# AVAILABILITY AND RELIABILITY ISSUES FOR ILC*

T. Himel[#], J. Nelson, N. Phinney, SLAC, Menlo Park, CA 94025, USA
M. Ross, FNAL, Batavia, IL 60510, U.S.A.

*Abstract*

The International Linear Collider (ILC) will be the largest most complicated accelerator ever built. For this reason extensive work is being done early in the design phase to ensure that it will be reliable enough. This includes gathering failure mode data from existing accelerators and simulating the failures and repair times of the ILC. This simulation has been written in a general fashion using MATLAB and could be used for other accelerators. Results from the simulation tool have been used in making some of the major ILC design decisions and an unavailability budget has been developed.

## INTRODUCTION

The International Linear Collider as presently planned will have over 20 km of superconducting linear accelerator, two 6 km circumference damping rings (DR), a complex Beam Delivery System to focus beams down to 5 nm, a polarized electron source and an undulator based positron source. Taken altogether, this will be one of the most complex machines ever built. It has over an order of magnitude more parts than most accelerators.

A typical high energy physics (HEP) accelerator currently has an availability of 75-85%. With so many more components that could potentially fail, the ILC availability would be unacceptably low unless significant attention is paid to component reliability.

Because of this concern, high availability design work started at an early stage of the ILC project. Much of this work depends on an availability simulation developed for the purpose. Results from this simulation have been used to make major design decisions for the ILC and also to develop an unavailability budget for the components and subsystems.

Many accelerators have estimated their availabilities during the design phase with a spreadsheet. SNS and APT are among the examples examined before embarking on the simulation. These spreadsheets used formulas to combine the availabilities of components to get the availability of the whole. There are also commercially available reliability software packages to perform such calculations. The approach taken here was to write a simulation which could allow several complexities to be handled that would have been nearly impossible in a spreadsheet and quite difficult in the commercial software packages. These complexities include the recovery and tuning time needed after a downtime, the complex redundancies built into the ILC design, the way in which accelerator physics experiments (Machine Development or MD) can be done when only part of the accelerator is available, and the way in which many devices are typically repaired during an access by a limited number of people. By writing a simulation tailored to the task, it was possible to incorporate knowledge derived from the authors' accumulated decades of experience in running real accelerators.

The simulation, named availSim, takes as input a list of components, their quantities, mean time between failure (MTBF), mean time to repair (MTTR), and the effect of their failure. It then simulates the failure and repair of components while tracking the integrated luminosity. The components include items identified as potential sources of failure from experience with existing facilities, such as klystrons, modulators, magnets, magnet power supplies, power supply controllers, vacuum pumps, pump power supplies, movers, diagnostics, water pumps, etc.

The rest of this paper covers the features of the simulation, some quantitative information mined from previous accelerators and used as input to the simulation, some implementation details, results, and conclusions,

## FEATURES INCLUDED IN AVAILSIM

Many features are included in the simulation to make it as realistic as possible.

Each component fails at a random time with an exponential distribution determined by its MTBF. When a component fails, the accelerator is degraded in some fashion. A klystron failure in the main linac simply reduces the energy overhead. The accelerator keeps running until this overhead is reduced to zero. Similarly there are 21 DR kickers where only 20 are needed so only the second failure causes downtime. Some components such as most magnet power supplies cause an immediate downtime for their repair.

Each component can be specified as hot swappable (meaning it can be replaced without further degrading the accelerator); repairable without accessing the accelerator tunnel, or repairable with an access to the accelerator tunnel. A klystron that is not in the accelerator tunnel is an example of a hot swappable device. If a BPM electronics module is housed in a crate that does not have to be turned off when the module is replaced, then it is hot swappable. These repairs are simulated to occur in a time MTTR after the failure. Devices which are not hot swappable, such as magnets and individual channels of multi-channel modules, are only repaired when the accelerator is down.

