CHAPTER 8

Reliability, Availability, and Operability

8.1 CHARGE AND ORGANIZATION

When ICFA commissioned the second ILC-TRC report, two working groups were formed, the first on Technology, RF Power and Energy Performance and the second on Luminosity Performance. Part of the charge to the Technology group was to "determine whether the machines can reliably reach their operating energy, [and] be tunable." The Luminosity group was charged to "analyze all those factors which affect the ultimate performance (both peak and integrated), including . . . tunability, and reliability." Because the issues of technology and luminosity reliability are so intimately coupled, a third joint Working group was formed with members from each of the primary working groups.

Working Group Members

Members from Technology Performance:
R. Pasquinelli, FNAL (Subgroup leader); C. Adolphsen, SLAC;
Y.H. Chin, KEK; H. Edwards, FNAL; K. Hübner, CERN; M. Ross, SLAC;
T. Shintake, KEK/Riken; N. Toge, KEK; H. Weise, DESY

Members from Luminosity Performance: **N. Phinney**, SLAC (Subgroup leader); R. Assmann, CERN; W. Kozanecki, CEA/Saclay;
D. Schulte, CERN; P. Tenenbaum, SLAC; N. Walker, DESY

The original goal of the Reliability, Availability and Operability Working Group was to estimate the expected machine availability for each project based on projected reliability of components and subsystems and overall operability considerations. Unfortunately, there was not sufficient data available for any of the projects to make plausible availability estimates. Instead this chapter merely documents the existing partial information on component failure and replacement times. It discusses the impact of such failures on machine uptime, the requirements on machine protection systems, and an evaluation of the tuning time required to reestablish luminosity after an interruption.

8.2 INTRODUCTION

To deliver high integrated luminosity over several years of operation, a linear collider must not only achieve the desired energy and peak luminosity, but also have a very high hardware availability and operating efficiency. The goal would be to have a hardware availability of 85% and deliver on average 75% of nominal luminosity, which is an extremely challenging task. All of the collider designs are very large, complex machines which are pushing the state of the art, with many thousands of components. Unlike a storage ring where the injector chain is only required for filling, the entire linear collider injector systems must operate on every pulse. High luminosity is only achieved when the collider produces and preserves beams of very small emittance. This requires precise alignment of magnets and structures using beam-based techniques, and a variety of sophisticated tuning procedures, which increase the time to recover full luminosity after an interruption. The relevant recovery time for an accelerator is not mean time to repair the component but mean time to reestablish luminosity, *i.e.*, Mean Time Beam Off (MTBO).

While studies analyzing mean time between failures (MTBF) and mean time to repair (MTTR) have been employed routinely in industry for many years, reliability has only relatively recently been given a top priority for accelerators. Most earlier high energy physics machines were innovative technological feats where the emphasis was on achieving breakthroughs in energy and luminosity, usually under tight cost constraints. They were designed and built by teams of skilled physicists, engineers, and craftsmen, who considered the accelerator itself as part of the experimental effort. Given the overhead of fills and ramping for storage rings, the luminosity uptimes achieved were in the range of 50%. This philosophy has evolved with the advent of accelerator user facilities such as the synchrotron light sources, and with the new generation of high energy physics "factories." The large energy-frontier machines such as the Tevatron at FNAL, HERA at DESY, LEP at CERN, and SLC at SLAC have achieved hardware availabilities in the range of 70–90%. In contrast, the B-factories at SLAC and KEK have closer to 95% availability. Synchrotron light or spallation sources have invested significant effort into reliability and now reach 98–99.5% [1]. Operational statistics from a variety of accelerator facilities are presented in the next section.

While the high reliability data is encouraging, great caution is advised in extrapolating the performance of synchrotron light sources (injected beam power <10 W) and cryogenic machines such as LEP and CEBAF to estimates of LC reliability. The design of modern proton linacs, such as the Spallation Neutron Source (SNS) at Oak Ridge, is overwhelmingly constrained by restrictions on beam losses (< 1W/m) that cause residual component activation. The primary reason for this is to facilitate safe hands-on maintenance of the accelerator components. Although protons cause about 100 times more activation per watt than electrons, linear collider designs call for up to 10 times more beam power than the 1 MW SNS (TESLA 11.3 MW/beam, NLC 6.9 MW/beam), so the fractional loss limits are quite severe. For example, the beam power in the TESLA injector complex where there is a large undamped beam is 226 KW, and in JLC-X/NLC only a factor of 4 smaller. A concentrated loss at the level of 100 W (only 0.05% for TESLA, 0.2% for JLC-X/NLC) will cause activation, and great care must be taken to localize these losses in suitably shielded areas. Experience at SLC (30 KW/beam) showed that, in addition to activation, radiation related component failure was a leading cause of problems.

Historically for each new accelerator project, a few of the major components such as magnets or rf power systems were carefully scrutinized with respect to reliability, but the same rigor was not applied to all systems. Significant resources are required for such in-depth analysis and testing, as well as for the added engineering or redundant solutions which would provide high reliability. A recent study for the European Spallation Source (ESS) estimated that a linac designed for extremely high reliability would cost 50% more than a conventional linac [2]. This is not directly applicable to a linear collider as the goals and criteria are different but it is clear that the desire to contain costs always interferes with the desire for high reliability. In a rigorous life cycle cost analysis, it is critical to balance the value of lost physics output against initial project costs.

Poor reliability can impact the peak luminosity achievable as well as integrated performance, as demonstrated by experience with the SLC and more recently with recommissioning the upgraded Tevatron and HERA. If the hardware interruptions are too frequent, then the machine is not up long enough to effectively make progress on the luminosity issues. It was only after the SLC achieved reasonable reliability that the many beam tuning challenges for a linear collider could be addressed. The more complex next generation of colliders must be designed for high availability so that the inevitable challenges can be addressed effectively. More discussion of many of these issues can be found in the NLC Zeroth-order Design Report (ZDR), Chapter 17 [3].

This chapter will attempt to contrast and compare the proposed collider designs and assess the associated reliability issues. It is not possible at this stage to estimate the availability each project may achieve as none of them have completed sufficiently detailed engineering studies. Even for the critical main linac rf components, no project has yet accumulated enough hours of operation at nominal parameters to reliably estimate MTBFs or fault rates for individual components, due to a lack of either final components or test facilities. In addition, the projects are at very different stages in the design process, making comparison difficult. The JLC-X/NLC has the most complete and detailed design for the entire accelerator complex, while TESLA has much more advanced manufacturing studies for the main linac rf, and CLIC is still at an R&D rather than engineering stage. Given the present lack of information, we will simply summarize the issues, list the reliability assumptions, and compare them with what has been achieved with existing technology. Where no information is available for a particular project, that project will simply not be mentioned. We will try to highlight areas where significant work on reliability issues has been performed, and identify areas of concern or where additional R&D is warranted.

Due to the inherently large power densities in the beams of a linear collider, there is a serious risk of damage to beamline components by an errant beam. An extensive machine protection system (MPS) is necessary, monitoring a large number of accelerator systems and beam parameters. The system must inhibit beam in case of a fault and automatically execute a recovery sequence, starting from a benign pilot beam and proceeding to full operating current and rate. Both NLC and TESLA have developed conceptual designs for these systems but much more work is required. The systems will be extremely complex and they must be both robust and redundant. Tuning and recovery procedures are of critical importance for these colliders and they will be discussed in a separate section. Finally, the impact of the machine configurations on commissioning and maintenance will be mentioned.

8.3 COMPILATION OF RELIABILITY DATA

8.3.1 Large Accelerators

To better understand the existing operations of current facilities, the committee has collected significant operational statistics from many of the world's large accelerator complexes. These data describe what has been achieved. They will be compared with the requirements of the linear colliders and thus indicate which components and subsystems require particular effort in order to reach a desired performance.

To predict the integrated luminosity, or integrated beam time delivered by an accelerator, there are three important quantities, HA, BE, and NL.

HA is the fraction of time the machine hardware is available to produce beam. Hardware downtime includes both unscheduled repairs (something critical breaks), scheduled repairs (either at regular intervals or when enough problems have accumulated), and all associated cooldown, warmup and recovery times. For an accelerator, one must consider not only how long it takes to repair a failed component but the total time the beam is off because of the fault, including time lost due to access and the time taken to retune the beam.

BE is the effective fraction of beam time actually delivering luminosity. Beam inefficiencies include Machine Development (time spent studying and improving the accelerator), the impact of tuning procedures, injection and the luminosity decay during a store (for storage rings), Machine Protection trips and recovery (for linacs), and last but not least, the simple fact that accelerators do not manage to deliver the same luminosity every pulse or every store.

NL is the nominal luminosity during a particular run. It may be more or less than design, but usually increases steadily as the accelerator becomes better understood. For a storage ring it is the typical luminosity at the beginning of a store. For a linear collider, it would be the luminosity when the beams are colliding well.

Table 8.1 illustrates the first two measures of performance for modern accelerator complexes, the hardware availability (HA) and/or the overall beam availability (BA), which is the product of HA and BE (BA=HA×BE). It should be noted that different facilities have slightly different methods of accounting for some inefficiencies. Typical synchrotron light source performance metrics use the full calendar duration of the run as scheduled hours so maintenance shutdowns and the associated recovery reduce availability. This is an appropriate way to treat such downtime as modern accelerators do not require routine maintenance and such interventions are only 'scheduled' when there is broken hardware that needs repair. However, for largely historical reasons, many of the numbers in the table include such repair downtime in BE. It should also be noted that the synchrotron radiation facilities are often more tolerant of variations in beam quality, which partially explains their higher achieved BA.

The analysis of recent PEP-II performance shows the relative magnitudes of these quantities [4]. PEP-II was intended to be a "B Factory," and accordingly, some special care

TABLE 8.1: Availability Performance of Various Accelerator Complexes. HA is Hardware Availability; BA is overall Beam Availability, or the product of HA and the Beam Efficiency BE. FNAL data from 2001–2002, SLC data from 1992–1996, PEP/SPEAR data from 2001, HERA data from 2000, LEP data from 1999–2000, KEK data from 2000–2001, APS data from 1999–2002, and CEBAF data from 1999 and 2001.

