

esa

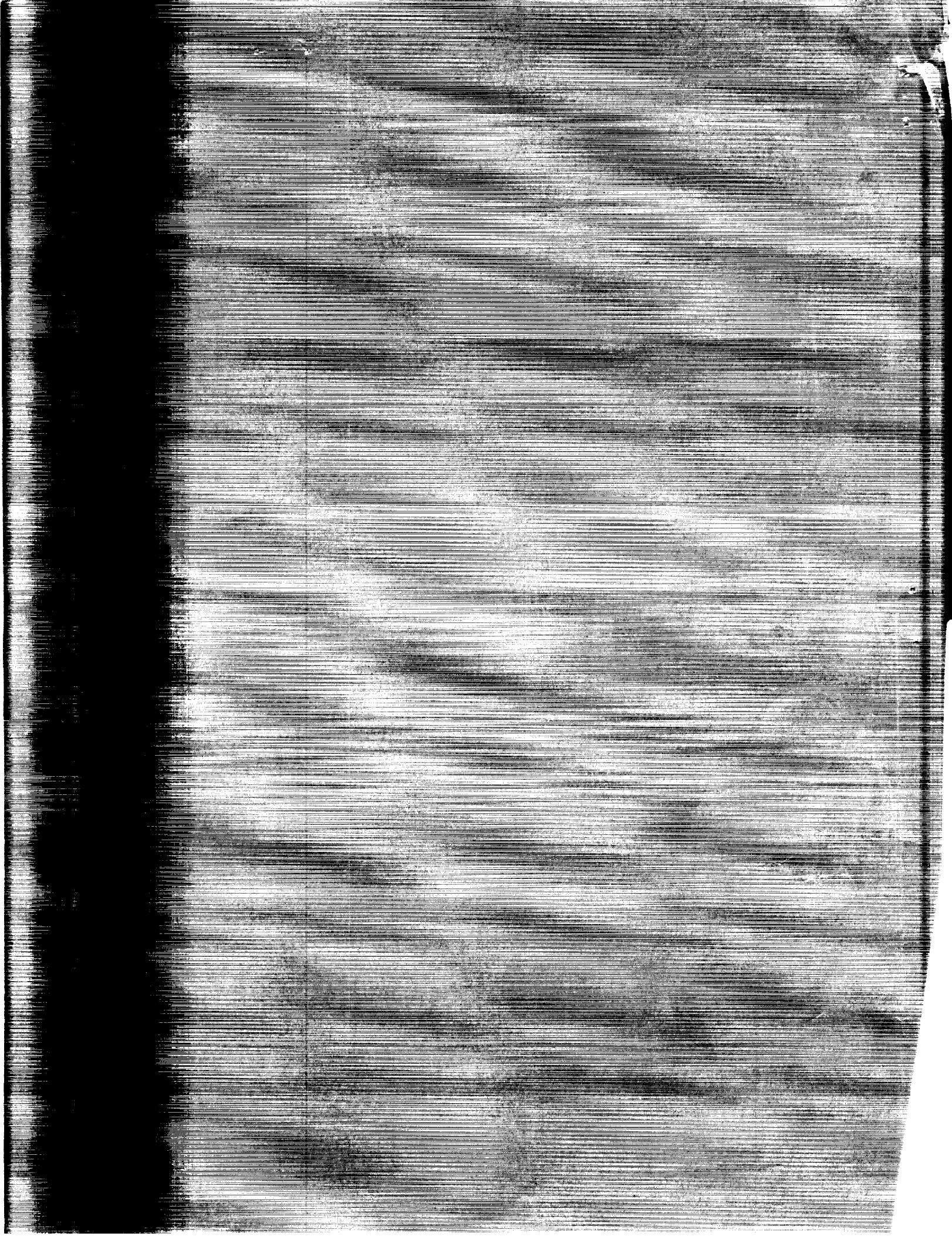
NASA-TM-110598

(NASA-TM-110598) STEP: SATELLITE
TEST OF THE EQUIVALENCE PRINCIPLE.
REPORT ON THE PHASE A STUDY (NASA.
Goddard Space Flight Center) 136 p

N95-26184

Unclassified

G3 ■/12 0046125



STEP

Satellite Test of the Equivalence Principle

REPORT ON THE PHASE A STUDY

SCI (93) 4
March 1993

J.P. Blaser	A.M. Nobili
M.Bye	H.J. Paik
G. Cavallo	R. Reinhard
T. Damour	R. Rummel
C.W.F. Everitt	M.C.W. Sandford
A. Hedin	C. Speake
R.W. Hellings	L. Spencer
Y. Jafry	P. Swanson
R. Laurance	P.W. Worden Jr.
M. Lee	



FOREWORD

STEP (Satellite Test of the Equivalence Principle) was proposed to ESA in November 1989 by an international team of scientists coordinated by C.W.F. Everitt, Stanford University, with P.W. Worden Jr. as technical leader. The proposal was submitted in response to a Call for Mission Proposals for the next Medium Size Project (M2), issued by ESA's Directorate of Scientific Programmes on 15 June 1989. After review of the competing proposals by ESA's scientific advisory bodies in early 1990, STEP was recommended for study as a joint ESA/NASA project at Assessment Phase Level. This study was conducted between April and June 1990, followed by a consolidation phase which ended in December 1990. The results of the Assessment Phase Study are described in document SCI(91)4. Of the six studies at Assessment Level, ESA's Space Science Advisory Committee (SSAC) selected in May 1991 four studies for a more detailed study at Phase A level, among them STEP. The Phase A Study was carried out from January until December 1992, and included an industrial study with Alenia as Prime Contractor. On the US side, JPL acted as the Lead Centre. The present document is the result of the joint ESA/NASA Phase A Study, carried out by a Science Team consisting of representatives of the European and US scientific communities and ESA and NASA staff.

The European members of the team were:

J.P. Blaser (Paul Scherrer Institut)
T. Damour (Institut des Hautes Etudes Scientifiques)
A.M. Nobili (University di Pisa)
R. Rummel (TU Delft)
M.C.W. Sandford (Rutherford Appleton Laboratory)
C. Speake (University of Birmingham)

The US members of the team were:

M. Bye (Stanford University)
C.W.F. Everitt (Stanford University)
A. Hedin (NASA / GSFC)
H.J. Paik (University of Maryland)
P.W. Worden Jr. (Stanford University)

The Study Scientists were R. Reinhard (SSD/ESTEC) and R.W. Hellings (JPL), the Study Managers R. Laurance (ESTEC) and P. Swanson (JPL). Y. Jafry (SSD/ESTEC) was the FSA Deputy Study Scientist. A. Atzei (ESTEC) served as ESA Study Manager until August 1992.

G. Cavallo (ESA HQ) was the ESA Headquarters Representative, L. Spencer (NASA HQ) and M. Lee (NASA HQ) were the NASA Programme Manager and Programme Scientist, respectively. D. Strayer (NASA HQ) served as NASA Programme Scientist until July 1992.

The Study team was supported by three sub-groups:

Theory

T. Damour (Chairman, IHES, Paris)
L. Blanchet (Observatoire de Meudon)
A. De Rújula (CERN)
C.W.F. Everitt (Stanford University)
G.W. Gibbons (Cambridge University)
D.J. Gross (Princeton University)
K. Nordtvedt (Montana State University)
N. Straumann (University of Zurich)
F. Wilczek (IAS, Princeton)
C.M. Will (Washington University)

Hardware

J.P. Blaser (Co-Chairman, Paul Scherrer Institut)
P.W. Worden Jr. (Co-Chairman, Stanford University)
M.C.W. Sandford (Secretary, RAL)
M. Bye (Stanford University)
P. Carelli (University of Aquila)
F. Fuligni (CNR, Frascati)
Y. Jafry (SSD / ESTEC)
J. Léon (LAEFF, Madrid)
N. Lockerbie (University of Strathclyde)
P. Mason (JPL)
H.J. Paik (University of Maryland)
R. Parmley (Lockheed, Palo Alto)
T. Patrick (MSSL)
C. Pegrum (University of Strathclyde)
T. Quinn (BIPM, Paris)
H. Seppä (VTT, Espoo)
C. Speake (University of Birmingham)
T. Sumner (Imperial College, London)
P. Touboul (Onera, Paris)

Geodesy

R. Rummel (Chairman, TU Delft)
A.J. Anderson (NASA HQ)
S. Lichten (JPL)
E. Pavlis (NASA / GSFC)
E. Schrama (TU Delft)
P. Schwintzer (GFZ, Potsdam)
M. Tapley (Stanford University)
T.P. Yunck (JPL)

During the course of the Phase A Study the Study Team met six times:

- 13 / 14 November 1991 (ESTEC)
- 6 February 1992 (Mürren, Switzerland)
- 20 May 1992 (ESTEC)
- 25 / 26 August 1992 (Stanford University)
- 15 / 16 December 1992 (JPL)
- 11 / 12 February 1993 (ESTEC)

Two scientific conferences on STEP were organised by members of the Study Team:

- 3 - 5 February 1992, STEP Workshop (Mürren, Switzerland)
- 6 - 8 April 1993, STEP Symposium (Pisa, Italy)

Technical contributions by the specialists of the ESA Technical Directorate at ESTEC, the Directorate of Operations in Darmstadt, Germany, and the Jet Propulsion Laboratory in Pasadena, California, are gratefully acknowledged.

Additional copies of this report are available from:

Mr. G. Whitcomb

or

Dr. R. Reinhard

ESTEC
Postbus 299
2200 AG Noordwijk
The Netherlands

Table of Contents

Executive Summary	x
1 The scientific significance of STEP	1
1.1 History of the Equivalence Principle	1
1.2 General Relativity : a basic ingredient of our present world view	2
1.3 Current status of the tests of General Relativity	2
1.4 Puzzles around the Standard Model of particle physics	4
1.5 Examples of forces violating the Equivalence Principle	5
1.6 Spin-couplings and Axion-like forces	7
1.7 The value of the Newtonian gravitational coupling constant, G	8
1.8 Summary	9
2 The STEP experiments	11
2.1 Experimental concepts and the advantages of space	11
2.1.1 Long-range test of the Equivalence Principle	11
2.1.2 Spin-Coupling experiment	13
2.1.3 Gravitational constant, Inverse Square Law and geodesy experiments	14
2.2 Superconducting linear accelerometer	15
2.2.1 Principle of operation	15
2.2.2 Magnetic bearing	16
2.2.3 Position sensing circuit	17
2.3 Superconducting differential accelerometer	18
2.3.1 Differential sensing circuit	18
2.3.2 Levitation and damping	19
2.3.3 Electrostatic control	20
2.4 STEP accelerometer configuration	20
2.5 Disturbing forces	22
2.5.1 Platform accelerations and drag-free control	22
3 Equivalence Principle experiment (EP)	25
3.1 The Equivalence Principle measurement concept	25
3.1.1 Accelerometer configuration	25
3.1.2 Experiment validation concept	25
3.1.3 Equivalence Principle measurement programme	26
3.2 Accelerometer design	28
3.2.1 Test mass materials	29
3.2.2 Superconducting sensing and levitation	31
3.2.3 Electrostatics	31
3.2.4 Caging	33
3.2.5 Differences between European and Stanford designs	33
3.3 Calibration and verification	34
3.4 Disturbance management	34
3.4.1 Gravity gradients and helium tide	35
3.4.2 Helium confinement system	39

3.4.3	Particle radiation and electric charging	41
3.4.4	Radiation sensor	45
3.4.5	Charge control system	45
3.4.6	Other disturbances	46
4	Spin-Coupling experiment (SC)	49
4.1	Outline of experiment design	49
4.2	Disturbances	54
5	Constant of Gravity G and Inverse Square Law experiment (G/ISL)	56
5.1	Hardware description	56
5.2	Design principles and experimental procedures	58
5.3	Instrument noise and experimental resolutions	59
5.4	Metrology and calibration requirements	61
5.5	Disturbance control requirements	62
6	Geodesy	65
6.1	STEP and Geodesy	65
6.2	Status of gravity modeling	66
6.3	Geodesy free-fall experiment	68
6.3.1	GPS receivers	68
6.3.2	Gravity field recovery	70
6.3.3	Alternative tracking concepts	71
6.4	Geodesy gravity gradiometry experiment	72
6.5	Applications in geodesy, solid-Earth physics, and oceanography	74
7	Aeronomy	77
7.1	Atmospheric science topics for STEP	77
7.2	Mission issues for aeronomy	78
8	Cryogenic Payload System	79
8.1	The probe assembly	80
8.1.1	Differential accelerometer package (DAP)	80
8.1.2	Probe	80
8.2	Dewar	81
8.2.1	Requirements	82
8.2.2	The helium constraint system	82
8.2.3	Lifetime	82
8.2.4	Thermal and structural performance analyses	82
8.3	Electronics	83
8.4	Integration and test	85
9	Spacecraft	86
9.1	Spacecraft requirements	86
9.2	Spacecraft configuration	86
9.3	Drag-free control and AOCMS	88
9.3.1	Sensors and thrusters	88
9.3.2	Control in translation	90
9.3.3	Control in rotation	90

9.3.4 Modes	90
9.4 Structure	93
9.5 Thermal control	93
9.5.1 Thermo-elastic analysis	94
9.6 Tracking, Telemetry & Command (TT&C)	96
9.7 Onboard data handling	96
9.8 Electrical power subsystem	96
9.9 Budgets	97
9.9.1 Power budget	97
9.9.2 Mass budget	97
10 Mission analysis	99
10.1 Sun-synchronous orbits, eclipses, and launch windows	99
10.2 The drag-free condition and the minimum orbital height	100
10.3 Nominal STEP orbits	103
11 Launch vehicle, ground segment, mission operations	104
11.1 Launch vehicle	104
11.1.1 Baseline launch vehicle	104
11.1.2 Titan II performance	104
11.1.3 Satellite interface	104
11.2 Tracking	104
11.2.1 Deep Space Tracking Network	104
11.2.2 Global Positioning System (GPS) tracking	105
11.3 Mission Operations System	105
11.3.1 Characteristics of STEP Mission Operations System	105
12 Pre-launch activities, early in-orbit setup and calibration, science operations	108
12.1 Payload calibration and mission timeline	108
12.2 Science operations	109
13 Science management, programmatic and schedule	111
13.1 ESA/NASA collaboration	111
13.2 Science management	111
13.3 Archiving	113
13.4 Management	114
13.5 Schedule	114

List of Figures

1.1	<i>Existing limits and resolution expected from STEP for a new composition-dependent force</i>	10
2.1	<i>Concept of the Equivalence Principle experiment on STEP</i>	12
2.2	<i>Concept of Spin-Coupling experiment on STEP</i>	13
2.3	<i>Concept of G and Inverse Square Law experiment on STEP</i>	14
2.4	<i>Concept of a superconducting accelerometer</i>	16
2.5	<i>SQUID operating principle</i>	18
2.6	<i>Sensing circuit of a superconducting differential accelerometer</i>	19
2.7	<i>SQUID energy resolution as a function of frequency</i>	20
2.8	<i>Levitation circuit of superconducting differential accelerometer</i>	21
2.9	<i>Overall STEP accelerometer configuration in the quartz block</i>	22
2.10	<i>Residual linear acceleration spectrum of the spacecraft under drag-free control</i>	23
3.1	<i>Spacecraft and orbit coordinates</i>	27
3.2	<i>EP differential accelerometer</i>	28
3.3	<i>Exploded view of EP differential accelerometer</i>	29
3.4	<i>Quartz block supporting each reference plate assembly</i>	30
3.5	<i>Electrodes for capacitive sensing and actuation of each test mass</i>	32
3.6	<i>Allowable helium tidal motion for EP accelerometers</i>	38
3.7	<i>Split dewar concept</i>	40
3.8	<i>Helium confinement: combined tidal and electrostatic equipotentials</i>	41
3.9	<i>Helium confinement: ratio of tidal to electrostatic forces</i>	42
3.10	<i>The number of significant solar events per year during the last three complete solar cycles</i>	43
4.1	<i>Putative spin-coupling force field</i>	50
4.2	<i>Exploded sectional view of the STEP spin-coupling differential accelerometer</i>	51
4.3	<i>Close-up view of one source/test mass annulus pair of the spin-coupling accelerometer</i>	52
4.4	<i>Putative spin-coupling acceleration as a function of range for the exact experimental geometry</i>	53
5.1	<i>Cutaway view of the G/ISL apparatus</i>	57
5.2	<i>Levitation for the G/ISL apparatus</i>	58
5.3	<i>Predicted signals for the G/ISL experiments</i>	59
5.4	<i>Existing limits and resolutions expected from STEP for a violation of the Inverse Square Law</i>	61
5.5	<i>Spectra of the calibration signal and uncertainties for the G experiment</i>	63
6.1	<i>Principle of the Geodesy free-fall experiment</i>	66
6.2	<i>Distribution of $1^\circ \times 1^\circ$ mean gravity anomalies with an accuracy below 10 mGal</i>	67
6.3	<i>STEP GPS antenna configuration</i>	69
6.4	<i>STEP gravity field sensitivity</i>	76
8.1	<i>Cryogenic Payload System</i>	79
8.2	<i>Cryogenic Payload System block diagram</i>	80
8.3	<i>Probe assembly diagram</i>	81
8.4	<i>Dewar lifetime plot</i>	83
8.5	<i>Dewar thermal transient</i>	84

9.1	<i>STEP spacecraft</i>	87
9.2	<i>STEP spacecraft functional block diagram</i>	89
9.3	<i>Block diagram for drag-free control in translation</i>	91
9.4	<i>Disturbance rejection ratio for drag-free control in translation</i>	91
9.5	<i>Block diagram for drag-free control in rotation</i>	92
9.6	<i>Disturbance rejection ratio for drag-free control in rotation</i>	92
9.7	<i>Sequence of AOCMS modes</i>	93
9.8	<i>Mass points distribution in the spacecraft model for the self-gravity stability verification</i>	95
9.9	<i>Structural model for thermo-elastic analysis</i>	95
10.1	<i>Shadow heights for circular Sun-synchronous orbits in spring</i>	100
10.2	<i>History and prediction for solar flux and geomagnetic activity index</i>	102
10.3	<i>Orbital heights below which the air-drag will exceed 1.2 mN</i>	102
10.4	<i>Nominal STEP orbit geometry</i>	103
11.1	<i>Mission Operations System overview</i>	105
11.2	<i>STEP EEIS/MOS architecture uplink</i>	106
11.3	<i>STEP EEIS/MOS architecture downlink</i>	106
13.1	<i>STEP science management structure during Phase A</i>	112
13.2	<i>STEP programme schedule</i>	115

List of Tables

2.1	<i>Summary of disturbances</i>	24
3.1	<i>Dimensions of European EP accelerometers</i>	37
3.2	<i>Dimensions of Stanford EP accelerometers</i>	38
3.3	<i>Temperature increase of the test masses for a single pass through the SAA</i>	44
3.4	<i>Induced charge on the test masses for a single pass through the SAA</i>	44
6.1	<i>Comparison of tracking systems</i>	71
9.1	<i>Drag-free requirements</i>	86
9.2	<i>Spacecraft power budget</i>	97
9.3	<i>Spacecraft mass budget</i>	98
10.1	<i>Summary of orbit selection requirements</i>	99
10.2	<i>Yearly launch windows for STEP orbits</i>	101
10.3	<i>Nominal STEP orbit elements</i>	103

Executive Summary

STEP (Satellite Test of the Equivalence Principle) is a mission in the discipline area of "Fundamental Physics". The fundamental nature of the STEP experiments has profound implications for the most important areas of modern physics. STEP would be ESA's first mission in Fundamental Physics, opening up a third discipline area in space science in addition to Solar System Exploration and Astronomy/Astrophysics. A very large community of physicists in numerous universities and science institutes in Europe and the USA would make use in their work of the results obtained from the fundamental physics experiments on STEP.

STEP is being studied as a cooperative venture with NASA. Essentially, NASA would provide the launch vehicle, one of the four fundamental physics experiments, overall payload system integration and testing, while ESA would provide the STEP satellite. Mission operations would also be shared. European science institutes would provide the other three experiments. During the Phase A, the original aim of STEP mission, to test the Equivalence Principle with the highest possible precision, was significantly widened, making use of the fact that the STEP satellite is an extraordinarily quiet laboratory for fundamental physics experiments in space.

Scientific Objectives

During Phase A, the STEP Study Team identified three types of experiments that can be accommodated on the STEP satellite within the mission constraints and whose performance is orders of magnitude better than any present or planned future experiment of the same kind on the ground. The scientific objectives of the STEP mission are to

- test the Equivalence Principle to one part in 10^{17} , six orders of magnitude better than has been achieved on the ground;
- search for a new interaction between quantum-mechanical spin and ordinary matter with a sensitivity of the mass-spin coupling constant $g_p g_s = 6 \times 10^{-34}$ at a range of 1 mm, which represents a seven order-of-magnitude improvement over comparable ground-based measurements;
- determine the constant of gravity G with a precision of one part in 10^6 and to test the validity of the inverse square law with the same precision, both two orders of magnitude better than has been achieved on the ground.

To achieve these objectives, the STEP model payload consists of nine differential accelerometers accommodated in a quartz block, which itself is accommodated in a cryogenic dewar. Six of the nine accelerometers are devoted to a test of the Equivalence Principle (EP), two accelerometers mounted at both ends of the quartz block are used to determine the constant of gravity G and to test the validity of the Inverse Square Law (G/ISL), and one accelerometer is used to search for any forces between quantum-mechanical spin and ordinary unpolarised matter (SC). All three types of experiments employ similar highly sensitive measurement techniques (superconducting differential accelerometers) but the driving force is different in each case. For the EP experiment, any differential motion between two test masses would be due to different materials being accelerated differently in the Earth's gravity field; for the SC experiment, any motion of the test masses would result from the hypothetical short-range force between spin-polarised and ordinary matter within the instrument; and for the G/ISL experiment, the motion of the test masses is the result of the gravitational force between two masses within the instrument. All test masses are levitated in the non-sensitive directions by superconducting magnetic bearings. In all three experiments, the signal is periodic, at a low frequency ($\sim 10^{-3}$ to 10^{-4} Hz) which is specific for each experiment. This enables spectral

separation from the noise sources. Differential sensing is employed to attenuate residual satellite motion by a factor of 10^4 .

Testing the Equivalence Principle (EP)

Gravity seems to enjoy a remarkable universality property: all bodies are experimentally found to fall approximately in the same way. Einstein raised this property of universality of free-fall to the level of a grand hypothesis that he termed the "Equivalence Principle". The principle states that there exist accelerated, freely falling local reference frames with respect to which both matter and its interaction fields become "weightless", i.e. apparently decoupled from the external gravitational field. Einstein used the Equivalence Principle as the basic postulate of General Relativity. As such, it deserves to be tested with the highest possible precision.

General Relativity and the Standard Model of particle physics are at the basis of our modern description of the Universe. Both disciplines have problems and limitations, notably the quantisation of gravity, the cosmological problem and mass hierarchies. Many attempts at solving these problems suggest the existence of new interactions between macroscopic bodies with amplitudes smaller than gravity and composition-dependent couplings. Testing the universality of free-fall appears to be the most sensitive way to search for such new interactions. A non-null result of this experiment would presumably constitute the discovery of a new fundamental interaction between macroscopic bodies.

Newton, using pendulums, determined the validity of the EP to one part in 10^3 . Using a torsion balance, Fötvös, in 1896, achieved a sensitivity of 5×10^{-8} . Since then others, using progressively more refined experiments, have achieved a sensitivity of about 10^{-11} . Experiments on the ground are limited by microseismicity and the small driving acceleration.

The EP experiment on STEP would achieve a factor 10^6 improvement over the best existing ground-based experiments. To achieve this very high accuracy, the test masses would be placed inside a satellite in a low-Earth orbit, where they "fall around the Earth". In this way the test masses never strike the ground, and any difference in the rate of fall can build up for a long time. The satellite must be *drag-free* so that the test masses are isolated from disturbing forces on the satellite. The experiment is roughly equivalent to a free-fall from a tower R_E/π km high, with the difference that the signal is periodic and that the experiment can be repeated several thousand times during the mission lifetime.

The experiment consists of two sets of three differential accelerometers which are mounted orthogonal to each other. Each set tests three materials. Ideally, the sum output from the three differential accelerometers $(a_A - a_B) + (a_B - a_C) + (a_C - a_A)$ should be zero. A non-zero sum indicates the level of systematic error.

One set of three accelerometers is provided by European experimenters, the other set by Stanford University. For the European experiment, zirconium, platinum and magnesium were chosen as test mass materials; for the Stanford experiment, copper, gold and magnesium. All six accelerometers use pairs of concentric hollow cylinders. The outer cylinders have a "belt" around the middle which is chosen such that the hexadecapole moments are reduced in addition to making the quadrupole moments disappear. In the European accelerometers, the hexadecapole moments are matched for the inner and outer test masses; in the Stanford accelerometers, the hexadecapole moments are made zero for the outer test masses and negligibly small for the inner.

This design, in conjunction with a tidal management system in the dewar vessel (two-chamber design with electrostatic constraint in the outer chamber) completely eliminates any disturbance (at the 10^{-17} g level) from tidal motions in the liquid helium. This had been identified as a critical problem for STEP during the earlier Assessment Phase study.

It is possible to test whether a signal is due to an EP violation or a disturbance by operating the STEP satellite in three different attitude modes, the *normal*, the *rotated* (rotated by an angle

about the satellite long axis) and the *turning* mode (slowly turning about the satellite long axis).

Spin-coupling experiment (SC)

This experiment searches for any coupling force between spin-polarised and ordinary matter to a sensitivity of $g_p g_s = 6 \times 10^{-34}$ (where g_p , g_s are dimensionless spin-coupling constants) at a range of 1 mm. The best comparable experiments to date have achieved a sensitivity of 2×10^{-27} at a range of 100 mm. In the SC experiment, the fixed source mass is a ferromagnetic material with a high saturation field. Inside the source mass is a coil to polarise the spins of the electrons within the atoms. The source mass exerts a force on an inner and an outer test mass of ordinary material (a platinum - iridium alloy was chosen because of its high density) in opposite directions and their differential acceleration is measured. The direction of the spin is reversed at 2×10^{-3} Hz to generate an oscillating motion of the test masses. To enhance the signal strength, a total of 16 toroidal source/test mass assemblies is used. Superconducting shields ensure that there is no magnetic interaction between the source and the test masses.

If the sensitivity of the instrument can be increased by two orders of magnitude, new limits will be placed on the spin-coupling interaction of the axion, a hypothetical, weakly interacting, massive particle which has been postulated to reconcile the theoretically allowed level of charge conjugation parity (CP) violation in the strong interactions, with the current upper limit to the electric dipole moment of the neutron. It has also been invoked as a possible candidate for the elusive "Dark Matter" in the Universe.

Constant of gravity G and Inverse Square Law experiment (G/ISL)

The Constant of Gravity G is the oldest known constant in nature, defined in Newton's Universal Law of Gravitation three hundred years ago. Today, G is only known to one part in 10^4 and is thus one of the least well-known fundamental constants in physics. Only the product GM is well known (one part in 10^9) for the Earth, Moon, Sun and other bodies. It follows that the mass of these bodies and their density is also only known to one part in 10^4 . To know G, M, and ρ with a high accuracy is not of fundamental importance for present-day physics but this could change as soon as a theory is developed which predicts G in terms of other quantities.

Cavendish made the first measurements of G with an accuracy of one part in 50, two hundred years ago. As a consequence, he was able to "weigh the Earth" for the first time with any precision. Since then, the knowledge of G has been improved only by two orders of magnitude. STEP would advance the present-day knowledge by another two orders of magnitude.

In the G/ISL experiment, a source mass is magnetically driven back and forth at a frequency of 3×10^{-3} Hz. The source mass exerts a time-varying gravitational force on a hollow cylindrical test mass surrounding the source mass. The amplitude of the signal is proportional to G. The shape of the signal curve provides a test of the Inverse Square Law (ISL) of gravity. The test mass is surrounded by another hollow cylindrical test mass (outer test mass) of different material to provide a composition-dependent test of the inverse square law. The addition of the outer test mass also provides a second measurement of G.

The accelerometers are calibrated by keeping the source mass stationary and observing the motion of the test mass in the varying Earth gravity field. The slight variations can be predicted with sufficient accuracy from the geopotential models, to be improved by the STEP geodesy experiment, in conjunction with satellite orbit altitude determination at cm-precision level.

Two identical accelerometers of this kind are mounted at either end of the quartz block and the two source masses are driven in antiphase to eliminate any motion of the combined centre of mass and thus any recoil on the satellite. In order to reject the background satellite acceleration, the inner two test masses are coupled to form a differential accelerometer (similarly the outer two test

masses). This also forms a highly sensitive gradiometer for the geodesy experiment.

Geodesy

The extreme demands of the fundamental physics experiments on STEP also offer an opportunity for a unique geodesy mission. With STEP's orbit being rather low (550 km) and almost polar ($i = 97^\circ$), the global gravity field can be mapped with considerable detail. The long wavelength features of the gravity field are extracted from the orbit which is determined with cm-level precision since the satellite is drag-free and continuously tracked in three dimensions using the Global Positioning System (GPS). Laser reflectors mounted on the STEP satellite remove ambiguities and convert this precision to accuracy.

The medium wavelengths (down to 130 km half wavelength) are derived from gravity gradiometry using the two G accelerometers. They will provide the out-of-plane gravity gradient component with a precision of $10^{-4} \text{ E}/\sqrt{\text{Hz}}$ ($1 \text{ E} \equiv 10^{-9} \text{ s}^{-2}$). This extremely high precision largely compensates for the natural attenuation with altitude. The two geodesy elements - gradiometry and GPS on a drag-free satellite - are unique and would result in a gravity field comparable to the proposed dedicated gravity mission GAMES and not much short of ARISTOTELES. The gravity field model, significantly improved both in terms of resolution and accuracy, would serve important applications in geodesy, solid-Earth physics and oceanography.