Without doubt the downtime planning is the most complicated part of the simulation. This should come as no surprise to anyone who has participated in the planning of a repair day. It is even harder in the simulation because computers don't get a gestalt of the situation like humans do. Briefly, the simulation determines which parameter

**Work supported by the U.S. Department of Energy under contract DE-AC02-76SF00515
[#]email: thimel at slac.stanford.edu

(e.g. e- linac energy overhead or e+ DR extraction kicker strength or luminosity) was degraded too much, and plans to fix things that degrade that parameter. Based on the required repairs, it calculates how long the downtime must be to repair the necessary items. It then schedules other items for repair, allowing the downtime to be extended by as much as 50 to 100%. Some other issues must also be taken into account:

- If an access to the accelerator tunnel is required, one hour is allowed for prompt radiation to decay before entry. One hour is also allowed for locking up, turning on and standardizing power supplies.
- The devices chosen for repair are those that give the most bang for the buck (most improvement in the parameter per hour of repair time).
- The number of people in the accelerator tunnel can be limited to minimize the chaos of tracking who is in the tunnel.
- There are no regularly scheduled maintenance shutdowns, except an annual 3 month shutdown. Interventions occur only when the accelerator is broken, which is the practice at most operating HEP accelerators. In real life, maintenance might be planned when the energy overhead was getting low without waiting to actually run out of energy. However, since the simulation does not add any penalty for unplanned or off-hours downtimes, this becomes a subtlety which does not really impact the results.
- Things which break during the downtime are just ignored (It is assumed they are immediately fixed). The long recovery time which is described in the next paragraph is intended to account for this.

The simulation assumes that all repairs are completed on schedule. It seemed an unnecessary complication to throw random numbers to distribute the repair times around the MTTR as the simulation integrates over a long enough time period to average out such variations.

Recovery of the beam is modeled in a crude fashion which matches the qualitative experience on many accelerators. This common experience is that it takes time to recover good beams after a downtime. In fact, the longer the down, the longer the recovery time. Contributions to the recovery come from myriad factors such as

- Hardware failures – devices such as pumps and power supplies which break because they were turned off during the shutdown or devices which just happen to break while the accelerator was down and were not detected
- Environmental factors – temperature changes caused by the access or ground motion over a few hour period which can be significant enough to require retuning
- Human error – mistakes made in doing the repairs (valves left closed, cables left disconnected…) or failure to restore settings after hardware or software tests
- Parameter drifts – multiple parameters which are continuously tracked and optimized during normal operation which all need to be found and retuned
- Commissioning – hardware or software improvements made during the shutdown which need to be tested, calibrated, etc.
- …

Rather than trying to model recovery procedures in detail, availSim simply assumes that the time it takes to get good beam out of a region of the accelerator is proportional to the time that region was without beam. The constants of proportionality used for each region were 10%, except for the DRs and interaction region, for which 20% was used. In real operation, the beam quality recovers gradually as each region is tuned up in succession, and the luminosity gradually ramps up to nominal. The simulation simplifies this by assuming that the machine goes from no beam at the end of a region to perfect beam at the end of the recovery time. Similarly, the luminosity jumps from zero up to the design value immediately at the end of the entire recovery/tuning time. While this is certainly an oversimplification, if the recovery time used in the simulation is considered to be the time it takes to get back to half the design luminosity, then the overall effect is reasonably well reproduced.

Machine Development (MD) is an essential tax on the operating efficiency of any accelerator. It is time used to better characterize the machine, develop new tuning procedures, and test possible future improvements. The amount of time spent on MD varies through the life of a project, with more MD required in the early stages or after a major upgrade. For this simulation, the ILC is assumed to have operated for a few years and to have settled into a nominal schedule of MD, which would occupy approximately 10% of the time. As with the recovery time, the MD was allocated to the individual regions of the machine. Each region was allocated 1% MD with the exception of the DRs, which were given 2%.