	FNAL		SLAC					DESY		
System	Tev		SLC [a]	PEP inj	PEP-II	PEP-II	SPEAR	HERA	TTF	TTF
	HA	BA	HA	BA [b]	HA	BA	BA	HA	HA	BA
Cryogenic plant [%]	98.8		na	na	na		na	98.6	97.5	
PS and magnets [%]	92.6		–	98.0	98.0		97.8	94.8	100.	
RF [%]	95.2		–	99.1	99.1		98.3	96.5	98.0	
Utilities [%]	98.3		–	98.0	98.0		99.3	99.5	99.0	
Vacuum [%]	98.1		–	99.7	99.7		98.7	99.1	99.8	
Controls [%]	99.1		–	97.8				99.7	98.8	
Other [%]	98.5							94.1	96.0	
Percent uptime [%]	82.0	50.0	81.1	92.8	94.8	71.5	94.2	83.4	89.5	75.0

	CERN			KEK	ANL	LBNL	TJNAF
System	LPI	SPS	LEP	KEKB inj	APS	ALS	CEBAF
	BA	BA	HA	BA	BA	BA	HA
Cryogenic plant [%]	na		97.8	na	na		98.8
PS and magnets [%]	99.7		99.7	99.5	98.8		94.2
RF [%]	98.7		98.5	98.6	99.0		95.7
Utilities [%]	99.7				99.8		97.3
Vacuum [%]			99.5		99.3		99.1
Controls [%]	99.7		99.7	99.0	99.5		99.5
Other [%]	99.6		93.3	97.5	98.4		81.9
Percent uptime [%]	97.4	55.0	89.2	94.0	95.0	96.0	70.0

---

[a] SLC numbers not broken down into systems

[b] BE weighted by the need for beam by PEP-II

was taken during the design and construction process to make it reliable. Even though the operation of the storage ring complex is subject to quite different constraints compared to those expected at a linear collider, it is useful to describe its performance record. PEP-II hardware is available 94% of the time (*i.e.*, not broken)and the linac, injector, transfer lines and damping rings are available when PEP-II needs them 92% of the time. Short shutdowns taken for maintenance (incipient failure or component repair) take another 12% and are typically classified as "scheduled" maintenance. Together this gives an overall HA of 76%. When PEP-II is running, the efficiency of delivering luminosity or BE is about 63%. This includes luminosity decay during a coast (12%), injection time (15%) machine development (6%), as well as the difference between achieved and nominal luminosity. The integrated luminosity over a run can be calculated by multiplying these quantities by the NL. For example, in the five months from Feb through June 2002, the peak luminosity was about 4.5×10^{33} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and PEP-II actually integrated 28.0 fb^{-1} over 150 days. This calculation gives [0.76×0.63×4.5×150 days×0.0864] = 27.9 fb^{-1}. Together the overall PEP-II efficiency HA×BE is [0.76×0.63]$\simeq$ 50%. For a linear collider where one can neglect the luminosity decay and the injection time, then HA×BE would become $\simeq$ 65%. A similar additional inefficiency beyond HA has been seen at both SLC and TTF, and is reflected in the TTF numbers in Table 8.1.

8.3.2 Extrapolation

Table 8.1 provides the basis for a rough extrapolation of BE and HA to a machine the size of the linear collider. The linear collider is about ten times the size of PEP-II, in terms of its footprint, number of components and total power usage. A very crude extrapolation, without consideration of hardware accessibility (worse for the LC than for PEP-II) and hardware redundancy (presumably better for the LC), gives an expected HA of 40%. As is known from the SLC, HERA and Tevatron commissioning, the BE drops faster than the HA, so the expected BA would be much worse, perhaps below 10%. The extrapolation can be validated in part by looking at routine HERA and Tevatron operation, where the BA, scaled by the actual luminosity compared to the peak, is perhaps 30–40%.

Therefore, if the typical reliability of existing HEP accelerators is simply scaled to the size of a 500 GeV linear collider, then the resulting uptime will be unacceptably low. This means that a new approach is required, and following previous practices will not be adequate. Reliability must be addressed up front by failure analysis, and appropriate remedies must be implemented. Adequate engineering margins for components are also essential. The key issue is the allocation, during development and construction, of sufficient engineering and financial resources to produce a reliable system.

8.4 RF COMPONENTS

Because of the large number of components and relatively short MTBF, the reliability and redundancy of the linac rf systems are key issues in operational performance. In particular, the klystrons (and some modulator components) must be replaced frequently and are considered a consumable expense. In addition, the modulators, klystrons, distribution system, and structures or cavities will experience brief faults or breakdown events where the hardware can be reset and continue operation after an appropriate timeout. All linear collider designs plan to include spare rf units which can be switched in when a unit faults or needs repair. Critical issues are the frequency and impact of faults, the adequacy of the spares overhead, and the accessibility and duration of repairs.

8.4.1 Main Linac RF Systems

Both TESLA and NLC have constructed test facilities for testing and prototyping the main linac rf components. TTF has over 13,000 hr operating experience with 2 cryomodules (16 cavities), a 1-cavity injector module, 2 klystrons and 2 modulators. Much of this time has been spent running beam for the FEL, which is good for operational discipline but was at relatively low gradients (14 MV/m) and 1 Hz repetition rate. Cavity testing has used another klystron and modulator with over 20,000 hr of operation. NLCTA has 3 modulators and 5 klystrons which initially powered 9 m of accelerator structures, but have recently been used for high gradient testing of up to 4 test structures at the same time. Since 1997, it has accumulated perhaps 20,000 hr of operation, testing a variety of structures at gradients from 35 to 90 MV/m, most of the time without beam. Although most components are not yet of the final design, this experience has been invaluable for both projects and provides an excellent indicator of where critical problems may arise and where specific attention is merited. Because of the prototype nature of the systems and studies, none of the projects has accumulated adequate running time with the final designs, making it difficult to estimate the MTBF for the rf components.

At SLAC and other installations with substantial quantities of klystrons, mature klystron designs have achieved in excess of 40,000 hr lifetime. Nonetheless, state of the art techniques in klystron construction are planned for all projects so there is no directly applicable experience base. The TTF 5 MW klystrons and the SLAC XL4 solenoid focused X-band klystrons have been used reliably for 5 years, but both projects plan to use more efficient klystrons for the collider. The first prototype TESLA multibeam klystron (MBK) achieved the requirements for LC operation, and has operated for about a year (5000 filament hours) at 3–4 MW. The first series production MBK accumulated only 2000–3000 hr in TTF before it was removed for repair. The KEK and SLAC X-band PPM klystrons are only now approaching nominal performance. Conventional modulator designs using thyratrons have frequent failures with an MTBF of 15,000 hr [5]. Modern solid-state designs are expected to have a much longer lifetime and some configurations can easily incorporate features to make them essentially fail-safe. Again there has not been sufficient operation time to date for reliable extrapolation.

In the JLC-X/NLC main linac, the rf unit is an 8-pack consisting of a single modulator powering eight klystrons, which in turn feed four girders of accelerator structures via SLED-II pulse compression systems. Modulator or klystron failures remove the entire 8-pack from service, as do most faults. Both modulators and klystrons are located in an accessible support housing so repairs can occur without interrupting operation. There are 13 spare 8-packs out of 245 (5%) per main linac. Of these, 3% are allocated for 8-pack failures and 2% for feedback and structure trips. The modulators are solid-state IGBT designs with extra boards included which provide excess capacity (4 spares out of 152 or 2.6%). Failed IGBT drivers are bypassed and the overall input voltage is adjusted automatically to compensate. The IGBT switch is believed to have a lifetime of 100,000 hr but, since repairs can be deferred until the next maintenance period, even a short modulator MTBF should not affect availability. Since the design is modular, the time to replace an IGBT board is short. There are also common mode failures which would remove the entire modulator from service. Assuming an 8 hour MTTR for the entire modulator, the MTBF required for these is only 1200 hr.

A variety of scenarios were studied to verify that 3% overhead was adequate with a conservative assumption of 20,000 hr klystron MTBF [6]. These were also used to estimate the number of crews required for klystron replacement. The scenarios included both start-up and nominal operation, with a variety of worst case assumptions, for a range of crew arrival times. The start-up cases assume that a subset of the klystrons (10% or 20%) will have infant mortality and only a 1000 hr life. Worst case scenarios included the situation where all failures occur on one linac, and the situation where four days a year the failure rate is three times the average. (This last scenario corresponds to an increased rate of failure after a downtime or power outage, as seen at the SLAC linac. It is also appropriate for a Poisson distribution of very low probability events.) Klystron replacement time was estimated at 8 hr with a delay of from 12 to 24 hr before the crew began repairs. Even with these pessimistic assumptions, the overhead available was adequate for all nominal operation scenarios and was exceeded only in the most extreme start-up cases, indicating that the planning is fairly robust. A summary of minimum acceptable MTBFs is given in Table 8.2.

TABLE 8.2
Required MTB Failures for NLC modulators and klystrons with 3% spares allocated.

	Recovery Time [h]	Number of units per linac	Required MTB Failures [h]	Achieved MTBF [h]	(basis)
Modulator	8	254	1200	n/a	
Klystron	8	2032	9200	>40000	SLAC S-band

When SLED-II or a structure breaks down, the power to the associated girder is inhibited, but the other three girders powered by that rf unit continue to operate. After the fault, the girder is ramped back up to full power over a period of 10 s. Modulator and klystron faults (likely arcs) differ in that they will shut off power to all four associated girders. The recovery time in these cases will be about 10 s as well. Table 8.3 lists reliability requirement estimates where, for the given recovery time, the Mean Time Between Faults (MTBF^{ault}) is computed such that the indicated overhead would be depleted only once a year on average due to that type of fault (in such cases, one would have to wait some fraction of the recovery time to be able to resume operation at full energy). Although the actual recovery times will depend on the type of fault, the values here set the scale for the minimum allowed MTBF^{ault}s. The 8 hr MTBF^{ault} listed here was achieved with a test structure operating at 90 MV/m, well above the design gradient of 65 MV/m. This structure did not incorporate the iris size and damping required for the final design, but indicates what should be achievable (see Section 6.5.2).

For TESLA, the rf unit is a modulator powering a single klystron, which in turn feeds three modules of twelve cavities each. A modulator or klystron failure removes a 36-cavity unit from service, as does any klystron, modulator or cavity fault. There are

TABLE 8.3
Required MTB Faults for NLC main linac rf components with 2% overhead allocated.

	Recovery Time [s]	Number of units per linac	Required MTB Faults [h]	Achieved MTBF [h]	(basis)
Modulator	10	254	4	n/a	
Klystron	10	2032	30	n/a	
SLED-II	10	1016	0.6	n/a	
Structure	10	6096	4	8	NLCTA

286 modulator/klystrons and 10296 cavities per linac. The modulators are located in the cryohalls and are accessible for repair. The klystrons, transformers and the high power cables from the modulator are located in the accelerator tunnel and can be replaced or repaired only on maintenance days. The klystron and transformer will be designed for quick connect and disconnect. The actual change out time should be a few hours.

The TESLA modulator uses solid-state (IGBT or IGCT) switches and a bouncer circuit that minimizes the stored energy required. The present design has only 15 switch units per modulator (as opposed to JLC-X/NLC with ~160) so reliability of the switch components is less of an issue, and methods for providing redundancy are under study. Three prototype modulators have been built, the last two with the planned IGBT design. These two have about 13,000 hr each of TTF linac operation at 1 Hz. The other modulator of an earlier design has been used primarily for cavity, coupler and waveguide testing, with almost 20,000 hr of operation at an average of 3.6 Hz. The TESLA modulator is specified at a MTBF of 2×10^9 pulses (100,000 hr). Repairs are estimated to take a few hours.

In TESLA, cavity faults can be due to breakdown in the couplers or a quench in the cavity itself [7], either of which requires all 36 cavities in the rf unit to be turned off. Coupler breakdown may under some conditions require disconnection of the coupler/cavity and proper termination of both the rf distribution and cavity systems. Since a quench is a thermal process, it is possible to detect most events as they develop and take action to reduce the klystron rf power either within the pulse or before the next pulse. During the high gradient test at TTF in early 2002, recorded trips averaged 11 events per day, for a MTBF^{ault} >17 hr per cavity operating at gradients between 19 and 22 MV/m. This number includes both cavity and coupler trips, and other events such as frequent gun rf trips. In addition, the conditions of the TTF run were far from optimum, with cavities operating very close to their limit and with unstable beam loading conditions.