Advantages of space, drag-free and cryogenics

The STEP space laboratory has a major advantage over any laboratory on the ground because the level of microseismicity is reduced by a factor of 10^4 . The STEP satellite compensates for any air drag so that the test masses essentially follow a purely gravitational orbit and the experiments can be performed under near-ideal zero-gravity conditions (10^{-13} g). Furthermore, in low Earth orbit, the driving acceleration for a test of the Equivalence Principle is much larger than on the Earth's surface. The experiment chamber is cooled to cryogenic temperatures (2 K) which has several important advantages:

1. A very stable and highly sensitive position detector is available, a SQUID magnetometer (Superconducting Quantum Interference Device), which can, in principle, detect in 1 s relative displacements of test masses of typically 100-200 grammes with a sensitivity of 10^{-15} m , the diameter of the nucleus of a hydrogen atom. In the STEP experiments, the superconducting sensor is optimised for maximum acceleration sensitivity and can detect a relative acceleration of $10^{-15} \text{ m s}^{-2}$ in 1 s.
2. Almost perfect magnetic shielding from the Earth's magnetic field is available by using superconductors (a thin lead bag around the experiment chamber attenuates the Earth's magnetic field by a factor 10^{10}).
3. Gas pressure can be greatly reduced. At 2 K all gases except helium are frozen and pressures less than 10^{-11} torr are feasible.
4. Radiation pressure disturbances due to temperature gradients are greatly reduced (decreases as the fourth power of the temperature).

Furthermore, the STEP satellite has no moving parts. It is three-axis stabilised with a pointing accuracy of 2 arc sec. By choosing a suitable launch window and a Sun-synchronous orbit, eclipses can be avoided, adding considerably to the thermal stability.

Satellite design and mission analysis

The STEP satellite has the shape of an octagonal box, 2.8 m high and 2 m wide. A solar cell array of 7.5 m^2 area providing 550 W of power is mounted on top of the satellite, continuously pointing towards the Sun in flight configuration. The spacecraft is open in the middle so that the dewar which is in the shadow of the solar array, can radiate freely to all sides.

In order to detect an Equivalence Principle violation of 10^{-17} g in 10^5 s (20 orbits), the STEP satellite residual acceleration must be less than $10^{-12} \text{ m s}^{-2}$ (rms) over the 10^{-5} Hz measurement bandwidth. This disturbance attenuation will be provided by the drag-free system which utilises proportional helium thrusters to null the drag forces from air drag and solar radiation. The boil-off from the dewar is used to feed the proportional thrusters. The dewar holds 200 l of superfluid helium which will last for at least six months with 100% margin.

The STEP satellite would be launched in the spring of 2002 into a circular, Sun-synchronous orbit. The lowest possible altitude is desirable to improve the resolution of the geodesy experiment at medium wavelengths and to avoid as much as possible the energetic charged particles (protons) in the South Atlantic Anomaly (SAA). The minimum altitude is determined by the helium boil-off rate in relation to the air drag at orbital altitude. The Earth's atmosphere expands and shrinks with the solar cycle. For a launch in 2002 (just after solar maximum), the minimum orbital altitude was determined to be 550 km under worst-case assumptions, i.e., an extremely active next solar cycle and with a 75% margin in the STEP thrust authority.

The high fluxes of energetic particles (protons with energies around 100 MeV) can disturb the measurements by causing momentum transfer, temperature rise, and electrostatic charging. To monitor this radiation originating from the SAA as well as from large solar flares, a bi-directional charged particle detector is included in the payload. Whereas the effects of momentum transfer are easily eliminated, charge compensation is required. Two approaches are considered, either discharging with a small UV light source, or grounding the test masses by a thin (5 microns) flexible wire which also removes heat. Therefore, the scientific objectives can be fully met at any time.

1. The scientific significance of STEP

The primary scientific objective of the STEP mission is to improve a crucial test of Einstein's theory of gravitation to a level a million times more precise than that presently achieved by ground-based experiments. A non-null result of the STEP Equivalence Principle experiment would presumably constitute the discovery of a new fundamental force.

Gravitation is one of the four known basic forces in our current understanding of nature. The other three are the electromagnetic, strong, and weak interactions. For gravity, as opposed to the other forces, the precision laboratory has always been space. Here most of the experimental basis for the theory has only come recently, due to the development of the space program and to the remarkable accuracy of modern spaceborne measurement technology. It is the general goal of the STEP mission to take advantage of this technology to advance the experimental knowledge of gravitation over a broad front.

The specific goals of the STEP mission in the area of gravitation are to improve by six orders of magnitude the verification that all test bodies fall at the same rate in a gravitational field, to improve by two orders of magnitude the test of the $1/r^2$ dependence of the gravitational force at distance scales r of a few centimetres, to search for a possible gravitational-strength coupling between spin and mass, and to improve by two orders of magnitude the measurement of the gravitational coupling constant.

1.1 History of the Equivalence Principle

Galileo and Newton were the first to point out that gravity seems to enjoy a very remarkable universality property: all bodies are experimentally found to fall in approximately the same way. Einstein raised this property of "universality of free-fall" to the level of a grand hypothesis, which he used as a foundation stone in building his theory of General Relativity. Einstein's basic hypothesis is termed the "Equivalence Principle" and states that the universality of free-fall is an exact law of nature which applies not only to all kinds of matter but also, in a sense, to all the fields that mediate the various physical interactions binding the matter. In other words, there must exist accelerated, freely falling local reference frames with respect to which both matter and its interaction fields become "weightless", i.e. apparently decoupled from the external gravitational field.

Starting from the Equivalence Principle, Einstein was led to make a number of new predictions for spacetime measurements and to propose a new theory of gravitation with a radically different intellectual basis from Newton's theory: General Relativity. This theory is essentially based on two independent postulates: (i) a precise mathematical formulation of the Equivalence Principle as the postulate that the response to an external gravitational field of all the elementary constituents of matter and their binding fields is described by a universal coupling to a curved spacetime metric $g_{\mu\nu}(x^\lambda)$. (ii) the assumption that the metric $g_{\mu\nu}(x^\lambda)$ suffices to describe the entire dynamics of the gravitational field.

From a historical standpoint, the universality of free-fall and its precise wording as the Equivalence Principle, constituted the basis of General Relativity. With the benefit of hindsight one may regard the universality of free-fall as the most precisely testable consequence of General Relativity, rather than a "First Principle". This point of view appears natural as one considers gravity along with the rest of the known fundamental forces. The weak, strong and electromagnetic interactions are all mediated by particles of spin one. The detailed dynamics of the corresponding theories follows from a "gauge principle": the statement that the local phases of the weakly, strongly or electromagnetically charged particles are unobservable. This gauge invariance is intimately related

to the fact that the corresponding charges are conserved. Similarly, General Relativity can be understood as a field theory of interactions mediated by massless particles of spin 2 (gravitons). The theory also follows from a Gauge Principle: invariance under general coordinate reparametrizations. The corresponding "charge" is energy-momentum or, in the case of a body at rest, its mass.

Whether "Equivalence" is regarded as a principle (the classical relativists' point of view) or as a key prediction (the particle-theorists' prejudice) is in practice immaterial. Testing Equivalence is the most sensitive way to search for putative non-electromagnetic long-range forces. If these were out of reach, a null result from STEP would still provide a deep confirmation of Einstein's startling idea that gravity is describable as a distortion of the four dimensional geometry of spacetime.

1.2 General Relativity : a basic ingredient of our present world view

For over forty years after Einstein proposed his theory in 1915, General Relativity remained quite isolated from the mainstream of science. This situation changed drastically starting in the 1960s because of the uncovering of new concepts within General Relativity, the discovery of new relativistic astrophysical objects, and the development of high-precision techniques to test various aspects of the theory. General Relativity now plays an essential role in our understanding of nature. On the one hand, it provides the theoretical basis of the present description of the macroscopic world: the big bang, the cosmological expansion, the large-scale structure of the Universe, the end points of the evolution of stars, gravitational collapse, neutron stars, black holes, gravitational waves, relativistic celestial mechanics of the solar system, the high-precision description of the motion of natural and artificial satellites, the definition of the international atomic time, etc. On the other hand, while for a long time the geometrical structure of General Relativity made the gravitational interaction appear very different from the other interactions in nature, progress in the 1970's on the understanding of the geometrical aspects of the three interactions ruling the microscopic world (gauge theories of electromagnetic, weak and strong interactions) has suggested that a unified description of nature may contain geometrical features generalising those of General Relativity and gauge theories.

In view of its essential role amidst modern science, General Relativity deserves to be tested with the highest precision possible.

1.3 Current status of the tests of General Relativity

General Relativity is the most appealing theory of gravity. It has already been tested to a high level of precision. Thus, a result in contradiction with one of its predictions would most conservatively be interpreted as the discovery of a new force of nature and not as a fundamental flaw of the original theory.

Most new forces would violate the universality of free-fall in the sense of involving couplings to quantities other than mass. A different rate of fall of two objects of the same mass but *different composition* would signal the presence of such a force. If the particles mediating a new force have a non-zero mass, the corresponding interaction between two bodies *A* and *B* extends over a finite range λ . Upon neglect of terms proportional to $(v/c)^2$ (Newtonian approximation), the total interaction potential of gravity plus the new force can be written as

$$V_{AB}^N(r) = -\frac{G_\infty m_A m_B}{r} [1 + \alpha_{AB} \exp(-r/\lambda)] \quad (1.1)$$

where α_{AB} (which is positive in case of an attractive force and negative otherwise) parametrizes the fractional effect of the new force. Its general form is

$$\alpha_{AB} = \tilde{\alpha} \frac{q_A}{\mu_A} \frac{q_B}{\mu_B} \quad (1.2)$$

where $\tilde{\alpha}$ is a constant, μ_A denotes the mass of body A in atomic units and q_A its macroscopic "charge", to which the new field is coupled.

A composition-dependent force ($q \neq \mu$) being significant for all separations not much bigger than λ , an observation at a distance r constrains $\tilde{\alpha}$'s for all ranges larger than r . The deviations from a pure $1/r$ behaviour and/or the $(v/c)^2$ corrections to the interaction potential are two other signatures of a new force. (Note that tests of the $1/r$ law at distance $\sim r$ significantly constrain $\tilde{\alpha}$'s only over ranges λ of the order of r .)

The most precise ground-based composition-dependent tests are the experiments of Roll, Krotkov and Dicke (1964), Braginsky and Panov (1972), Niebauer, McHugh and Faller (1987), Adelberger et al. (1990) (see Adelberger et al. (1991) for a review). Their work has established that, for several different pairs of bodies (A, B), the long-range component of gravity acts proportionally to the total mass of the bodies with a fractional accuracy $\sim 2 \times 10^{-11}$, at the 95% confidence level. STEP's primary objective will be to improve this limit, by six orders of magnitude, down to the $\sim 10^{-17}$ level.

The *Post-Newtonian* approximation, the inclusion of $O(v^2/c^2)$ sub-leading terms, makes many more tests of deviations from General Relativity possible. In the late 1960s, Nordtvedt and Will, extending earlier ideas of Eddington, developed a phenomenological parametrized Post Newtonian (PPN) formalism aimed at giving a general parametrization of the structure of the metric $g_{\mu\nu}$ in the quasi-stationary weak-field post-Newtonian limit appropriate to a description of the Solar System. This formalism has helped in compiling a long list of post-Newtonian non-Einsteinian effects that distinguish General Relativity from its possible competitors. The ensemble of gravitational experiments performed in the Solar System has enabled one to determine with an accuracy uniformly better than 2×10^{-3} (1σ level) all the PPN parameters (see Will, 1992 for a review). They were all found to be consistent with the values predicted by General Relativity.

The two key PPN parameters are $\gamma - 1$, the coefficient of the $O(v^2/c^2)$ deviations from General Relativity in the two-body potential, and $\beta - 1$, the size of anomalous three-body effects. Among all the existing tests, two stand out as leading to the most precise determination of γ and β . One is the separate determination of γ through the measurement of the gravitational time delay of electromagnetic signals passing near the Sun. This effect was first proposed by Shapiro in 1964, and the Viking mission result is $|\gamma - 1| < 2 \times 10^{-3}$ (1σ level). The second test was proposed by Nordtvedt in 1968, and gives access to the combination $\eta \equiv 4\beta - \gamma - 3$. It consists in looking for a possible additional 28-day oscillation in the Earth-Moon distance as the two bodies orbit each other in the gravitational field of the Sun. The effect would be there if the gravitational binding energies of the Earth-Moon system "fell" towards the Sun in an anomalous fashion (in violation of the "Strong" Equivalence Principle). The measurements are still being made with very high precision by laser ranging to several passive retroreflectors placed on the Moon by the Apollo astronauts and a French retroreflector placed by the Russian Lunar Lander. The analysis of present data finds no violation of the strong Equivalence Principle but rather that the fractional differential acceleration of the Earth-Moon system towards the Sun is smaller than 2.2×10^{-12} (1σ level). Given the gravitational binding energy of the Earth ($-4.6 \times 10^{-10} m_E$) and the Moon ($-0.2 \times 10^{-10} m_M$) this yields $|4\beta - \gamma - 3| < 5 \times 10^{-3}$. Combining this result with the above determination of γ gives $|\beta - 1| < 2 \times 10^{-3}$ at the 1σ level. Note that the precise measurement of β and γ has been made possible by space techniques.

The above result means that "gravity pulls on gravity" in the way predicted by General Relativity. This is a test of a very specific non-linear property of the theory; one would be well advised to eliminate its only known weakness. After all, the iron-dominated Earth and the silica-dominated Moon may accelerate differently toward the Sun simply by virtue of their different material compositions, an effect that may accidentally cancel against the effect of a non-vanishing η . This offers a

strong motivation to improve the accuracy of Equivalence Principle tests from the present $\sim 10^{-11}$ level (1σ) down to at least a few parts in 10^{13} . More precisely, the limits on composition-dependent forces obtained by Dicke et al. (1964) and Adelberger et al. (1990), when combined, still allow for an anomalous fractional differential acceleration of the Earth and the Moon in their fall towards the Sun of $\approx 5 \times 10^{-12}$ at the 2σ level. This is larger than the present 2σ observational uncertainty $\approx 4.4 \times 10^{-12}$. One expects that in the coming years, with more and better data, the precision of the lunar experiment will be improved by one or two orders of magnitude. Hence, a clear need exists for an improved experimental confirmation of the universality of free-fall rates for bodies of different material at the $\approx 10^{-13}$ level.

A complete list of General Relativity tests ought to include the gravitational-wave radiation damping and strong-field effects that have been studied via the high-precision timing data of several binary pulsar systems (Damour and Taylor, 1992).

Given that General Relativity has been shown to be consistent with experiment and observation at the 0.1% level, why should one continue to test its predictions and foundations? General Relativity contains no free dimensionless parameters and any improvement in the precision of a measurement can be a lethal test of the theory. Beyond the desire to improve the verification of any basic theory, there are several puzzles in fundamental physics that lead one to expect that General Relativity may not be the complete description of how apples fall. These reasons are reviewed in the next two sections.

1.4 Puzzles around the Standard Model of particle physics

The Standard Model of particle physics is based on an $U(1) \times SU(2) \times SU(3)$ quantum gauge field theory of the electromagnetic, weak and strong interactions. The model successfully accounts for all existing particle data, but has its own shortcomings and puzzles.

A main demerit of the Standard Model is its suspiciously complicated structure: a non-simple gauge group, repetitive particle multiplets and some seventeen free parameters. We do not yet have a rationale for this considerable complexity. Moreover, the Standard Model has its own internal deficiencies (concerning CP violation and the mass-hierarchy problem) as well as peripheral troubles (related to the cosmological constant and to quantum gravity). We proceed to recall how some of these limitations naturally suggest extensions that lead to violations of the Equivalence Principle:

- A) *The strong CP puzzle.* The strong interactions, described by the $SU(3)$ gauge theory of coloured quarks and gluons, are such that parity (P), time reversal (T) and charge conjugation (C) are automatically conserved in perturbation theory. Non-perturbative effects, however, induce a strong violation of P and CP, parametrized by a dimensionless angle called θ . This parameter is observationally found to be smaller than 10^{-9} , while the a-priori expectation is of the order of unity. The only attractive resolution of this problem involves the introduction of a new force carried by the *axion* field, to which we shall return.
- B) *The hierarchy puzzle.* The scale at which weak and electromagnetic interactions are “unified” into an $SU(2) \times U(1)$ theory is ~ 250 GeV. A possible “Grand” unification with the strong $SU(3)$ interactions cannot occur below some 10^{15} GeV. This large ratio of scales must be put in by hand and, still worse, to keep these scales separate, the quantum theory must be fine-tuned at each order of perturbation. The only known solution to the second of these problems involves the introduction of a *supersymmetry* between particles of integer and half-integer spin. In some supersymmetric models the existence of a new force mediated by a spin-one field is unavoidable.
- C) *The cosmological constant.* In General Relativity, the mass density of the vacuum (the cosmological constant Λ) has a definite meaning, since it necessarily couples to gravity. Cosmological

observations show that Λ is smaller than 10^{-17}GeV^4 (in units where $\hbar = c = 1$). As the Universe expanded and cooled, it ought to have suffered various phase transitions described by the Standard Model of micro-physics. During the electroweak transition, Λ changed by some 10^{+8}GeV^4 , while during the QCD “deconfining” transition, it changed by some 10^{-1}GeV^4 . Overall, the naive expectation is $\sim 10^{+71}\text{GeV}^4$, if the zero-point vacuum quantum fluctuations are cut-off at the Planck mass scale $m_P \simeq (\hbar c/G)^{1/2} \sim 10^{19}\text{GeV}/c^2$. Why is the current value of the cosmological constant ~ 0 on any of its natural micro-physical scales? How did the newly-born Universe know that it had to home in on a vanishing Λ ? Once again, attempts to solve this behemoth problem often imply the existence of new long-range forces.

- D) *The quantisation of gravity.* How to combine the (perturbatively renormalizable) quantum theory of electroweak and strong interactions with General Relativity, the dimensionality of whose coupling constant $G = \hbar c/m_P^2$ precludes the application of perturbative quantisation? Practically all solutions to this most challenging problem involve the existence of novel long-range forces.

1.5 Examples of forces violating the Equivalence Principle

Most puzzles just discussed involve the concept of mass: mass hierarchy within the standard model, the mass density of the vacuum, the mass scale ($m_P \sim 10^{19}\text{GeV}$) defined by the gravitational constant. The origin of mass is one of the least understood issues in present-day physics. Perhaps, as in the examples to be now reviewed, the solutions to these puzzles will involve additions to the interaction which couples to mass, i.e. gravity.

Following the labelling of the problems discussed in the previous section:

- A) The “axion” solution of the strong CP problem involves a new field a , with a non-vanishing mass m_a . Its dominant coupling to matter is pseudoscalar, with a tiny scalar admixture. Through its scalar couplings to quarks, $ga(\bar{u}u + \bar{d}d + \dots)$, the axion mediates an attractive force between macroscopic bodies of finite range $\lambda_a = \hbar/m_a c$, which current observations constrain to $2 \times 10^{-4}\text{cm} < \lambda_a < 20\text{ cm}$ (Moody and Wilczek, 1984). The magnitude of the force can be comparable to gravity for the smallest allowed ranges, but is expected to be at least ten orders of magnitude weaker than gravity for $\lambda_a = 20\text{ cm}$. This force violates the Equivalence Principle.
- B) The extra vector field U that appears in some supersymmetry-inspired extensions of the standard model (Fayet 1986, 1990) generates a repulsive interaction between macroscopic bodies. Its range λ_U is arbitrary and its magnitude relative to gravity is constrained to be $|\alpha_U| \lesssim 0.16(1m/\lambda_U)^2$. Once again, the new force violates the Equivalence Principle because it acts (on neutral matter) proportionally to

$$B - L \equiv \text{baryon number minus lepton number} = N \quad (1.3)$$

where N denotes the number of neutrons ($B = N + Z$, $L = Z$ for neutral matter, with Z the number of electrons, i.e. the atomic number).

- C) In an attempt to solve the cosmological constant problem some authors (Peccei, Solà and Wetterich, 1987) have suggested the existence of a new scalar particle (the cosmon) linked to the breaking of dilatation symmetry. Such a field would generate an intermediate-range, attractive Equivalence-Principle-violating force between macroscopic bodies. The estimated mass of the cosmon is $\geq \Lambda_{QCD}/M$, with $\Lambda_{QCD} \sim 300\text{MeV}$ the Quantum Chromo-Dynamics mass scale, and M the mass scale of the breaking of dilatation symmetry (taken to be $\sim m_P$).

This yields a macroscopic range, $\lambda_c \leq 10 \text{ km}^1$. The magnitude of this extra force is at most three times weaker than gravity. The new force couples to a combination of mass, baryon number and atomic number approximately given by

$$\mu = 0.05B + 0.002L \quad (1.4)$$

where $\mu = m/(1 \text{ amu})$ is the mass in atomic units.

- D) Attempts to solve the challenging problem of quantising gravity also generically predict the existence of new fields generating Equivalence-Principle-violating macroscopic interactions. A well-motivated attempt is supergravity. In particular, within the framework of extended supergravity, it was noted (Scherk, 1979) that the usual (spin-2) graviton is accompanied by lower-spin partners, notably a spin-1 particle (the graviphoton) giving rise to repulsive Equivalence-Principle-violating interactions. From an estimate of the mass acquired by the graviphoton via spontaneous symmetry breaking of a scalar field, ϕ , coupled in a $U(1)$ -gauge-invariant way to the graviphoton ($m_g = g_\phi < \phi >$ with $g_\phi = \sqrt{4\pi G} m_\phi$), one finds a macroscopic range of typical kilometric scale

$$\lambda_g \approx \left(\frac{1 \text{ GeV}}{m_\phi} \right) \left(\frac{1 \text{ GeV}}{< \phi >} \right) \text{ km} \quad (1.5)$$

This force couples with gravitational strength to the mass-current of the quarks : $m_u \bar{u} \gamma^\mu u + m_d \bar{d} \gamma^\mu d + \dots$. Since these quark masses represent only a small non-universal fraction of the mass of the nuclei of atoms, this gives rise to a force between macroscopic bodies which is about $[(m_u + 2m_d)/m_N]^2 \sim 3 \times 10^{-4}$ weaker than gravity, and which couples (approximately) to the combination

$$B = 0.17L \quad (1.6)$$

The most ambitious theory attempting to unify gravity with the other interactions is string theory. In this approach the original (ten-dimensional) tensorial gravitational field $G_{\mu\nu}$ has two partners: a scalar field Φ (called the dilaton) and an antisymmetric tensor field $B_{\mu\nu}$. These fields are coupled to the other fields in the theory with gravitational strength, and in ways which generally violate the Equivalence Principle (Taylor and Veneziano, 1988). One does not know at present how to connect in detail the field content of string theory to the four-dimensional "low-energy" world described by the standard model and General Relativity. Many scalar and pseudo-scalar partners of the graviton could survive as massless particles in the four-dimensional low-energy world (dilatons, axions, moduli fields, etc.). A mechanism has been suggested wherein the cosmological expansion drives the matter couplings of these fields below, but not unreachably below, the current Equivalence Principle limits (Damour and Nordtvedt, 1992; Damour and Polyakov, 1993).

In conclusion, though none of the above-discussed models is truly convincing, they all point to the existence of new interactions between macroscopic bodies, with amplitudes smaller than gravity, and Equivalence Principle violating couplings. Neither the range, nor the intensity, nor the precise way in which the Equivalence Principle is violated are known in advance. STEP is an exquisitely precise tool to search beyond the Standard Model for new physical phenomena of the types we have discussed.

¹Note that a very general argument (Weinberg, 1989) shows that if the cosmological constant problem is to be solved dynamically through some self-adjustment mechanism driven by a field ϕ , then this field must be very light, $m_\phi < 10^{-3} \text{ eV}$, i.e. it must have a macroscopic range : $\hbar/m_\phi c > 0.1 \text{ mm}$.

1.6 Spin-couplings and Axion-like forces

It is a very general property of field theories that the spontaneous breaking of continuous global symmetries generates massless spin-zero particles (Goldstone bosons). When these parental symmetries are only approximate, the (pseudo-)Goldstone bosons acquire a small mass (inversely proportional to the energy scale of the symmetry's breakdown). Though the best-motivated particle of this breed (after the pion) is the axion, it is quite conceivable that various types of such particles do exist and induce macroscopic forces.

Axions or axion-like particles have scalar and pseudo-scalar couplings to matter; they couple both to a “monopole charge” and to a magnetic-like “dipole charge”. Only spin-polarised bodies would possess a macroscopic charge of the dipole type. Axion exchange results in monopole-monopole, monopole-dipole and dipole-dipole forces. Moody and Wilczek (1984) have argued that the monopole-dipole forces are the easiest to detect. The interaction potential between an electron with spin polarised in the direction $\hat{\sigma}$, and an unpolarised nucleon, extends over the range λ , and is of the form

$$V_a = g_s g_p \left(\frac{\hbar^2}{8\pi m_e} \right) \hat{\sigma} \cdot \hat{r} \left(\frac{1}{\lambda r} + \frac{1}{r^2} \right) \exp(-r/\lambda) \quad (1.7)$$

where g_s and g_p are dimensionless coupling constants referring to the scalar and pseudo-scalar vertices, m_e is the mass of the electron and \hat{r} is the unit position vector of the nucleon relative to the electron. The axion interaction is described by two quantities; a spontaneous symmetry-breaking energy scale, F , and the magnitude of CP violation suppression, θ . One property of the axion is that its mass is related to the energy scale F as

$$m_a \simeq 10^{-5} eV (10^{12} GeV / F) \quad (1.8)$$

Alternatively, the range of the axion-mediated interaction is given as

$$\lambda \simeq 2 \text{ cm} (F / 10^{12} GeV) \quad (1.9)$$

Moody and Wilczek's theory leads to

$$g_p g_s = \frac{\theta}{\lambda^2} \times 6 \times 10^{-33} \quad (1.10)$$

with λ in metres.

There are well accepted upper and lower limits to the range: In order that the Universe should not have a density much larger than the closure density, the relic axions must be more massive than around $10^{-6} eV$ (the range must be less than about 20 cm). An upper limit to the mass of the axion can be derived from observations of neutrinos from the supernova SN1987a. If axions coupled sufficiently weakly then a significant amount of thermal energy could have been transported from the collapsing core and this would have shortened the observed neutrino signal. The axion is therefore thought to be lighter than $10^{-3} eV$, which gives a minimum range of around 0.2 mm (these limits are discussed in Kolb and Turner, 1990). Constraints on the value of θ are derived from the electric dipole moment of the neutron which is limited to a maximum value of $5 \times 10^{-25} e \text{ cm}$ (e is the electron charge) (Pendlebury et al., 1984). Values of θ are model dependent with a maximum of about 10^{-8} but the majority of models give an upper limit of 10^{-9} . There have been some attempts at deriving a theoretical value for θ , notably Wilczek estimates its value as 10^{-14} . In practice it is difficult to examine ranges as small as 0.2 mm so, if we restrict ourselves to a minimum range of 1 mm, we see that the maximum value that can be expected for the product of the coupling constants is

$$g_p g_s = 6 \times 10^{-36} \quad (1.11)$$

This extremely small coupling is specific to a specific axion theory. Our current understanding of hypothetical scalar particles is very modest and definitely allows for the existence of axion-like forces with stronger couplings, in a large domain that has not been explored.

This is a fairly new area of experimentation and there are, as yet, only a handful of results for the limits on the product of $g_p g_s$. Experimental results up to the end of 1990 are described in the comprehensive review of Adelberger et al. (1991). More recently Venema et al. (1992) employed NMR techniques with mercury atoms to establish an upper limit of 7×10^{-35} for an interaction with a range of 10^6 m. Ritter et al. (1993) performed a torsion balance experiment and have placed an upper limit on $g_p g_s$ of 5×10^{-27} for ranges larger than 10 cm. It should be noted that, for a given acceleration sensitivity, a test for a long-range coupling using the Earth will place smaller upper limits on the coupling constant product simply because the long-range test can make use of the whole Earth as a source. It is difficult to predict the progress that can be made with these ground-based experiments over the next decade, but it seems likely that the sensitivities of the mechanical experiments will increase by about 1 order of magnitude when techniques are developed for suppressing the influence of seismic noise on the apparatus. One would expect a short-range test using NMR techniques to be done in the near future with a sensitivity of about 10^{-28} at 10 cm. It appears then that no experiments have yet approached the sensitivity required to place restrictions on axion theory.

The exceptionally quiet STEP environment can be put to use to search for spin-coupling interactions with a sensitivity of $g_p g_s$ of 6×10^{-34} at a range of 1 mm. This represents a seven-order-of-magnitude improvement over the existing ground-based measurements.

1.7 The value of the Newtonian gravitational coupling constant, G

Newton's gravitational constant, G, is one of the least precisely measured fundamental constants in physics (~ 130 ppm as opposed to 0.045 ppm for the fine structure constant α , or 0.6 ppm for the electron mass m_e , to cite but two examples). As it happened for the electroweak mixing angle θ_W , this lack of precision could become significant² as soon as a theory is capable of predicting G in terms of other quantities by stating, for instance, that $G = [(7\pi)^2 \hbar c / 5m_e^2] \exp(-\pi/4\alpha) = 6.6723458 \dots \times 10^{-8} \text{ cm}^3 \text{ g}^{-1} \text{ s}^{-2}$ which is currently compatible with experiment.

As any other fundamental constant, G ought to be measured with state-of-the-art precision. STEP may do it.

The gravitational constant is the oldest known constant in nature, defined in Newton's Universal Law of Gravitation three hundred years ago. In fact, the attempt to determine G has become the cornerstone of modern precision experiments. Cavendish started the tradition of precision laboratory measurements when he "weighed the Earth" using the torsion balance apparatus (Cavendish, 1798) which had been invented earlier by Mitchell for the same purpose. His result, when interpreted in terms of the gravitational constant, gives a value of $(6.754 \pm 0.123) \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$, which is within 1% of the presently accepted value, $(6.67259 \pm 0.00085) \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$. Two hundred years of dedicated effort on the part of experimenters has, therefore, improved the value of G by only a factor of 100, with the uncertainty still standing at a level of about one part in 10^4 .

