GEMINI
8-M Telescopes
Project

SPE-S-G0046

Reliability and Maintainability Plan

Version 1.0

Glen Herriot

Canadian Gemini Project Office

National Research Council Canada
Herzberg Institute of Astrophysics
Dominion Astrophysical Observatory
5071 W. Saanich Road
Victoria, BC V8X 4M6
CANADA

June 10, 1994

GEMINI PROJECT OFFICE 950 N. Cherry Ave. Tucson, Arizona 85719
Phone: (602) 325-9329 Fax: (602) 322-8590

GEMINI
8-M Telescopes
Project

SPE-S-G0046

Reliability and Maintainability Plan

Approved by:______________________________ ____________
Jim Oschmann, Date
Systems Engineer

Approved by:______________________________ ____________
Richard Kurz, Date
Project Manager

Science Review:______________________________ ____________
Matt Mountain Date

Table of Contents

RELIABILITY AND MAINTAINABILITY PLAN

1.0 INTRODUCTION

Reliability for Gemini is specified in terms of overall availability. The Gemini Science Requirement is to lose no more than 2% of the scheduled observing time to equipment failure, with a goal of losing no more than 1%.

The purpose of the Gemini Reliability and Maintainability Plan is to help make trade-offs and identify candidates for more detailed reliability engineering and redesign, or more extensive, on-line Predictive Diagnostics, Preventive Maintenance and spares stocking.

"The problems you anticipate are not the ones that get you." By recognizing a potential problem early, you tend to have a solution in hand.

Designing in reliability from the start is always much easier than trying to fix problems after the fact. While preventive maintenance can partly compensate for unreliable equipment, the operating cost of preventive maintenance could exceed the cost of building it right the first time.

There is an overall systems budget for availability which is analogous to the image quality budget. Top down partitioning of downtime among sub-systems will be compared with bottom up estimates of actual performance to see where potential problems lie. One soon realizes which parts of the reliability budget need the most attention and whether there is any hope at all of reaching the goal.

We do not expect a full-blown military style campaign with life-cycle testing of every component and complete fault trees. But, each Work Package will contain a few deliverable items to ensure Gemini telescopes achieve reliability. "Best efforts" alone, as experience on many other projects has shown, just don't suffice to achieve reliability goals. Without a quantifiable, systematic approach, you might as well just delete the availability specifications from the Science Requirements.

There is one central reason for this: Reliability is essentially a statistical discipline — and people's intuition about probability is notoriously poor. This problem applies to component selection and to overall system architecture design (and design review) and to making best use of operational experience. This document describes the Gemini Reliability and Maintainability Plan which will ensure the highest practical availability for the telescopes.

2.0 FAILURE MODES

Since we are not doing a full analysis of the system, we have to choose the most important items for consideration by carefully thinking about what could go wrong and how serious the result would be. Approximately 6 -10 key sub-systems per Work Package are worth considering because the projected loss of observing time is high. Items merit attention for four reasons:

1. When a single, somewhat unlikely fault would cause a lot of downtime;
2. When a moderately likely recurring failure will accumulate downtime;
3. When repairs are expensive;
4. When a frequently occurring failure annoys observers no matter how quickly it is repaired; the consequent loss of data, confidence, and reputation is unacceptable.

This process of choosing where to pay attention is known by the daunting acronym: Failure Modes and Effects Criticality Analysis (FMECA).

Each subsystem will include a FMECA table starting with 6-10 key faults. Each row contains the estimated likelihood, the resulting damage, and the time to repair. The amount of detail and analysis required will vary between subsystems. This is negotiated on a case-by-case basis between the Gemini Project Office and each subcontractor or work package.

As well, the table should continue with other potential faults which a layman may feel are important, but are not severe according to the four criteria listed above. The purpose of extending the FMECA table is to encourage thinking about reliability and to demonstrate that the Work Package Manager (WPM) and his team has done a plausible analysis of his equipment.

3.0 THE AVAILABILITY MODEL

The Gemini Availability Model combines failures, repairs, and scheduled maintenance to determine overall up-time for the telescope.