At 500 GeV TESLA cavities should be some MV/m away from their quench level (assuming higher gradient cavities installed). Unlike normal structures which can breakdown below their operating gradient, superconducting cavity quenches disappear when they are operated a few % below their limit. The LLRF will also incorporate a soft inhibit based on measured beam intensity and individual cavity gradients. Extrapolating this limited TTF experience to 20,000 cavities, 1 trip/cavity in 30 hr would be an acceptable rate (*i.e.*, would deplete the spares once per year on average) only if there was 2% overhead available for faults and the recovery time was less than 3 s. This 2% overhead would need to be in addition to the overhead required to cover klystron failures. No data is available on fault rates for klystrons or modulators during this run, but these rates were low.

RF power input couplers have historically been a severe problem both at CEBAF and at LEP. Considerable R&D has gone into coupler design, including computational tools for analysis of multipacting, implementation of a dc bias between the antenna and the outer coax, and minimization of the warm rf surface along with increased vacuum pumping. The couplers have two ceramic windows between the cavity and external wave guide. This is to greatly reduce the probability of a vacuum failure. A coupler problem that would make a cavity-coupler system inoperable, such as coupler breakdown, would require disconnecting the rf waveguide feed to that cavity and detuning the cavity. One effect that takes time (∼year) to develop is the desorption of gases from the warm rf window area and re-condensation in the cold window area, which then can cause electron emission or coupler events. In the recent run at TTF, coupler events were not a problem (few if any events were recorded) but the coupler operated somewhat below nominal power. Continued long term operational tests of modules at TTF-II will provide extended experience and failure statistics.

RF overhead is required to cover all units with failed klystrons, transformers or low-level rf until the next access for repairs. It must also cover failed modulators until the repair is complete and brief klystron, modulator or cavity faults. In the TDR, the design was to have 2% overhead or 5 stations per linac out of 286. Some of this overhead has since been taken up with fine tuning of the accelerator design by BNS damping and other phase offsets, leaving only 1.3 GeV or less than 2 klystrons per linac. This certainly appears too small even with a 40,000 hr klystron MTBF, so some modification of the parameters will be required. At 500 GeV c.m., use of piezo tuners for Lorentz force compensation would easily allow increasing the gradient to 23.8 MV/m to restore overhead, assuming higher gradient cavities have been installed. Another option is to lower the bunch current and accept a slight reduction in luminosity.

In the JLC-C main linac design, an rf unit consists of a pair of klystrons powered by a single modulator, which in turn feeds four 1.8-m accelerator structures. There are 5% spare units out of 848 per main linac. The modulators use an updated PFN design and a thyratron as switching device, with an estimated lifetime of 10,000–15,000 hr. A klystron lifetime of 50,000 hr is assumed. Most failures remove both klystrons from service, and both modulators and klystrons are located in an accessible support housing.

For CLIC, the klystrons and modulators in the drive beam accelerators are L-band. Each klystron has its own modulator, and pairs of these units power a single accelerator structure. A failure of either klystron/modulator of the pair removes the unit from service. The two accelerators are located in a cut-and-fill tunnel near the surface, with two galleries on either side of this tunnel housing the klystrons while the modulators are on the surface. In this arrangement the elements of the power sources are in accessible areas. There are 112 pairs of klystron-modulator units per drive linac which provides a margin of about 10% since the nominal klystron power is 50 MW but only 44 MW are needed. There are additional units for the rf deflectors of the combiner rings (2 × 2 of 50 MW at 937 MHz and 2 × 2 of 20 MW at 3.75 GHz) and for the deflectors of the delay loops (2 × 1 of 1 MW at 468 MHz). Spare units must also be provided for these parts. The modulators use a PFN design, which is expected to have about the same lifetime as achieved for the other schemes but no prototype has been constructed. A klystron lifetime of 30,000 hr was specified in the klystron design study.

A fault in any main linac structure or in any Power Extraction Transfer Structure (PETS) of one 625 m long section of the drive-beam decelerator presently takes that whole section of linac and decelerator off-line, *i.e.*, approximately 450 transfer structures and two times as many CLIC Accelerating Structures (CAS). Some solution must be developed for isolating only a few structures in the case of a trip. First ideas have been tested at CTF2 and gave promising results, but have not been pursued for lack of resources. The solution would be implemented on the PETS where it would suppress the power in the pair of accelerator structures fed by one PETS as well as in the PETS itself, regardless of whether the fault occurred in the main linac or decelerator structures. Faults in the main linac structures are much more probable than in the drive-beam decelerator because the peak surface field in the PETS is only around 100 MV/m whereas it is 300 MV/m in the CAS.

8.4.2 Other RF Systems

For NLC, the pre-linacs, damping rings, and bunch compressor rf systems all have klystrons and modulators located in an accessible support housing to allow for non-invasive repair [8]. In the injector linacs, the klystrons are powered in pairs (2-packs) by one modulator. In the 6 GeV pre-linacs and e^- drive linac, there are 2 spare 2-packs out of 18 or 19 total (12%). In the booster linacs, there is 1 spare 2-pack out of 6 (20%). The S-band klystrons are assumed to have a lifetime of 40,000 hr, a conservative estimate given SLAC klystron lifetimes. The same lifetime is assumed for the L-band klystrons. Some thought has gone into redundancy for the specialized rf systems but not all details are complete. For the second bunch compressors, the X-band rf has 4 klystrons of which only 3 are required. There is no explicit redundancy yet included for the first bunch compressors and injector capture sections.

The NLC main and pre-damping rings have no spare cavities as they would add unwanted impedance. The ZDR called for three 350 KW tubes per main ring and 2 tubes for the pre-damping ring, but present plans are to use a single 1 MW tube per ring. The klystrons are similar to those used for PEP-II or KEKB with an estimated lifetime of 30,000 hr. This is a reasonable number for well engineered devices (LEP had a klystron MTBF >25,000 hr). The actual average lifetime of the PEP-II klystrons has been ∼6000 hr, dominated by a number of early tube failures. Later tubes have reached the target 25,000–30,000 hr [9]. At KEKB, the 1.0 MW klystrons have an 18,000 hr lifetime and the 1.2 MW Toshiba tubes about 47,000 hr (excluding initial failures in <1000 hr) [10]. It would be possible to configure the system with extra klystrons as hot spares, but this is not presently included.

For the TESLA damping rings, all klystrons and modulators are located in a hall, accessible for repair. There are a total of 4 klystrons and one spare for the e^+ ring, and 2 with 1 spare for the e^- ring. A MTBF of 40,000 hr is assumed. For the injector linacs and bunch compressors, the klystrons are located in the tunnel and the modulators in a nearby cryohall, as for the main linacs. Each 5 GeV linac has 9 klystrons each of which feeds 2 cryomodules operating at a gradient of 20 MV/m. Failure of up to 2 klystrons could be compensated by raising the gradient of the remaining modules to 25.7 MV/m. The compressors have 4 rf stations each of which feeds 3 cryomodules; only three rf stations are required for operation (1 spare). Both the electron and positron pre-accelerators use warm rf. For the positrons, nine 10 MW standard MBK klystrons (40,000 hr MTBF) are required to deliver 250 MeV. The klystrons together with the modulators will be located in a shielded area, and will be accessible during operation. The electron source has a total of 4 standard MBKs, three of which drive the first normal conducting cavities, while the fourth drives 2 cryomodules. All klystrons and modulators are accessible in a separate hall, along with spares.

In the CLIC scheme, the primary electron linac and the pre-injector linacs have 28 L-band klystrons operating with 80 MW peak power; the injector linac has 24 L-band klystrons at 40 MW; the booster linac after the damping rings has 52 S-band klystrons at 40 MW. These klystrons and their associated modulators are expected to have lifetimes equivalent to the mature SLAC components. The rf systems for the pre-damping and damping rings are not yet designed.

8.4.3 Low Level RF

All of the proposed linear colliders will require extensive, complicated Low Level Radio Frequency (LLRF) systems. For years, LLRF systems have been implemented with analog hardware, as digital techniques did not have the computing speed to handle the required frequencies and bandwidths. With the advent of cellular wireless telephony, significant improvements have been made in the digital rf technology sector. As such, many existing accelerators have adopted digital techniques for the implementation of new LLRF systems.

In a pulsed cold linac, the LLRF system is much more complex than in a normal conducting linac where neither fast intra-pulse feedback or intra-pulse exception handling is possible. Older CW superconducting rf systems, such as those at LEP and CEBAF, have relied on simpler analog implementations. At TTF, the LLRF is done digitally [11]. A further complication with the TESLA design is that many cavities are driven by one klystron and there is substantial Lorentz force detuning.

The tight tolerances due to the very high Q of the superconducting cavities and to variations in beam loading make it necessary to use high speed, 100 KHz bandwidth, feedback as well as feedforward to stabilize the accelerating field. Modern electronic ADCs and DACs are quite capable of the speeds required ($\sim$100 MHz), but the high speed signal processing has yet to be done cleanly. Laboratories that adopted the new digital techniques have consistently greatly underestimated the level of professional manpower it takes to get a system up and running. The field has become so specialized that individuals with the combined rf and computer knowledge are in short supply.

There is very little high intensity operational experience with large-scale digital LLRF systems and gaining this experience should be an urgent priority for TTF. A digital rf system was installed at the PEP-II B factory and one is under construction for the SNS. Even though PEP-II is an e^+/e^- storage ring complex, many of the challenges are similar to those faced in the TESLA design, including the needed high level integration into the control system and the diagnostics that allow precise system optimization at the highest intensities. Experience with the PEP-II system has been very sobering and has shown how important it is to have a well-designed and well-supported system.

In the normal conducting linac the LLRF must program the phase transitions to control the pulse compression. Since the rf pulse is too short to allow beam intensity variations to be compensated by feedback, a feedforward scheme based on a beam intensity estimate from the damping ring will be used. This scheme was tested and used at the SLC. In all proposed linear colliders, the LLRF is a key component of the machine protection system. The system must be designed with an appropriate level of redundancy and without possible simple single-point failures in order to avoid beam damage in the linac cavities.

8.4.4 High Power Microwave Components

All of the collider designs will deal with high microwave power, tens to hundreds of megawatts. Since the required power levels have not yet been fully demonstrated, the reliable operation of high power components is a concern. Another issue is the segmentation of the power distribution systems. All designs have relatively large regions connected by a common vacuum or cryogenic system, with the result that a failure of a single component can potentially have a major impact and require a lengthy recovery after repair.

The TESLA design has the lowest power requirements, with only 10 MW peak power from the klystrons to 36 cavities and no pulse compression required. For 500 GeV operation, many components are being tested at TTF. The higher gradient 35 MV/m cavities required for energy upgradability will require higher drive power and further R&D on components. Even for these, the power required is modest and problems are not expected. The JLC-X/NLC designs have 510 MW peak power delivered to each 6 structure unit. The CLIC design has 460 MW peak power delivered to the structures, the highest rf power per structure. Due to the distributed two beam nature of the CLIC design, the rf power has only a very short distance to travel between the primary and secondary beams. The critical components are the power extractor and structure coupler.