The difficulty in improving the value of G arises fundamentally from the weakness of the gravitational coupling, 10^{-10} of that of the electromagnetic interaction. However, modern space technology, coupled with the exquisitely sensitive STEP accelerometers, gives us an opportunity to improve the

²Up to the early seventies there seemed to be no particular point in measuring with precision $\sin \theta_W$ nor α_s , the strong-interaction analog of the fine structure constant α . With the discovery of "Grand" unified theories of the electroweak and strong interactions, the situation changed, for it became possible to predict $\sin \theta_W$ in terms of α and α_s . History repeated itself as the prediction of $\sin \theta_W$ turned out to disagree with improved data and supersymmetry came to the theorists' rescue.

value of G by two orders of magnitude, a feat representing two hundred years of progress, if extrapolated linearly from the history of G experiments.

Closely related with the measurement of G is the inverse square law nature of gravity. If the inverse square law should be violated, G might not be a *universal constant*, or, at least, it should be redefined. In fact, Cavendish, in his classic experiment, remarked that an “objection, perhaps, may be made to these experiments, namely, that it is uncertain whether, in these small distances, the force of gravity follows the same law as in great distances.” STEP will also investigate the distance dependence of the gravitational force at short range, improving the value of α (Eq. 1.1) by two orders of magnitude, to 10^{-6} at $\lambda \approx 1$ cm.

1.8 Summary

The Equivalence Principle, the inverse square law, and the strength of the interaction itself are the three foundation stones on which both Newton and Einstein built their gravitational theories. STEP will test all three of these basic properties of gravity. The STEP satellite borrows its name from its main experiment to test, with unprecedented precision, the Equivalence Principle, a pivotal ingredient of Einstein’s theory of gravitation. Even a null result from this experiment would remain, for years to come, a significant constraint on future theories of fundamental physics.

General Relativity and the Standard Model of particle physics are at the basis of our modern description of the Universe. Both disciplines have problems and limitations, that we have briefly reviewed (the quantisation of gravity, the cosmological constant, masses and mass-hierarchies...) Scores of purported solutions to these problems have a point in common: the introduction of new, collective, macroscopic forces that would violate the Equivalence Principle by implying a non-universality of free-fall. STEP will probe a large domain of the parameter space of these possible theories that no ground-based experiment has access to. This is illustrated in Fig. 1.1 which shows the existing limits of a new, hypothetical composition-dependent force (assumed to couple to baryon number) and the resolution expected from STEP. STEP will improve α by six orders of magnitude in the range from 10^3 km to infinity.

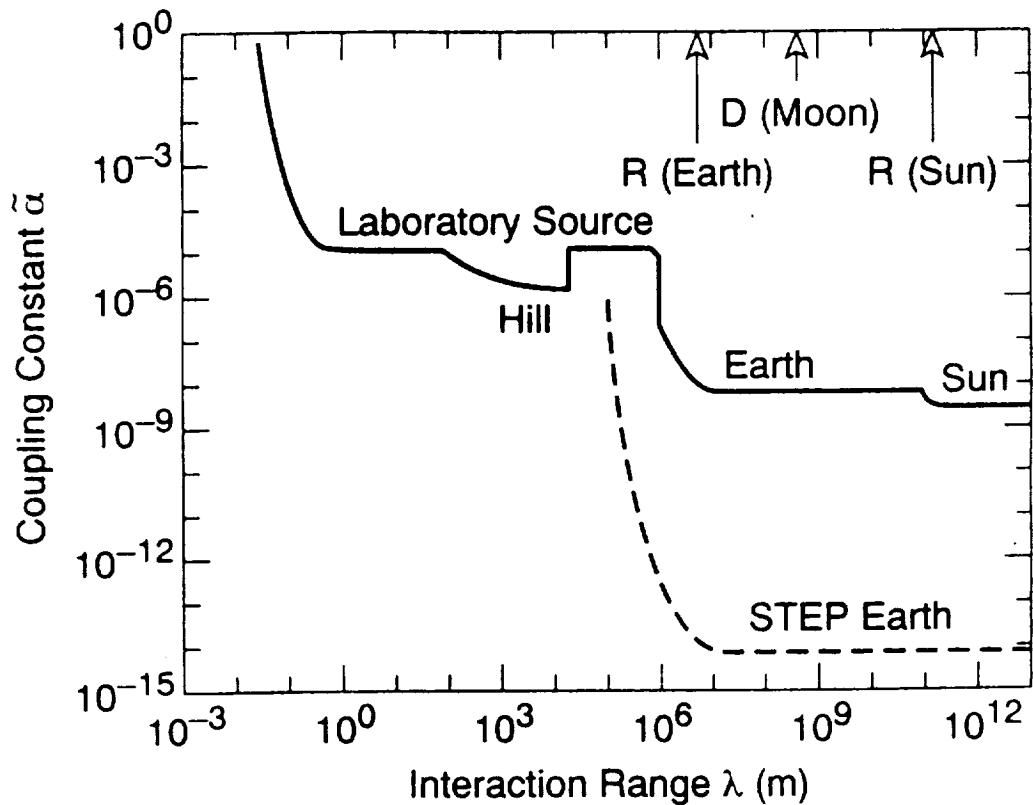


Fig. 1.1. Existing limits and resolution expected from STEP for a new composition-dependent force. The solid curve represents the existing limits in $\tilde{\alpha}$ as a function of λ , adapted from Adelberger et al. (1991). The charge for the new macroscopic force is assumed to be baryon number. The dotted line represents the resolution expected from STEP. The Equivalence Principle experiment, which uses the Earth as the source, improves $\tilde{\alpha}$ by six orders of magnitude in the range from 10^3 km to infinity.

2. The STEP experiments

2.1 Experimental concepts and the advantages of space

2.1.1 Long-range test of the Equivalence Principle

The *inertial mass* M_i of an object measures its acceleration a in response to an applied force according to Newton's formula $F = M_i a$, whereas the gravitational force on the object (its weight) is proportional to its *gravitational mass* M_g times the gravitational field g . The Equivalence Principle says that M_i and M_g always have the same ratio regardless of the nature of the object. If this were not so, different objects would fall with different accelerations, proportional to M_g/M_i , in the same gravitational field, and we would have to regard M_g and M_i as independent variables. The ratio M_g/M_i can be measured for different materials by measuring their rates of fall, much like the charge/mass ratio can be measured for different elementary particles. The experimental fact that all objects fall with the same acceleration (as nearly as we can measure) causes us to regard M_g and M_i as measuring one variable, *mass*, in two different ways.

On the ground, Galileo's classic free-fall test of the Equivalence Principle (EP) has the advantage that it uses the full acceleration of the Earth's surface gravity (9.81 m s^{-2}), to drive any difference in the rate of fall, but it suffers disadvantages from the short fall even without air drag. The most accurate experiments searching for any differences in the ratio M_g/M_i have therefore used torsion balances to convert any difference in acceleration into a rotation.

In the Eötvös experiment, the horizontal component of the centrifugal force (produced by the rotation of the Earth acting on the inertial mass of a test object) is balanced against a component of the Earth's gravity (acting on the test objects' gravitational mass). The maximum value of the driving force occurs at a latitude of 45° and corresponds to an acceleration of 0.013 m s^{-2} . In experiments of the type performed by Dicke and Braginsky (Roll et al., 1964; Braginsky and Panov, 1972), the centrifugal acceleration produced by the Earth's orbit around the Sun (0.007 m s^{-2}) is used to provide the driving force. Since its direction is modulated by the rotation of the Earth, it is unnecessary to reverse the force by rotation of the experiment. This removes one of the principal sources of error in the Eötvös experiment, the hysteresis of the torsion fiber. Experiments of Dicke's type are the most sensitive that can be made on the Earth but they are limited by changes in the local Earth's gravitational field and seismic fluctuations which produce a torque on what is essentially a dumbbell suspended on a pendulum. Many years of progress have reduced the influence of gravity gradients and noise, for example by increasing the symmetry of the balance to reduce its low-order gravitational moments. The resulting present disturbance noise is about $10^{-13} \text{ m s}^{-2}$, corresponding to an EP measurement of two parts in 10^{11} with a 95% confidence level.

By conducting a Galileo free-fall experiment with a satellite in low-Earth orbit, one immediately gains a factor of more than 1000 in sensitivity. This comes from using as driving acceleration the full gravity of the Earth, 8.4 m s^{-2} at orbit height, compared to 0.007 m s^{-2} for the solar gravity. A second factor of 1000 comes from replacing seismic noise in the torsion balance experiment with the disturbances in orbit. Disturbances from fluctuations in the Earth's gravity gradient along the orbit can be made completely negligible by using small concentric test masses. The largest remaining disturbing forces are those produced by the spacecraft, principally from its residual motions induced by air drag. The *drag-free* control system on STEP can reduce the disturbances down to the level of the accelerometer sensitivity by feeding back the acceleration signal to proportional helium thrusters which apply a force equal and opposite to the external disturbance in the bandwidth of interest. This principle of disturbance compensation has been demonstrated in a satellite called TRIAD.

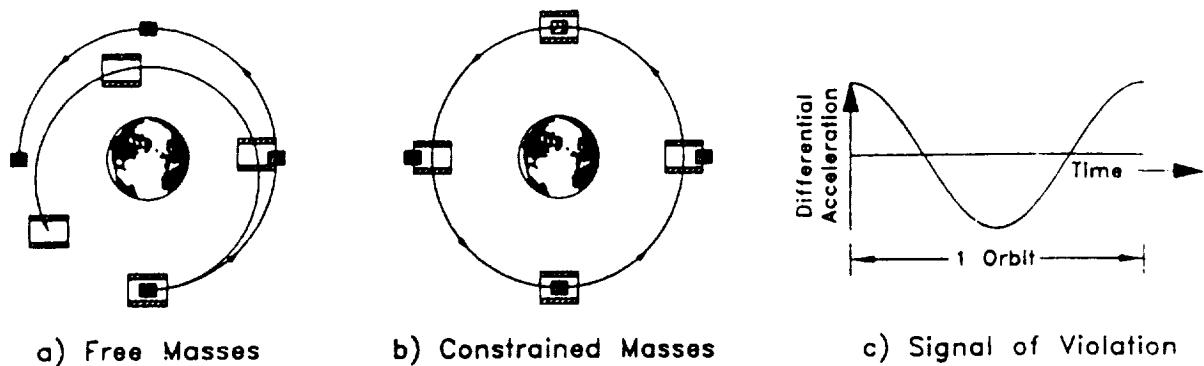


Fig. 2.1. Equivalence principle violation: (a) Shows the relative orbit of free masses where the ratio of inertial mass to gravitational mass depends on the composition of the masses. (b) Shows the configuration in STEP where the masses are constrained by linear bearings and sensing circuits. The relative displacement of the masses is a measure of the differential acceleration. (c) Equivalence Principle violation signal appears at the orbital frequency.

To take full advantage of the low disturbance level in a drag-free satellite, and indeed to provide that level, it is necessary to have sufficient acceleration measurement sensitivity. In STEP the measurement is made with a superconducting accelerometer, which is described in outline in Sections 2.2 and 2.3. With this instrument, STEP can resolve an acceleration of $2 \times 10^{-17} \text{ m s}^{-2}$. When this is divided by the full gravity signal of 8.4 m s^{-2} , the corresponding EP resolution becomes better than one part in 10^{17} , a factor of 10^6 improvement over ground-based experiments.

Each differential accelerometer used in the EP experiment contains two test masses of different materials. These test masses are held in all three directions by weak magnetic springs. Concentricity of the masses is guaranteed by iteratively using the Earth's gravity gradient to measure their offset, and adjusting the magnetic springs to correct it. In the inertially oriented configuration shown in Fig. 2.1, an EP violation appears as a differential acceleration at the orbit frequency. The full acceleration sensitivity of $2 \times 10^{-17} \text{ m s}^{-2}$ is achieved by integrating the signal for about 20 orbits. Other configurations and rotation states will be used to test whether any EP signal is real or not.

One EP experiment comprises three differential accelerometers comparing three pairs of masses of three different materials. The present design tests materials of widely differing atomic number, such as magnesium or aluminum for the low values, versus copper or zirconium for intermediate, and gold or platinum for high atomic number.

Two EP experiments are included in the satellite to help confirm or disprove the result. Since both experiments are subject to the same disturbances, this is possible only if the two designs differ in their response to the disturbances. Part of this difference is provided by orienting the two experiments at right angles to each other, so that any EP signals will differ in phase by 90° . The present concept is that the two experiments will be provided independently by European and Stanford experimenters. Chapter 3 describes in detail the approaches being considered by the Stanford team and by the experimenters contributing to the European study. Although it is desirable that some test materials will be in common in the two experiments to provide a direct confirmation of the result, a larger number of accelerometers also makes it possible to include a wider range of materials and so to distinguish more readily the source of the violation, e.g., the ratio of baryons to leptons.

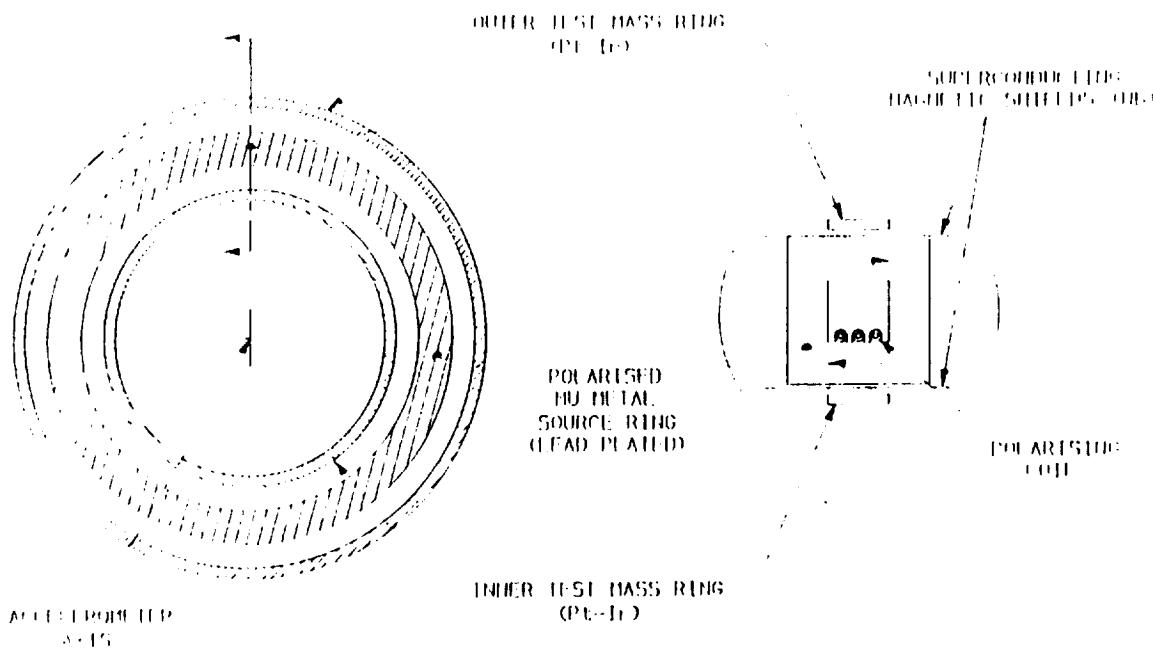


Fig. 2.2. Concept of Spin-Coupling experiment: the spin-polarised source ring is fixed to the spacecraft. The test mass rings are radially constrained but are free to move along the common axis. They are magnetically shielded from the source. Any force due to spin-coupling will penetrate the magnetic shield, causing a differential acceleration of the test masses.

2.1.2 Spin-Coupling experiment

The Spin-Coupling (SC) experiment is designed to take advantage of the extraordinary sensitivity of the differential accelerometer technology developed for the EP experiment and the very low noise environment provided by STEP. Figure 2.2 shows the principle of the experiment, which in concept consists of three concentric rings of material. The inner and outer are platinum-iridium test masses, which are free to move along their common axis of symmetry. Between these a source mass, constructed of cryogenic mu-metal, is held fixed to the spacecraft. This source mass contains a polarising coil which carries an AC current at 2×10^{-3} Hz in order to polarise the spins of the electrons in the mu-metal. A spin-coupling interaction will produce differential motion of the test masses. This motion is measured using SQUID sensors. Superconducting shields ensure that there is no magnetic interaction between the source and the test masses.

At the frequency chosen, the disturbances due to drag-free control of STEP are 10^{-3} of the seismic noise on Earth, so the experiment will be at least this factor more sensitive than any experiment that can be performed on the Earth. In practice, the improvement may be substantially greater because this field of experimentation is relatively new. For example, the best experiment to date (Ritter et al., 1993) used a torsion balance at a range of 100 mm to achieve a measurement of the coupling term, $g_s g_p$, of 2×10^{-27} . STEP makes it possible to achieve at least 7 orders of magnitude improvement to 6×10^{-31} and to extend the measurement to the much shorter range of 1 mm.

An alternative approach for ground-based experiments would be to use a much higher frequency, say 10 Hz, and utilise a seismic isolation platform that can be constructed to give high attenuation at that frequency. Such an experiment is likely to be limited by the Brownian motion noise in the suspension. The Q ("quality factor", see Eq. 2.4) of such a suspension is not likely to be any greater

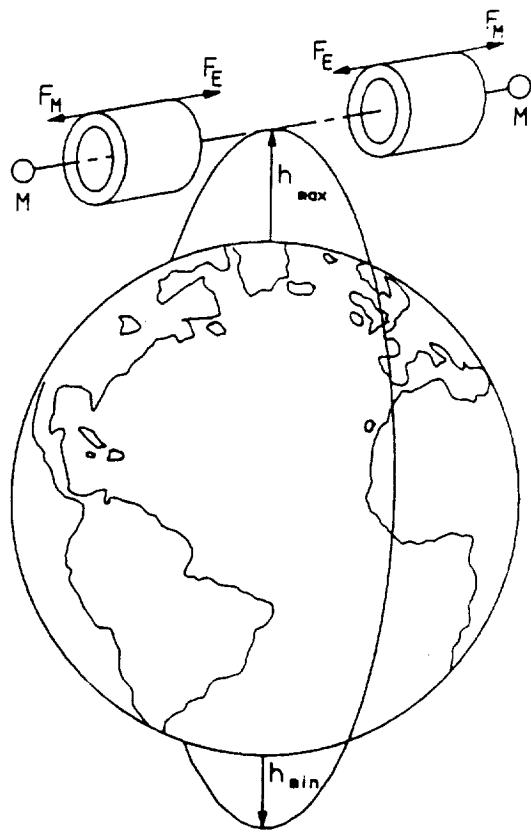


Fig. 2.3. Concept of G and Inverse Square Law: In the G experiment, source masses are used to apply gravitational forces, F_M , on the test masses. The resulting accelerations can be measured absolutely to one part in 10^6 , hence G can be determined to the same resolution. In the short-range ISL experiment, F_M is examined as a function of the distance between the test mass and the source mass. Coupling the two accelerometers in differential mode ensures attenuation of spacecraft accelerations by 10^4 , and provides a gravity gradiometer for geodesy. The spacecraft orbit is accurately known from GPS tracking, and so the local Earth gravity field, F_E , can be accurately predicted (from geopotential models). The accelerometers are calibrated against this known Earth signal.

than that which will be achieved by STEP at a much lower frequency. The formula of the noise in a differential accelerometer is given below in Eq. 2.4. For the same Q and the same integration time, STEP will gain in acceleration sensitivity by the square root of the ratio of the frequencies, i.e. by about 100.

2.1.3 Gravitational constant, Inverse Square Law and geodesy experiments

These experiments also use superconducting differential accelerometers. The difference is that, unlike the EP and SC experiments, the device is configured to maximise rather than minimise its sensitivity to gravity gradients.

The G and the Inverse Square Law experiments can be performed, in principle, with a single test mass and a single source mass. However, by adding a second test mass and making a differential measurement between the two test masses, the residual spacecraft acceleration noise can be rejected by 10^4 . So the experiment becomes 10^4 times more sensitive than a single accelerometer experiment.

The differential accelerometer is formed by two identical cylindrically symmetric test masses, separated by about 0.75 m along their common axis, as shown in Fig. 2.3. These act as a gradiometer sensitive to the gradient of the horizontal component of the Earth's gravitational field. Since STEP

is in a near-polar, Sun-synchronous orbit, the gradiometer will carry out a global gravity survey. Because of the high sensitivity of the gradiometer and the drag-free spacecraft, the survey will be of unprecedented accuracy for the low-order harmonics up to degree 150, where a limit is set by the 550 km altitude of the satellite. The role of these measurements in geodesy is addressed in Chapter 6. Clearly such a global survey of gravity can be performed only from a space platform.

To measure the gravitational constant (G) and to test the Inverse Square Law (ISL), two identical source masses are accurately moved through each test mass along the common axis at a frequency of 3×10^{-3} Hz. The movements are in antiphase in order to eliminate any motion of the combined centre of mass and thus any recoil of the spacecraft. The maximum differential signal occurs when the source masses are near the ends of the test masses and this maximum is proportional to G . The absolute accuracy of the measurement of G (Luther and Towler, 1982) depends on the metrology, for which 10^{-6} can be achieved, and on the absolute accuracy of the calibration of the gradiometer sensitivity. The best measurement of G on the ground was limited to 10^{-4} and the dominant error source was the absolute calibration of the torsion balance.

In the STEP experiment, the sensitivity calibration will be performed in an elegant way using the Earth's gravity. As the altitude of the spacecraft changes due to some eccentricity of the orbit, the gravity gradient will be modulated at the orbit frequency, producing a well-known calibration signal.

The geocentric gravitational constant GM_E , where M_E is the mass of the Earth, is known to 10^{-9} from the orbital dynamics of artificial satellites. The altitude can be determined to 0.1 m using the Global Positioning System (GPS). This can be further refined to a precision of 0.01 m by using the Earth's known gravity field and the dynamics of STEP's orbit. With these accuracies of measurement, G will be determined to 10^{-6} , a 100-fold improvement over the best ground measurement.

The ISL is made by measuring the differential acceleration as a function of source mass position. This test, which is performed at a range of 2 cm, will also be limited by metrology errors to 10^{-6} in α , the dimensionless coupling constant of an ISL violating force (see Chapter 1). The G and ISL experiments would be impossible without the benefit of the low-noise environment of STEP. However, the improvement is not the full thousand-fold noise reduction compared with terrestrial seismic noise, since metrology and calibration errors limit the improvement to a factor of 100.

Just as in the SC experiment, it is not practical to design a high-frequency experiment on a seismically isolated platform, since the Brownian motion noise will then dominate. Furthermore, in order to generate a high-frequency signal, the source mass may have to be rotated. It is then questionable whether the null geometry, utilising a source mass moving along the symmetry axis to reduce the metrology error in the separation between the source and the detector, as proposed for STEP, could be achieved. Also, it seems unlikely that the calibration error of the accelerometer could be improved much beyond 10^{-4} for ground-based experiments.

2.2 Superconducting linear accelerometer

2.2.1 Principle of operation

The sensitive accelerometers required for STEP can only be realised by using liquid helium temperatures to lower the Brownian motion noise of the test masses and by employing highly sensitive superconducting electronics to detect the small forces involved. As shown in Fig. 2.4, the accelerometer consists of a test mass moving on a linear magnetic bearing. The position of the test mass is sensed by coils coupled to a SQUID magnetometer.

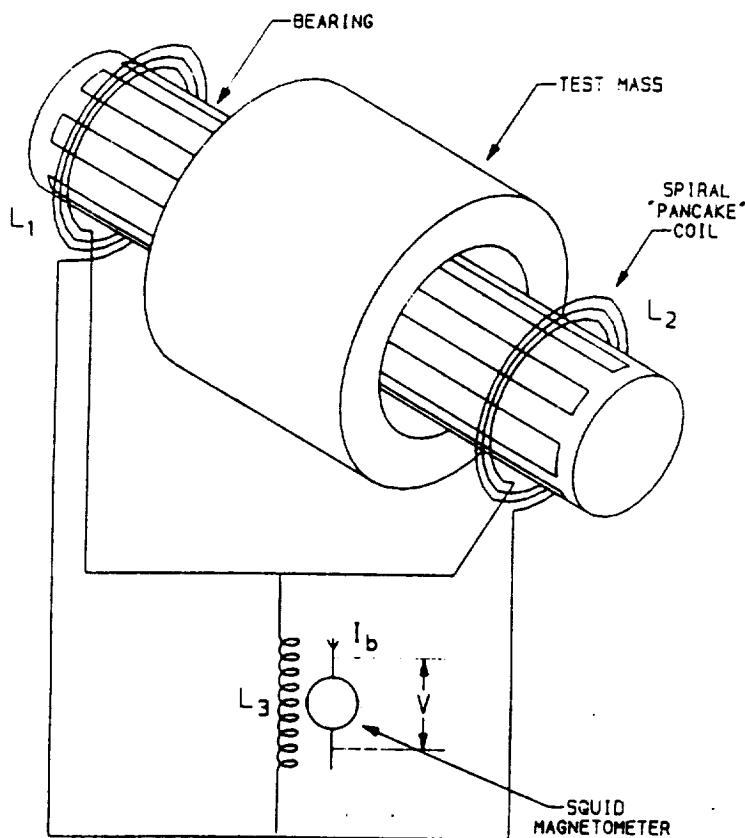


Fig. 2.4. Concept of a superconducting accelerometer: the test mass has cylindrical symmetry. Its detailed profile depends on the requirements of the experiment. In particular, it can be made more or less sensitive to different moments of the gravitational field by adjusting its longitudinal cross section. The test mass is coated by a thin film of superconducting niobium. This superconducting layer prevents changes in external fields from penetrating into the mass. The resulting repulsive force is used to constrain the test mass by a linear magnetic bearing which is aligned with its symmetry axis. The linear position is measured by detecting the change in magnetic flux in spirally-wound pancake-shaped detection coils, arranged opposite the end faces of the test mass. The change of magnetic flux is detected by a Superconducting QUantum Interference Device (SQUID) magnetometer.

2.2.2 Magnetic bearing

The linear bearing, shown in Fig. 2.4, is made from a quartz rod which carries a thin film of niobium forming a superconducting coil. The coil is wound in a meander pattern which cancels its magnetic field to high order. The magnetic field falls off exponentially in the radial direction, thus minimising the cross-coupling to the SQUID position sensing coils.

The bearings constrain the test masses to one-dimensional motion along the axis, and also provide the ability to perform an important gravity gradient centring procedure. In addition, it is possible to use the bearings to test for certain disturbing forces from trapped flux and magnetisation of the test masses.

The magnetic bearings are briefly described as an array of superconducting wires stretched parallel to the sensitive direction of the accelerometer. These are arranged in pairs, with antiparallel currents in each pair, to cancel any net magnetic moment (Worden, 1976). This minimises interaction with the SQUID. These wires lie on the surface of a cylinder and give a radial spring constant to a surrounding test mass. The number of wires in each bearing is chosen to give an optimised

spring constant. This occurs when the spacing between wires is about 1.6 times the spacing between the wires and the mass. Each bearing is divided into quadrants. The wires in each quadrant are connected with those in the diametrically opposite quadrant, in a circuit similar to the position detector sense-coil circuit, but with a transformer replacing the SQUID. Each circuit provides adjustment of the centre of force along a line perpendicular to the cylinder axis. This is needed to be able to move the centres of mass of the test masses into coincidence.

All loops in the circuits will be fully superconducting except when the currents are being charged. This guarantees that the forces from the bearings will be as stable as the supercurrents themselves, so there will be no disturbance to the experiment. Actually winding the bearings with wires, which was done for the preliminary experiments on the ground, is not appropriate for the satellite experiment. Instead we propose to use the more mechanically robust and reproducible thin film technology. Since little stiffness is required in orbit, thin films can carry enough current to manage the masses in the orbital experiment. They could not on Earth.

The magnetic bearings must provide radial restraint against the residual spacecraft motion of about $7 \times 10^{-11} \text{ m s}^{-2}$. They will give the 200 g test masses a radial period, nominally 200 seconds, which will be adjusted to ensure that the masses do not resonantly couple to each other. Longer periods run the risk of coupling the uncertainty in radial position into an uncertainty in force in the axial direction (due to background forces). Much shorter periods risk disturbing the axial motion because trapped flux and imperfections in manufacture cause axial forces. Based on laboratory prototypes we expect one part in 10^6 of the restraint force to appear in the axial direction due to manufacturing errors, but this increases dramatically if there is much trapped flux. The longer periods decrease the risk of large trapped flux levels.

While providing a tight radial constraint, the bearing ideally provides no force along the longitudinal direction and thus allows an extremely low frequency suspension in the measurement direction. The mass also has radial rocking modes, but these can be damped so that they do not interfere with the axial mode.

2.2.3 Position sensing circuit

The two sensing coils L_1 and L_2 form a superconducting loop which carries a persistent current I_0 . As the test mass m moves towards L_1 , it compresses the magnetic flux in the space between L_1 and m and expands the flux between L_2 and m . The effect of this is to modulate the inductances L_1 and L_2 , which causes a balancing current to flow in a third coil L_3 . L_3 is coupled to a SQUID magnetometer to measure this current.

A (DC) SQUID (Fig. 2.5) is a single superconducting loop containing two very narrow insulating gaps called Josephson junctions. Electrons from a bias current I_b can tunnel through these junctions by the Josephson effect and the electron-pair wave functions in the two arms interfere with each other to produce a voltage V which regularly repeats as the magnetic flux linking the SQUID loop changes. The magnetic flux in a perfectly superconducting circuit is quantised in units of $\hbar/2e = 2.07 \times 10^{-15}$ Weber, where \hbar is Planck's constant and e is the unit of electric charge. By counting the oscillations in V , it is possible to count the magnetic flux quanta that link the SQUID loop.