Work Package Managers (WPMs) will submit reliability information about their equipment and software as outlined below. The Gemini project office will incorporate this information into a model, and flag problem areas.

The model works by accumulating downtime from all sources and finally subtracting it from the scheduled observing nights.

In general, failures occur in three distinct phases during the lifetime of equipment: infant mortality, random failure during useful life, and wear-out. Properly running in ("burning in") equipment at the vendor can eliminate infant mortality caused by factors like improper assembly. Of course, one does not want to burn-in too long and waste some of the useful life of the equipment. Preventive maintenance and replacement can eliminate wear-out. The Gemini availability model assumes that equipment is operated in the random failure phase of its life cycle.

As a sanity check on the downtime model, the maintainability budget tracks all of the proposed servicing and engineering time. The goal will be to do servicing in the daytime, with at most 10% of the nights set aside as engineering nights. Any servicing time in excess of the engineering time budget will count as downtime against the availability budget.

The remainder, after engineering time, will be scheduled observing nights; any loss of these due to a failure (as defined below) will count as downtime.

3.1 Definitions

FAILURE. A fault which shuts down the telescope, costing any observing or engineering time or which degrades the imaging by more than 100% TBR of the Science Requirements.

This definition of failure assumes that imaging may degrade gracefully up to an arbitrary threshold which will be consistently applied for the life of the Gemini project. Beyond this threshold, we must take a hit for downtime and record it as such. We have to be consistent in our definition of failure so that we can quantify progress towards our goal as set out in the Gemini Science requirements. To do otherwise would be equivalent to a company changing accounting practices to make its finances appear sound.

MTBF. Mean time between failure in hours. Suppose one were to run a large number of identical units for a long time. As units fail, we repair them and put them back in service. Further suppose that we perform preventive maintenance according to the manufacturer's recommendations. During this hypothetical experiment, log the total hours of operation for all units and the total number of failures.

The MTBF is the hours of operation divided by the number of failures. However, see Section 3.4, Estimating MTBF, for several other more practical ways to determine MTBF.

MTTR. Mean Time To Repair. After the above thought experiment, divide the total unplanned repair time by the number of failures. Be sure to include unplanned time to reboot, run diagnostics, get parts to the summit, exchange units, qualify (re-test), and get back on line.

To analyze the impact of a fault we need to know whether the night crew can fix a problem; hence, the repair time falls into two categories:

MTTR(n) Mean time to repair at night. If it is not possible for the night crew to fix a fault, **MTTR(n)** is infinity.

MTTR(d) Mean time to repair in the day.

HOURS OF OPERATION. Operating hours are logged regardless of when and whether observing actually occurs. What is important is the duration that equipment is energized or experiencing at least normal stress. If some standby or idle mode prolongs life substantially (say ten fold), then do not include idle standby time.

We need to know both duration and time of use. Examples:

a) 24 hours a day. Cooling Plant.

b) 12 hours a day. Night only, e.g.. Mirror surface heating.

c) 12 hours a day. Day only, e.g. Dome cooling.

d) n hours per use. No. of uses per year. E.g. coating plant

e) ...

3.2 Downtime Model

The downtime model is an intermediate calculation step between the information from WPMs and the final estimate of availability.

The model calculates expected lost time for a sub-system from the product of MTTR and Operation Time divided by MTBF.

Then the downtime model segregates lost time into three categories depending on whose time is lost:

* scheduled observing time;
* scheduled engineering night time;
* day time.

Once the low priority engineering time is used up, then repairs impact the observing availability.

On average a fault will occur half-way through the night. Thus, a failure of equipment at night will typically cost 6 hours of observing unless the telescope operator can fix the problem. E.g. if MTTR(n) is <6 hours.

If the day crew takes more than a shift to repair the damage, (MMTR(d) >12 hours), then a fault will begin to impact the next nights observing.

Inputs to Downtime Model:

Using the definitions from above, the downtime model requires this information for each subsystem:

1. Hours of operation

2. Time of day for operation
3. MTBF Mean time between failure in hours.
4. MTTR Mean Time To Repair. Can this repair be done by night crew? e.g. MTTR(n) or MTTR(d)
5. Preventive Maintenance Schedules.