8.4.5 Cables

The TESLA design has long cables between the modulator and the klystron, up to 2.5 km in length. For reliability, these should be continuous cables for the full length. Repair or replacement of cables in the tunnel may be difficult and spare cables are planned. It is believed that several manufacturers can produce the needed cables, but a field test of powering a klystron with such a long pulsed cable is essential. A test cable (few km) will be delivered and tested at TTF beginning in 2003. In the other LC designs, the klystron is in close proximity to the modulator and long power cables are not an issue.

8.4.6 Evaluation—RF Components

As stated earlier, critical reliability issues for the rf systems are the frequency and impact of faults, the adequacy of the spares overhead, and the accessibility and duration of repairs. It is difficult to estimate the MTBF for any of the components as none of the projects has accumulated adequate running time with the final designs. JLC-X/NLC have taken a conservative approach to the design of the rf system, allocating a large 5% spares overhead for the main linac and providing redundancy elsewhere. All klystrons and modulators are accessible for repair during normal operation. For TESLA, the klystrons are located in the linac tunnel and require access for repair. The original 2% spares allocation has been reduced due to accelerator design modifications, but at 500 GeV this reserve can be recovered by a slight gradient increase. For CLIC, a structure fault would presently take a whole 625 m section of linac off-line and a solution for isolating smaller segments must be found. The warm rf machines require very high peak rf power which has not yet been demonstrated. This is a high priority technology R&D issue. Issues related to LLRF have been covered in Section 6.3.

R&D issues:

- A comprehensive assessment of the MTBF and MTBO for rf components is required, and an adequate fraction of hot spares must be included in the final design.
- CLIC needs to develop a mechanism to turn off only a few structures in the event of a fault.

8.5 OTHER COMPONENTS

This section discusses a number of areas of concern either because they are potential single point failures or because of the large number of components.

8.5.1 Sources

The electron and positron sources for a linear collider are complex systems which in some cases require regular maintenance and for which a high availability is difficult to guarantee because of the large number of different components. All of the designs plan some level of redundancy for the sources. For the positron target and collection systems, redundancy is more crucial as these systems will have such high radiation levels after use that a prolonged cooldown will be required before repairs. Even with multiple targets, a detailed analysis and optimization of maintenance scenarios may indicate that quick disconnects or robotics are desirable.

The NLC plans to have two identical polarized electron sources including the photocathode gun and laser, either of which can provide the required beam. Similarly, the electron source for the positron drive beam will contain two redundant guns. NLC had originally foreseen two redundant positron target and collection systems, separated by adequate shielding to allow work in one area with beam in the other. Recent evidence for a lower target damage threshold indicates that multiple targets will be required for normal operation. The NLC baseline now contains four targets, of which any three are required for operation and one can be under repair. An undulator-based source is also under study where the layout would include two target assemblies for redundancy. JLC has proposed having two electron sources, either both polarized or possibly with one of them a simpler unpolarized source for commissioning. The positron production system is similar to that of NLC.

TESLA plans for two independent sources, one of which would be an unpolarized rf gun for commissioning and initial operation, and the other a polarized source. The rf gun could be changed later to a polarized source to improve the system availability. For the undulator-based positron target, the radiation levels with a thin target will be much lower than for a conventional target. A standby target system with appropriate handling will be available locally so that it can be moved in place if required. Also planned is an electron source to be used for positron commissioning. Both electrons and low intensity positrons should be available.

8.5.2 Magnets and Power Supplies

The reliability of magnets and their power supplies is a potential concern for a linear collider because of the large number of devices, even if individual MTBFs are long. For example, the NLC design includes about 4600 dipole, quadrupole and sextupole magnets, essentially all of which are required to operate the machine. A failure of any component takes the collider down. There are also nearly 10,000 correctors, BPMs and movers but, in contrast, the collider can continue to operate with a handful of these out of service, as long as the failure is passive (*i.e.*, not runaway). These components need not have a significant impact on reliability as long as they have moderate MTBFs and as long as the control system provides the capability for rapid diagnosis and response to failures.

For the essential magnets, failures of a magnet itself can be minimized by careful engineering. At NLC, much effort was devoted to Failure Modes and Effects Analysis (FMEA) and an improved prototype quadrupole has been built. The major remaining risk is failure of the power supplies or of the controls and interlocks. To minimize the total number of power supplies, NLC has adopted adjustable permanent magnets wherever possible. Where electromagnets are required, reliability can be enhanced by providing redundancy and by stringing magnets to reduce the number of power supplies.

Permanent magnets offer several advantages for reliability. They require no cooling water or high power electrical cable and connections, all vulnerable to failures. They do require a movable adjustment mechanism but in general, the collider can continue to operate if some magnets are no longer adjustable. Of course, they introduce new potential failure modes such as degradation of the PM material. Given the radiation environment, the lifetime and reliability of both the adjustment mechanism and PM material must be verified. Permanent magnets have been successfully used in wigglers and undulators for light sources. The Recycler storage ring still being commissioned at Fermilab is also composed mostly of permanent magnet dipoles and quadrupoles.

In the present NLC layout, one third of the dipoles and two thirds of the quadrupoles are specified to be permanent magnets. These are used in the injectors, bunch compressors, and main linac beam and bypass lines. They are not used in the damping rings because of the radiation load and they are not used in the beam delivery for energy flexibility. Their use in the main linac beamline is still under discussion due to the desire for energy flexibility and the tight beam-based alignment tolerances. Most of the dipoles, sextupoles and damping ring quadrupoles would be powered in strings, reducing the number of supplies needed from 1000 to perhaps 200.

In the linacs, each quadrupole requires an adjustable strength which scales with the beam energy. To enhance reliability, several linac quads can be strung together on a common supply with individual trim supplies to provide the fine adjustment. This has the advantage that it is possible to match around a failed trim supply, whereas it is difficult if not impossible to match around a quad that is off. Where individual supplies are used, full redundancy will be provided with a spare supply for each quadrupole, or in some cases, by including a hot spare for a group of several supplies. The first choice automatically recovers from a failure, the latter requires manual intervention.

TESLA considers mainly three types of supplies in order to provide reliable operation and cover the majority of requirements [12]. Low current supplies for steering dipoles and quadrupoles (50 or 100 A, less than 1 kW) are to be built in a modular fashion (up to six power units per supply) so that if one or two units fail the others can take over immediately. These supplies are located in the tunnel and failed units can be replaced on maintenance days. They cause downtime only when the supply can no longer deliver the required current (*e.g.*, when more than two modules out of six fail), or in the event of a failure of the regulation board, which will not be redundant. MTBF of a supply is estimated to be 200,000 hr. There are 1900 of these supplies in the main linacs and 1260 in the damping rings. On the average it is expected that six main linac supplies will fail over a 4 week period and require retuning of adjacent quads or correctors.

In case of a quad or power supply failure, it is assumed that the optics can be rematched by an automated procedure with some slight degradation to the luminosity. The question of how many failures can be tolerated with what penalty in luminosity needs to be simulated. The quads in the upstream end of the linac where the energy spread is large will be particularly sensitive and may warrant redundant supplies. Failure of a corrector magnet or BPM causes a degradation of the orbit in that region, which can usually be compensated by a tuning knob. As with the linac quads, there may be a gradual degradation of the luminosity, which should be simulated to quantify the sensitivity. A few components are more critical, such as the BDS sextupoles, and they probably require some provision for redundancy. The BPMs associated with the fast IP feedback, as well as some others, will also be essential for operation and require some provision for redundancy.

The medium sized power supplies (up to 600 A, 200 V, 120 kW) will be switch mode supplies, which may be pulse-width modulated. MTBF is estimated at 40,000 hr but the supplies will be arranged in groups of a few with spares that will be remotely switchable in case of the failure of one supply. Mean time to determine the fault, switch and put the spare into operation is estimated at 3–5 minutes. The supplies for the beam delivery and extraction lines are located in the tunnels, but switchable spares will be located in service halls with cables running the length of the BDS. Supplies and spares for the damping rings are located in accessible halls and can be changed out during operation. There are about 120 of these supplies for the beam delivery and extraction lines.

There are a few ($\sim$20) big supplies exceeding 120 kW. These will use SCR technology and will be located in accessible halls. MTBF is estimated at 40,000 hr. Some may have spares that can be switched in with high current switches; others will probably require one-to-one replacement. These supplies are used mainly in the extraction line or damping rings.

8.5.3 Cryogenic Systems

Only the TESLA project will have an extensive cryogenic system. Operational statistics from the large cryogenic installations at CERN, DESY, FNAL and KEK demonstrate that such systems can be very reliable. Availability was generally lower during the first year after installation while crews were being trained and design errors corrected [1]. Subsequently, reliability of over 99% was achieved with >100,000 hr of operation. At TTF, the cryogenic system reliability is 97.5%, the poorest performing subsystem. To achieve a reliability over 99%, redundancy of critical or fragile subcomponents is essential. The proposed TESLA cryogenic facility should be able to achieve the necessary reliability with appropriate engineering.

The TESLA linac is divided into 2.5 km-long cryogenic units. This segmentation choice is influenced by the cryoplant layout with sparse access points, cost considerations and cryostat design constraints. If a part of the linac must be warmed up, for example to fix an internal problem or change a cryomodule, then the whole 2.5 km unit will be warmed up (and a 0.5 km section of insulating vacuum vented). Manual valves on each module can isolate individual beam tube vacuum sections. It is expected that the process, including warmup, repair and cooldown, may take about a month. This may be an availability risk that deserves further study. HERA has 1.6 km vacuum segmentation that has not caused problems. Any risk analysis must also address just what sort of failure would require immediate repair as opposed to a temporary fix and repair in long maintenance periods. Failures that might require immediate repair include: failure of both rf windows in an rf coupler, helium vessel to beam vacuum rupture, large local heat leak. It is important to note that unlike circular machines with magnets, linacs can operate with unpowered modules. In addition, in the cryogenic linac, there are no flange interfaces between the beam tube and helium. Any beam tube to insulating vacuum leaks are at a very low differential pressure.

8.5.4 Vacuum

The vacuum systems of a linear collider pose a potential reliability risk because of their total volume and complexity. There are a large number of pumps for the normal conducting linacs, most of which must be operational; their power supplies and controls may be situated in the tunnel for cost savings. The systems also require a large number of valves to provide segmentation. In TESLA, the consequences of loss of cavity vacuum can be very severe, because of the sensitivity of the cavities to contamination. TESLA has three vacuum systems: the cavity beamline system, the input coupler system, and the cryostat insulating vacuum. The TESLA cavity system has few lumped pumps because of the cryogenic pumping (a total of 200 ion pumps for the linac cavities and one ion pump per module (total 858) for the input coupler vacuum). For the TESLA insulating vacuum there are in total 120 installed turbo pumps. Additional movable pump stations can be added every 50 m for initial pump down or in case of a helium leak to the insulating vacuum.