Although the SQUID measures the *displacement* of the test mass *relative* to the coils, it is important to realise that, when the test mass and suspension are included, the entire device forms an *accelerometer*. The persistent current I_0 provides a stiffness along the sensitive axis, forming a spring-mass system with resonant frequency ω_0 which converts an acceleration $a(\omega)$ into a displacement $x(\omega)$ as follows

$$x(\omega) = \frac{a(\omega)}{|\omega^2 - \omega_0^2|} \quad (2.1)$$

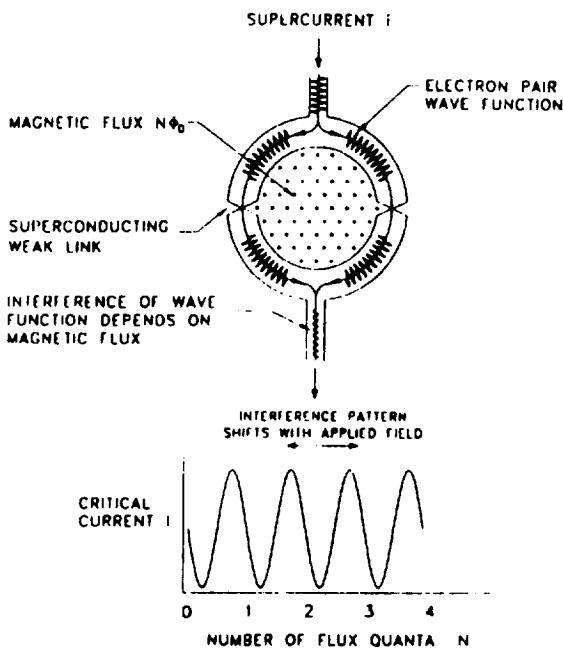


Fig. 2.5. SQUID operating principle.

The displacement results in a voltage across the SQUID given by

$$V(\omega) = \left(\frac{\partial V}{\partial i} \right)_{\text{SQUID}} i(\omega) \quad (2.2)$$

where

$$i(\omega) = \frac{2}{1 + \gamma} I_0 \frac{x(\omega)}{d} \quad (2.3)$$

The displacement sensitivity of the SQUID is proportional I_0 , as is seen from Eq. 2.3. However, a larger I_0 increases the resonant frequency and thus decreases the acceleration sensitivity according to Eq. 2.1, if ω_0 is increased beyond the signal frequency ω .

2.3 Superconducting differential accelerometer

2.3.1 Differential sensing circuit

Two accelerometers of the type described in Section 2.2 can be combined in the circuit shown in Fig. 2.6 to form a differential accelerometer. The two test masses can be concentric as shown in Figs. 2.1 and 2.3 or separated over a baseline as in Fig. 2.3.

In the symmetric case with $m_1 = m_2 = m$, $I_1 = I_2 = I_0$, and the equilibrium values of L_{11} , L_{12} , L_{21} , and L_{22} equal to $2L_0$, Eqs. 2.1 through 2.3 hold if $x(\omega)$, $a(\omega)$ and ω_0 are now interpreted as the differential displacement, differential acceleration, and differential mode frequency, respectively (Chan and Paik, 1987).

The intrinsic noise of the differential accelerometer is given by an acceleration power spectral density

$$S_a(f) = \frac{8}{m} \left[k_B T \frac{\omega_0}{Q} + \frac{\omega_0^2}{2\eta} E_A(f) \right] \quad (2.4)$$

where m is the mass of each test mass, k_B is Boltzman's constant, T is the temperature, ω_0 is the angular resonant frequency, η is the electrical coupling efficiency of the circuit, Q is defined as

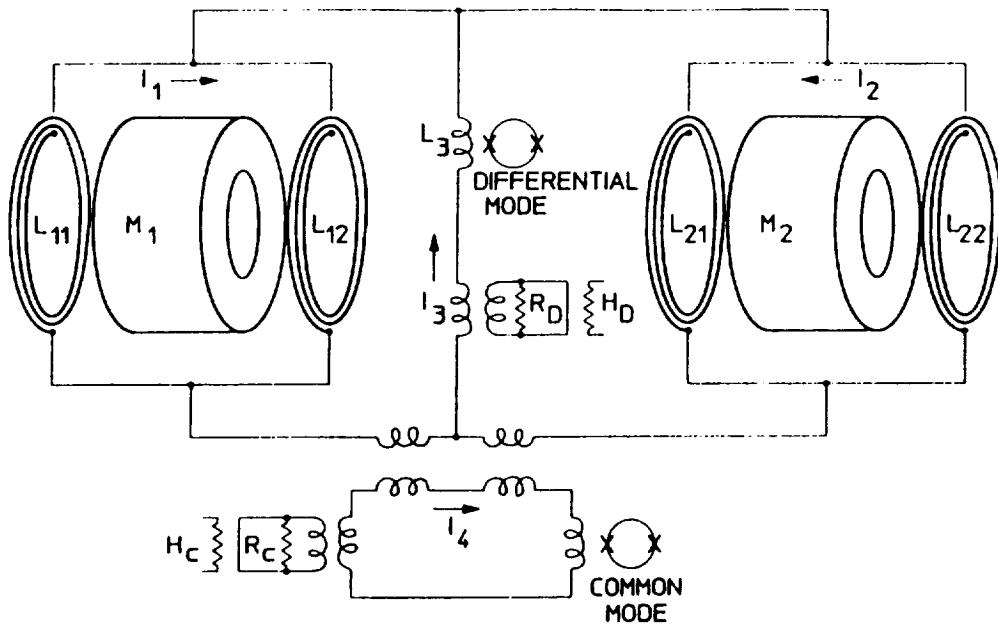


Fig. 2.6. Sensing circuit of a superconducting differential accelerometer. The signal currents from the two test mass positions are summed to give a common-mode signal in one SQUID and are differenced to give a differential signal in a second SQUID. A common-mode balance can be achieved by adjusting the ratio of the persistent currents I_1 and I_2 while shaking the whole device. In orbit, this shaking is effected by the drag-free thrusters during the initial setting up of the accelerometers. The limit on the common-mode rejection will then be set by the alignment of the accelerometer axis. A rejection of 10^{-4} can be achieved with realistic alignment accuracies. The centres of mass of the two masses can be matched by adjusting the persistent current I_3 . The optional damping resistors, R_D and R_C , are discussed in Section 3.2.5.

$m\omega_0/c$ where c is the damping coefficient (hence, Q is inversely proportional to the level of damping in the system, and is thus called the "quality factor"). $E_A(f)$ is the SQUID noise expressed as an input energy resolution at frequency f . Eq. 2.4 applies for $f \leq \omega_0/2\pi$.

The SQUID noise spectrum for the best currently available commercial device (Quantum Design DC SQUID) is shown in Fig. 2.7. In the frequencies relevant for the STEP experiments, $f \leq 0.1$ Hz,

$$E_A(f) = (10^{-31}/f) \quad \text{J/Hz} \quad (2.5)$$

For the EP design parameters: $m = 0.2$ kg, $T = 2$ K, $\omega_0/2\pi = 3 \times 10^{-3}$ Hz, $Q = 10^6$, $\eta = 0.5$, and at $f = 1.7 \times 10^{-4}$ Hz, the nominal EP signal frequency, Eq. 2.4 gives $S_a^{1/2}(f) = 6 \times 10^{-15} \text{ m s}^{-2}/\sqrt{\text{Hz}}$. When averaged over the measurement period of 10^5 s, this gives an rms acceleration noise of $2 \times 10^{-17} \text{ m s}^{-2}$.

The intrinsic noise of the common-mode output is given by an equation similar to Eq. 2.5. The common-mode resonant frequency is chosen to be 10^{-3} Hz for the EP accelerometers. In principle, the common-mode readout can be as sensitive as the differential mode. However, because of the dynamic range requirement over a wide bandwidth, we intentionally keep the sensitivity low by choosing a weak coupling to the common-mode SQUID, $\eta \approx 10^{-6}$.

2.3.2 Levitation and damping

As in the case of a single accelerometer, the test masses of the differential accelerometer are constrained to move along the common symmetry axis by magnetic bearings. Each test mass has

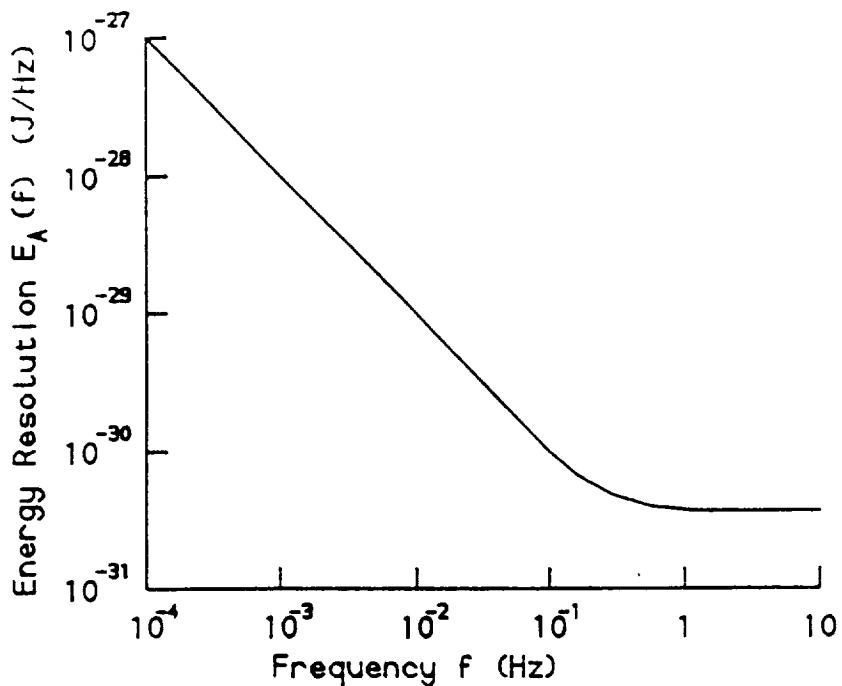


Fig. 2.7. SQUID energy resolution as a function of frequency.

six resonant modes. In addition to the axial linear mode, which is used for sensing, there are five unwanted modes: two radial linear, two radial angular, and one axial angular. The low resonant frequencies of these modes coupled with low loss in the superconducting circuits will result in very high values for Q . Thus oscillations produced during uncaging, setup, or other undesirable events can persist for years. The modes must therefore be damped, but without adding a significant amount of noise into the sensing circuit.

In principle, there are two options for damping: passive or active. Passive damping increases the Brownian motion noise of the mode and therefore must be switched off after the test mass is brought to rest, if it is applied to the axial linear mode. The advantage of active or "cold" damping is that it does not add noise as long as the noise temperature of the amplifier is low compared to the ambient temperature. The disadvantage is that it requires additional instrumentation; each mode must be detected and a phase-shifted signal must be fed back to control the mode. Each option is used in different STEP experiments.

Figure 2.8 shows the levitation circuit which forms a linear bearing for one test mass.

2.3.3 Electrostatic control

As mentioned above, active damping can be applied to the test mass modes by using a capacitive pickoff and electrostatic positioner. Further, electrostatic control is needed to measure the electric charge on the test mass. These topics are discussed in greater detail in Chapter 3.

2.4 STEP accelerometer configuration

The nine differential accelerometers for the various STEP experiments are configured to form an orthogonal set from which signals for drag-free and attitude control of the spacecraft can be derived. Figure 2.9 shows the STEP accelerometers housed in a quartz block. The orbit plane coincides with the xy -plane. The three Stanford EP accelerometers are spaced evenly with the sensitive axis along the x -axis. The three European EP accelerometers and the SC accelerometer are spaced evenly with the sensitive axes along the y -axis. The concentric test masses of the two G/ISL gradiometers,

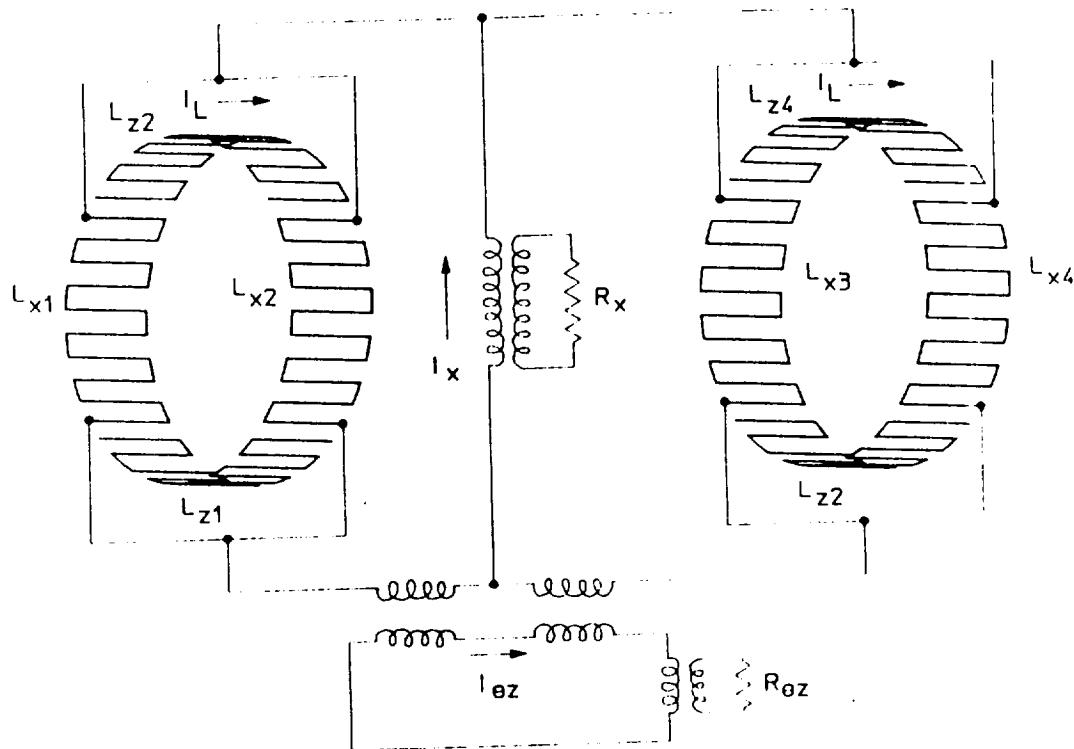


Fig. 2.8. Levitation circuit of superconducting differential accelerometer. This circuit provides levitation of one test mass along the x -direction and applies passive damping to the x -linear and z -angular degrees of freedom. In order to be able to adjust the centre of mass along the two radial linear degrees of freedom, the levitation coils are divided into four quadrants. In order to control the test mass orientation in the two radial angular degrees of freedom, the coils are subdivided into front and back sections. The persistent currents I_x and I_z can be adjusted to displace the mass along the x -axis and rotate it about the z -axis, respectively. An identical circuit using the flow coils L_{z1} to L_{z4} provides levitation in the z -direction.

separated by a baseline of about 0.75 m, occupy the two ends of the quartz block and point along the z -axis.

The orthogonal orientation in the orbit plane of the Stanford and European EP accelerometers permits two-phase detection of the EP violation signal. The orbit normal, the z -axis, is the optimum orientation for the G/ISL accelerometers since this axis experiences least disturbance from the Earth's gravity gradient. The z -axis is also the symmetry axis of the spacecraft and remains fixed in space even when the spacecraft is in the "turning mode". This makes the centrifugal acceleration a second-order error. The horizontal alignment of the z -axis also makes the pointing error a second-order effect in gravity gradient. As a result, a very sensitive geodesy experiment can be carried out with the G gradiometers without increasing the attitude control requirement beyond the level imposed by the EP experiment.

The common-mode outputs of the three Stanford EP accelerometers (a_{x1} , a_{x2} , a_{x3}) can be summed and differenced to derive signals for drag-free control in the x -axis and attitude control about the y -axis. Likewise, the common-mode outputs of the three European EP accelerometers (a_{y1} , a_{y2} , a_{y3}) can be combined for drag-free control in the y -axis and attitude control about the x axis. The common-mode outputs of the two G/ISL accelerometers (a_{z1} , a_{z2}) can be combined for drag-free control along the z -axis. Notice that there is one redundant acceleration output in each of the three orthogonal directions. These redundancies can be used to identify effects of local gravity.

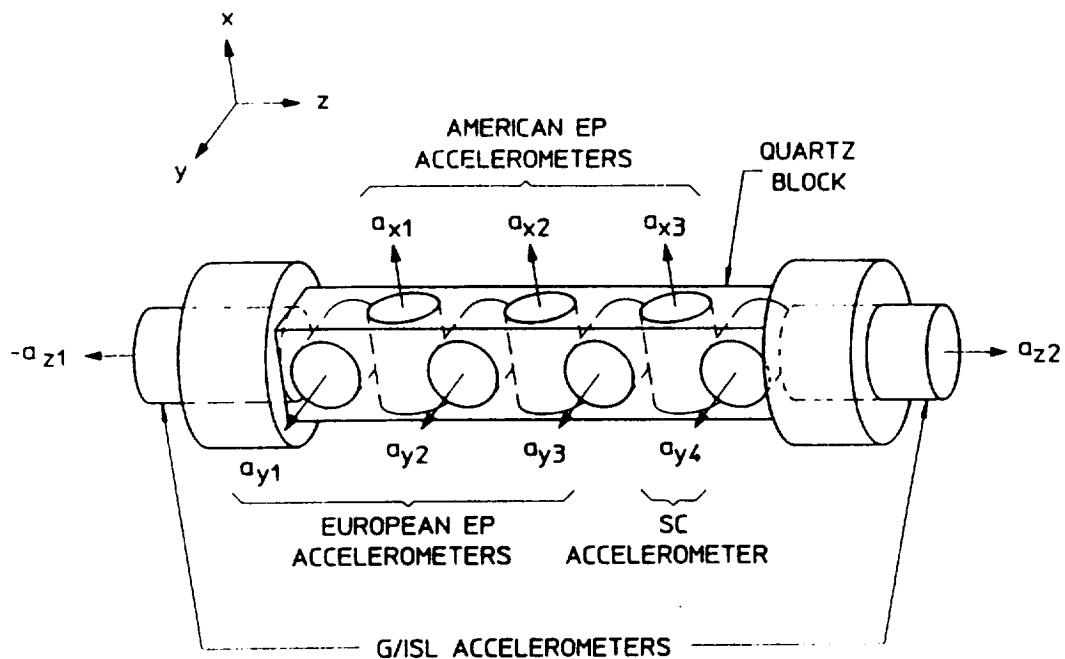


Fig. 2.9. Overall STEP accelerometer configuration in the quartz block. The common-mode outputs of the three Stanford EP accelerometers are used for drag-free control along the x -axis and attitude control about the y -axis. The common-mode outputs of the three European EP accelerometers are used for drag-free control along the y -axis and attitude control about the x -axis. The common-mode outputs of the two G/ISL gradiometers are used for drag-free control along the z -axis.

disturbance such as the helium tide.

2.5 Disturbing forces

The stochastic measurement noise given by Eq. 2.4 is not the limit to the sensitivity because it can always be reduced by averaging for a long time. Other random disturbances may increase the averaging time required, while systematic disturbances may imitate or mask an EP violation. The STEP philosophy is to reduce all of these disturbances to the minimum, well below the proposed EP sensitivity, and use cross-checks to prove that the requirements have been met. The planned cross-checks, which are discussed further in Chapter 3, all together can determine whether any putative EP violation is or is not due to any of the known disturbances, and can identify the electrical, magnetic, or mechanical nature of any unknown disturbance. The principal disturbing forces are summarised in Table 2.1. They are treated in greater detail, where appropriate, in subsequent chapters.

2.5.1 Platform accelerations and drag-free control

Fluctuations in air-drag and solar radiation produce linear and angular motion of the spacecraft. These accelerations couple to the differential accelerometers through small but finite departures of the sensitive axes from parallelism and concentricity. In the STEP spacecraft, a drag-free control is applied to reduce the spacecraft acceleration level. Figure 2.10 shows the residual linear acceleration spectrum of the spacecraft when under drag-free control. In order to reduce the angular acceleration of the spacecraft in the roll and yaw, a second drag-free control loop can be closed at another point

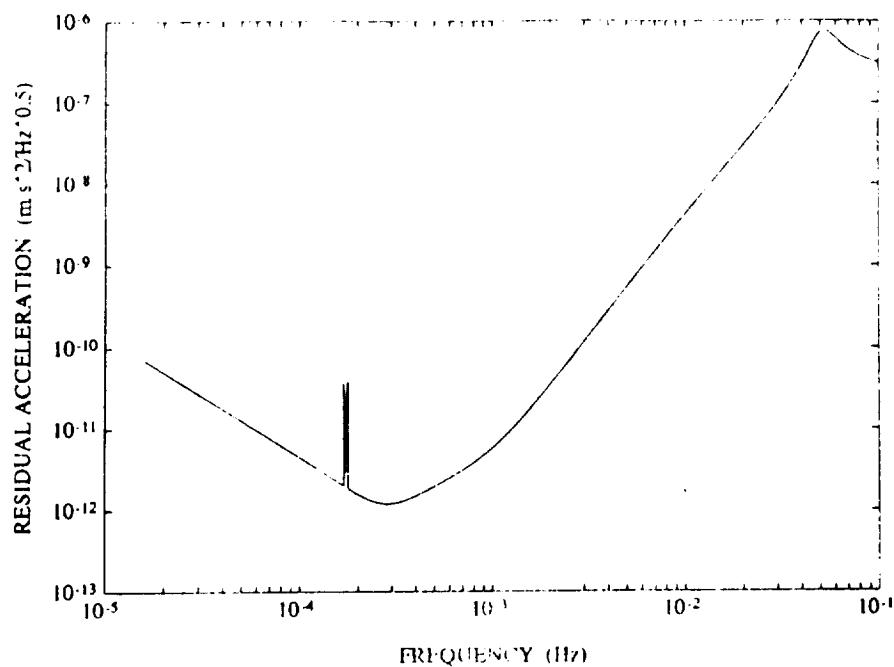


Fig. 2.10. Residual linear acceleration spectrum of the spacecraft under drag-free control. At low frequencies, the performance is limited by the weakly coupled ($\eta \approx 10^{-6}$) common-mode SQUID sensor noise. The sharp peaks at the orbit frequency ($\approx 1.7 \times 10^{-4}$ Hz) correspond to the steady drag component modulated at the orbit rate with respect to the inertially pointing accelerometers. The broadband disturbances arise from low amplitude atmospheric density and wind variations, and from thruster noise. The bandwidth of the control system is kept below 0.1 Hz to avoid excitation of the spacecraft structural modes. Attitude control is effected by applying drag-free control to a second point in the spacecraft. An approximate residual angular acceleration spectrum can be obtained by dividing the vertical scale by the baseline between the two control points, which can be anywhere between 0.1 and 0.5 m.

of the spacecraft. Star trackers provide a long-term attitude reference of the spacecraft.

From Fig. 2.10, it can be seen that the residual rms linear acceleration noise is $10^{-13} \text{ m s}^{-2}$ in the EP measurement bandwidth of 10^{-5} Hz. Since the differential accelerometers are aligned to have a common-mode rejection of 10^{-1} , the resulting differential noise is $10^{-17} \text{ m s}^{-2}$, an order of magnitude below the required level.

Table 2.1. Summary of disturbances.

Force Category	Source of Disturbance	Coupling Mechanism	Required Control Level for EP	Control Method
Platform Accelerations	Fluctuations in air drag and solar radiation	Misalignments of sensitive axes	$5 \times 10^{-11} \text{ m s}^{-2}/\sqrt{\text{Hz}}$ at EP signal frequency	Drag-free and attitude control
Gravity Gradients	Helium tide caused by the Earth's gradient	Non-vanishing multipole moments of test masses	$10^{-17} \text{ m s}^{-2}$ at EP signal frequency	(1) belted test mass (2) divided helium tank (3) electrostatic confinement of helium
	Common-mode acceleration caused by the Earth's gradient	Misalignments of sensitive axes	$10^{-17} \text{ m s}^{-2}$ at EP signal frequency	Drag-free control switched between accelerometers
Electric Charge	Patch effect	Modulation by test mass motion	10^{-9} N/m force gradient	Large gap ($d = 1 \text{ mm}$)
	South Atlantic Anomaly protons	Asymmetry in sensing coils	10^{-13} C	Discharging by UV or wire
Magnetic Fields	Trapped flux	Modulation by test mass motion	10^{-3} Gauss	Mu-metal shield
	Earth's magnetic field	Penetration into the cryostat	10^{-13} Gauss at EP signal frequency	Mu-metal and superconducting shields
	Magnetic field from the source (SC and G/ISL)	Direct pickup by sensing circuit	10^{-13} Gauss at SC and G/ISL frequency	High quality superconducting shield
Temperature	Common-mode temperature fluctuation	Penetration depth modulation and asymmetry in coils	1 mK at EP signal frequency	Sun-synchronous orbit
	Differential-mode temperature fluctuation	Radiometer effect	1 mK/quartz block at EP frequency	Low pressure ($\leq 10^{-11} \text{ torr}$)

3. Equivalence Principle experiment (EP)

3.1 The Equivalence Principle measurement concept

The objective of the STEP Equivalence Principle (EP) experiment is to compare the rates of fall of several test objects to an accuracy of one part in 10^{17} . The test objects fall continuously in Earth orbit and are protected from outside disturbances by the surrounding spacecraft. In STEP, the test masses are concentric hollow cylindrically symmetric bodies which are free to move relative to each other along the symmetry axis (Worden and Everitt, 1974). A pair of such masses constitutes a differential accelerometer. A violation of Equivalence would produce a harmonic differential motion between the outer and inner mass of different composition, as the pair orbits the Earth (Fig. 2.1). The EP experiment utilises six differential accelerometers.

In STEP's measurement concept, ideally the only disturbing force is from the device performing the measurement. When measuring a tiny force, additional forces applied during the measurement must be small if the measurement is to be significant. Measuring Equivalence to 10^{-17} requires being able to resolve an acceleration of $10^{-17} \text{ m s}^{-2}$ in about 20 orbit's integration time, with negligible disturbance to the test mass.

In the EP experiment, the forces on the test masses are compared by a differential accelerometer system attached to the spacecraft. We described its operation in Section 2.2. This system, based on SQUID magnetometers, will be very sensitive to differential motions of the test masses, but at least 10^4 times less sensitive to their common-mode motion.

3.1.1 Accelerometer configuration

The EP experiment consists of twelve test masses in six differential accelerometers: three Stanford and three European (Fig. 2.9).

Only the pair of masses in each differential accelerometer can be directly compared, but if we have three accelerometers, making measurements between materials A, B, and C, the sum of the acceleration differences

$$(a_A - a_B) + (a_B - a_C) + (a_C - a_A) \quad (3.1)$$

must add to zero if the measurements are valid. A non-zero value indicates that a systematic force is disturbing the measurement. This *cyclic condition* is an important check for error and is the reason for having a multiple of three EP accelerometers.

As described in detail later, all EP accelerometers have "belted" outer cylinders. This design was suggested by the European team as a way to reduce sensitivity to gravity gradients, and was subsequently adopted also by the Stanford team.

3.1.2 Experiment validation concept

Any credible Equivalence Principle signal must be clearly distinguishable from other effects. The most important characteristic of an Equivalence Principle violation in STEP is that it comes at a completely determined frequency and phase, which are entirely under the control of the experimenters. It can therefore be separated from disturbances by changing its frequency and phase. The *signal frequency* is the difference between the frequency of the orbit and the frequency of the turning mode, and it ranges from DC to four times the orbit frequency. The amplitude of any Equivalence Principle violation must be independent of the signal frequency, and its phase must be such that the signal is a maximum when the sensitive axis of the accelerometer points to the centre of the Earth. Those disturbances which always come at the signal frequency will change amplitude or phase as the

frequency changes, because they depend on dynamic processes which have characteristic response times, such as thermal distortion or helium tide. The well-defined Equivalence Principle signal can be detected with an optimal filter, so that the major source of uncertainty will be from unknown systematic disturbances.

To validate the Equivalence Principle measurement, a number of checks for systematic effects and other disturbances need to be performed. If the experiment is to be at all credible, all known disturbances must be understood and shown to be under precise control. It is important to understand that these tests and checks are intended only to confirm that the requirements for reaching 10^{-17} have been met. In most cases they cannot be used to "correct" the Equivalence Principle data, although some modelling may be acceptable if an effect is understood and well-behaved. It is unrealistic to expect that the experiment will not be limited by some residual disturbance at the 10^{-17} level, even though we may have designed the system for no disturbance at all; the tests give a way to show that a signal is definitely not any of the known effects. The planned tests cover the known forces, provide a lot of data not directly related to the Equivalence Principle measurement, and therefore have a good chance of identifying even unexpected disturbances.

The tests include, for example, changing the orientation and rotation of the spacecraft, selecting another proof mass or control law, putting known electric charge on the masses, and adding known thermal gradients or centre of mass displacements. Environmental factors which are under our control will be exaggerated to see if they produce measurable accelerations when they are increased. The measured response of the accelerometers to the exaggerated disturbance will be used to estimate how much the disturbance could have affected the measurement. In addition to deliberate tests, environmental factors such as spacecraft temperature distribution, radiation environment, and helium level will be measured and correlated with the accelerometer outputs.

Other tests can better define the source of a disturbance; for example, the effect of magnetic disturbances can be separated from all other effects by making measurements with different currents in the superconducting bearings, and electric patch effects can be identified by changing the charge on the test mass. Changes in spacecraft gravity fields can be identified by making gradiometer measurements between the separate accelerometers, which makes it possible to localise the disturbance in the spacecraft.

Each set of tests will take about as long as the original EP measurement; most of the mission time will be spent verifying the experiment. Fortunately, most of the disturbances are easily distinguishable from the signal.