3.3 Maintenance (schedules, spares, diagnostics)

Every work package delivered to Gemini will include a preventive maintenance program. As much as possible such maintenance will be done in the daytime. However, some maintenance will extend into the night and will be planned for in the Engineering Time budget. However, if for example, the sum of all preventive maintenance time exceeds the scheduled engineering time budget, then it will hurt the scheduled observing availability.

For example let us assume that 10% of nights are engineering nights. If mirror re-coating is planned monthly and takes the telescope out of operation for 3 nights, then there would be no time left for any other engineering work. If in fact, the coating plant failed and had to be repaired before continuing, then this fault would prolong the maintenance time and cause a loss of Availability.

3.3.1 Scheduled Maintenance

You may be able to increase reliability through preventive maintenance (PM) such as:
* lubrication
* off-line (day-time) diagnostics
* scheduled replacement of parts
* rotating refurbishment of sub-assemblies

Work Package deliverables will contain suggested Preventive Maintenance Schedules. Here is an outline of an example schedule:

Category	Mean Service Time
Monthly items	3 hours
Semi-annual items	6 hours
Annual items	8 hours

Notice that this schedule has a potential problem; once a year, all three categories are due at the same time. All the work cannot be done in one daytime shift and would continue 5 hours into the night. It may be okay to use up part of the engineering time this way; but, the operations staff may wish to split the work over several days. However, the WPM may have assumed that the annual maintenance takes 8 hours by piggy-backing on assembly-disassembly or requalification procedures contained in the semi-annual Preventive Maintenance items.

Hence, the PM schedule should also indicate service time for stand alone maintenance not done in conjunction with any more common procedures. There should also be recommendations about which PM categories should be grouped together.

Category	**Mean Service Time**	**Standalone Time**
	Continuous Shutdown	Individual Service
Monthly items	3 hours	3 hours
Semi-annual items	6 hours	6 hours
Annual items	8 hours	10 hours

It is ***NOT THE INTENT*** of the Gemini Project to specify recommended service intervals! The purpose of the above example is to show WPMs what type of information they should provide to the project.

While thinking about and documenting what PM servicing will be done and how often, some procedures may appear excessive. Ideally the designers may rethink their design or procedures to reduce the frequency of failure, or duration of maintenance, both scheduled and unscheduled. For example, oiling ports, or a more automated test procedure, may make all the difference to the downtime.

3.3.2 Service Procedures and checklists

WPMs shall deliver service procedures, diagnostics and checklists to aid both scheduled and unscheduled maintenance. A written procedure with a few key illustrations dramatically shortens PMs.

3.3.3 Requalification Tests

After maintenance, and periodically, some specified system level tests should be run. The WPM should describe how they would quickly verify that their equipment is working properly.

3.3.4 Suggested Spare Parts Lists

- custom
- off-the-shelf

Part description, recommended quantities, cost, part number, manufacturer, address, phone number.

The project will not necessarily procure all of every spares list. Instead the project will collate these lists and amalgamate them if possible. The list should also include quantities actually used in the subsystem. A disproportionate recommended quantity may flag unreliable parts or vendors.

3.3.5 Rotating Spares

The WPM should include recommendations for rotating spare assemblies where appropriate.

Some items, such as actuator diaphragms may take too long to replace in situ. However, by replacing whole actuators, then the telescope may be brought back on line more quickly. The WPM may recommend several rotating spares which are refurbished off-line.

This idea may be extended to the PM procedures where a number of actuators may be swapped at frequent intervals, and reworked in time for the next scheduled maintenance episode.

It may not be advisable to replace the entire stock of spares during one PM, but to hold one in reserve in case a unit fails immediately after a PM and all your spares need rework.

3.4 Estimating MTBF

Instead of running a life-test program, there are a number of ways to estimate MTBF which are less costly, but still substantially better than a wild guess.

As a rule of thumb, a properly designed 19" rack of electronics as high as your head has an MTBF of 30,000 hours. Some manufacturers can be persuaded to provide FITS (= Failures in time) data, which are expected failures per billion hours of operation. A 140-pin chip has FITS $\geq$ 140; the same chip in a socket $\geq$ 280.