The number of primary beam line valves in JLC-X/NLC is about 1900 in total for both linacs. The SLED-II system can be easily isolated into 28 m long accelerator vacuum sections. TESLA has a relatively small number of vacuum sectors (16 total). In TESLA there are two manual valves per module in order to allow installation of the modules using a double-lock system so that the cavities remain under vacuum. The manual valves are only operated when the system is warm and the insulation vacuum opened. The TESLA superconducting cavities place constraints on the design of auxiliary components nearby, such as gate valves, BPMs and bellows, so as to minimize particulate matter generation. TTF experience, where all auxiliary components were made and prepared to exacting vacuum dust specifications, shows some degradation in cavities closest to the warm sections. It is not clear if this is due to assembly, inadequate magnetic field shielding or particulate matter from the warm sections.

The danger of a serious disruption of accelerator operation in TESLA by accidental venting of the cavity vacuum is being carefully considered [13]. Following an accidental vent within the cold section of TESLA, such as that caused by errant beam burning a hole in the niobium cavity wall, by input coupler double window failure, or other failure or error, some cavities may become contaminated by the inflow of helium or air. In the case of a helium leak, even though the differential pressure between the helium and vacuum is small, helium will enter some cavities and form liquid. Furthermore, during the warmup that follows, warm gas may carry dust to adjoining modules. In order to minimize this, TESLA is considering automating some of the module manual valves. The present TESLA design includes only one beam line valve per 2.5 km but it would be possible to segment to as short as 150 m sections by automating some of the manual module valves. Even though beam induced holes seem unlikely with the 70 mm diameter cavity aperture and an appropriate machine protection system, both a failure analysis and actual tests seem warranted.

On the other hand, NLC structures have come close to the design performance with little serious concern for dust contamination. The processing of NLC structures is hastened by an *in-situ* 220°C bake. Vent tests have been done with the NLC structures at NLCTA. The recovery is relatively quick, with a three day *in-situ* bake to speed up the processing. Tests of the vacuum performance of the SLED-II components have not been done.

The vacuum specification for the TESLA damping ring straight sections has been tightened to 0.1 nT to prevent collective instabilities (the previous spec was 1.0 nano-torr). This would require an *in-situ* bake, even though there is no synchrotron radiation in the straight sections. The TESLA design calls for 5000 pumps for the damping rings. The NLC damping rings probably also require *in-situ* bake, but the vacuum limit is 10 times higher. Other areas have less stringent requirements and do not pose a concern.

8.5.5 Controls

The control system for a linear collider poses a significant reliability risk itself, unless a serious effort is devoted to careful engineering. There will be a very large number of components distributed over a length of more than 30 km. The distances alone significantly increase the mean time for even the most minor repair, and require that components be remotely diagnosable to a very high degree, well beyond standard practice. This is true whether the remote distance is tens of km at the site or thousands of km in a Global Accelerator Network. This requirement combined with currently available technology lends itself to a model where device control is done by smart network appliances connected by Ethernet. Unfortunately, this model is unlikely to produce the required system availability given the large number of components and typical MTBFs.

Beyond the sheer number of devices, the LC tuning, feedback and machine protection requirements are extremely challenging. These are described in later sections. To reliably deliver high luminosity, the collider depends on extensive feedback and automated tuning procedures, operating continuously. The high beam density implies that an errant beam can easily damage beamline components, necessitating a complex machine protection system involving a large number of components for safe operation. Both the automation and the machine protection require fast communication paths linking distant parts of the machine. Together these requirements make the system intolerant of controller failures and demanding of very high reliability.

NLC has enumerated and evaluated the controls subsystems that form critical single point failures which would jeopardize accelerator component safety. For example, a large error in the overall linac phase probably causes beam loss, which could damage the structures. The timing system, damping ring subsystems (kicker and rf), software systems and machine protection system are all areas of concern. While the cost of redundant networked systems is not large, a careful evaluation is not trivial and must be considered in the design of the machine protection system.

None of the projects have really begun to address these issues in sufficient depth. A few years back, NLC developed a conceptual design for a highly reliable, redundant communication network that minimized exposure to fragile components in the field. At the very least, this design would need to be updated in light of advancing technology. Given the central role of the control system and the complexity and sophistication of the algorithms envisaged, a major effort will be required in this area. It is an important topic for future R&D and eminently suited to international collaboration as most demands are common across designs.

8.5.6 Evaluation—Other Components

All projects envisage some level of redundancy for the sources. NLC has taken a very conservative approach with fully operational spare systems, as has JLC. TESLA has a combination of redundant systems and relatively easy-to-replace hot spares. Redundancy is also foreseen for magnet power supplies, particularly where they are inaccessible as for TESLA. Here the key issue is the allocation, during development and construction, of sufficient engineering and financial resources to produce a reliable system. The vacuum system size and segmentation is an issue for all designs but is also amenable to engineering solutions. TESLA requires a large cryogenic plant which should not itself be a reliability concern. However, the segmentation is such that a very large 2.5 km length of the accelerator must be warmed up together, making repairs very lengthy. A flexible, sophisticated control system is essential for any linear collider, and providing adequate reliability with such a large, complex system will be extremely challenging.

In summary, while there are many areas requiring attention to ensure reliability, there are no fundamental technical reasons why the systems cannot be made reliable. The only clear difference between the projects is the more limited accessibility for TESLA in a single tunnel, which makes the requirements more stringent.

R&D issues:

- A detailed evaluation of critical subsystem reliability is needed to demonstrate that adequate redundancy and MTBFs have been achieved.
- A model for communications and controls with adequate reliability must be developed.
- The TESLA cryogenic and vacuum system segmentation should be reevaluated.

8.6 ENGINEERING MARGINS

The reliability of accelerator components is improved when they are operated well below their maximum rated parameters, *i.e.*, sufficient engineering margins have been allocated for their environment and their support utilities. Perhaps the best example is the question of semiconductor electronics installed in the beam line enclosure near the accelerator, where the air temperature might be well over 40°C, the radiation dose rate might be over 10 rad/hour and the electricity supplied might contain high voltage spikes. Components other than electronics also have engineering margins against which their environment can be rated, but since both the TESLA and NLC designs include tunnel electronics, a specific comparison is appropriate in this case.

For both NLC and TESLA, a substantial lifetime derating factor is required unless controls are used to mitigate the harsh environment. Structure cooling water, nominally at a temperature of 45°C, heats the NLC tunnel. Radiation in the NLC linac tunnel, generated by beam-gas scattering, may amount to 2 W/m. Extrapolating from present NLCTA performance at 70 MV/m and 400 ns, the radiation dose generated by dark current will be similar, but with a softer energy spectrum [14]. The TESLA tunnel is heated by the klystron water which has a nominal return temperature of 80°C. The return water pipes will be insulated, so it is expected that tunnel temperature will remain below 30°C and not change dramatically during short maintenance periods. This moderate and relatively constant temperature has little impact on the cold linac but is probably very important for stable operation of the damping rings. Electronics and power supplies in the TESLA tunnel will be housed in environmental enclosures which must be carefully engineered. Prototypes have been built and will be tested in TTF. Input air and water cooling to the electronics and power supplies will be at 30°C or less.

Radiation in the TESLA tunnel will be from linac dark current, and from the damping ring and other secondary beam transport systems (such as the e^+ target to ring line). Dark current will cause an average load on the cryogenic system via ionization loss to the cavity and helium (cold mass) so the cryogenic capacity effectively sets an acceptable limit for these losses. The goal is 0.1 W/m radiation loss, which is $\sim$15% of the static and dynamic rf load at 2 K. As the cryogenic load is accumulated over 2.5 km intervals, local spots can have higher loss levels and many sections will have less. Losses from damping rings and transport lines will also contribute, even though specific hot areas will be appropriately shielded. Electronic equipment located in NLC and TESLA klystron enclosures and tunnels may be subject to supply voltage spikes due to the nearby pulse cables and klystrons. Proper attention must be given to electromagnetic and rf noise.

8.6.1 Radiation Damage

The lifetime performance derating associated with radiation dose depends on the type of radiation and the nature of the semiconductor components. For a variety of semiconductors, damage is predicted for gamma doses between 10^3 rad (10 Gy) and 10^7 rad and for neutron integrated fluences (1 MeV equivalent neutron damage per cm^3) between 10^{12} and 10^{16}. For the NLC design, the dose rates have been modelled for small, shielded, 0.6-m deep tunnel wall electronics housing enclosures, which would need to be water-cooled. The modelling indicates that dose rates from radiation sources other than dark current are low enough (0.5 rad/hr and 3×10^7 n/cm^2/hr of 1 MeV neutrons) to allow the use of non-radiation hard electronics in the enclosures.

Modelling of radiation in the TESLA tunnel is based on 0.1 W/m deposited in the cold mass. For this loss, the radiation dose rate is 1–3×10^{-2} Gy/hr (1–3 rad/hr) in the tunnel at the electronics location. This dose rate decreases by about two orders of magnitude with the planned 50 cm of concrete shielding. For 100,000 hr operation (20 years at 5000 hr), the dose would amount to 10–30 Gy (1–3×10^3 rad). Neutron fluence is estimated to be 10^7/hr with no shielding and a factor of 10 less with 25 cm of shielding (resulting in 10^{11} in 20 years). It is likely that the radiation distribution along the TESLA tunnel will not be uniform (probably more intense downstream from the quadrupoles) and that the more sensitive electronics can be put in lower radiation areas. Dark current levels follow the Fowler-Nordheim law; a change of a few MV/m gradient can change the dark current generation by an order of magnitude. Thus some management of individual cavity gradients may be advantageous in minimizing dose.

8.6.2 Evaluation—Engineering Margins

Experience at SLC showed that the most serious cause of component failure was exposure to radiation. Most radiation exposure was due to steady, relatively low losses, in congested parts of the machine, not in the bulk of the linac. In TESLA, where all machines coexist in a single tunnel, great care must be taken to shield the linac cryostats from radiation generated by other subsystems. In both designs, further study is needed to evaluate the exposure to radiation of the tunnel electronics.

Tunnel air temperature and humidity are the next most important concern. TESLA, with only one tunnel, has a much greater density of delicate equipment which must be housed in environmentally controlled and shielded containers. JLC-X/NLC can separately control the environment of the support tunnel in order to keep it closer to a nominally benign condition, but may require water-cooled tunnel enclosures.

8.7 MACHINE PROTECTION SYSTEM (MPS)

The Machine Protection System (MPS) is responsible for protecting machine components from beam related damage [15]. It automatically controls changes in beam power, both by halting operation when a fault is detected and by restoring operation when the fault is cleared. The minimum response time, the interval between the occurrence of a fault and the termination of the beam sequence, is one full interpulse period for the short train machines (JLC-X/NLC and CLIC) and about 1/10 of the train length ($\sim$100 μs or 300 bunches) for TESLA. Since it is not possible to stop a given beam bunch once extracted from the damping ring and since a single beam bunch is capable of causing substantial damage, a permit signal indicating the readiness of the downstream systems is required before extraction from the ring is allowed. The permit signal is derived from beam data taken on the previous pulse and from a system that monitors the performance of all devices whose state can change substantially between pulses and which are strong enough to steer the beam into a vulnerable machine component. Before operation can be resumed, the MPS provides for the production of a sequence of pilot and low power pulses that prove the fitness of the downstream systems for high power operation.