3.1.3 Equivalence Principle measurement programme

The STEP spacecraft must be able to execute some simple orientation changes to confirm or deny any apparent violation of Equivalence. These manoeuvres allow a separation to be made between the few remaining disturbances and a violation. To better define these orientations, we have defined two sets of coordinates: orbit coordinates and spacecraft coordinates (Fig. 3.1). Orbit coordinates are fixed in inertial space (ignoring the slow 1° per day drift of the orbit plane), with the X_0 -axis taken in the direction of motion as the spacecraft passes over an arbitrary reference point such as the North Pole. The Y_0 -axis is then radially toward the centre of the Earth, and the Z_0 -axis is positive toward the Sun. Spacecraft coordinates are fixed in the spacecraft and are identical with orbit coordinates when the spacecraft is in normal orientation. In this case, the accelerometer sensitive axis X is parallel to the direction of motion, and the line through the out-of-plane masses (Z-axis) is parallel to the orbit Z_0 -axis. At least two other operational modes are necessary to do the experiment properly.

Baseline measurements are made in the "normal" orientation. This orientation is used as a standard reference mode to which the others may be compared. The signature of a violation in the

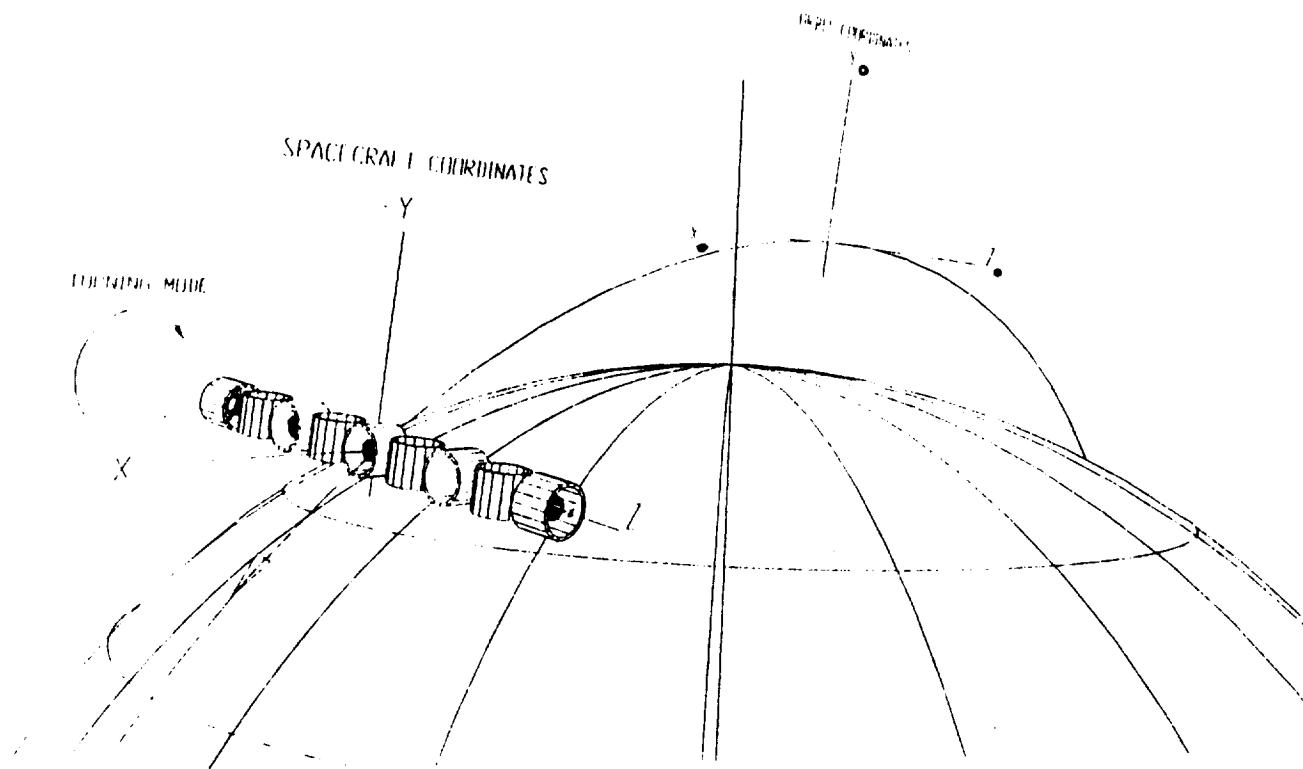


Fig. 3.1. Spacecraft and orbit coordinates.

normal mode is a differential acceleration at orbital frequency ω_o with an amplitude significantly greater than that at frequencies just above and below ω_o , and "zero" phase such that the acceleration is a maximum when the sensitive axis points to the Earth.

To confirm that the signal originates in the spacecraft and not in the environment, the spacecraft can be re-oriented to a "rotated" position by turning around the Z-axis by an arbitrary angle θ . Environmental disturbances will keep constant phase with respect to the orbit, but a violation, or any disturbances onboard, will now have time dependence $\sin(\omega_o t + \theta)$. Several values of θ should be used.

A signal passing these tests may be caused by any of several onboard effects including helium tide, thermal distortions and mass shifting, as well as Equivalence Principle violations. To distinguish these, the spacecraft is now rotated slowly about the Z_o -axis at a speed ω_1 comparable to the orbital frequency. The Equivalence Principle signal will now come at a frequency $\omega_0 - \omega_1$ and will have unchanged amplitude and zero phase. In this *turning mode* the disturbance from thermal distortion of the spacecraft (for example) will change both amplitude and phase. This distortion is driven from outside by temperature gradients which tend to recur at the signal frequency, and results in a change in the internal gravity fields of the spacecraft. The spacecraft has an unknown, finite response time due to its thermal inertia and conductivity, so it will respond differently if the driving frequency is changed. If the thermal response time is very short, the spacecraft changes shape uniformly and cannot cause big changes in gravity gradients; if it is very long, the spacecraft cannot distort significantly. In between, the distortion can be distinguished from a violation, although it cannot be

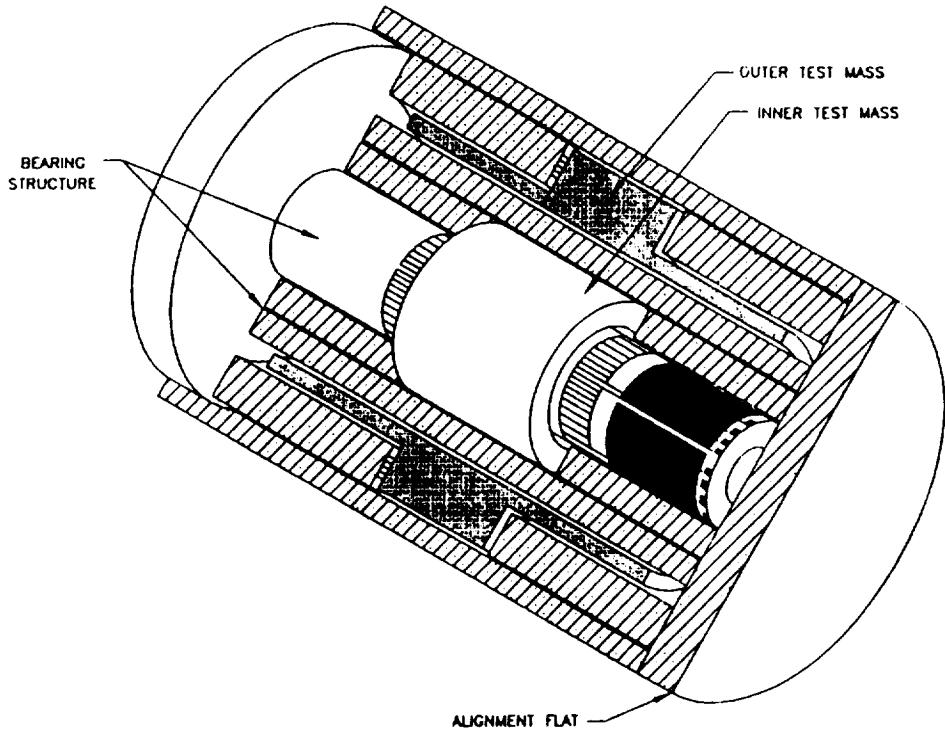


Fig. 3.2. EP differential accelerometer with "belted" outer test mass.

accurately modelled. Several rates $\omega_1, \dots, \omega_n$ should be used, to map out the frequency dependence of the amplitude and phase of the signal. Tides in the liquid helium will show similar behavior, but should have a richer frequency spectrum, including possibly resonances of wave motion with the driving gravity gradient.

The other checks are internal. These include offsetting the mass centres to measure internal gravity gradients; mapping background electric and magnetic forces; introducing deliberate bias accelerations, temperature changes and gradients, and electric charge; and changing control gains and modes of operation.

3.2 Accelerometer design

Differential accelerometers form the basis of all STEP instruments. Each differential accelerometer has a sensitivity of $3 \times 10^{-15} \text{ g}/\sqrt{\text{Hz}}$, or better, in its differential mode, and common-mode sensitivity better than $10^{-12} \text{ g}/\sqrt{\text{Hz}}$. Each accelerometer contains two cylindrically symmetric test masses constrained to move along their common axis by superconducting magnetic bearings (Fig. 2.4). The cylinder or x-axis is the sensitive direction of each accelerometer. The motion of the masses in the x-direction is measured by a SQUID position detector.

An engineering drawing and an exploded view of the EP differential accelerometer are shown in Figs. 3.2 and 3.3. The accelerometers each comprise two test masses, SQUID position sensors to measure them, magnetic bearings for constraint along the cylinder axis, and a set of electrodes for a capacitance pickoff, electrostatic positioner, and charge control system. Associated parts which are not actively involved in the measurement include a caging mechanism, superconducting shields, setup and adjustment circuits for the SQUIDs and bearings, and an alignment flat which is common to all accelerometers in each experiment.

The test masses are isolated from each other and from the environment by superconducting shields. An electrostatic positioning system, based on a set of electrodes surrounding the masses, is used for manipulating them. This system is also used to measure and control the charge on

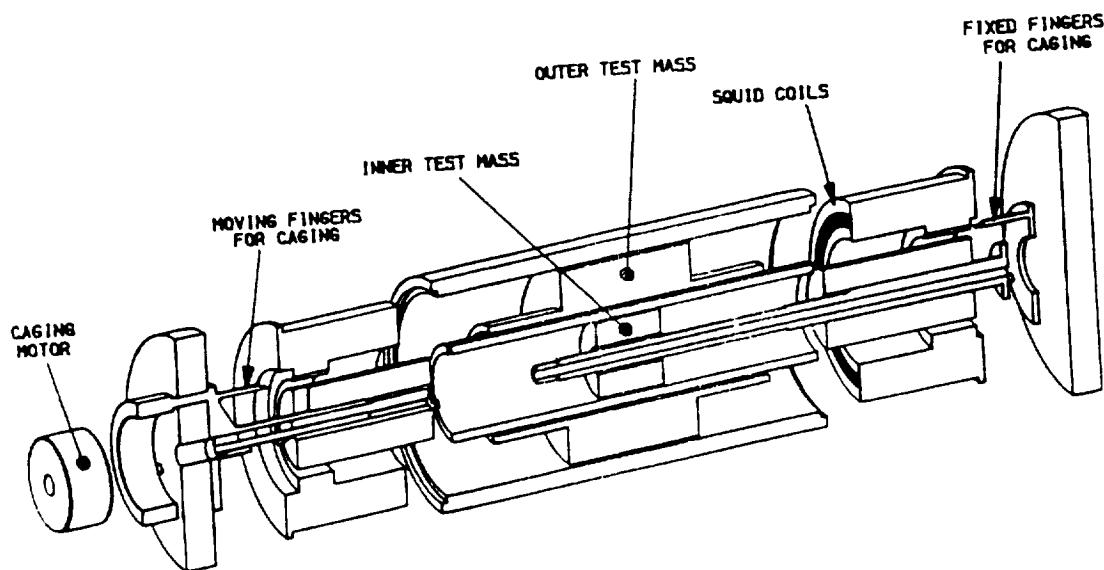


Fig. 3.3. Exploded view of EP differential accelerometer.

the suspended masses. Alignment and calibration of the accelerometers are essential and critical procedures, which must be done in orbit to get the necessary performance (although testing is possible on Earth).

The accelerometers are made from fused quartz, which has a very small thermal expansion from 0 to 300 K. Other materials (i.e. silicon) have smaller thermal expansion and better thermal conductivity at 2 K, but larger overall expansion. Each accelerometer comprises several nested cylinders which support SQUID coils, electrodes, magnetic bearings and shields. Each set of accelerometers is optically contacted to a quartz reference plate, which is simply an optical flat that accurately aligns and supports them (visible in Fig. 3.2). A quartz block manufactured to relatively low precision surrounds and supports each reference plate assembly (Fig. 3.4). By building the instrument from simple shapes (flats, cylinders) we increase the accuracy of the instrument while avoiding difficult and expensive manufacturing procedures. The mounting for the reference plate and optical contacting methods are derived from existing GP-B technology.

The entire quartz block assembly is contained in a vacuum vessel with integral superconducting shield. This vessel is heat sunk to the helium bath. A dewar probe assembly provides the interface to the controlling electronics which are at room temperature.

3.2.1 Test mass materials

From a theorist's point of view, properties such as atomic number, fermion number, and binding energy are the most important criteria for selecting test mass materials since different theories indicate that these fundamental parameters might be important. There are, however, practical properties such as density, electrical conductivity, thermal expansion, homogeneity, and machinability, which tend to outnumber the theoretical motivations for selection.

For completeness, the greatest possible variety of materials and properties should be tested. We are limited to a total of twelve masses by the available cryogenic volume of the spacecraft and complexity of the apparatus. However, good experimental design requires redundancy of measurement,

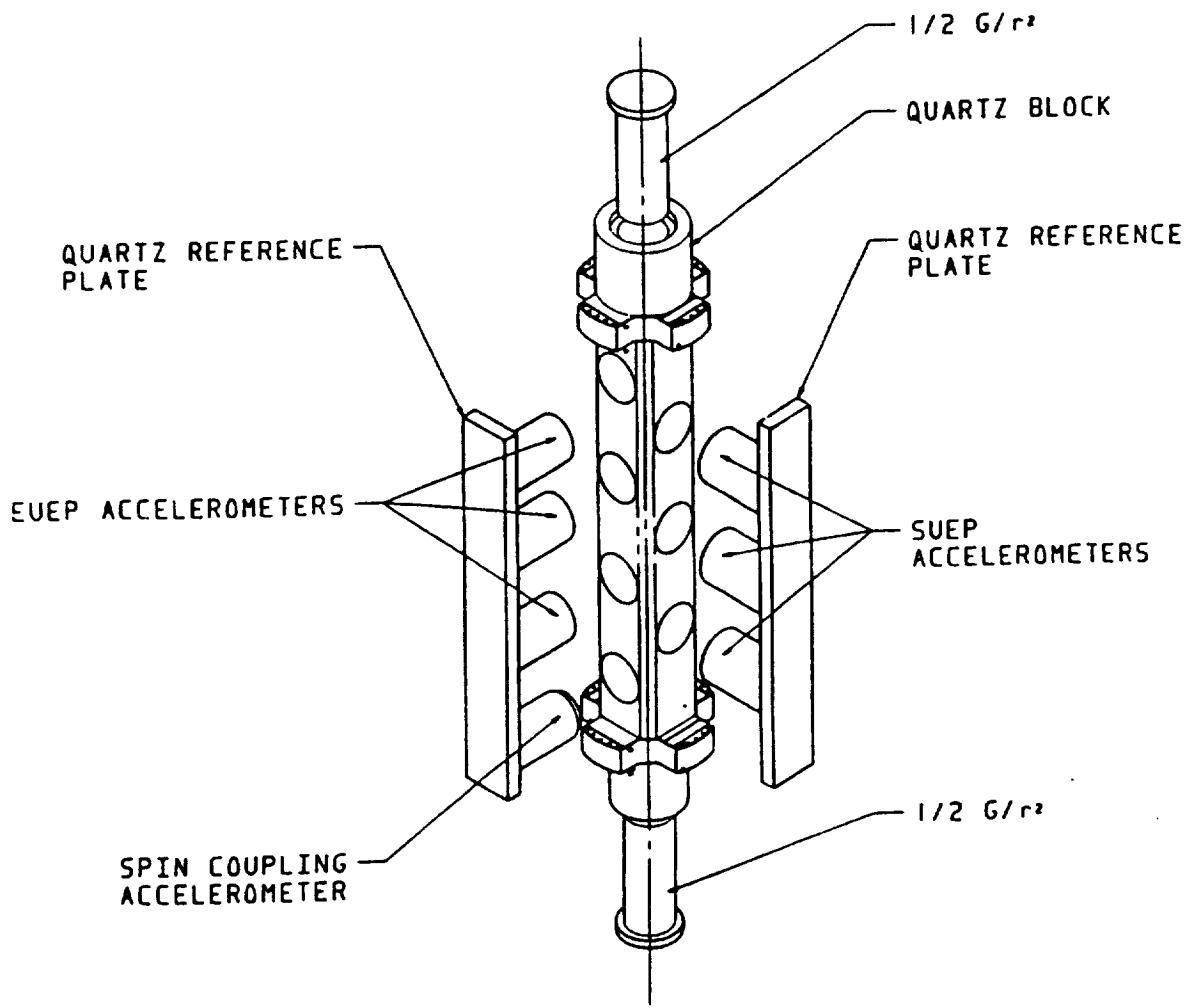


Fig. 3.4. Quartz block supporting each reference plate assembly.

which further limits the number of materials.

The test mass materials must be selected before proceeding with detailed design because the magnetic bearings and position detectors need to be optimised for density and size variation. The test masses should be as heavy as possible to reduce their response to non-gravitational disturbances. On the other hand, the masses should be as small as possible to reduce the coupling of higher mass moments to the spacecraft and especially to possible motion of the helium refrigerant. Materials with high density are therefore preferred.

Our materials were selected from various parts of the periodic table. For the Stanford instrument, the chosen materials are copper (Cu), gold (Au), and magnesium (Mg). Copper is a good choice for a binding energy test. It is close to the top of the binding energy curve, and is preferable over iron or cobalt since it is much less magnetic. Elements at the upper end of the periodic table are favourable for contrasting atomic number and neutron/proton ratio differences with the low end of the table. Gold, lead, iridium and platinum, as well as rare earths, were candidates. Gold was chosen for its mechanical properties. Finally we selected magnesium as a representative of the low

end of the periodic table. In practice these metals will need to be alloyed with a small fraction of other materials to improve their handling characteristics, coated with a thin film of superconducting material (lead or niobium), and overcoated with an inert material with uniform surface properties (gold), in order to meet other requirements. These additional materials will result in a small dilution of any Equivalence Principle violation. For the European instrument, similar considerations led to the choice of zirconium alloy (Zr), platinum-iridium (Pt), and magnesium (Mg). The common use of a material (Mg) in the Stanford and the European instruments provides an important cross-check.

3.2.2 Superconducting sensing and levitation

The SQUID position sensor for the masses was discussed in Section 2.3. The common-mode rejection of the accelerometers will be better than 1% by design, that is, the scale factors for the masses will be matched and the test masses weigh the same to that accuracy. During the calibration of the apparatus the acceleration scale factors will be matched to better than one part in 10^4 by adjusting the supercurrents in the SQUID circuits. This may result in a slightly larger mismatch in the resonant periods of the two masses, which does not significantly affect the measurement. A superconducting transformer coupled differentially allows adjustment of the equilibrium position of the test masses without destroying the superconductivity of the measurement circuit and without significant disturbance to the scale factor matching.

The position measurement must be very linear to allow the necessary sensitivity matching and prevent frequency conversion, and must be insensitive to anything but differential motion along the sensitive axis. The last requirement can be met by careful attention to coil design. Two types of coil are suitable, which we call the pancake and solenoid designs. The solenoid design is a short section of solenoid surrounding each end of the test mass. The pancake design is a flat spiral coil near each end of the test mass (Fig. 2.4). Pancake coils are preferable to solenoids or Helmholtz coils owing to the fact that they are easier to model, and hence their performance is easier to predict. The coils and ancillary circuits are formed on quartz cylinders using thin-film technology.

In keeping with the general philosophy that any forces applied to the test masses be kept small, we have made the axial force from the differential measurement circuit as small as we think is prudent, giving the masses a nominal period of 1000 seconds. With a much longer period there is a risk of instability in the turning mode. There is also a limit below which increased periods gain no more acceleration sensitivity, which is set by uncontrolled background forces (Fig. 2.10).

The axial position of the test masses is controlled by the magnetic force from the SQUID sense coils. If this were the only constraint, the test masses would be unstable radially because of the field configuration. A radial constraint system is therefore required. This function is provided by superconducting magnetic bearings (see Section 2.2.2).

3.2.3 Electrostatics

The concept of the electrostatic sensing and positioning system was introduced in Section 2.3.3. Electrostatic measurement of the test mass position is required when the SQUIDs are inoperative. In addition to this function, during the Equivalence Principle measurement the electrodes will be used for measurement and control of modes that the SQUIDs cannot sense, for control of the test masses, and for charge measurement. Their charge measurement function will be discussed in Section 3.4.5.

The electrostatic positioner also provides a backup radial positioner (with poorer performance) for the magnetic bearings. The system will also be used during the checks for systematic effects and other disturbances. Appropriate reversals of the potential on the electrodes and deliberately charging the masses can give unambiguous measurements of the forces from electric patch effect, in a way analogous to that used with the magnetic bearings to test for trapped flux forces.

Fig. 3.5 shows how the electrodes are arranged for both test masses. This arrangement has

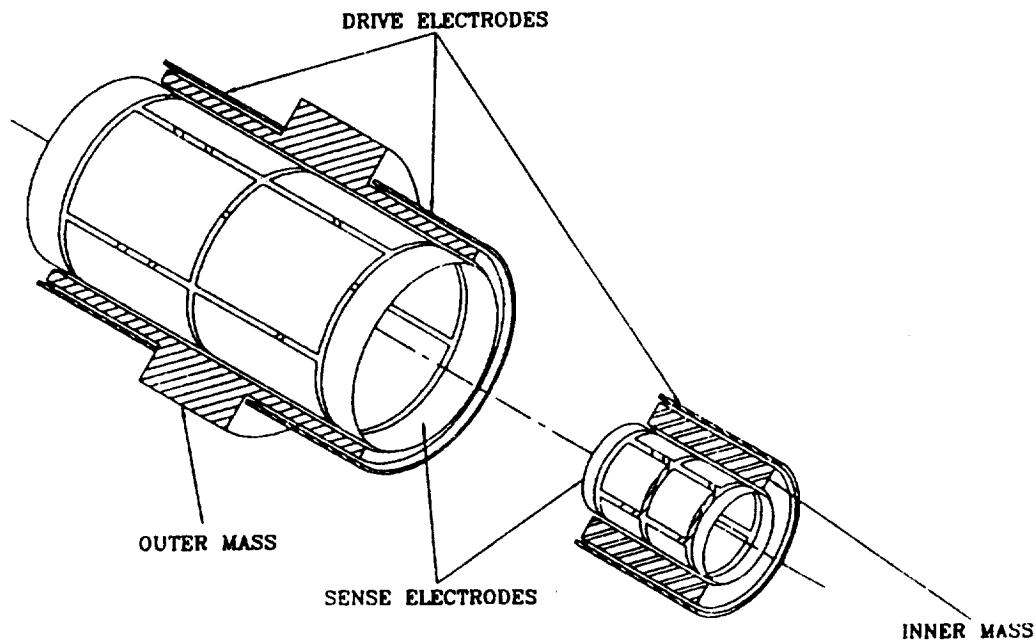


Fig. 3.5. Electrodes for capacitive sensing and actuation of each test mass.

some useful features. The axial force it applies to the mass is independent of position in first order because the end effects are small. This means that spacecraft residual motion will not be coupled to the test mass. Likewise, the axial position measurement with the capacitance sensor is almost perfectly linear. With a 200 gram test mass the total capacitance might be about 150 pF, which would change by about 1 pF/cm with mass motion. A one volt potential difference would produce an acceleration of 8×10^{-10} m s⁻². We therefore expect that the voltages needed to position the masses will be small (a few tens of volts).

The capacitance position measurement is much less sensitive than the SQUID, but it is almost perfectly linear along the axis of the experiment, and can be accurately calibrated on the ground or from the known dimensions of the electrodes and mass. This provides an easy way to precisely calibrate the SQUID accelerometers, which otherwise could only be done indirectly or by a more complicated procedure.

Used as a controller, the system will position the masses before and during SQUID setup. The mass positions must be accurately fixed during the setup procedure if the common-mode balancing procedure is to be successful, because the position detector inductances depend on mass position. Control voltages applied to the electrodes can exert forces on the test masses due to images or charge on the mass. This permits the masses to be manipulated.

The capacitance bridge also gives a way to measure other degrees of freedom of the test mass which the SQUIDs cannot measure. This is especially important if there are background forces in addition to those needed for the measurement process, because the gradient of these forces directly causes an unknown disturbance if the mass position is uncertain.

During Equivalence Principle measurements, the electrostatic positioning system adds damping to axial, radial and angular motions. This limits the time and amplitude of displacement in these modes, reducing the uncertainty in force due to the masses being offset from their nominal positions. It is important that the electrostatic positioner add only damping (velocity-proportional force)

because this cannot result in static or low-frequency offsets. Measurements of the radial and angular motions can only be made by the capacitance sensor/electrostatic positioning system. These measurements are an important component of the Equivalence Principle data, because they confirm that the masses have remained within the required distance of their equilibrium position and that the equilibrium in these modes has not drifted. If there are background forces such as electrostatic patch effect which can change dramatically with position, any uncertainty in mass position becomes magnified into an uncertainty in the force on the mass. The measurements of the radial and angular modes limit this uncertainty.

Test mass rotation

The remaining degree of freedom which is hard to measure or control is rotation of the test mass around the axial direction. This becomes worse as the mass is more accurately made. Because of the large motion, the uncertainty from background forces is also large, and we might expect the axial equilibrium position to shift as the mass rotates. If the background forces are small enough, this causes no problem other than signals at the rotation frequency and its harmonics, which of course are easily distinguished from a violation.

A more prudent course than letting the mass rotate is to stop it. The rotation can be sensed by the SQUIDS if it is causing trouble. A set of superconducting stripes parallel to the magnetic bearings (and a few microns high) can provide enough "cogging" to couple to the rotational mode. This limits the possibilities for rotation, and allows some control. During the initial setup the bearings can be pulsed to reduce the already small rotational energy.

3.2.4 Caging

The test masses in each experiment must be supported against launch loads on the order of 50 g. Mechanisms are proposed using stepping motors (with some soft ferromagnetic alloy but no permanent magnets) which drive three fingers against each test mass (visible in Fig. 3.3.) Although details vary, the same design of cage is, in principle, applicable to all nine pairs of test masses, with nine identical motors. For the G experiment, a tenth step motor rotates a ballscrew to translate a nut which liberates (or re-cages) the source mass.

3.2.5 Differences between European and Stanford designs

Although the Stanford and European EP accelerometers are very similar, there are some functional differences. These are summarised as follows:

- the radial stiffness of the superconducting magnetic suspension for the European design will be about 10^3 times greater than the Stanford design (leading to frequencies of $\approx 3 \times 10^{-2}$ Hz). This reduces undesirable cross-couplings between modes, and allows relaxing the radial frequency matching requirements. A simplification of the electrostatic system, to one degree-of-freedom, can also be achieved.
- the European accelerometers will utilise "permanent" passive damping for the radial and rotational modes. The resistors R_x and $R_{\theta z}$ in Fig. 2.8 provide the damping. One can show that R_x and $R_{\theta z}$ can be chosen to critically damp the modes while keeping $Q \geq 10^6$ for the axial linear mode, as long as $f_R \geq 30 f_0$. In order to avoid coupling to the axial mode through an edge effect, the levitation coils should be located well within the edges of the test mass. In the axial direction, electrostatic feedback and "switchable" passive damping options will be available for both the common and the differential modes. Figure 2.6 shows the damping resistors R_D and R_C which can be coupled to the sensing circuit by turning on heat switches H_D and H_C .

- discharging of the European test masses will be performed using either a permanent gold wire, contacting whiskers, ultraviolet light, or field emitters. The choice has not yet been made. The Stanford test masses will be discharged using ultraviolet light.

3.3 Calibration and verification

Calibration of the Equivalence Principle accelerometers comes from two sources. The absolute or common-mode calibration is from ground calibration of the setup currents for the SQUID and calibration of the capacitance pickoff. This only needs to be accurate to a percent or so, although better than 10^{-4} is possible. The important calibration is the differential mode balance. This can only be done in orbit by a combination of accelerations and rotations of the spacecraft, adjusting the accelerometers until the differential acceleration vanishes.

Pre-flight verification of the operation of the accelerometers is possible to the extent that their performance is limited by gravity and seismic noise. If they could be tested to their full sensitivity on the ground, there would be no need for going into space. The operation of the superconducting circuits in each accelerometer can be verified at any time after the apparatus is cold, in much the same way as any ordinary electronic circuit. The performance of the accelerometers can be tested (to the seismic and gravitational limit) in several ways before final assembly. These include levitation of very light test masses and using a fiber to support most of the weight of a real mass. After assembly and cooldown, we can use the caging mechanism to manipulate the masses and verify that they are free and can be sensed by the SQUIDs. Most of the procedures for verification of operation are not too different, except in degree, from those used prior to cooldown of any large cryogenic system.

With the present design of the accelerometers, an Equivalence Principle violation at the 10^{-17} level can be measured in about 10^5 seconds, a little over a day. In fact we plan to take a week to 10 days on each measurement. Because this measurement time is determined by random instrument noise, small improvements in noise figure may result in large improvements in measurement time.

3.4 Disturbance management

The STEP test masses can be directly disturbed by electrical, magnetic, mechanical, or gravitational forces (see Table 2.1). The spacecraft greatly modifies the effect of the first three, by shielding the masses from external drag and electromagnetic fields, and by replacing these external forces with a controlled environment. Two disturbances are known to penetrate this barrier: gravitation and high-energy particle radiation.