Counting or estimating connections sets an upper limit on the MTBF of a system as a kind of sanity check on your MTBF. Here connection means each solder joint, or pin of a connector, i.e. a 25-pin bulkhead connector with a cable arriving at both the male and female halves would count as 75 connections. Each connection causes roughly one failure in a billion hours of operation. 10,000 connections would therefore have an MTBF of 100,000 hours. This is true providing connectors are not abused, are properly strain relieved, and do not exceed the manufacturer's recommended number of insertions. Some connectors endure a surprisingly low (e.g. 25) number of matings.

Alternately, manufacturers may be persuaded to give warranty return information. Given the length of the warranty, the hours of operation, and the percentage of units repaired under warranty, then MTBF may be estimated as follows:

$$\text{MTBF} = \frac{\textit{Hours of operation during warranty}}{\ln(\%\ \textit{returned} \div 100)}$$

3.5 Redundancy

The term redundant has two separate meanings: that which is superfluous, and that which provides a backup. In reliability engineering, these are called serial redundancy and parallel redundancy. In general, we wish to reduce the total number and variety of components. For example if there are three computers, the failure of any one of which will prevent observing, then they are serially redundant. Conceptually, the path to successful operation goes through each in series. Hence the phrase serial redundancy — a negative term.

As well, we would prefer not to design in extra quantities of anything because it will increase the chances of something failing, and increase labour for maintenance. However, if some component has a short service life, and we cannot find an alternate, the overall observatory as a system may be made more reliable by having a backup-online. This is parallel redundancy, because there are parallel logical paths to keep the telescope running. The observatory uptime is aided, but at the price of increased capital costs and daytime repair costs for failed units.

Serial redundancy is the baseline assumption for all elements in the top level Availability Model, because of the definition of failure: any single fault which puts the telescope out of service or makes the image too poor to be useful.

There will likely be only a few items where continued discussion between the WPM and the Project can identify particularly unreliable sub-systems that can be upgraded cost-effectively via parallel redundancy. The redundant pair then form a sub-system which in turn is serially redundant with the rest of the observatory.

3.6 Combining MTBF To Give A Total Subsystem Estimate

A thousand equally reliable components together make a system which will fail a thousand times more often than the individuals. MTBF's of serially redundant components are combined by the same process as adding resistors in parallel, which implies that a system is less reliable than its weakest link, just as parallel resistors have less net resistance than the smallest of them. As well, if the reliability of one component is much smaller than the others, then it will dominate the overall MTBF.

Assuming a single point failure, by definition, puts the telescope down, (serial redundancy), then combining lower level items into a higher level system:

$$1/\mathrm{MTBF(system)} = \mathrm{sum}\ (\ 1/\mathrm{MTBF}(i)\ )$$

Where MTBF(i) are individual components.

Because of the nuisance of adding reciprocals in a multi-level hierarchy, it is convenient to convert MTBF hours into units of Failures in Time. FIT = 10^9/MTBF, i.e., how many failures you can expect in a billion hours of operation of a single component or subsystem. Some manufacturers provide reliability data in FIT units.

$$\text{FIT system}) = \text{sum} (\text{FIT(i)})$$

Having added up all the FIT for the system, then the overall MTBF is just:

$$\text{MTBF (system)} = 1/\text{FIT(system)}$$

3.7 Complexity Is the Enemy of Reliability

A major manufacturer of laser printers has advertised that their latest printers have 30% fewer parts than their original printer. To a customer, this translates to a lower manufacturing cost and hence a low purchase price; but, the real payoff to the manufacturer, is dramatically reduced warranty costs.

This is a direct consequence of the fact that any single component failing will put the printer out of action, i.e., all components are serially redundant. By reducing the total number of parts, the total FITS are reduced and hence the MTBF is increased.

4.0 FAULT TRACKING AND ONLINE DIAGNOSTICS

During commissioning and operations, the engineering data handling system will include record keeping and data analysis tools to help Gemini meet its availability goals.