8.7.1 Component Vulnerability

Beam initiated damage can result either from low losses lasting many pulses or from high power losses of a single (or a few) bunches or bunch trains. The two effects are known as average power damage and single pulse damage, respectively. The MPS is intended to protect the rf structures, collimators, special instruments and simple vacuum chambers. It is not known for any of the designs if MPS will be responsible for limiting the dose to beamline and support equipment such as permanent magnets and electronics. By their nature linear collider beams have a much greater potential for single pulse damage than beams of present day machines. Tests done at SLAC [16] have shown that a single pulse with charge density of more than 1 $\mathrm{pC}/\mu\mathrm{m}^2$ will perforate a 1.4 mm thick copper iris on which it is normally incident. Calculations done by the CLIC group support these measurements [17]. It is expected that Nb damage will be similar. Simulations done for the NLC structure indicate that, in many failure scenarios, an oscillation large enough to drive the beam into the structure irises does not necessarily enlarge the beam size sufficiently to prevent damage.

Protection of collimators is a special case handled elsewhere in this report (Section 7.4.5). Special instruments may have an associated collimation system specifically for their protection. For the most part, simple vacuum chambers will be hit by grazing incidence, $\sim$1 mrad impact angle beam pulses. Estimates done at SLAC show that an aluminum vacuum chamber will not be perforated by a full NLC train at 1 mrad or lower grazing incidence.

The TESLA beam aperture in the linac is large (70 mm diameter versus 10 mm for NLC), thus it is considerably less likely that the beam can be steered into the cavity walls. Even so there are doubtless failure modes that would destroy some cavities and contaminate others. Full failure mode analysis has yet to be carried out to determine just what the possible fatal scenarios might be. While it might appear that linac quad and steering failures cannot steer the beam sufficiently to reach the aperture, this must be verified by detailed calculations. The pre-linac collimation system is key to assuring that misaligned beam is never injected. For CLIC, simulations with alignment errors indicate that no significant beam loss occurs if a complete drive beam sector fails to deliver rf power, unless the lost sector is one of the first two. In this case, there is severe beam loss due to over-focusing [18].

8.7.2 Permit System

In all designs, the MPS is segmented into three zones: the bunch compressor system, the main linac and the beam delivery, separated by the pre- and post-linac collimation sections which are capable of dissipating the full power. The collimation is not perfect, however, and failures of the bunch compressor system, for example, must be considered a threat to the main linac components.

For both the long bunch train machine, TESLA, and the short bunch train machines, data from position monitors and loss monitors are used to allow operation to continue. In the case of NLC, for example, a beam trajectory outside predetermined limits detected on pulse n-1 would stop operation before the extraction of pulse n from the damping ring. The beam from pulse n would remain in the ring or would be locally aborted. In the case of TESLA, an aberrant trajectory of the first few bunches in the train would be used to stop extraction of the rest of the train from the damping ring. The minimum response time could be as long as 300 bunches, due to beam and signal propagation delays. In neither design is linac damage completely prevented in all scenarios by the above system.

There are a number of fast devices whose field can change substantially in a time less than the minimum response time and which are strong enough to deflect the beam outside acceptable trajectory limits. The list of fast devices is not complete for any of the designs. For the NLC, for example, the fast device list for the main linac includes focusing magnets and diagnostic section dipoles. It is probable that the magnetic field in these devices can change significantly during the interpulse interval. The list of fast devices also includes elements that can cause simple common-mode failures such as programming controls for linac phasing. Consideration of damping ring beam intensity transients is still to be done.

8.7.3 Power Restoring Sequence

Following a fault, or any interruption lasting a long time compared to the inter-pulse interval of the warm structure machines, operation must re-commence with a completely benign pilot pulse. This is because of the slow devices whose field is only indirectly monitored via beam data. Once the complete fast device list is formulated and evaluated, it may be decided to include all accelerator components in the fast device list. The pilot pulse must be only a single bunch (or a few for TESLA) and must have a charge density less than 1 pC/μm^2. For JLC-X/NLC, this means a reduction of between 10^{-4} and 10^{-5} with respect to the nominal charge density, to be accomplished either by an emittance spoiler, reduced intensity or by a combination of the two.

All critical feedback loops and basic beam position monitors must function on the pilot pulse. This is important especially for TESLA where beam loading signals from the cavities themselves will be used to fine adjust the cavity phases relative to the beam and verify that the beam loading compensation is ready for the rest of the sequence. JLC-X/NLC will have cavity phase monitors that will pick up the phase of the beam to check the sector phasing, and these should have little problem detecting the signal from a very small intensity pilot beam. Once the parameters of the pilot bunch's path through the machine have been checked and found within tolerance, the power ramp sequence can begin. For JLC-X/NLC the sequence begins with pilot bunch operation at the full repetition rate, followed by the removal of the emittance spoiler and restoration of nominal bunch charge and finally by ramping up the number of bunches to nominal.

8.7.4 Average Power

For uncooled components, a beam loss of a few watts/m can be a concern. For a cryogenic system, a few tenths of watts/m loss is a cause for concern. The lost fraction of the beam is 10^{-7} and 10^{-8} in each case, respectively. Average power beam loss is detected using conventional loss monitors. Because small average power losses depend on the extrema of the beam distribution, the power restoration sequence processor will estimate average power losses expected at each transition before allowing it.

8.7.5 Other Concerns

Aside from direct damage from impact, the presence of the beam can cause unexpected behavior in other ways, primarily from the electromagnetic fields it leaves in its wake. Two good examples of this are the field left in a tuned, unpowered superconducting cavity and the field in a backphased, powered structure (warm or cold). The LLRF controllers must be programmed to deal with such events appropriately to avoid damage.

8.7.6 CLIC Drive Beam MPS

The power in each CLIC drive beam is 63 MW, about eight times the primary beam power of JLC-X/NLC and five times that of TESLA [19]. In addition to the power handling steps described above, there is serious concern for the mechanical deformation that may be caused by a relatively small, 0.1% steady beam loss. The drive beam power extraction structure efficiency will degrade when there are beam losses in the area and it is possible that will feedback on the beam and create additional beam losses. In order to mitigate this problem, the CLIC drive beam power ramp sequence will include intermediate steps where the peak power is nominal but the average power is greatly reduced. Further study is needed to devise protection sensors and explore the beam power managing sequence.

8.7.7 Evaluation—Machine Protection System

The small, very intense, beams in a linear collider require a new approach to machine protection untested at any existing or soon to be completed machine. Furthermore, the pulsed time structure of the beam, as opposed to the CW nature of the Tevatron or LHC is an additional difficulty. There is a proposed protection system scheme that is feasible for both TESLA and JLC-X/NLC that relies heavily on the use of a pilot bunch and a fast permit system. The scheme needs further design and evaluation.

Since a single, nominal intensity bunch will damage almost any accelerator hardware it happens to strike, there is little fundamental difference in the TESLA and JLC-X/NLC MPS exposure or design strategy. The long inter-bunch interval in TESLA allows the beam to be switched off somewhat more quickly than in JLC-X/NLC. Steady, low intensity losses are also important but these can be handled with a conventional MPS strategy.

R&D issues:

- A detailed Machine Protection System design that meets requirements must be developed, including a careful study of failure modes.
- CLIC needs a detailed design of an adequate MPS system for the drive beam.

8.8 RECOVERY AND TUNING IMPACT

When considering reliability in terms of integrated luminosity delivered, it is necessary to estimate the effect of invasive and non-invasive tuning, where invasive refers to interrupting luminosity. Optimization of performance then becomes a balance between peak luminosity and time lost to tuning. In much of the collider, the component alignment tolerances are extremely tight and cannot be achieved by traditional survey techniques. All of the designs foresee extensive use of beam-based alignment. In addition, the tight tolerances make the machines very sensitive to slow drifts due to temperature and ground motion effects. As a result, beam-based feedback systems are mandatory, and both invasive and non-invasive retuning will be required at intervals. A quantitative estimate of the impact requires detailed simulations of the tuning algorithms and feedback in the presence of errors and time-dependent effects, such as ground motion. The tools to perform fully integrated simulations of a linear collider are only now becoming robust and complex enough to address these questions, and hence any estimates are only a first approximation.

As a framework for assessing the operational impact of tuning procedures, five scenarios were considered:

- Tuning during normal luminosity operation
- Beam-trip recovery (seconds to minutes)
- Recovery after a repair not requiring access to the accelerator housing (hours)
- Recovery after a repair requiring access (shifts)
- Recommissioning after a long shutdown (weeks to months)

The shorter the duration of the outage, the less tuning should be needed to recover adequate luminosity. It is however important to remember that the impact of trips (such as rf unit faults) can be significant because of their high frequency, even though the integrated loss per trip may be small.

In making estimates of tuning durations, it is assumed that feedback systems stabilize the beam position and energy at key locations throughout the complex, that machine protection trips and recovery sequences are completely automated and generally require no operator intervention, and that tuning procedures are essentially completely automated. The most invasive procedures required after a long downtime may require operator monitoring, but must be automated to minimize execution time. Those which occur during routine operation should be completely hands off, including optimization procedures both for the linac emittance and for the beam size at the IP. Needless to say, these conditions will not be met for initial commissioning. The level of automation described goes far beyond what has been achieved at existing accelerators, but it is essential. If one simply extrapolated the time required for more manual tuning procedures to the scale of a linear collider, then the beam availability would be unacceptably low. It should also be emphasized that the time estimates give the minimum time required when everything executes perfectly. Experience with existing machines would indicate that typical times for recovery or tuning are considerably longer than the minimal estimate, even when procedures are automated. We first briefly review the various algorithms.

8.8.1 Tuning Procedures

The tuning procedure for each part of the collider starts with beam-based alignment of structures, quadrupoles, and any higher order multipole magnets [20]. This is a multistep procedure where the first step is to find the relative offsets between the quads and the nearby BPMs by means of quad shunting or ballistic techniques. These offsets define a reference trajectory. The beam is steered through the center of the quads using either dipole correctors or movers. Dispersion-free steering can then be used to find a reference orbit with even lower emittance growth. Closed bumps are applied to cancel residual dispersion or wakefield distortion. Finally, damping ring or IP tuning knobs are optimized. Efficient operation of the collider will require many more feedback and tuning procedures than described here but we have tried to focus on those with the largest perceived impact on efficiency.

Beam-Based Alignment (BBA): Quad shunting (or ballistic alignment) is a slow, time-consuming, invasive procedure which is predicted to take a few shifts for each main linac, even assuming completely automated procedures and data taking at multiple locations in parallel. It would be done throughout the complex as part of initial commissioning and redone locally after a hardware change. It is hoped that the entire procedure would not need to be repeated, possibly ever and certainly not more often than after a yearly shutdown.

Trajectory Correction: TESLA has steering dipoles at each quadrupole which are used to simultaneously center the beam in the quadrupoles, minimize the average offset in the cavities and compensate for cavity tilts. JLC-X/NLC and CLIC have movers on the individual quadrupoles and structure girders. The correction algorithm moves the components onto a smooth trajectory for sequential sections of the main linacs, as well as for the rings, pre-linacs, and beam delivery. For initial commissioning and after a long shutdown when large moves are anticipated, the procedure would be invasive and likely take one or more shifts for the whole complex.