The spacecraft produces its own set of disturbances, although these internal disturbances are much smaller than the unattenuated external disturbances. The spacecraft is coupled to the test masses mechanically (by residual gas), electromagnetically (by the measurement system, and by residual background fields including thermal radiation), and gravitationally. If the couplings change, they will directly disturb the masses. If something outside disturbs the spacecraft, it may indirectly affect the test masses through one or more of these couplings. If any of this is at the signal frequency, the net effect may tend to mimic an Equivalence Principle violation. The STEP design philosophy is to make these couplings as small as possible. This reduces the possibility of a false signal. Even the reaction force on the masses from the measurement process must be minimised.

We expect the cryogenic environment to be extremely stable, so that any changes in the couplings will be minimised and much less than the room temperature value. Thermal expansion, the major cause of "drift" at room temperature, decreases roughly as the square of the temperature, and will have a negligible effect at 2 K. The same or similar result is true of many other temperature dependent disturbances, but not, for example, the so-called "radiometer effect" described below.

An example of an external disturbance which can indirectly disturb the masses is thermal radiation from the Earth or Sun. If the spacecraft is heated from outside, it will change shape because of

thermal expansion differences, especially if the source is changing direction. This redistributes the mass in the spacecraft, changing its internal gravity gradients. This couples directly to the masses. Similar indirect disturbances might come from drag variations (residual spacecraft motion coupling to the masses electromagnetically), Earth's gravity gradient (raising tides in the superfluid helium refrigerant), and the thermal heating from particle radiation (mechanically coupling through gas pressure and the temperature gradient). When these effects are large, they put requirements on the spacecraft or payload. For example, thermal heating of the spacecraft puts a requirement on its thermomechanical stability, and the helium tide effect determines the allowable motion of the helium surface.

The remainder of this section describes briefly the strategies for controlling the most important (or most visible) remaining disturbances.

3.4.1 Gravity gradients and helium tide

If the weights in Galileo's experiment were dropped simultaneously from different heights, they would be expected to fall at different rates. This is because the Earth's gravity weakens with height, so the masses would not be in exactly the same gravity field. The same thing happens in orbit, but it is relatively more important because we have a much more sensitive experiment. For this reason we make each pair of masses concentric with one another (one mass inside the other). Still, even with the most accurate manufacture, this would be an impossible requirement if we did not measure the centre of mass offset and correct it. Correcting this offset reduces not only the disturbance from the Earth's gradient, but also eliminates most of the gravitational disturbance from the spacecraft.

Additional gravitational disturbance comes from the interaction of the higher mass moments with higher derivatives of the gravitational field (the effects of "non-sphericity" of the test masses). After the centre of mass offset, the quadrupole moment term is most important for disturbances from the Earth's and the spacecraft's field. This leads to a requirement, discussed below, that each mass have its principal moments of inertia equal. Higher moments interact mostly with the spacecraft and result in the spacecraft's mass motion limits.

It is necessary to eliminate these gravity gradient forces, not only from Earth but also those originating in the spacecraft, because they are large and tend to be at harmonics of the Equivalence Principle signal frequency. For ordinary centre of mass offsets that might result from machining or alignment tolerances, these forces are large on the scale of $10^{-14} \text{ m s}^{-2}$. Additionally, if the orbit is not quite circular, there is some modulation of the Earth's gravity gradient. This comes from the nonlinear height dependence of gravity and mimics the signal from an Equivalence Principle violation, although only at orbit frequency. This can be completely distinguished from an Equivalence Principle violation by being phase locked to the orbital height. It is still essential to eliminate the centre of mass offset because of the spacecraft gradients. To reduce these effects to an insignificant level, the masses need to be centred on average to within 10^{-8} cm , and there are restrictions on allowable mass distribution changes in the spacecraft as well.

We eliminate the effect of the Earth's gravity gradient by measuring the effect of the disturbing gravity gradient itself, which is an acceleration proportional to the centre of mass offset. This acceleration has twice the frequency of any Equivalence Principle signal and so can be measured independently any violation. After measuring this acceleration, we can calculate the offset and reposition the masses. When the gravity gradient acceleration becomes too small to measure, it will also be too small to disturb the experiment.

The very important centring procedure for the test masses uses the sense coils and magnetic bearing to passively centre the test masses. In this procedure, a fast controller uses added damping (via the electrostatic suspension) to bring the test masses to rest with respect to one another and to the spacecraft. (The drag-free system, referenced to the common-mode of the masses, participates

in this controller). With the distance between its two masses fixed, each accelerometer acts as a gradiometer. The differential acceleration measured is proportional to the product of the gravity gradient and the displacement of the masses. Since we know the orbit of the spacecraft, we know how big the Earth's gravity gradient is, and since it recurs with twice the Equivalence Principle signal frequency, it is easily distinguished from other disturbances (for example spacecraft gravity gradients). A slow controller uses the orbital position and the amplitude and phase of this component of the acceleration to calculate the centre of mass displacement. The centring is accomplished by adjustments to the current in each quadrant of the bearing, and to a differential input to the position detector coils. The axial position of the test masses is controlled by the equilibrium position between the SQUID sense coils, and the radial position is controlled by the magnetic bearings. The controller adjusts the supercurrents until the centres of mass are superimposed on each other. This may take several orbits if the noise level is high. The trapped supercurrents should not need further adjustment after the initial setup.

This centring procedure effectively eliminates the most important disturbance from gravity gradients.

Test mass shape and gravity gradient coupling

The axial acceleration, a_z , experienced by an extended cylindrical body (rather than a point mass), due to a unit perturbing source mass at position (R, θ) , can be expanded in terms of Legendre polynomials $P_{n+1}(\cos \theta)$ as:

$$a_z = G \cdot \sum_n (n+1) \cdot k_n \cdot P_{n+1}(\cos \theta) / R^{n+2} \quad (3.2)$$

where $n = 0, 1, 2, 3, \dots$

where G is the gravitational constant ($6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2$), the z -axis is the axis of cylindrical symmetry of the body, and spherical polar coordinates are assumed. The factors k_n are simply geometrical factors determined by the shape of the body: they are the n^{th} moments of the (homogeneous) density integrated over the entire volume of the body, divided by its mass, and have dimensions of (length) n .

The first term in the expansion has k_0 identically equal to unity, and $P_{n+1}(\cos \theta) / R^{n+2} = \cos \theta / R^2$: this is the "monopole" coupling term which decreases with increasing distance as $1/R^2$. It makes the major contribution to the acceleration a_z experienced by the body and acts as if all the mass were concentrated at its centre of mass. The terms which follow in the expansion give a measure of the body's departure in gravitational behaviour from that of a point-mass, uniform sphere, or spherical shell. For a spherical body only the monopole term survives, and $k_n = 0$ ($n > 0$).

Ideally one would use spherical test masses, whereby the effects of gravity gradients would be drastically reduced, and the disturbance would couple to the measurement only as a consequence of density inhomogeneity in the test masses. However, the practicalities of building nested spherical test masses with suspension and sensing systems for both the inner and the outer mass, render the proposal unfeasible.

However, if the non-spherical body has mirror-symmetry about a mid-plane perpendicular to the z -axis then the odd-order terms $n=1$ (dipole), $n=3$ (octopole), $n=5$, etc., vanish since they have geometrical factors k_n mentioned above that integrate to zero over the volume of the body. This leaves, in addition to the monopole term, just the even coupling terms $n=2$ (quadrupole), $n=4$ (hexadecapole), $n=6$ (64-pole), $n=8$, etc. (from potential field theory these terms are known respectively as "q2", "q4", "q6", etc.) Clearly q2, q4, q6, etc. need to be minimised so that the body is as much like a point-mass as possible. If each of a pair of proof-masses possess mirror-symmetry, and their centres of mass coincide, then the differential acceleration they experience due

Table 3.1. Dimensions of European EP accelerometers.

	material	mass	R_1	R_2	R_3	L_1	L_2
Accelerometer #1							
	outer mass	Mg	200g	16.76	19.06	31.47	47.99
	inner mass	Pt	200g	4.5	12.25	—	11.30
Accelerometer #2							
	outer mass	Mg	100g	18.85	21.15	28.49	40.52
	inner mass	Zr	100g	4.5	14.35	—	13.03
Accelerometer #3							
	outer mass	Zr	200g	16.75	19.05	26.29	32.8
	inner mass	Pt	200g	4.5	12.25	—	11.30

to a source-mass at (R, θ) is due just to differences in their coupling terms q_2 , q_4 , q_6 , etc., since their monopole-induced accelerations are identical, and cancel.

A good shape for the proof-masses is therefore one which minimises q_2 , q_4 , q_6 , etc., and for straight-cylindrical bodies it is always possible to make $q_2=0$, whilst simultaneously having q_6 , q_8 , q_{10} , etc. very close to zero. However, q_4 , though small, cannot be nulled with this geometry, and is always much larger for the outer than the inner mass of the test mass pair. The condition $q_2=0$ requires accurate dimensioning of the test masses (to better than $10 \mu\text{m}$), but in practice its fulfillment can be checked to high accuracy by testing for equality of the moments of inertia of each test mass about its three principal axes.

A particularly good configuration for the the test masses has a belted outer cylinder, and a straight inner cylinder geometry, where both test masses have $q_2=0$, their q_4 's are equalised, and q_6 for the outer mass is minimised. The dominant gravitational coupling to the proof-masses is therefore via q_6 , and this decreases very rapidly with increasing distance as $1/R^8$. This is the configuration adopted in the European experiment, and in terms of differential acceleration it is about 100 times less susceptible to gravity gradients, such as helium tidal influences, than is a conventional straight cylinder pair. Table 3.1 gives the dimensions of the 3 European EP accelerometers.

Recently, the Stanford team has adopted a similar design for their baseline. The corresponding dimensions are listed in Table 3.2. The Stanford design makes $q_4=0$ for the outer mass. The remaining q_4 of the inner mass results in an acceptably small sensitivity to gravitational disturbances. In Tables 3.1 and 3.2, for the outer belted test masses R_1 is the inner radius of the main cylinder; R_2 is the outer radius of the main cylinder; R_3 is the outer radius of the belt; L_1 is the half-length of the main cylinder; and L_2 is the half-length of the belt. (For the straight inner test masses, R_3 and L_2 have no meaning.)

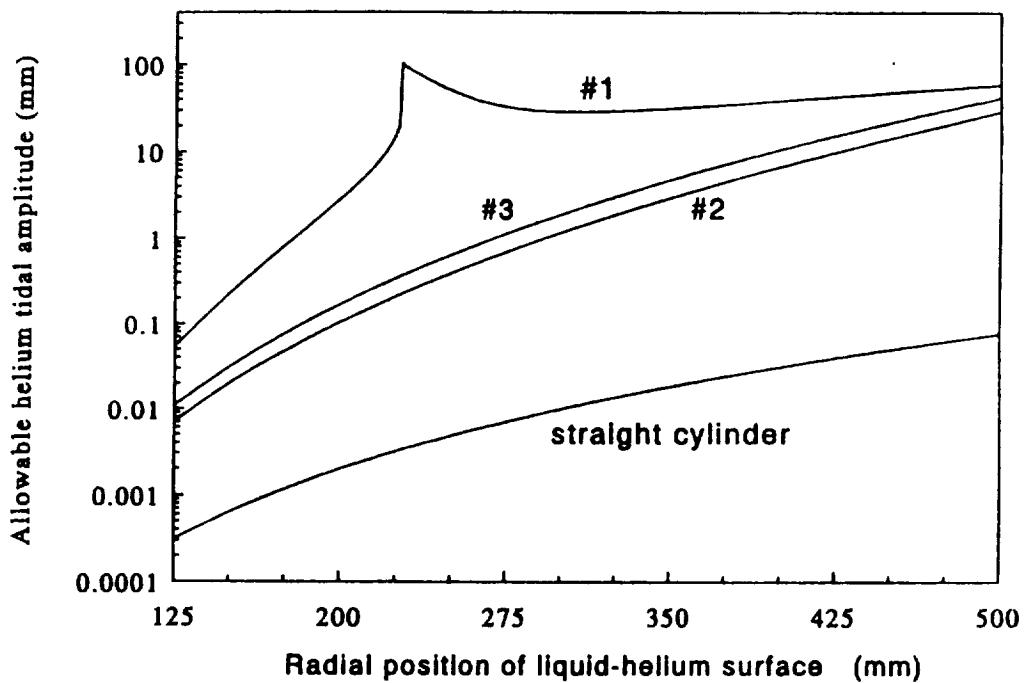
Helium tides

These refinements to the shape result in test masses which behave as much like a point mass or uniform sphere as is practical. The remaining extra gravitational coupling is not negligible, however, especially for disturbing masses closer than 25 cm. This puts a limit on the allowable tidal motion of the superfluid helium.

The expected sensitivity of the three European accelerometers to helium tidal effects is shown

Table 3.2. Dimensions of Stanford EP accelerometers.

	material	mass	R_1	R_2	R_3	L_1	L_2
Accelerometer #1	Mg	211.53 g	26.00	29.07	39.00	53.49	14.77
	Cu	211.53 g	13.00	18.88	—	19.85	—
Accelerometer #2	Mg	211.53 g	26.00	29.07	39.00	53.49	14.77
	Au	211.53 g	13.00	16.27	—	18.04	—
Accelerometer #3	Cu	956.66 g	28.49	30.55	39.00	58.65	16.49
	Au	956.66 g	13.00	22.67	—	22.63	—

**Fig. 3.6.** Allowable helium tidal motion for $10^{-18} g$ sensitivity in the Equivalence Principle accelerometers. (Labels #1, #2, #3 refer to the accelerometer test mass designs from Table 3.1.) Also shown is the result for a straight cylinder design.

in Fig. 3.6 (for the Stanford designs, the curves are similar). Here, a very much “worst case” tidal scenario has been assumed, in that the tidal bulge is imagined to be unilateral (whereas it will actually appear on both sides of the body of the helium), and in the form of a distorted spherical shell equal in thickness to the tidal amplitude. Moreover, it is assumed that this shell translates from one side of the dewar to the other in the period of the signal, thereby doubling its gravitational influence on the test masses. The position of the inner surface of this shell is given in the figure in

millimetres from the centre of the helium dewar as the "radial position of liquid-helium surface"; and that tidal thickness of helium shell which will generate a differential acceleration of 10^{-18} g in the pairs of test masses has been calculated at each radial distance for the 3 pairs of European test masses. The performance of a straight-cylinder design is shown for comparison. Note that a disturbance of 10^{-18} g is an order of magnitude smaller than the desired EP experiment sensitivity of 10^{-17} g.

It is seen that for the most susceptible design, a helium tide amplitude of 6.5 mm at a radial distance of 40 mm will exceed the 10^{-18} g requirement, whilst at 50 mm a 28 mm limit on the tide is required. This should be compared with the expected disturbance which can be estimated from a balance between surface tension and gravitational forces. On the scale of the dewar, surface tension is a large force. The "capillary length" $\sqrt{2\sigma/\rho g}$ measures the scale below which surface forces dominate gravity; using the largest gravity gradient acceleration within the tank, $5 \times 10^{-7} \text{ m s}^{-2}$, with the density of helium $\rho = 0.125 \text{ g/cm}^3$ and surface energy $\sigma = 0.12 \text{ dynes/cm}$, the capillary length is about 2 metres or greater, somewhat larger than the diameter of the tank. This means we can expect the helium tide to be a small perturbation to the shape of the surface, which is largely determined by surface tension. We can estimate the size of the tide by calculating the eccentricity of a drop of helium in a gravity gradient dg/dr . This is roughly $\sqrt{\rho a^3(dg/dr)/19\sigma}$ in which a is the semimajor axis of the drop. The resulting distortion amounts to a few millimetres for a drop 50 cm across, and is much less for smaller drops. The effects of the two tidal bulges will cancel to a large degree.

In order to guarantee a sufficient margin of safety from tidal effects, a helium confinement system has been devised.

3.4.2 Helium confinement system

Split dewar concept

Since the higher moments of the test masses are almost insensitive to disturbing masses farther away than about 25 cm, it is proposed that the dewar (inner radius 26 cm, outer radius 50 cm) be divided into two equal concentric cylindrical volumes with a dividing wall at 40 cm (see Fig. 3.7). During the first half of the experiment, the free surface of the liquid helium will be in the outer half of the dewar, at radial distances greater than 40 cm. The helium will then be transferred completely from the inner to the outer volume via a superleak and the fountain effect, so that for the second half of experiment the free surface of the liquid helium will be once again at distances of at least 40 cm from the test masses.

Electrostatic constraint

A further degree of control of the helium tide in the outer chamber will be achieved electrostatically.

Applying an electric field to the helium produces a force per unit volume

$$\frac{1}{2}(\epsilon - 1)\epsilon_0 \nabla E^2$$

The problem is to find an optimum configuration of electrodes which maximises the volume in which ∇E^2 is large while minimising the peak magnitude of E to prevent electrical breakdown. A single cylindrical electrode (Fig. 3.7), should be sufficient. Figure 3.8 shows the equipotential surfaces for the combined tidal and electrostatic forces. For the calculations, the field gradient was chosen to be only 10^5 V/m in order that the effect of the tide would be visible on the plot. At a radius of 40 cm, the tidal bulge is about 15 mm. For a field gradient of $2.5 \times 10^5 \text{ V/m}$, this would be reduced to 3-4 mm which, referring to Fig. 3.6, is within the requirements for 10^{-18} g

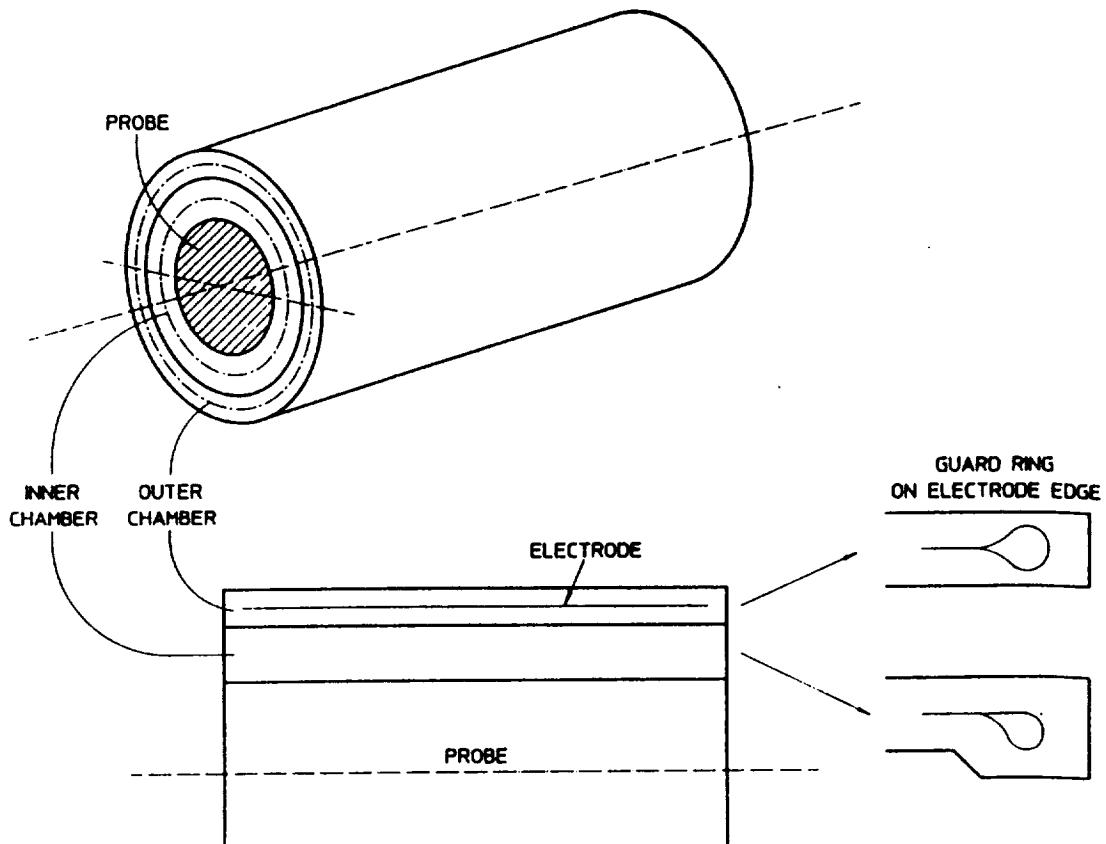


Fig. 3.7. Split dewar concept. The inner chamber is kept full during the first half of the mission, and empty during the second half. In this way, any helium tidal motion is sufficiently remote from the test masses. Also shown is the cylindrical electrode for electrostatic suppression of the tide in the outer chamber.

sensitivity (and well within the requirements for 10^{-17} g). Figure 3.9 shows the ratio between tidal and electrostatic forces along a circumferential arc at a radius of 40 cm. For field gradients of 10^5 and 2.5×10^5 V/m, it is confirmed that the electrostatics effectively suppress the tides. Further calculations incorporating the real eccentricity of the dewar and the effects of surface tension (so far omitted) should be carried out to determine the actual margins obtained in tidal suppression at the orbit rate. Electrical breakdown at the edges of the electrode can be avoided by folding the edges to form smooth, rounded lips (guard ring). These are shown in the insets in Fig. 3.7. For such a configuration, the maximum field strength at the edge is limited to about three times the nominal field strength (midway along the dewar). This configuration has the added benefit of attracting the helium to the ends of the dewar when the tank is almost empty, thereby reducing the influence on the accelerometers.

From the engineering viewpoint, the single electrode has the advantage of simplicity, requiring only one feedthrough and four to six insulated supports. Furthermore, the guard ring provides sufficient structural stiffness.

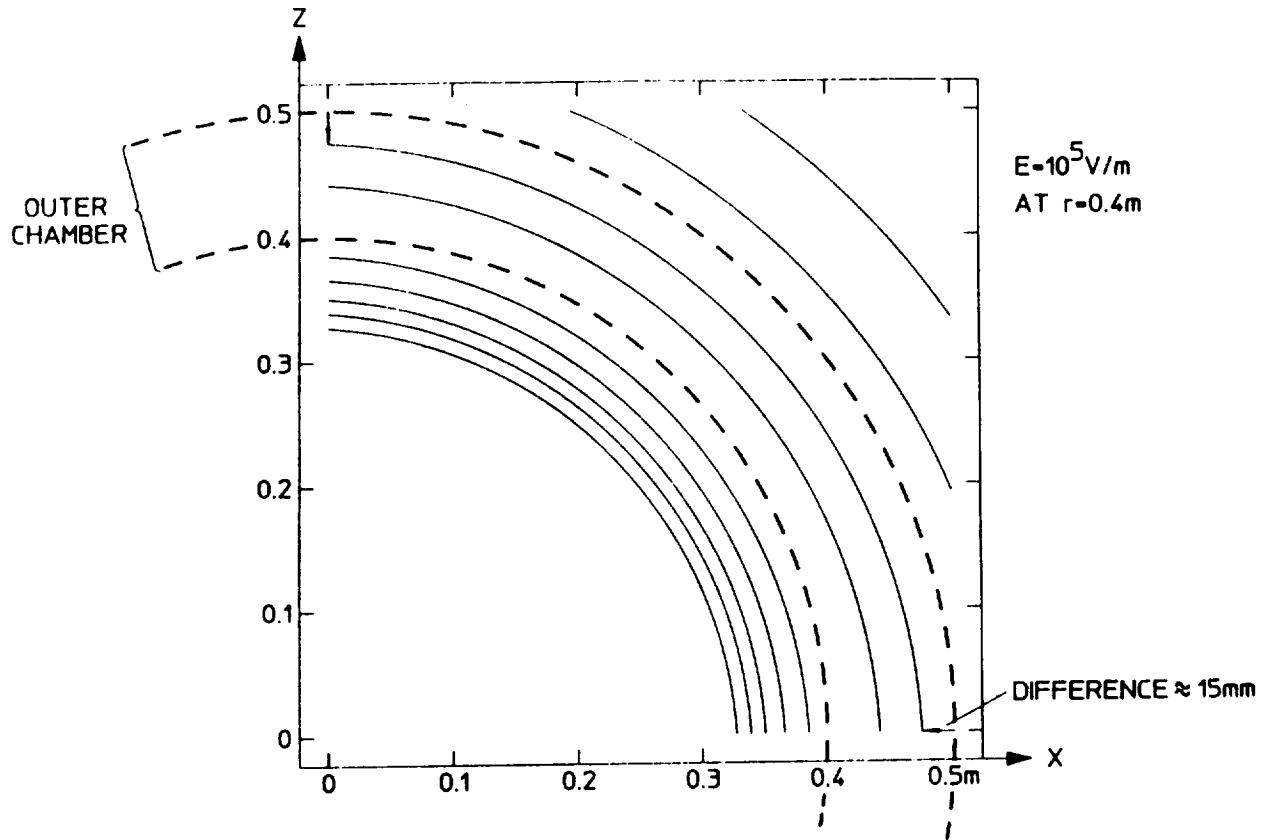


Fig. 3.8. Combined tidal and electrostatic equipotentials for the single cylindrical electrode with an applied field gradient of 10^5 V/m. It can be seen that the tide is suppressed.

3.4.3 Particle radiation and electric charging

Another disturbance which can penetrate the shielding is particle radiation, especially protons above 100 MeV which are common in the radiation belts and solar wind. Particles in this energy range can pass through the spacecraft structure and impact the test masses. It is impractical to consider shielding against them. Effective shielding would require tens of centimetres of lead.

The particles reaching the test masses produce three disturbing effects:

1. energy deposition causing the test masses to heat up
2. electrostatic charging of the test masses
3. direct momentum exchange resulting in forces on the test masses

In order to assess the magnitude of these radiation effects, extensive modelling was carried out. Modelled trapped particle and solar proton spectra were used to define ambient radiation populations, and a sophisticated 3-D Monte-Carlo code (GEANT, Brun et al., 1984) was employed to transport the incident radiation through a model of the STEP geometry.

Simulations reveal that particles below about 70 MeV are effectively shielded by the spacecraft. The peak energy deposition occurs for ≈ 130 MeV particles. Higher energy particles pass right through, depositing some energy on the way.

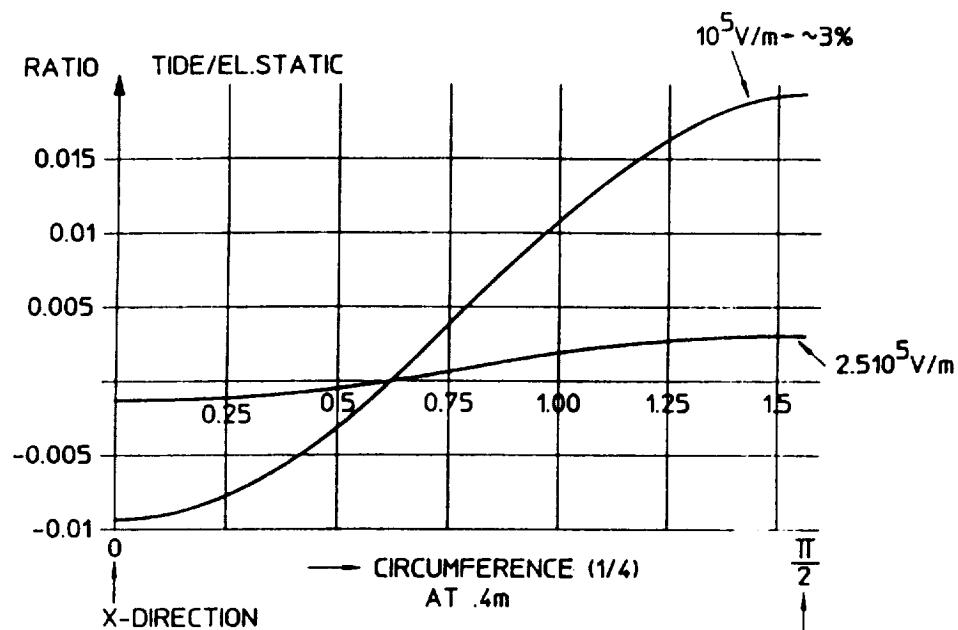


Fig. 3.9. Ratio of tidal to electrostatic forces along a circumferential arc at a radius of 40 cm. For electrostatic field gradients $\geq 10^5$ V/m, it can be seen that the electrostatic forces dominate.

The South Atlantic Anomaly (SAA)

The orbit will take the satellite through the South Atlantic Anomaly (SAA) which is a region of intense geomagnetically trapped particle fluxes at low altitude, resulting from the offset and tilt of the geomagnetic axis with respect to the geographic north-south axis. Protons provide the predominant charged particle population in the SAA since the peaks in the electron population occur at higher altitudes.

At low altitudes, fluxes during solar maximum are *lower* than those during solar minimum. This is due to the heating and expansion of the atmosphere during periods of high solar activity which results in the enhanced absorption of protons. The energy range of the trapped proton population is between 100 keV and a few hundred MeV.

For certain orbits, STEP will miss the SAA altogether, while for others it will cross the heart of this region of intense fluxes. Passes through the fringes of the SAA result in moderate fluxes, and the full daily profile is clearly modulated. A pass through the centre of the SAA will typically last for about 20 minutes, though the majority of the radiation will be encountered in about 5 minutes.

The SAA could be avoided altogether if STEP employed an equatorial orbit. However, such a trajectory would take the spacecraft in and out the Earth's shadow, thereby interfering with the measurements due to the effects of important temperature fluctuations on the experiment.

Trapped electrons do not create as severe a radiation problem for STEP as do protons since they have much lower energies than their positive counterparts, and are unable to penetrate the spacecraft structure.