A proper fault tracking system is a key part of a systematic approach to reliability.

4.1 Systematic Approach

Borrow some ideas from Total Quality Management (TQM) used by high-quality manufacturers like Motorola.

TQM includes:

1. Setting quantifiable goals.
2. Measuring performance in meeting those goals.
3. Continuous improvement by identifying weaknesses and fixing them.

We have a quantifiable goal: a tight goal for telescope "uptime": 99%

We need the tools to measure performance.
We need the tools to identify shortcomings and show that they have been fixed.

4.2 Why is a traditional bug tracking system inadequate?

Anecdotal intuition about frequency/severity of bugs is often faulty.

* People stop reporting recurring bugs because they assume that "head office already knows."
* Without a painless way to report good information, faults are not logged properly.
* Head office assumes that a bug doesn't occur often enough to merit attention.

Intuition about down total down time is frequently distorted. Telescopes need a good data base showing time spent:

* Observing
* Down due to failure/repair/reboot
* Engineering and commissioning
* Scheduled down time (e g. mirror recoating)

Initial diagnosis of a problem is often misleading.

* One primary fault may manifest itself in many symptoms.
* Hardware/software/systems problems are hard to sort out when an observer is eager to get going. The temptation is to just re-boot and carry on while logging a guess about the cause (i.e. blame software).
* Bug trackers often force pigeon-holing of each fault in a single category.

4.3 What is required in bug reporting tools?

* A convenient user interface to report faults. An e-mail based/command line approach like GNATS may suffice during subsystem development.
* A more robust system is desired for operations (and commissioning). A possible candidate is TkGNATS which includes a graphical user interface. Another possibility is to extend the automated fault logging in the engineering database to include screens so that staff may enter event records and comments. Other considerations are:

- A logging system for faults of all types, mechanical, electronic etc., not just software.
- Records of all events, whether re-booting solved them, or whether repairs were necessary.
- Accurate measurements of telescope down time, whatever the cause.
- Additional open fields in the fault records to allow subsequent tagging of the data in multiple ways which may not be apparent at the outset. (e.g. EPICS A&G, Motor Driver).

4.4 System Fault Analysis Tools

On line at the telescopes and headquarters etc. should be software to access the fault data base and display performance and assist in identifying weaknesses. The most leverage occurs by noting which "unrelated" events really have a root cause.

* Provision in the data base for sober second and even third diagnoses and for sorting the data by new categories added as experience and insight develops.

* Trend charts for up time. A plot of up time per month, will show how close Gemini is coming to the 99% goal.

* Pareto charts are essential (vital!) tools. Pareto charts are Histograms of fault types sorted by frequency of occurrence or by hours of resulting loss of observing time. They are created from logs of fault data. Typically the top 5 - 15 items are shown (plus "miscellaneous"). The bar on the left is biggest, and should get the most attention if you want to make serious improvement. By comparing a Pareto chart for one time period with another, you can see if you really have resolved problems which were identified earlier.

The WPM should provide an initial list of a dozen sub-systems for which he would like to see individual tallies kept on his equipment. This list will be form part of the functional specifications for the engineering data handling system. As a first guess, these items are largely the same as the inputs to the availability model.

4.5 Trend Charting

- mean value
- RMS

Failures due to end-of-life wear out may also be prevented by continuously monitoring the health of key parameters. A long term trend towards some unsafe limit can be identified before the failure occurs. Examples include increasing motor torque (current) or temperature. WPMs should provide printed lists of such parameters, their safe limits, and suggested frequency of monitoring. Where possible, these should also be on-line such as in the EPICS data base of the subsystem controllers.

Another useful diagnostic is trend lines for the RMS of a parameter. A load cell may stay within expected limits, but become erratic, which will show up on a trend chart of RMS.

4.6 Primary Faults vs. Secondary Faults

Very frequently some part fails and takes other equipment with it. For example, a power supply may produce too high a voltage, and damage other equipment. The service staff may replace the failed equipment but fail to notice that the power supply is faulty. Again the repaired equipment will fail and quickly gain a reputation for being unreliable.