Dispersion-Free Steering (DFS): DFS is essentially a back-up BBA technique for residual errors left after initial correction. The incoming energy is varied for sequential sections of the linacs and both the on-energy and off-energy orbits simultaneously minimized. To achieve adequate resolution, the energy change must be large and therefore the procedure is invasive. DFS is also sensitive to various systematic errors and hence is not suitable as the only method of BBA [21]. NLC has estimated it would take 0.5–2.0 hr per main linac, depending on the number of iterations required. Like quad shunting, it is assumed to be needed rarely, but since it is relatively fast, it can be applied as a tuneup method during a run. Some method of dispersion correction will also be needed in the damping rings.

Emittance Bumps: Closed-orbit bumps will be used to compensate for dispersion or wakefield distortions, and are most effective when orthogonal control is possible. (JLC-X/NLC and CLIC have the ability to move structure girders which allows clean wakefield control.) Optimization of the bumps is a procedure which would occur as needed non-invasively during normal operation, although initial tuning after a long outage might be done before delivering luminosity, if the beam quality warranted.

Orbit Correction and Feedback: The beam-based alignment procedures and bumps establish a “golden orbit” which must be maintained to a given precision. Without trajectory control, the orbit would quickly deteriorate due to time-dependent environmental effects resulting in an unacceptable loss of luminosity. All designs foresee pulse-to-pulse trajectory feedback and periodic orbit correction. JLC-X/NLC and CLIC plan to use the orbit correction procedure quasi-continuously during normal operation. They estimate that each main linac could be completely corrected every half hour, in parallel with correcting other parts of the complex as needed. TESLA would use an automated procedure to adjust the corrector dipoles, but simulations are not complete enough to specify how frequently this is needed.

Trajectory feedback would be used to stabilize the beam downstream during the corrections as well as to minimize orbit drifts between correction passes. At SLC, feedback was developed to perform the final optimizations of the beam size at the IP. Emittance bump feedback was also attempted but was never fully successful. It is expected that feedback systems similar to those developed for the SLC will be used extensively for optimization as well as energy and orbit stability. Eventually, non-invasive procedures may not adequately restore the luminosity, and invasive DFS or even beam-based alignment will be needed to compensate for drifts in the diagnostics. Simulations of slow diffusive ground motion have shown that dispersion correction might be stable for months in the linacs, but could be required every few days in the damping rings.

Energy Tuning: To achieve the required beam energy at the IP, the amplitude and phase of thousands of linac rf cavities/structures must be adjusted correctly. For TESLA, the Low Level rf (LLRF) controls of each station adjust the phase and amplitude using only a few pulses. Because of potential energy mismatch problems, the procedure can only be applied step-wise down the linac but should still only take minutes. The JLC-X/NLC LLRF system will measure the phase of the beam induced rf and power source rf at the output of one of the structures on each girder. To cleanly separate these signals, the rf for each girder may have to be shut off for one pulse. The measured phase difference will then be used to adjust the source drive phase to achieve the desired beam-to-rf phase for that girder. Fine energy adjustment and stabilization will be achieved by a feedback system which uses a spectrometer to measure the beam energy and adjusts a set of klystrons. NLC has included provision in their klystron complement for such a feedback, while TESLA has not and will need to add overhead for energy stabilization.

8.8.2 Evaluation of Tuning and Recovery Scenarios

Even without any beam interruption, the various tuning procedures will invariably decrease the delivered luminosity. Orbit correction in the damping rings and linacs will cause some emittance degradation during the procedure, if executed non-invasively, and will require lost beam time, if invasive. Optimization of tuning bumps also impacts the luminosity while the best setting is being determined. While a detailed quantitative estimate cannot yet be made, the NLC luminosity projections include a 5% derating to cover the effects of this tuning and CLIC includes 10%, but even this is probably an underestimate. Other projects can expect a similar reduction from their ideal luminosity.

An essential requirement for minimizing the impact of tuning is that each system be designed with sufficient overhead that it can accept incoming beams with somewhat worse than nominal parameters. As an example, the NLC damping rings have sufficient damping to reach the specified equilibrium emittance even if the injected beam is 50% larger than nominal. Likewise both electron and positron sources are designed to produce 50% more current than required. Overheads are particularly critical in the injector complexes in order to allow these systems to be tuned while delivering acceptable beam. Ideally, the collimation and masking should also be designed to reduce backgrounds to an acceptable level even when the emittance or tails are not yet fully minimized. Before finalizing any of the designs, all systems should be reviewed to assure that appropriate overheads have been included.

Generally after any beam interruption, the shorter the outage, the shorter the recovery time. For a brief beam trip, the automated recovery procedures as outlined in Section 8.7 should quickly reestablish colliding beams with relatively little loss in luminosity. Frequent trips will still have a cumulative impact even if beam recovers in a few seconds. After a short beam-off for a repair which does not require access to the accelerator housing, recovery will only require resteering back to the gold orbits with some final IP tuning. This should be less than 0.5 hr, some of which might be compatible with starting to deliver luminosity. If the beam-off time is long enough, then time may be required for thermal stabilization; an example would be the damping rings, where the absence of beam (and therefore synchrotron radiation) would allow the ring to cool. A rule of thumb for such interventions would be that the recovery time is roughly equal to the beam-off time.

When access is required to the accelerator housing, there are significant implications for the TESLA one-tunnel approach. For JLC-X/NLC and CLIC, where the various subsystems are in separate housings, unaffected systems could remain up, perhaps even with beam upstream of the housing accessed. These systems would remain in stable operation and be available to deliver beam as soon as required with no additional tuning. For TESLA, access to the main linac tunnel immediately requires that the damping ring and positron source be shutdown. After the access, a large part of the machine has to be restarted from a “cold state.” This will almost certainly extend the recovery time by a duration which still needs to be evaluated. A prolonged access also implies recovery time for thermal stabilization. The linac tunnels are expected to run with an equilibrium temperature of 30–35°C. For an access of 1–2 shifts, in principle only trajectory smoothing and bump optimization would be required. Since these procedures are nominally compatible with delivering luminosity, the invasive recovery time could be on the scale of hours, depending on the areas affected.

During an extended shutdown of a few months, one would expect extensive maintenance, installation of new hardware and other upgrades. Recovery typically involves relatively major recommissioning of the machine. Components will likely have moved enough that the full beam-based alignment procedures would have to be performed, as well as linac phasing. In general, a subset of the initial commissioning procedures would be required. Recommissioning could ideally take a minimum of several shifts, but more likely several days to weeks, allowing for the inevitable problems.

A special note here should be made of the impact of the proposed undulator-based positron source for TESLA. Clearly if there is no electron beam, then there can be no positron beam. For all scenarios outlined previously (and including the initial commissioning), recovery of the positron beam implies recovery of the electron system, and these scenarios need more detailed study.

8.8.3 Commissioning

Any collider design where the injectors and damping rings are in separate housings from the main linac has a significant advantage in staging the commissioning of the machine, as these sections can be fully studied long before the main linac installation is complete. All of the JLC-C, JLC-X, NLC, and CLIC designs have this configuration. CLIC plans to share some components between the positron and electron injectors, necessarily coupling their commissioning. Some site layouts considered for the NLC have the injectors in a central location, which requires long transport lines to bring the beams to the low energy end of the linacs. These lines would be in the same tunnel as the main linac, and in this case, only the beamlines up through the damping rings can be commissioned independently of the main linacs. To set the scale, the NLC schedule includes about 3 years of commissioning for the injector and damping ring systems.

In the TESLA design, the injector linacs, damping rings, bunch compressors, and positron source are all located in the main linac tunnel, making beam tests of those areas incompatible with linac installation. An added complication is that the positron source requires a high energy electron beam and so cannot be tested until commissioning of the electron systems are fairly well advanced. A low current conventional positron source is foreseen to allow some commissioning of the positron systems before the full power source is available. Commissioning with a high power electron beam would also be possible with this source, but would require magnet polarity switching. Overall, the constraints imposed by the single tunnel for all subsystems and linkage between positron and electron beams is likely to have the most detrimental impact during the final installation and commissioning phases, when frequent access for repairs or upgrades can be expected, most of which will interrupt all operation. A working group on TESLA commissioning strategy has been set up to address these issues [22].

8.8.4 Maintenance Model

TESLA plans to have routine tunnel access for maintenance about every 3–4 weeks in order to replace failed klystrons. The actual frequency will be determined by the klystron lifetime achieved and the number of spares available for overhead. As noted in Section 3.1.1, the gradient must be increased to 23.8 MV/m to provide 2% spares, but this increase may not be sufficient to cover both faults and failures with an access interval of 3–4 weeks. To achieve the desired goal, TESLA will need to select a set of operating parameters which provides an adequate overhead, even if it requires slightly lower energy or luminosity.

Before access, a 3 hr cooldown period for radiation is anticipated before tunnel air can be exchanged with outside air. The klystron and transformer will be designed for quick connect and disconnect. Klystron replacement has a target time of 3 hr per tube, a very aggressive goal compared with present experience (8–12 hr replacement time at SLAC). It is assumed that there are sufficient crews to replace all of the expected average of

5–6 klystrons in parallel. Allowing a few hours for luminosity recovery, each such access can be estimated to take a bare minimum of 2 shifts, more likely considerably longer. If the target availability for the collider is 85%, then these activities can potentially already absorb a sizeable fraction of the downtime budget, even under optimistic assumptions.

All of the other designs include utility housings for the modulators, klystrons and power supplies which are fully accessible for maintenance during collider operation. This is partly because the X-band and C-band machines have a factor of 6 (JLC-X/NLC) or seven (JLC-C) more klystrons than TESLA, although CLIC has fewer, about 75% as many. More importantly, the designers of these projects believe that it is essential to be able to maintain, repair and possibly even upgrade key rf components without interfering with beam operation, granting that this activity entails some risk. The experience at most accelerators has been that interruptions are very costly and should be minimized to reliably deliver luminosity. In this context, the additional cost of the support housings would seem to be a worthwhile investment in order to ensure good system availability and maximize integrated luminosity. Other tunnel configurations could certainly be considered for a superconducting machine.

8.8.5 Evaluation

Clearly there is not enough detailed information available at this time to extrapolate from the tuning and recovery durations and estimate the percentage of time during which each project might actually deliver luminosity. NLC has the most complete plan for tuning procedures, with extensive diagnostics and controls included in the design. Detailed simulations of these procedures have allowed estimates of durations and frequencies for retuning. TESLA is in the process of revising the tuning scenario which may result in design changes. Information was not available to evaluate the other projects.

The most significant differences between the projects of importance for tuning, recovery, commissioning or maintenance are the choice of a single tunnel configuration, and the strong coupling of the positron production to the electron operation, for TESLA. The present TESLA design calls for access to the tunnel every 3–4 weeks to replace failed klystrons. There are a large number of other components located in the tunnel, which are inaccessible during normal operation. Because the injectors and damping rings share the linac tunnel, access necessarily affects most systems, prolonging the recovery time. This appears to pose a significant risk to reliably delivering a high integrated luminosity. The shared tunnel also constrains the initial commissioning of the machine. The single tunnel configuration was chosen for cost savings and to meet the demands of the DESY site. It is in no way fundamental to superconducting technology, and could certainly be reconsidered.