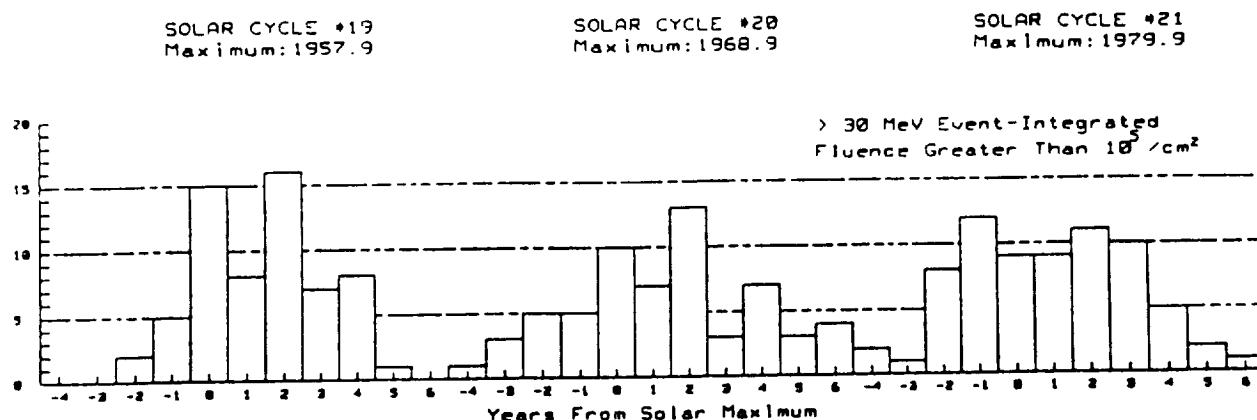


Fig. 3.10. The number of solar events per year with event-integrated fluence of >30 MeV protons greater than 10^5 particles/cm 2 during the last three complete solar cycles (numbered 19, 20 and 21). The yearly intervals are measured relative to solar maximum, as defined by the 13-month running average sunspot number. The data are taken from Feynman et al. (1990).

Solar flare protons

The high inclination of the STEP trajectory does, however, mean that the satellite will be exposed to solar flare protons for an appreciable fraction of the orbit (about 25% for a 500 km orbit). Instead of deflecting incoming charged particles which have entered the magnetosphere close to the polar regions, the Earth's magnetic field at high latitudes has the opposite effect of funnelling protons down towards the magnetic poles. Figure 3.10 shows the number of events with integrated fluences of >30 MeV protons exceeding 10^5 particles/cm 2 for the past three complete solar cycles. The data were obtained from the JPL model data set. For reference, the >30 MeV proton fluence for the August 1972 flare (anomalously large event) was about 8×10^9 /cm 2 . This figure clearly shows that the number of large events is highly variable from cycle to cycle, but peaks around solar maximum. At solar minimum, the occurrence of large solar flare events can be neglected.

If STEP flies during solar maximum, it will be impossible to discount the possibility of a flare of the magnitude of the August 1972 event ever occurring within the 6 month mission lifetime.

Heating due to radiation

Protons (either trapped or of solar origin) which have sufficient incident energy to penetrate through to the test masses will deposit energy through ionisation. This energy will cause bulk heating of the test masses, which can, in principle, disturb the temperature stability of the experiment.

Bulk heating calculations at solar maximum as a function of altitude are tabulated for the outer and inner test masses in Table 3.3.

The heating values represent near worst-case passes of the SAA. There is a significant increase in the level of heating with higher orbit altitude. As described earlier, during periods of low solar activity, the SAA proton fluxes increase. This causes the test mass temperatures to correspondingly increase to about twice the values shown in the table.

The heat production is almost uniform throughout the volume of the masses which have short thermal time constants. Since it is thermal gradients which drive gas pressure and other disturbances, this heating has a negligible effect on the experiment (see Section 3.4.6). Similarly, calculations reveal that the heating of the quartz block is essentially uniform (even in the presence of input flux anisotropies), so there is not a problem with gradients across the sensing circuits.

If we now consider solar protons instead of trapped protons as the ambient radiation source, bulk

Table 3.3. Temperature increase (mK) of the test masses for a single pass through the SAA centre as a function of orbit altitude at solar maximum.

	Orbit Altitude (km)						
	350	400	450	500	550	600	650
Outer Mass	0.14	0.31	0.63	1.17	1.96	3.24	5.01
Inner Mass	0.12	0.29	0.58	1.08	1.81	3.01	4.68

Table 3.4. Induced charge ($\times 10^{-13} \text{ C}$) on the STEP test masses for a single pass through the SAA centre, as a function of orbit altitude at solar maximum.

	Orbit Altitude (km)						
	350	400	450	500	550	600	650
Outer Mass	0.10	0.22	0.43	0.77	1.25	2.00	3.03
Inner Mass	0.09	0.20	0.39	0.70	1.13	1.82	2.76

heating becomes a much more serious problem. The temperature rises over the duration of a solar flare event are about 10 mK for an ordinary event, and about 10 K for an anomalously large event. Hence, if STEP were to experience an anomalously large event, the test masses would have to be re-caged to conduct the heat away, otherwise they may well lose their superconductivity. The plan is to cool the masses occasionally by re-caging them or touching them to a wall. If superconductivity is not actually lost, only a few tests should be needed to confirm that this operation has had no effect on the masses or calibrations. A test of common-mode rejection, and measurement of the periods to show that they have not changed, may be enough.

If superconductivity is lost temporarily, a subset of the setup and calibration procedures must be repeated. This includes at least the common-mode balance of the differential accelerometer, the common-mode calibration, and background force measurements.

Charging due to radiation

Charging is the most serious consequence of the particle radiation since it does not take very much charge to disturb the experiment at the $10^{-17} \text{ m s}^{-2}$ level.

Electrostatic charging can result from the large number of secondaries generated by incident particles colliding with the test mass nuclei.

Individual SAA protons produce relatively few secondaries due to their moderate energies. However, if these particles are integrated over a complete pass, they can be responsible for a significant build-up of charge.

Table 3.4 summarises the analysis of electrostatic charging of the test masses resulting from complete passes of the SAA at varying altitudes during solar maximum conditions. As with bulk heating, there is a significant increase in charging at higher altitudes due to the greater trapped proton populations. Similarly, the charging due to SAA particles is about twice as large during solar minimum because of the increase in the SAA fluxes. A charge of about 10^{-13} C is enough to cause a disturbing acceleration of $10^{-17} \text{ m s}^{-2}$. From Table 3.4, it is clear that, at our nominal altitude of 550 km, the charging is large enough that it must be controlled. This is the subject of Section 3.4.5.

Solar events generate a large number of protons which are also capable of inducing electrostatic charging effects on the test masses. For an ordinary event, the charge deposited on a test mass will be about 10^{-12} C which creates a disturbance an order of magnitude larger than our target sensitivity. For an anomalously large solar flare, the resulting charge is about 10^{-9} C.

Momentum transfer due to radiation

If there is a large enough asymmetry in the transfer of momentum along the sensitive axis of the test masses by the penetrating radiation, the STEP measurements will be compromised. Such an asymmetry may result from the geometry of the experimental design, but is more likely to be associated with any anisotropy of the incident particle population, as is the case in the SAA. The peak amplitude of the acceleration of the test masses due to radiation is about 10^{-15} m s⁻². However, although this is an order of magnitude greater than the desired measurement sensitivity, it does not cause a problem because the time history of the disturbance appears "impulsive" over the relatively short duration of the pass. This time signature will be most apparent from the data analysis. It is true that some fraction of this disturbance will appear at the orbit rate, but Fourier analysis suggests that this fraction is orders of magnitude below 10^{-16} m s⁻². Furthermore, even if this disturbance came at exactly orbital frequency it could be clearly distinguished from an EP signal at the same frequency by the fact that its phase is locked to the orbit rather than to spacecraft attitude.

3.4.4 Radiation sensor

Monitoring the radiation environment of the STEP accelerometers allows us to invalidate data when the radiation rises above a critical level. Sufficiently detailed monitoring could allow the effects of heating, charging, and momentum exchange to be modelled using radiation transport codes such as GEANT. This is unnecessary within the baseline sensitivity because the contaminated data can be simply cut out. However, a radiation measurement could provide a significant improvement in the response of the charge control system, once the charging response of the individual masses is calibrated.

Two existing radiation sensor packages which would be suitable for STEP have been identified. These are the Charged Particle Monitor (CPM) used for the OSSE program and the Radiation Environment Monitor (REM) developed in Europe under an ESA contract and due to fly for the first time in 1993 on STRV. Both packages incorporate dual detectors which separate electron and proton components in the radiation environment by using different thicknesses of passive shielding material in front of each detector. The CPM uses plastic scintillators viewed by photomultipliers whilst the REM uses thick fully depleted silicon detectors.

3.4.5 Charge control system

Perhaps the most important feature of the electrostatic suspension is that it can be made sensitive to electric charge on the test mass. This makes it possible to measure and control the electric charge on the test mass. Without this system, particle radiation would soon charge the mass to a level that would interfere with the science measurement.

A charged mass interacts with its enclosure through the gradient of the capacitance. The force on a test mass in an enclosure is approximately

$$F_Q = \left(\frac{\delta}{2C_{tot}^2} \right) \frac{dC}{dx} Q^2 \quad (3.3)$$

where Q is the disturbing charge, C_{tot} is the total capacitance, and δ is the fractional difference between the derivatives of the capacitances dC/dx from each end of the mass to the enclosure. To

prevent the charge interfering with the EP measurement we must control it to be less than the value $Q_d \sim 10^{-13}$ C which gives a force roughly equal to the design sensitivity, 10^{-17} m s⁻² times the mass m .

To make a measurement of the charge on a suspended mass, we modulate the voltage on the electrodes with a dither voltage ΔV having known amplitude and frequency. The mass will respond with an acceleration proportional to its charge, which we can synchronously measure with the SQUID differential accelerometer. The sensitivity of the charge measurement depends directly on the degree with which this acceleration can be distinguished from the spacecraft's residual acceleration a_s . The dither must produce a measurable force $ma_s\eta_c$ where η_c is the common-mode rejection ratio of the differential accelerometer. The smallest measurable charge is

$$Q_m = \frac{ma_s\eta_c C_{tot}}{\Delta V(dC/dx)} \quad (3.4)$$

We can expect a common-mode rejection (η_c) of only 10^{-2} at frequencies above the resonant frequency. This can be extended to 10^{-3} or better if we apply further compensation of the common-mode in real time. The two masses in each accelerometer can have different dither frequencies, so independent measurement of each mass is possible.

The best way to reduce Q_m is to decrease the effective noise $a_s\eta_c$. The spacecraft noise a_s decreases rapidly below the drag-free control bandwidth of 0.1 Hz and is low enough at 0.01 Hz that $Q_m < Q_d$ can be achieved (Fig. 2.10). We therefore expect that the charge can be controlled to the required level in a few minutes – comparable to the length of a passage through the SAA.

Closed-loop control of the charge of a suspended gyroscope rotor has been demonstrated for GP-B using ultraviolet light and a supplementary electrode. The UV light produces photoelectrons from both the suspended mass and the electrode. A bias potential on the electrode determines whether the net current is toward or away from the mass. Currents many times the expected charging rate are possible without exceeding the requirements on disturbances to the test mass. We expect to be able to achieve closed loop control of the mass charge on a time scale of 100 seconds.

3.4.6 Other disturbances

The test masses are isolated from each other and from the environment by superconducting shields. These attenuate external electromagnetic disturbances, such as the Earth's magnetic field, by a factor of 10^{10} . Gravitational effects are largely eliminated by design of the test masses and helium control, while particle radiation is a disturbance at known times, easily distinguishable from an Equivalence Principle violation. The largest remaining disturbances are internal.

Residual gas pressure

The largest internal disturbances will probably come from residual gas inside the spacecraft. Gas flows can be caused by several processes inside a closed system, especially by temperature differences. The best solution is to work at the lowest possible gas pressure. Experimentally, the lowest pressures are obtained just above absolute zero, where most gasses are frozen out and the residual helium molecules are moving very slowly and can be easily adsorbed. A “low temperature bakeout” procedure will reduce the helium pressure in the sealed instrument chamber to the required level of about 10^{-11} torr. This procedure raises the temperature to about 6 K to remove adsorbed gas; in terms of change in absolute temperature, that is like baking a room-temperature apparatus at 900 K.

The gas pressure requirement is set by considering several effects which are proportional to some portion of the surface area of the mass and a pressure difference. For example, a pressure difference could be caused by one end of the chamber being transiently warmer than the other, expanding the

gas; or by outgassing from a real or virtual leak. In all cases, it is the disturbance caused at signal frequency that counts. Thermal gradients might be expected to vary at orbital frequency because of changing external heat loads on the dewar. Using a closed-loop system with temperature sensors (and heaters, if necessary), we will regulate the temperature to 1 mK per orbit and the gradients to 0.1 mK per cm per orbit. If there is a need for more cooling at a given time, the thrusters will be opened to vent additional helium.

The "radiometer effect" is the worst offender among gas pressure effects. This is due to molecules being emitted from a hot region with greater velocity than from a cold region. With a long mean free path, the molecules simply transfer momentum from the hot to the cold region, and to any test mass that happens to be between. This causes an acceleration $P(dT/dx)/2\rho T$ where P is the pressure, T the temperature, dT/dx the temperature gradient, and ρ the test mass density. We have been able to show that at 10^{-11} torr, the temperature gradient variations in the dewar will be disturb the test masses by less than $10^{-17} \text{ m s}^{-2}$.

Thermal effects

Operation at cryogenic temperature turns out to reduce several other disturbances in addition to those caused by gas pressure. Size and shape changes from thermal expansion are a major problem in precise room temperature measurements, but low temperature reduces the thermal expansion of most materials very significantly—approximately as the square of the temperature. The heat capacity of crystalline materials, at low temperature, decreases as the cube of the temperature, while the heat conductivity tends to decrease linearly or as the square. This implies that the relaxation time for thermal gradients can become very short, so that objects tend to heat or cool as a unit rather than on one side at a time. This further reduces thermal distortion at low temperature, so that extremely stable structures are possible. Another disturbance that is reduced is thermal radiation pressure, which depends on the fourth power of temperature. The superconductivity of heat in superfluid helium makes certain that the apparatus is at a uniform temperature. These considerations are what lead us to operate the STEP instrument at less than 2 K.

The ordinary thermal expansion of fused quartz near 2 K is about 2×10^{-9} . Since the temperature regulation will be 1 mK or better, any scale factor or shape changes will be about 2 parts in 10^{12} . These are quite insignificant.

Another important temperature effect is the change in superconducting penetration depth with temperature. This causes a change in the average position of the supercurrent and consequently a drift in the equilibrium position of the test mass. The depth has changes by $d\Lambda/dT = 1 \times 10^{-10} \text{ m/K}$ for niobium at 2 K. The temperature control to 1 mK per orbit keeps this disturbance to below $2 \times 10^{-19} \text{ m s}^{-2}$ at the EP signal frequency, which is completely negligible.

Patch fields and trapped flux

Other internal disturbances include coupling from the spacecraft to the test masses by electric charge patches or magnetic fields from within the spacecraft. These forces will be nearly constant in time, whereas the EP signal is periodic. However, they might vary from point to point within the apparatus, which produces a time-varying force on the masses because of the inevitable small jitter of the spacecraft around them. This places limits on the gradients of these background forces, for a given vibration level of the spacecraft. For a spacecraft jitter of 3×10^{-11} , the background electromagnetic forces should not vary by more than about 10^{-9} N/m .

Forces from the electrostatic patch effect and trapped flux depend on the areas involved. But more importantly, they depend on the separation between surfaces. A randomly distributed patch effect produces a force which decreases as the fifth power of the separation, because of statistical averaging and scaling laws. It is therefore important to maintain the largest possible gap between

the masses and the surrounding chamber. Based on available data, we estimate that the 1 mm gap in the present design is enough to meet the force variation requirement. Additional care must be taken to ensure that the patches (or trapped fluxes) are not correlated between surfaces on a scale of millimetres to centimetres. As mentioned in Section 3.2.3, the effects of trapped flux and patch effect can be separately measured, and to some extent the disturbance can be calibrated out, provided both radial and axial mass motions are measured.

There are remaining questions about the magnetic flux trapping properties of thin films. Thin films are usually in a highly stressed condition, and may have embedded impurities. This leads us to expect them to trap flux nonuniformly and irreversibly under some conditions. Achieving small levels of trapped flux, and keeping it uniform, depends strongly on the choice of materials, their preparation, and the operating field level. So long as the trapped flux remains constant and below the level required to make the EP measurement, its effects can be measured by a procedure developed for the ground experiment. The effect of trapped flux is that it increases the random force gradients felt by the masses. If the mass positions are uncertain, this causes an increased uncertainty in the total force and a corresponding degradation of the Equivalence Principle measurement.

The flux measurement procedure measures the force the bearing produces with forward and reversed current in the bearing. The difference in force between the two cases is due to trapped flux, which adds to or subtracts from the total field according to its sense. The sum of the two cases is due to the other background forces, possibly only coming from the mechanical shape of the bearing. This measurement should be made at a series of positions covering the volume each test mass might be in, to map the entire force gradient. A similar method—using charge instead of current—will be used with the electrostatic positioning system to test for patch effect forces. Together, the two systems—magnetic bearing and electrostatic suspension—provide tests for the major non-gas pressure forces coupling the test masses to the spacecraft.

4. Spin-Coupling experiment (SC)

The aim of the STEP Spin-Coupling (SC) experiment is to make use of the low-noise, zero-g environment on the STEP spacecraft to search for a new interaction between quantum-mechanical spin and matter.

We present a design of a spin-coupling experiment which shows that a sensitivity of $g_p g_s$ (spin-coupling constants) of 6×10^{-34} at a range of 1 mm is feasible. This represents a seven-order-of-magnitude improvement over the existing ground-based measurements. Axion coupling increases as the inverse square of the range, so this level of sensitivity will be 11 orders smaller than has been up to now tested by a weak force experiment. We must note, however, that for the experiment to be competitive with the constraints on θ derived from measurements of the electric dipole moment of the neutron, we need to find another two orders of magnitude of sensitivity. We intend to investigate the possibilities of further reducing the noise sources due to spacecraft residual motion, patch fields and trapped flux, during Phase B.

4.1 Outline of experiment design

In Chapter 1 we gave an expression for the putative axion-like coupling (Eq. 1.7). Now we must address the question as to how we can design an experiment which both optimises the spin-coupling signal and eliminates, as far as possible, systematic effects due to the electromagnetic forces which will arise in association with spin-polarised sources. Moody and Wilczek (1984) proposed an experiment which used a source of high magnetic permeability which could be polarised with a solenoid. A sapphire crystal placed close to the source would then resonantly detect a spin-coupled stress due to the modulation at acoustic frequencies of the spin directions within the source. They proposed to use superconducting shielding to eliminate magnetic forces on the crystal due to the leakage of magnetic field from the source. The STEP experiment has adopted this basic experimental arrangement. However, we intend to replace the crystal with a superconducting differential accelerometer. The differential accelerometers will work at a much lower frequency than the sapphire and this will not only improve sensitivity but will also enable us to use commercially available high permeability materials which cannot be polarised at acoustic frequencies.

If we have a polarised source with an electron spin density of ρ_s and a test mass of nucleon density ρ_N , then we can use Eq. 1.7 to calculate the spin-coupled force between them as

$$f_a = \frac{\theta}{\lambda^2} \times 7 \times 10^{-61} \rho_s \rho_N \frac{dI}{dx} \quad (\text{mks}) \quad (4.1)$$

where I is the integral of the potential from Eq. 1.7

$$I = \frac{1}{8\pi} \int \int (\hat{\sigma} \cdot \hat{r}) \left(\frac{1}{\lambda r} + \frac{1}{r^2} \right) \exp \left(\frac{-r}{\lambda} \right) dV_1 dV_2 \quad (4.2)$$

Figure 4.1 shows the putative spin-coupling force field for a single electron (source) interacting with a single nucleon (test body). In designing the experiment, we must consider the geometry of the field, and configure the three dimensional sources and test masses to provide the largest signal.

We have considered possible designs where the test mass and source were cylindrically symmetric and coaxial and where the spin alignment axis was parallel to the axis of the cylinders. For these coaxial geometries, an expression for dI/dx was derived which reduced the six-order integration to a single numerical integration (Juan Léon, private communication) and this was used to optimise the acceleration signal. As the dimensions of the source and test mass, and the gap between them,

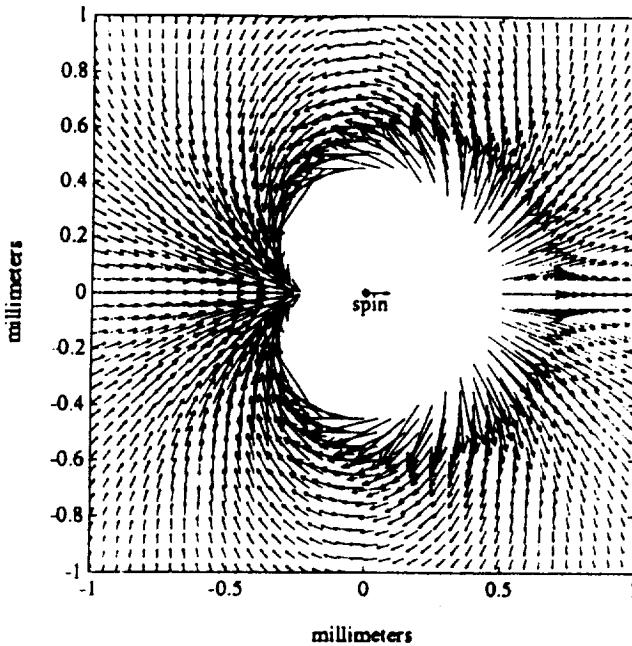


Fig. 4.1. Putative spin coupling force field. The directionality of the force demonstrates the vector-scalar coupling. For clarity, the strong force field in the close vicinity of the fermion is not shown.

are reduced, the acceleration increases because the axion coupling constants vary as $1/\lambda^2$. However, as the size of the test mass becomes smaller, many experimental problems arise. At present, we believe that the smallest range which can be designed for is around 1 mm.

Figure 2.2 illustrates the concept of the instrument, and Fig. 4.2 shows the design of the instrument. We employ 16 square-section toroidal polarisable sources which are each coated with superconductor. On the inside of the source assembly, there is a test mass comprising 16 annuli mounted on a former which is supported by a superconducting bearing. A similar composite test mass encloses the source, and the test mass pair forms a differential accelerometer. Superconducting wires provide current to the sources and a net polarisation of spins is created within the sources close to the test mass annuli. The annuli in each of the inner and outer test mass assemblies will then experience a force along the axis of the superconducting bearing; but in opposite directions. The ensuing differential motion is detected by the SQUID coils. Active dynamical charge control is achieved by applying appropriate voltages to capacitor plates, as for the EP accelerometers (Section 3.4.5). Figure 4.3 shows a close-up of one end of the instrument (showing two-test mass annuli, two source toroids, and the sensor/actuator arrangement).

In order to discuss the magnitudes of potential noise sources, we will focus on a nominal noise level goal of $4 \times 10^{-18} \text{ m s}^{-2}$. This corresponds to a sensitivity to the coupling constant product ($g_p g_s$, Eq. 1.7) of 6×10^{-34} or a value of 10^{-7} for θ . Figure 4.4 shows a plot of the putative spin-coupling acceleration signal as a function of range for the exact experimental geometry. We see that, if the axion were to have a range of 1 mm, then we might expect to see an acceleration of $4 \times 10^{-20} \text{ m s}^{-2}$ ($4 \times 10^{-21} \text{ g}$), which, alas, is two orders-of-magnitude below our target sensitivity. Future efforts will aim at improving the sensitivity so that a true search for the axion may become within reach.

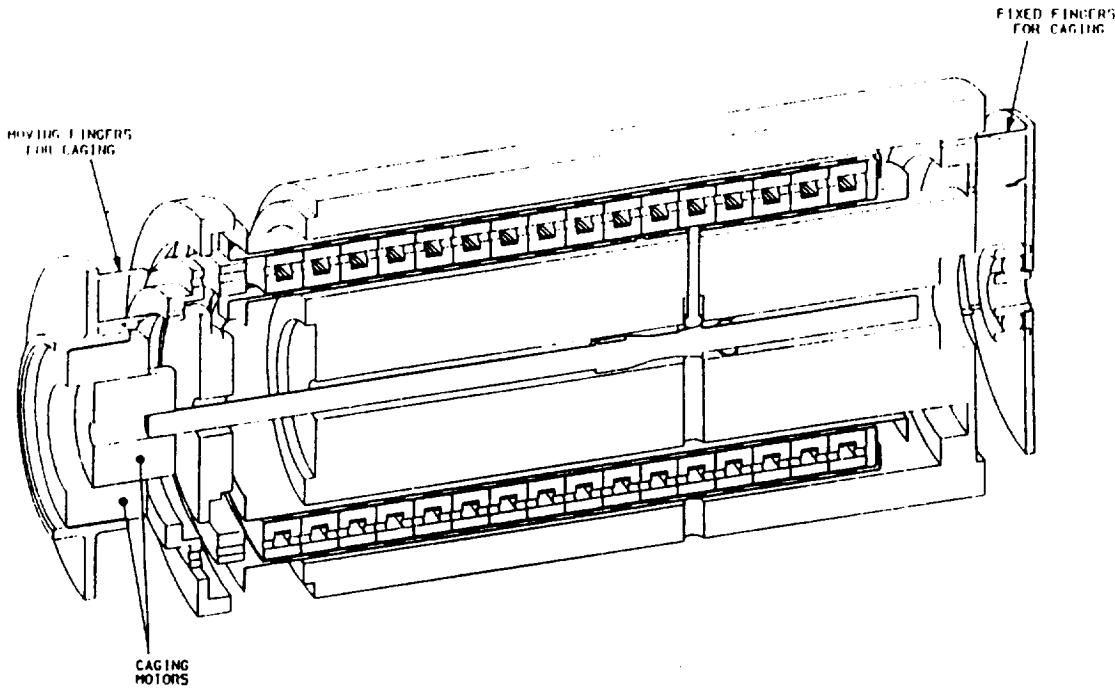


Fig. 4.2. Exploded sectional view of the STEP spin-coupling differential accelerometer.

The polarisable source

When a ferromagnetic material is placed in a magnetic field of intensity H , a magnetic induction B is produced within the bulk of the material according to the relation

$$B = \mu_0 \mu_r H \quad (4.3)$$

In iron, the increase in the internal B field is due to electronic spin magnetisation and not due to alignment of orbital angular momenta. The angular momentum contribution to the magnetic moment of the iron atom is said to be quenched. Thus, if we know the value of the relative permeability of the source and the external H field, we can calculate the density of spins as

$$\rho_s = (\mu_r - 1)H/\mu_B \quad (4.4)$$

or

$$\rho_s = B_{\text{int}}/(\mu_B \mu_0) \quad (4.5)$$

where $\mu_B = e\hbar/2m_e$ is the Bohr magneton. We evidently require a material with a large relative permeability. Further, we can calculate the current turns product, NI , for the source as

$$NI = \int H \cdot dl \quad (4.6)$$

where $\int dl$ is the path integral around a magnetic circuit. For Cryoperm 10 (a product of Vacuumschmelze), B_{int} reaches a saturation value of 0.8 T at a magnetic intensity of 8 A/m and the internal field reaches 75% of its saturation value at an intensity of about 2 A/m. The maximum value of the current turns product for each source is about 0.2 Amp-turns. There are some glassy ferromagnetic materials which have a higher saturation field (absolute maximum of 2 Tesla) and the feasibility of these materials should be investigated.

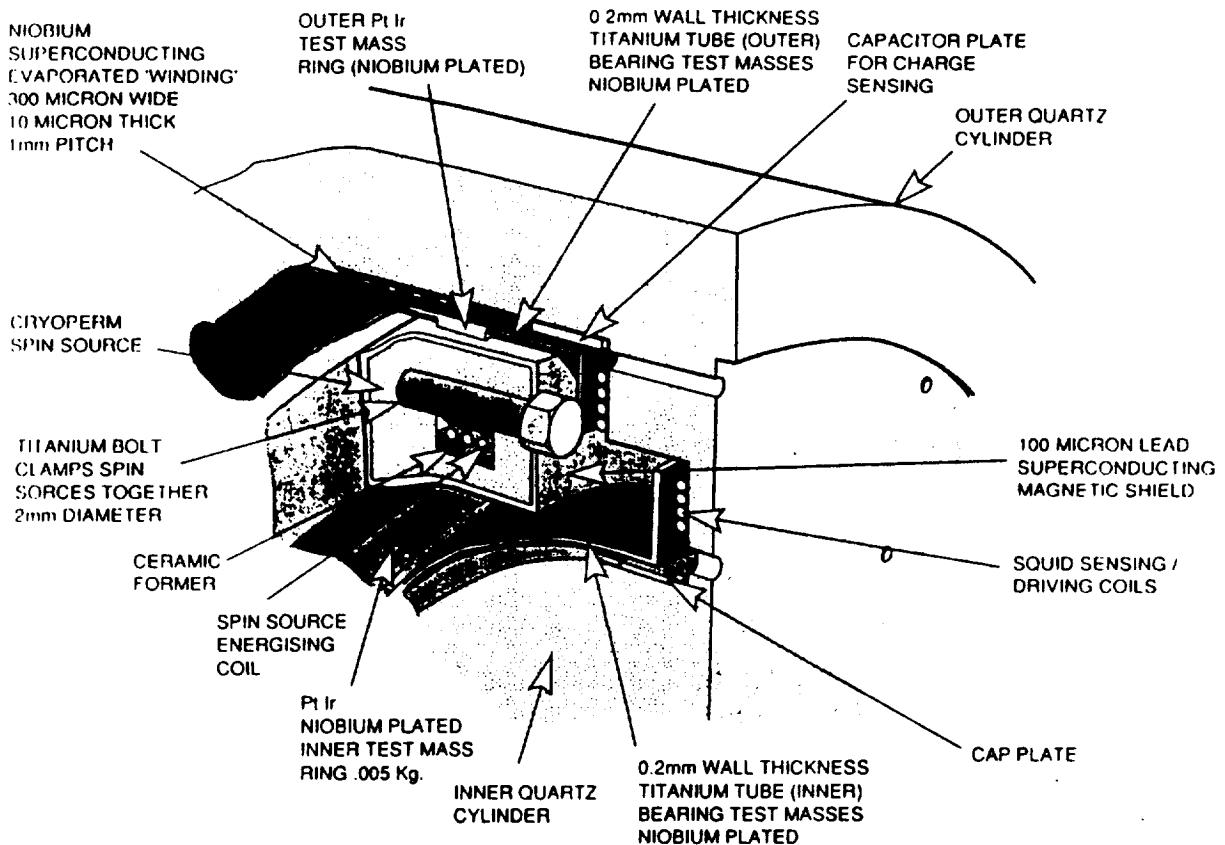


Fig. 4.3. Close-up view of one source/test mass annulus pair of the spin-coupling accelerometer.