1-4244-0917-9/07/$25.00 ©2007 IEEE

Table 1: MTBFs and MTTRs used in the simulation and where they came from

| Component | MTBF (hr) | MTTR (hr) | comment |
|---|---|---|---|
| Water cooled magnet | 1,000,000 | 8 | Average from SLC. There have been magnet families with MTBF > 13,000,000 |
| Air cooled magnets | 10,000,000 | 2 | SLC |
| Super conducting magnet | 10,000,000 | 472 | MTBF given is 10 times that of Tevatron dipole magnet as the SC quads in ILC are much lower current. We assumed a failed SC quad would be tuned around in 2 hrs as a kludge repair |
| Kicker pulsar | 10,000 | 2 | SLC |
| Magnet Power supplies | 50,000 | 2 or 4 | SLAC and FNAL average. The larger MTTR is for large not easily replaceable supplies |
| Electronics modules | 100,000 | 1 | This is a crude average over many types of electronics modules |
| Water flow switch | 250,000 | 1 | SLAC |
| Movable collimators and stoppers and valves | 100,000 | 8 | SLAC |
| DR klystron | 30,000 | 8 | SLAC |
| Linac Modulator | 50,000 | 4 | SLAC |

To more fully mirror the complexity of operation of a real machine, the simulation assumes that some of the required MD can be done on an *opportunistic* basis. Typically repairs may be completed in some regions earlier than others. As an example, there could be 16 hours of repair work needed in the e- linac, 2 hours in the e- injector and none anywhere else. In this case, there will be time after the e- injector is repaired when beam has been tuned up into the e- DR while repairs are still going on in the e- linac. If this is more than two hours, the simulation assumes useful MD can be done in the e- injector and DR. This *opportunistic* MD time is tracked by the simulation. It then assumes that sometime during the running period enough scheduled MD is done in each region to bring the total of opportunistic + scheduled MD up to the desired levels.

*Kludge* repairs can be simulated. This is done where a proper repair would take too long to complete, so a quicker work-a-round is implemented and the proper repair is done later when there is more time. An example of this is when a power coupler to a superconducting cavity starts breaking down; its input can be disconnected to allow the rest of the cavities powered by the klystron to run. Replacing the coupler requires a warm-up which would be done during the annual long down time.

## AVAILABILITY DATA FROM PREVIOUS ACCELERATORS

The features of the simulation reflect in a qualitative manner the authors' knowledge of accelerator operation.

More quantitative information was developed for use in the simulation by analyzing historical data from accelerators at SLAC and FNAL.

Table 1 shows the MTBFs and MTTRs for some components. These were derived from databases of problems that caused downtime, which were compiled by the laboratories. These databases contained many errors so it was necessary to go through thousands of entries by hand and correct the errors. Errors typically occurred because the cause of the downtime was entered when a problem first occurred, before it was fully understood. After the repair, the true cause would often be recorded in a comment but not in the database fields which were used to calculate MTBFs. There was also some discretion as to which component was assigned the fault. We also chose to omit a family of components which were known to have a design flaw that caused frequent failures. For these reasons, the data in Table 1 are approximate. Also, note that MTBFs for similar components can easily vary by an order of magnitude due to seemingly minor design differences.

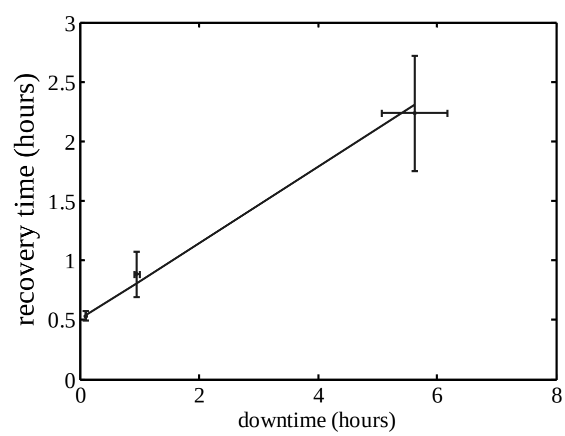


Figure 1: Measured recovery time as a function of the downtime. The fitted line has a slope of 0.32.

