. Five key components are tested: the pit, the secondary, the detonator assembly, the high explosives, and the gas bottle system. The D-test unit is not rebuilt and is, therefore, not returned to the stockpile. The remainder of the samples can be reconstructed and returned to the stockpile if replacement components are available for rebuild. If components are not available for rebuild, those warheads are eliminated from the stockpile. These reductions are called QART consumption in the national-level stockpile planning documents. (See Chapter 2: *Stockpile Management, Processes, and Organizations* for a discussion of national-level stockpile planning documents.)

E.3.4 Stockpile Flight Testing

Flight testing of nuclear delivery systems is accomplished using warheads with inert nuclear components known as joint test assemblies (JTAs). JTAs use non-fissile nuclear components (surrogates) and/or instrumentation that replace the fissile components in the tested weapon. This precludes any possibility that the JTA can produce a nuclear detonation while providing critical information about performance in the actual combined environments experienced in flight. Typically, two to four JTA flight tests per weapon family are planned each year. The JTAs may be either high-fidelity JTAs (HF-JTAs) or instrumented JTAs (I-JTAs).

HF-JTAs replicate actual WR warheads as closely as possible, with the exception that the fissile material (plutonium and highly enriched uranium) and the tritium are removed and replaced with surrogates. HF-JTAs provide some data concerning the system as a whole, while I-JTAs provide more instrumented data about individual components and sub-assemblies. HF-JTAs demonstrate the functioning of the warhead in as complete a configuration as possible without a nuclear test. I-JTAs use data-recording instruments to record the in-flight performance of certain components. Normally, I-JTAs provide much more component and sub-assembly performance data than HF-JTAs. However, in order to have these data-recording instruments embedded in the warhead, the instruments

may replace selected warhead components. Therefore, any one I-JTA will have selected warhead components replaced with data-recording instruments, while another I-JTA for the same weapon-type may have a different set of warhead components replaced with other instruments. As much as possible, the data-recording instruments are designed to have the same physical dimensions (height, width, length, weight, center-of-gravity, etc.) as the components they replace.

The Non-Nuclear Assurance Program (NNAP) ensures that actual nuclear weapons are not accidentally used in flight tests in place of the JTAs. The verification process includes inspections of tamper-evident seals and other indicators in conjunction with measurements taken by radiation detection instruments. For joint tests with the DoD, the NNSA provides the test assemblies with permanent "test" markings, tamper-evident seals, signature information, and radiation test equipment.

Flight tests are conducted at various locations in the United States including the Tonopah Test Range in Nevada, the Utah Test and Training Range in Utah and Eastern Nevada, Vandenberg Air Force Base (AFB) in California, and Eglin AFB in Florida. Stockpile flight tests (SFTs) involve JTAs built with components from WR weapons that have already experienced stockpile handling. These tests demonstrate the continued compatibility between the warhead and the delivery vehicle and verify weapons system function throughout the stockpile-to-target sequence.

E.3.5 Component and Material Evaluations

The NNSA is also undertaking a set of activities to baseline margins at a component level and to detect and assess changes in these components over time. Component and material evaluations provide a basis for assessing and projecting aging effects on system performance. These programs utilize components from production, from the stockpile, and from stores to understand failure mechanisms, margins, and the effects of aging on component and system performance, including safety, security, and reliability effects. These activities also provide a knowledge base that can inform planning for future stockpile modifications or LEPs, design decisions, and investigations of anomalies. Hardware utilized in these tests is not typically reintroduced into the stockpile.

E.3.6 Safety Validation and Reliability Estimates

Safety and reliability are evaluated based on the results of the stockpile laboratory testing (SLT), SFT, CME, other surveillance, computer analyses, and when required, the scientific

and engineering judgment of the QA experts. The safety of each warhead-type in the stockpile is validated each year to ensure that it meets established safety standards. (Safety standards and certification are discussed in detail in Chapter 5: *Nuclear Safety and Security.*) Reliability is the probability that a warhead-type will function properly if employed as intended. Reliability estimates for each warhead-type are evaluated twice per year. They are estimates, not solely statistical calculations, because the sample size is not sufficiently large to preclude the possibility that scientific and engineering judgment may be included. Reliability is estimated for each mode of operation (e.g., surface bursts and laydown).

E.4 Conclusion

Though the program has evolved throughout the years, the United States has always performed quality assurance tests on its nuclear weapons. In the past, QA was composed of a combination of nuclear and non-nuclear testing. Since 1992, however, the United States has observed a self-imposed moratorium on nuclear testing. In order to continue QA on U.S. weapons, scientists have used existing test data (collected from those nuclear tests conducted prior to 1992) to develop models and simulations to evaluate nuclear weapons. For the past two decades, those models and simulations, when combined with the judgment of scientists and engineers, have been sufficient to certify the continuing effectiveness of nuclear weapons in the stockpile. Unfortunately, due to the age of the U.S. nuclear stockpile, the nation is moving beyond the era in which there exists past nuclear tests with which to compare current-day stockpile weapons using simulations and models. (For more information on the past life-cycle of nuclear weapons, see Chapter 1: *Nuclear Matters History and Policy*.) When coupled with the fact that many of the scientists and engineers who participated in nuclear testing have retired, this movement into an era beyond what is known (because it was verified by past UGT data) creates a situation in which it becomes harder to certify the effectiveness of the stockpile. Because of this, efforts to ensure that options are in place for life-extension programs are being pursued at all levels of government to ensure that the U.S. nuclear deterrent remains safe, secure, and effective for as long as nuclear weapons exist. (For more information on the life-extension options being considered, see Chapter 2: *Stockpile Management, Processes, and Organizations.*)