R&D issues:

- The performance of tuning procedures should be simulated in the presence of a wide variety of errors, both in the beam and in the components.

- A comprehensive reevaluation of design parameter overheads throughout the complex is required.

- A comprehensive evaluation of the frequency and impact of tunnel access is needed, including the impact of all repairs.
- TESLA needs a detailed analysis of the risk of a single tunnel on operability.
- TESLA needs to evaluate the operational impact of coupled electron/ positron production.
- TESLA needs a detailed analysis of the operational impact of structure faults.
- JLC-X/NLC need to evaluate the impact on operations of the interleaved positron target system.
- JLC-X/NLC and CLIC need to evaluate the systems aspects of a large interrelated alignment system.
- CLIC needs to evaluate the impact on reliability, stability and operations of drive beam faults coupling to the entire linac.

8.9 SUMMARY

Reliability considerations must be a high priority in the final design of a linear collider. While the linear collider is probably not more complex than the full chain of machines required to run the large colliders at CERN or FNAL, these facilities have the advantage that the injector subsystems have typically operated for other uses and had years of debugging to weed out poorly performing systems. The linear collider will need to understand and debug all of these systems almost simultaneously. The storage rings also have a major advantage in that the injectors are not required between fills and so can often be repaired or studied without impact. In a linear collider, all systems from injectors to beam dumps must be fully operational on every pulse. While overhead can be built into systems which deliver the energy, redundancy against single-point failures is crucial for luminosity. As the size of a linear collider and the number of its components increases, the short-term role of feedback and feedforward loops becomes essential for operability, and their failure is totally unforgiving. These conditions set the stage for the design and operation of all the machines studied in this report.

While there appears to be no technical obstacle to achieving the desired reliability in principle, none of the designs have yet developed complete plans. Considerable resources will need to be given to reliability issues in developing the final design. A list of concerns and needed R&D follows. Most of the issues are common to all designs, with a few project-specific topics at the end. The significance of the rankings can be found in Chapter 9.

Reliability Concerns and R&D

Items common to all designs:

- If the typical reliability of existing HEP accelerators is scaled to the size of a 500 GeV c.m. linear collider, then the resulting uptime will be unacceptably low. This means that a new approach is required, and following previous practices will not be adequate. Reliability must be addressed up front by failure analysis, and appropriate remedies. Adequate engineering margins for components are also essential. The key issue is the allocation, during development and construction, of sufficient engineering and financial resources to produce a reliable system. Areas of concern either because they are potential single point failures or because of the large number of components include:
    - Sources and Bunch compressors
    - Electronics located in the tunnel, and other inaccessible components
    - Magnets and power supplies, especially where inaccessible
    - Diagnostics, movers and rf tuners, LLRF, kickers
    - Crab cavities (machines with crossing angles)

  R&D (Ranking 2): A detailed evaluation of critical subsystem reliability is needed to demonstrate that adequate redundancy and MTBFs have been achieved.

- All designs require extensive beam-based tuning procedures to ensure that the magnet and structure alignment tolerances are met. Tuning can have a major impact on achievable luminosity if the required precision is not achieved.

  R&D (Ranking 2): The performance of tuning procedures should be simulated in the presence of a wide variety of errors, both in the beam and in the components.

- The low emittance beams can potentially damage any component they strike, making an extensive Machine Protection System essential. The projects have preliminary concepts of these systems but none have developed an MPS design in sufficient detail. This is an important R&D topic. A related issue, where more attention is needed, is proper protection of components against radiation generated by such high power beams.

  R&D (Ranking 3): A detailed Machine Protection System design that meets requirements must be developed, including a careful study of failure modes.

- Critical reliability issues for the rf systems are the frequency and impact of faults, the adequacy of the spares overhead, and the accessibility and duration of repairs. Here JLC-X/NLC and JLC-C have tried to take a conservative approach and have carefully analyzed the overhead requirements. This is primarily a cost issue for all machines.

  R&D (Ranking 3): A comprehensive assessment of the MTBF and MTBO for rf components is required, and an adequate fraction of hot spares must be included in the final design.

- A critical operability issue for all parts of a linear collider is providing adequate overhead in machine parameters. Each system should allow for some deviation from nominal input parameters.

  R&D (Ranking 3): A comprehensive reevaluation of design parameter overheads throughout the complex is required.

- All machines should evaluate the frequency of access to the tunnel based on expected MTBFs for all components located in the tunnels. It is also critical to evaluate the recovery time, considering regulation of air temperature after access for damping rings, final focus, sources, and main linacs.

  R&D (Ranking 3): A comprehensive evaluation of the frequency and impact of tunnel access is needed, including the impact of all repairs.

- A linear collider requires extensive feedback systems, machine protection, precision diagnostics, and automated tuning procedures. Failures of these systems cannot generally be tolerated. Providing adequate reliability with such a large, complex system of devices and controls will be extremely challenging. This is a common issue which has not yet received the necessary attention.

  R&D (Ranking 3): A model for communications and controls with adequate reliability must be developed.

TESLA:

- The TESLA single tunnel configuration appears to pose a significant reliability risk because of the frequency of required linac access and the impact of access on other systems, particularly the damping rings. This tunnel layout choice is possible because of the smaller number of klystrons but is not inherent to superconducting technology. The final design must balance initial cost against the life-cycle cost of delivered luminosity.

  R&D (Ranking 2): TESLA needs a detailed analysis of the risk of a single tunnel on operability.

- The proposed undulator-based positron source for TESLA requires an electron beam before there can be a positron beam. This affects recovery scenarios as well as commissioning.

  R&D (Ranking 3): TESLA needs to evaluate the operational impact of coupled electron/positron production.

- The 2.5 km segmentation of the TESLA cryogenic system is a concern because of the time required for warm up and cool down. The vacuum system size and segmentation is also an issue.

  R&D (Ranking 3): The TESLA cryogenic and vacuum system segmentation should be reevaluated.

- Recovery scenarios from structure faults need better definition.

  R&D (Ranking 3): TESLA needs a detailed analysis of the operational impact of structure faults.

JLC-X/NLC and JLC-C:

- The interleaved target system for positron production is complex and requires a thorough evaluation of its operational impact.

  R&D (Ranking 3): JLC-X/NLC need to evaluate the impact on operations of the interleaved positron target system.

- For the linacs of the warm rf machines, the tuning procedures require structure alignment as well as quadrupole centering. The systems aspects of this large interrelated system need to be studied in depth.

  R&D (Ranking 3): JLC-X/NLC need to evaluate the systems aspects of a large interrelated alignment system.

CLIC:

- In the present CLIC design, an entire drive beam section must be turned off on any fault. This makes operation impractical unless the fault rate is extremely small.

  R&D (Ranking 1): CLIC needs to develop a mechanism to turn off only a few structures in the event of a fault.

- The extremely high power in the CLIC drive beam requires an extensive and complex Machine Protection System, above and beyond what would be required for the production beams.

  R&D (Ranking 2): CLIC needs a detailed design of an adequate MPS system for the drive beam.

- For the linacs of the warm rf machines, the tuning procedures require structure alignment as well as quadrupole centering. The systems aspects on this large interrelated system need to be studied in depth.

  R&D (Ranking 3): CLIC needs to evaluate the systems aspects of a large interrelated alignment system.

- A variety of common mode problems can arise due to the fact that faults in the CLIC drive beam complex affect the whole acceleration chain of the main linac.

  R&D (Ranking 3): CLIC needs to evaluate the impact on reliability, stability and operations of drive beam faults coupling to the entire linac.

BIBLIOGRAPHY for CHAPTER 8

[1] L. Hardy, "Accelerator Reliability-Availability," Proc. of EPAC02 (Paris, France, June 3–7 2002).

[2] R. Ferdinand, *et al.*, "The European Spallation Source (ESS): Reliability aspects and related budget considerations," presented at Accelerator Reliability Workshop (ESRF, France, February 4–6 2002).

[3] Zeroth-order Design Report for the Next Linear Collider, SLAC Report 474, May 1996.

[4] J. Seeman, SLAC, Private communication.

[5] S. L. Gold, "5045 Klystrons & Thyratron Reliability" presented at LC2002, SLAC, 2002
`http://www-conf.slac.stanford.edu/lc02/wg2/WG2_Gold_3602.pdf`

[6] J. Cornuelle, SLAC, personal communication.

[7] H. Weise, "Operational Experience with the Test Facilities for TESLA," Proc. of LINAC02 (Korea, August 2002).

[8] 2001 Report on the Next Linear Collider, SLAC-R-571, June 2001.

[9] J. Cornuelle, SLAC, personal communication.

[10] S. Isagawa, KEK, personal communication.

[11] S. Michizono, M. Liepe and S. Simrock, "Design of low level rf control for the TESLA superstructure," DESY-M-01-06V, Proc. of PAC 2001, Chicago, Illinois, 18–22 Jun 2001.

[12] J. Eckoldt, "Magnet Power Supplies for TESLA, TESLA Report 2000-37.

[13] M. Seidel, D. Trines, and K. Zapfe, "Failure Analysis of the Beam Vacuum in the Superconducting Cavities of the TESLA Main Linear Accelerator," TESLA Report 2002-006, February 2002.

[14] S. Roesler, J. C. Liu and S. H. Rokni, "Radiation damage to electronics in the beam tunnel of the next linear collider," SLAC-PUB-8490, Proc. of SATIF-5, Paris, France, 17–21 Jul 2000.

[15] C. Adolphsen. *et al.*, “The Next Linear Collider Machine Protection System,” SLAC-PUB-8130, Proc. of PAC99, New York, NY, 29 Mar-2 Apr 1999.

[16] M. C. Ross, *et al.*, “Single pulse damage in copper” SLAC-PUB-8605, Proc. of LINAC00 (Monterey, California, 21–25 Aug 2000).

[17] S. Fartoukh, *et al.*, “Heat Deposition by Transient Beam Passage in Spoilers,” CERN-SL-2001-012 AP, Paper submitted to Physical Review ST-AB.

[18] D. Schulte and F. Zimmermann, “Failure Modes of CLIC,” Proc. of PAC01 (2001) 4068; also CLIC note 492.

[19] C. Johnson and P. K. Klöppel, “Examination of the CLIC drive beam pipe design for thermal distortion caused by distributed beam loss,” CLIC-Note-328, Pres. at: American Health Physics Society Topical Meeting on the Health Physics of Radiation Generating Machines, San Jose, CA, 5–8 Jan 1997.

[20] P. Tenenbaum, “Beam-based Alignment of the NLC Main Linac, Part II: Dispersion-free Steering and Dispersion Bumps,” LCC-0071, September 2001.

[21] T. Raubenheimer, “Comments on Linear Collider Beam-Based Alignment,” August, 2002.

[22] P. Castro, “First Thoughts on Commissioning of the TESLA Collider,” TESLA Report 2002-009, September 2002.