Data from PEP-II was used to quantify the relation between recovery time and downtime. Downtimes were found by searching control system records for when the luminosity went to zero. If refilling of PEP started in less than 0.5 hours (which was usually the case), the time between loss of beam and start of injection was taken as the downtime. The time from start of injection to declaration of stable beams to the BaBar detector was the recovery time. For cases where it was more than 0.5 hours to start of injection, the log book was examined to determine what broke and when its repair was finished. The recovery time started from this time. In this way extra downtime caused by things which broke during the original repair were counted in the recovery time in a manner similar to the simulation. The results are shown in Figure 1 where the hundreds of downtimes are averaged into 3 points. There is a large scatter in the recovery to downtime ratio. The measured slope is 0.32 ±0.08. This compares to the 20% that was used to simulate the ILC DRs.

## IMPLEMENTATION DETAILS

The simulation was written in the MATLAB scripting language. It was written in a general way with nothing ILC specific in the code. All of the machine dependent input data was contained in a spreadsheet which was read by the MATLAB program. The spreadsheet included a set of macros which aided in handling the large amount of data needed to describe the number, locations, MTBFs, etc. of the components. The MATAB program output results to another spreadsheet which allowed one to examine in detail which devices caused the downtime. It also provided summaries of which type of device caused the downtime and how much each region of the accelerator contributed to the downtime.

This software could be useful (and has been used) in the design of other accelerators. The MATLAB code, spreadsheets and instructions are available at http://www-project.slac.stanford.edu/ilc/acceldev/ops/avail/source_code.htm.

## SOME ILC RESULTS

Simulation results have been used to help make several ILC design decisions and to establish an unavailability budget for the various systems and components.

### *Use for Design Decisions*

AvailSim was used to compare the uptimes of versions of the ILC with slightly different designs. We could then decide whether the loss in uptime was worth the cost savings. For example, putting both DRs in a single tunnel decreased the uptime by less than one percent while saving the cost of a >6 km tunnel. This was cost effective so the baseline ILC design has both DRs in a single tunnel. In contrast, if all the klystrons, modulators, and power supplies were in the same tunnel as the linac instead of in a separate tunnel, the uptime was reduced by about 14%. This availability was too low to be acceptable, but it was judged too expensive and too large a risk to try to improve the reliability of individual components sufficiently to regain the 14% loss. Hence, the baseline ILC design has two tunnels for the linac.

Many similar comparisons have been mad to aid in other design decisions.

### *Unavailability Budget*

For a given set of MTBFs and MTTRs, availSim gives the total downtime and the downtime caused by each type of component. Using this information the MTBFs were tuned to improve the downtime from the 30% originally predicted by using the MTBFs in Table 1 to the desired 15%. One could have accomplished the same availability improvement by increasing all the MTBFs by a factor of 2, but instead the MTBFs of the devices that were dominating the downtime were increased by a larger ratio while other device MTBFs were left unchanged. The result is a set of MTBFs that if achieved will allow the ILC to have an 85% uptime. Water cooled magnets and their power supplies and power supply controllers are the components which need the largest improvement (a factor of 10-20). This is not surprising as these devices are typically single points of failure and there are thousands of them in the ILC design. Hardware R&D projects have been started to develop high availability versions of some of the devices that need significantly improved MTBFs.

## CONCLUSIONS

The availability simulation has been a valuable tool in the design of the ILC. It has been used to make major design decisions and to determine what components need to have reliabilities much greater than has typically been achieved in operating accelerators. The tool will continue to guide the ILC design and help determine where efforts should be concentrated on design and high availability R&D. It is a general purpose tool that can be used for other accelerators.