The major problem to be considered is the leakage of magnetic field from the source, which will exert forces on the test masses. If the leakage field has a magnitude B_L , then a gradient in the field will create a force on the test mass due to its magnetic susceptibility. Since the test masses are coated with niobium they will be perfectly diamagnetic. In this case they will experience an acceleration of

$$da_{m1} \approx \frac{B}{\mu_0 \rho_{Pt}} dB/dx \quad (4.7)$$

where ρ_{Pt} is the density of the platinum-iridium alloy test mass annuli. We have been careful to choose a geometry where the magnetic flux path is closed to reduce the leakage.

The leakage field is estimated to be about 10^{-5} T. A lead coating of 0.1 mm isolates each individual source from its neighbour and from the test mass. Given that the London penetration depth of lead is 37 nm, a thickness of lead shielding of only 0.5 μm (about 12 penetration depths) would be sufficient to attenuate the field leakage from the Cryoperm by a factor of about 4×10^5 , resulting in da_{m1} smaller than the nominal sensitivity goal of $4 \times 10^{-18} \text{ m s}^{-2}$. However, we have chosen a minimum thickness to eliminate pin-holes in the coating. It is well established that superconducting shielding can attenuate time dependent magnetic fields by factors of 10^{11} (Vitale et al., 1989).

It is worth noting that the magnitude of the force given in Eq. 4.7 does not change sign as the leakage field oscillates. This means that such a perturbation will create disturbance forces on the test masses at twice the modulation frequency, thus enabling the disturbances to be spectrally

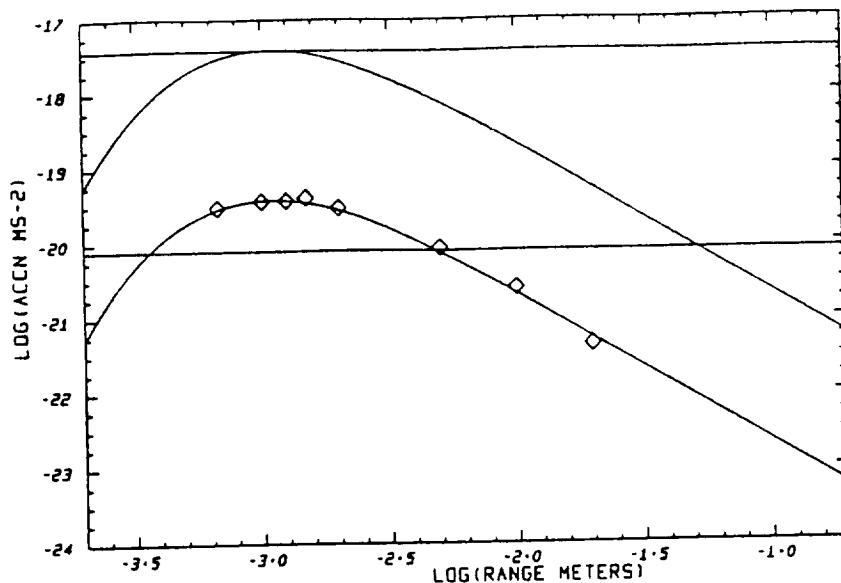


Fig. 4.4. Putative spin-coupling acceleration as a function of range for the exact experimental geometry. The lower and upper curves represent accelerations, calculated analytically following Léon, which would be expected from an axion spin-coupling with values of θ of 10^{-9} and 10^{-7} , respectively. The diamonds are values calculated using Monte Carlo integrations. The numerical results are more accurate at longer ranges as the analytical model ignores the radial spins in the sources. The lower horizontal line represents the thermal noise limit due to gas damping at 10^{-11} Pa and 2 K after an integration time of 5×10^5 s. The upper horizontal line is the thermal noise limit for a Q of 2×10^6 .

distinguished from a true signal. Despite the above arguments, it is absolutely necessary that there be no magnetic coupling between the test masses and the sources. We have therefore decided to add thin niobium shields between the sources and the test masses (not shown in Fig. 4.3). These will be niobium sheaths, $125 \mu\text{m}$ thick and separated from the sources by 0.1 mm, which will be mounted independently of the source assembly.

Another possible magnetic perturbation may come from trapped flux penetrating the wall of the source, interacting with trapped flux in the test mass assembly. It can easily be seen, however, that the forces due to trapped flux will generate accelerations which are negligible. This is due to the exponential attenuation of the magnetic field with the characteristic length which is of the order of the separation of the flux quanta.

Mechanical design

Each source will be electron-beam welded from 2 mm plate. Before the outer rim is welded in place, 10 turns of niobium wire will be wound on to insulating ceramic formers. The wires will exit the source via a ceramic insulated hole. Then, after the final weld, the complete unit will be annealed at around 1000°C . Each source is then coated in lead and held in a stack by two threaded titanium rods (not shown in figures). The entire assembly can then be electroplated with lead. The sources will be wired in parallel in two groups which can be energised independently. One group will comprise sources 1, 3, 5 etc., and the other, 2, 4, 6 etc. The current required to energise the sources is $\pm 0.18\text{A}$.

The source stack is supported on a titanium mount with a quartz spacer which carries capacitor plates for charge and force compensation. The titanium mount is connected to the top end of the quartz block and will be inserted into the test mass and suspension assembly in the final stage of manufacture. In this way, we are mechanically and electrically decoupling the source from the test masses and SQUIDS. On launch, the free end of the source stack will be radially constrained by a

caging mechanism. The total mass of the 16 sources is about 1.1 kg.

There will be a change in the dimensions of each source associated with the magnetisation modulation due to magnetostriction. The radius and the length of a source element will undergo a fractional change by 2.5×10^{-5} . The dimensional change of the source is proportional to the square of the internal magnetic field and therefore the dimensions will not vary at the signal frequency but at its harmonics. (The fundamental component of B_{int} will have a phase lag of about 25° with respect to the drive current.) However, it will be important to ensure that the effects due to magnetostriction do not cause saturation in the SQUID detectors. We will mount each source in such a way that it is mechanically decoupled from its neighbour and pinned at its centre to the titanium rods. The maximum displacement of the surfaces of the sources will then be approximately $0.1 \mu\text{m}$. Any flux which is trapped on the surfaces of the sources will be displaced by the magnetostriction. The niobium shield, which is mechanically decoupled from the sources, will prevent this motion from producing spurious forces on the test mass.

The changes in the length of the source assembly due to magnetostriction will create a gravitational acceleration on each test mass. We have numerically evaluated the differential gravitational acceleration that the expansion produces on the test masses is about $10^{-16} \text{ m s}^{-2}$. This is a factor of 25 times larger than our nominal signal but should not present any difficulties because it occurs at the second harmonic.

In order to eliminate the possibility of systematic forces due to spurious effects such as inductive voltage pick-up and magnetostriction, a double modulation scheme will be employed whereby the two drive currents to the coils will be sinusoidal at slightly different frequencies. In this way, the sources will produce a putative spin-coupling force at the 'carrier' period of 500 s, but this signal will be modulated at the beat period of about 2000 s. Thus, changes in any spurious noise sources taking place over time scales of longer than 2000 s can be eliminated.

The annuli and support tubes (Fig. 4.2) are manufactured from platinum-iridium alloy and titanium, respectively. The tubes are completely coated with niobium. Titanium and platinum alloy are chosen because they optimise the ratio of useful test mass (material within the spin-coupling field) to the total mass of the test mass assemblies. The titanium and platinum have similar integrated thermal expansion coefficients.

The two test mass assemblies form a differential accelerometer. The two pairs of pancake coils face the end flanges of the test mass assemblies (Fig. 4.3). The inductance of the coil versus the test mass displacement was calculated numerically (Sumner, 1987) and using these values of inductance we can calculate the acceleration sensitivity of the accelerometer. If the period of the differential mode of the accelerometer is around 500 s, then we obtain an acceleration noise of $10^{-15} \text{ m s}^{-2}/\sqrt{\text{Hz}}$ from Eq. 2.4. This noise level is limited by the thermal Brownian motion noise.

4.2 Disturbances

The spin-coupling experiment is subject to the disturbances common to all the STEP experiments (see Sections 2.5, 3.4), plus a few which are unique to this experiment.

The largest source of random noise will be the residual acceleration of the spacecraft (see Fig. 2.10). At a frequency of $2 \times 10^{-3} \text{ Hz}$, the acceleration noise of the spacecraft will be about $4 \times 10^{-11} \text{ m s}^{-2}/\sqrt{\text{Hz}}$. If the common-mode rejection ratio of the differential accelerometer is 10^4 , then this residual spacecraft acceleration should limit the experiment to the nominal acceleration sensitivity of $4 \times 10^{-18} \text{ m s}^{-2}$ after about 10^6 s.

The background spacecraft displacement will produce a background random noise acceleration due to common-mode motion of the test masses relative to trapped flux elements and patch effect fields. At the signal frequency of $2 \times 10^{-3} \text{ Hz}$, the spacecraft displacement noise will be approximately $3 \times 10^{-10} \text{ m}$ after averaging for 5×10^5 s. We note, however, that the residual rms SQUID noise

after a similar integration time, will be 10^{-14} m. Thus, in principle, we can extract the forces due to spacecraft displacement by cross-correlating the differential acceleration output with the large common-mode signal. These noise sources are discussed in Section 3.4.6 in the context of the EP experiment.

A more fundamental noise source than spacecraft residual acceleration comes from fluctuations in any damping forces of thermal origin. We can assume that the damping of the accelerometer will be greater than that due to the residual gas pressure ($P \approx 10^{-11}$ torr) in which case the minimum thermal noise spectral density at 2 K will be

$$S_a^{1/2}(f) = [4(k_B T m_{He})^{1/2} (PS/m_t^2)]^{1/2} \text{ m s}^{-2}/\sqrt{\text{Hz}} \quad (4.8)$$

where m_{He} is the mass of a helium atom, m_t is the mass of the test mass assembly, and S is the total surface area. After 10^6 s of integration, this gives a sensitivity limit to the experiment of 8×10^{-20} m s $^{-2}$ which is far below the target sensitivity.

Another likely source of thermal noise will be eddy current damping as leakage fields from the SQUID detectors interact with non-superconducting metal components. We have been careful to place capacitor plates away from the SQUID detectors and we will construct them from niobium if necessary.

In order to achieve our design goal of 4×10^{-18} m s $^{-2}$, the Q of the differential mode must be at least 2×10^6 . We believe that this is a feasible goal.

5. Constant of Gravity G and Inverse Square Law experiment (G/ISL)

In order to decrease the uncertainty in the value of G and improve the limits on deviation from the Inverse Square Law (ISL) at short range by two orders of magnitude, STEP will carry a dedicated gradiometer system composed of two concentric pairs of accelerometers. This gradiometer system will also allow a valuable short-range test of the EP and high-resolution mapping of the Earth's gravity.

The density inhomogeneity of materials and metrology errors limit the ultimate precision of short-range gravity experiments. Single crystals of dielectric materials can be grown with density homogeneity as high as 1 part in 10^8 . For simple shapes with dimensions of centimetres, standard metrology would determine the shapes to 10^{-5} . For transparent materials, this error could be reduced to 10^{-6} or better by using optical measurement techniques. We set 10^{-6} as a common goal for the G and ISL experiments. Terrestrial torsion-balance experiments for G and the ISL at cm ranges have been limited to 10^{-4} (Luther and Towler, 1982; Spero et al., 1980).

In addition to the metrology requirements for mass density and dimensions, the G experiment requires an *absolute calibration* of the accelerometers to the accuracy desired for G. In the experiment by Luther and Towler (1982), which yielded the presently accepted value of G, gravitational acceleration was calibrated by measuring the resulting increase in the resonant frequency of the torsion fibre. The authors report, however, that the measurement of this frequency shift was their dominant error source and is the one most difficult to improve.

In space, one finds an elegant solution for acceleration calibration. Modern satellite geodesy has advanced the knowledge of the Earth's gravity to the point of measuring the geocentric gravitational constant, GM_E , to 10^{-9} . The GPS technology can determine the geocentric position of the spacecraft to 10^{-8} . These can be combined, along with the improved gravity model of the Earth that STEP will produce, to calibrate the STEP gradiometers to the required accuracy of 10^{-6} .

5.1 Hardware description

The gradiometer system for the G/ISL experiments is composed of two identical instrument packages separated over 75 cm along the length of the quartz block, as shown in Fig. 2.9. Figure 5.1 shows a cutaway view of one of these packages. The source mass is made out of platinum-10% iridium alloy (density of 21.6 g/cm^3) and is driven along the symmetry axis to produce time-varying gravity signals. The inner and outer test masses are made out of single-crystal BGO ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$, density 7.1 g/cm^3) and single-crystal sapphire (Al_2O_3 , density 4 g/cm^3), respectively. These crystals are transparent and their average atomic numbers are well separated for a composition-dependence test. A single-crystal lead fluoride (PbF_2 , density 8.2 g/cm^3) is another candidate for the inner test mass material.

The source mass is a circular cylinder with a hole through the centre to accommodate a threaded rod for caging. Each source mass weighs 93.6 g and its dimensions are chosen to make all of its quadrupole moments vanish. It is levitated on a superconducting magnetic bearing, formed by six meander-pattern coils with alternating taper, located on its outer surface (see Fig. 5.2). The purpose of the taper is to produce an axial component of the restoring force in addition to the main radial component. This axial spring will centre the source mass with respect to the test masses along the axis. The radial and axial resonant frequencies of the source mass may be tuned to about 0.1 and 0.01 Hz, respectively. The levitation coils are deposited on top of a copper film to provide

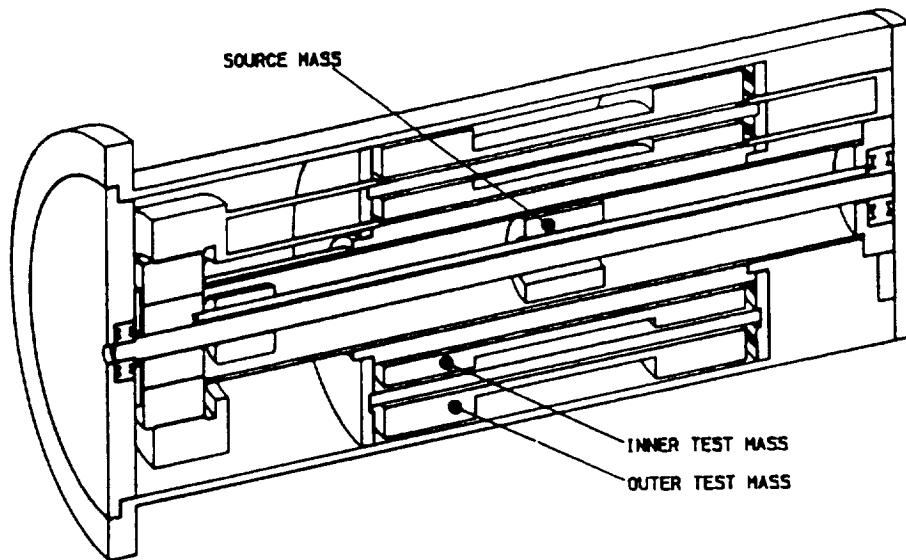


Fig. 5.1. Cutaway views of one half of the G/ISL apparatus. The other half of the apparatus, 75 cm away along the symmetry axis (not shown), contains a mirror image of the instruments shown here. The inner test masses on the two ends form a gradiometer. The outer test masses form a separate gradiometer. The source masses are constrained to move along the axis by magnetic bearings. The gravity signals from the moving source masses are detected by the two gradiometers to perform the G and ISL experiments.

passive damping for the modes.

To drive the source mass, an AC current is injected externally to the levitation circuit. The axial position of the source mass is sensed electrostatically. The position sensing circuit for the source mass is similar to the axial sensing circuit of the EP test mass except that the electrodes are repeated along the axis to cover the entire range of the source motion. The required position resolution is $0.01 \mu\text{m}$ for a small $\pm 2 \text{ mm}$ range from the centre and $1 \mu\text{m}$ for the full range of $\pm 6 \text{ cm}$. Although the two masses will be driven symmetrically for an optimum result, they will be controlled by separate circuits in order to maintain freedom to drive one mass at a time.

There are four test masses in the system. Two inner masses are connected together by the standard differencing circuit, shown in Fig. 2.6, to form a gradiometer. The outer masses form a separate gradiometer. The inner and outer test masses weigh 301 and 334 g, respectively. Since these gradiometers will also be used to obtain geodesy data with a bandwidth from near DC to about 0.02 Hz, we keep the axial differential-mode frequencies above the signal bandwidth, at around 0.1 Hz. The axial common-mode frequencies are kept a little lower, at about 0.03 Hz. The levitation, sensing, and mode damping schemes for the G/ISL accelerometers are similar to those for the European EP accelerometers. In order to make the test mass a one-dimensional system and thus minimise coupling between the axial and radial degrees of freedom, the radial springs are much stiffer than the axial spring; the radial resonant frequencies are about 1 Hz.

Efficient electromagnetic shielding is extremely important to prevent a direct electrical cross talk between the driving circuit for the source mass and the gradiometer detection circuits. The gradiometer circuits also need to be shielded from each other to maintain mutual independence. The ability of superconductors to perfectly screen electromagnetic fields allows us to achieve the requisite high degree of isolation between the circuits.

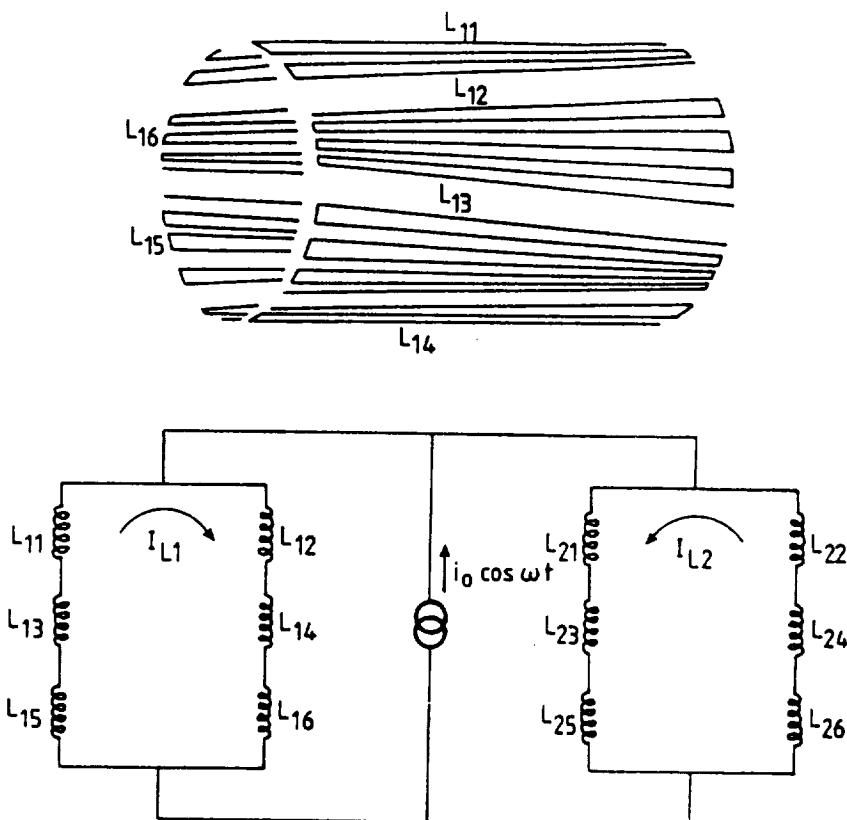


Fig. 5.2. (a) Arrangement of levitation coils for one source mass. Six tapered meander-pattern coils are used to provide radial as well as axial confinement of the source mass. **(b)** Levitation and driving circuit for both test masses. The persistent currents I_{L1} and I_{L2} provide restoring forces in the radial and axial directions. The AC current $i_0 \cos \omega t$ provides symmetric driving forces to the two source masses.

5.2 Design principles and experimental procedures

Even in the extremely quiet drag-free environment, the seismic background noise is many orders of magnitude larger than the acceleration resolution needed for the experiments. The gradiometers evade this problem by rejecting the common-mode acceleration by 10^4 , while maintaining full sensitivity to a differential acceleration. The gradiometer configuration also permits the geodesy experiment and the absolute calibration of the gradiometer scale factors.

The double gradiometer configuration provides redundancy and valuable cross checks for the G, ISL and geodesy experiments, as well as for the common-mode acceleration signal for drag-free control normal to the orbit plane. Cross checks are essential for credibility of such drastically improved, new scientific results as we are seeking in STEP. Furthermore, constructing the two gradiometers with different materials, composition dependence can be tested for a force violating the ISL. Thus the scientific merit of the double gradiometer configuration outweighs the increase in complexity.

The symmetric movement of the two source masses generates additive gravity signals at the gradiometers while producing zero net linear and angular momentum. The common-mode signals are unaltered by the source motion and thus can continue to be used for drag-free control while the G/ISL experiments are performed. During the G experiment, the source masses exert time-varying gravitational accelerations as high as $3 \times 10^{-10} \text{ m s}^{-2}$ on the EP accelerometers but at a frequency well above the EP signal frequency, so it can be easily filtered out. Furthermore, the source masses

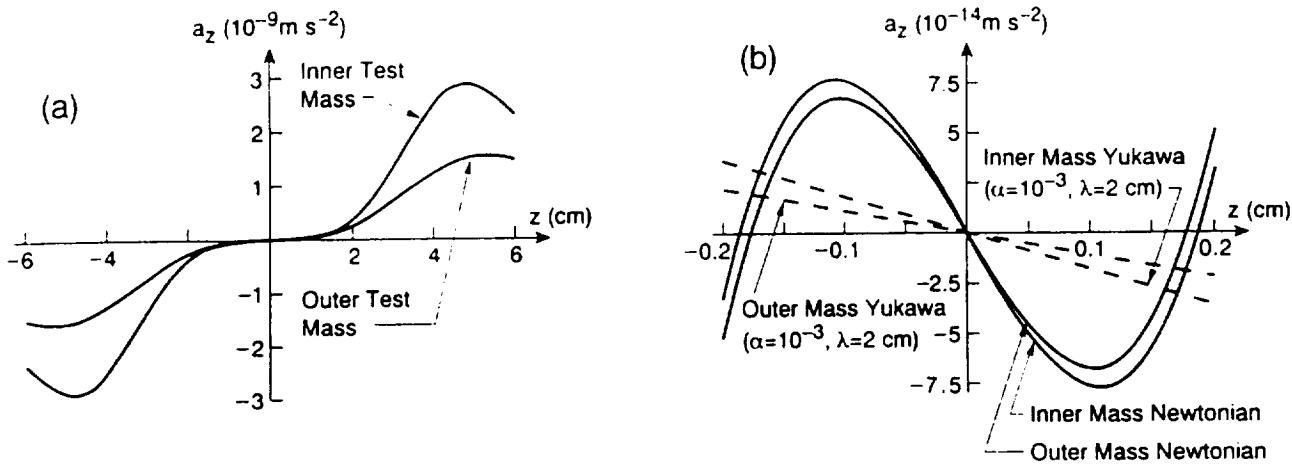


Fig. 5.3. (a) Newtonian accelerations on the test masses as a function of the source mass position. The peak values of these signals are measured and compared with the computed values to determine G . (b) Newtonian and Yukawa accelerations on the test masses near the centre as a function of the source mass position. The Newtonian signals are averaged out by modulating the source mass position with a proper amplitude. The Yukawa signals are detected at the modulation frequency.

will be locked down when the G/ISL experiments are turned off.

The cylindrical symmetry employed in the experiment reduces several important metrology errors. By constraining each source mass to move along the symmetry axis of the cylindrical test masses, the gravity error arising from the uncertainty in the radial position of a test mass relative to the source mass becomes second order, relaxing the radial centring requirement from $0.01 \mu\text{m}$ to $10 \mu\text{m}$. The geometry gives two axial positions, symmetric with respect to the centre, where the test mass experiences maximum gravitational attraction from the source (see Fig. 5.3(a)). At these positions, the error due to the uncertainty in the axial position of the source mass also becomes second order, reducing the axial positioning requirement also to $10 \mu\text{m}$. The source position is modulated between these points and the resulting acceleration is used as the G signal. The particular shape of each test mass enables a *near-null* test of the ISL. According to Newton's law, the gravitational field vanishes everywhere inside an infinitely long cylindrical shell. In the test mass, the rings on each end of the short cylinder correct for the missing mass and simulate the effect of an infinite shell near its centre, creating a small region where the Newtonian field almost vanishes (Fig. 5.3). This *near-null* geometry reduces the scale factor linearity requirement considerably for the ISL test.

5.3 Instrument noise and experimental resolutions

The intrinsic noise of a superconducting accelerometer deteriorates at low frequencies due to the $1/f$ noise of the SQUID (Eqs. 2.4 and 2.5, Fig. 2.7). On the other hand, the spacecraft acceleration noise improves as the frequency is lowered toward 10^{-4} Hz since the drag-free controller is optimised for the EP experiment (Fig. 2.10). Therefore, there is an optimum signal frequency one can choose for the G/ISL experiments, which is about 0.003 Hz . Substituting $T = 2 \text{ K}$, $Q = 10^6$, $f_C = 0.1 \text{ Hz}$, $f = 0.003 \text{ Hz}$, $\eta = 0.5$, and $m = 0.3 \text{ kg}$ into Eqs 2.4 and 2.5, we find an intrinsic acceleration noise level of $3 \times 10^{-14} \text{ m s}^{-2}/\sqrt{\text{Hz}}$ for the gradiometers. In order to take full advantage of the low noise level, we require that other disturbances be kept well below these levels. The disturbance control requirements will be discussed in Section 5.5.

The Newtonian accelerations that the near source mass exerts on the inner and outer test mass are plotted in Fig. 5.3(a) as a function of the source position. The far source mass exerts an acceleration about 10^3 times smaller, which can easily be corrected for. The maximum Newtonian accelerations occur at around ± 4.7 cm for the inner mass and ± 5.2 cm for the outer mass. The peak-to-peak values of these accelerations can be computed in terms of G from the known dimensions and densities of the source and test masses. These values will be equated with the measured peak-to-peak accelerations, which will then be solved for G. Clearly, the gradiometers must be calibrated *absolutely* to the same accuracy as required for G.

Since the two test masses in each gradiometer experience equal and opposite accelerations, the peak signals on the inner and outer gradiometers are 5.8×10^{-9} and $3.0 \times 10^{-9} \text{ m s}^{-2}$, respectively, twice the values given in Fig. 5.3(a). Since one is interested only in the peak-to-peak amplitudes, ideally one would like to apply a square-wave modulation to the source position. A practical substitute would be a truncated triangle-wave modulation with the corners smoothed. The modulation can be designed to have the source mass spend roughly half of the time while stationary at the two maximum points and the other half moving between them. It is then straightforward to show that the gradiometers can resolve the above signals to better than 3×10^{-7} in a total time of only 10^4 s.

Although the ISL test does not require an absolute calibration, the gradiometer resolution is more critical since the strengths of the Yukawa signals have been compromised in the interest of creating a near-null Newtonian region. Figure 5.3(b) expands the central region of Fig. 5.3(a), within ± 2 mm of the centre, and also shows the Yukawa components corresponding to $\alpha = 10^{-3}$ and $\lambda = 2$ cm. The test mass dimensions have been adjusted to make the Newtonian components oscillate about zero as the source is moved back and forth. Therefore, by modulating the source position between two carefully selected points, the Newtonian signals can be averaged out, while the Yukawa components are detected at the modulation frequency.

Figure 5.3(b) shows that, after averaging, the Newtonian signals must be smaller than 3×10^{-4} times their maximum values in order to resolve α to 10^{-6} . This defines the scale factor linearity requirements for the gradiometers and the source position detectors. The near-null nature of the experiment has reduced the requirements by two orders of magnitude.

The strength of ISL violation depends on λ . For the source-detector geometry chosen, the Yukawa force becomes a maximum when λ is approximately 2 cm. According to Fig. 5.3(b), the amplitudes of the differential Yukawa signals for $\lambda = 2$ cm are $7.0 \times 10^{-11}\alpha$ and $4.0 \times 10^{-11}\alpha \text{ m s}^{-2}$, respectively, for the inner and outer gradiometers. Assuming an integration time of 10^6 s, α can then be resolved to 5×10^{-7} and 7×10^{-7} , respectively, by the inner and outer gradiometers. For different values of λ , the resolution of α deteriorates.

The $\alpha-\lambda$ plot for the STEP ISL experiment is shown in Fig. 5.4, along with the existing limits. The orbit fundamental and second harmonic components of the Earth's gravity gradient, which are measured as part of the geodesy experiment, can be analysed as functions of altitude to obtain a test of the ISL at $\lambda \approx 300$ km. The expected resolution from this experiment is also plotted in Fig. 5.4.

The results obtained with the two gradiometers can be compared to check the composition dependence of the new force. Three scenarios are possible. (1) Both gradiometers obtain null results: This implies that the ISL is valid to 10^{-6} at 2 cm; therefore, there cannot be an EP violation to the same resolution in α at 2 cm. (2) Both gradiometers register positive results, but with the same value of α : This means that the ISL has been violated, but *without* composition dependence; therefore, the EP is still valid. (3) The two gradiometers register two different results (one could indicate null): This indicates that both the ISL and the EP have been violated. In this case, the possibility of a systematic error in one or both gradiometers must be examined more carefully.

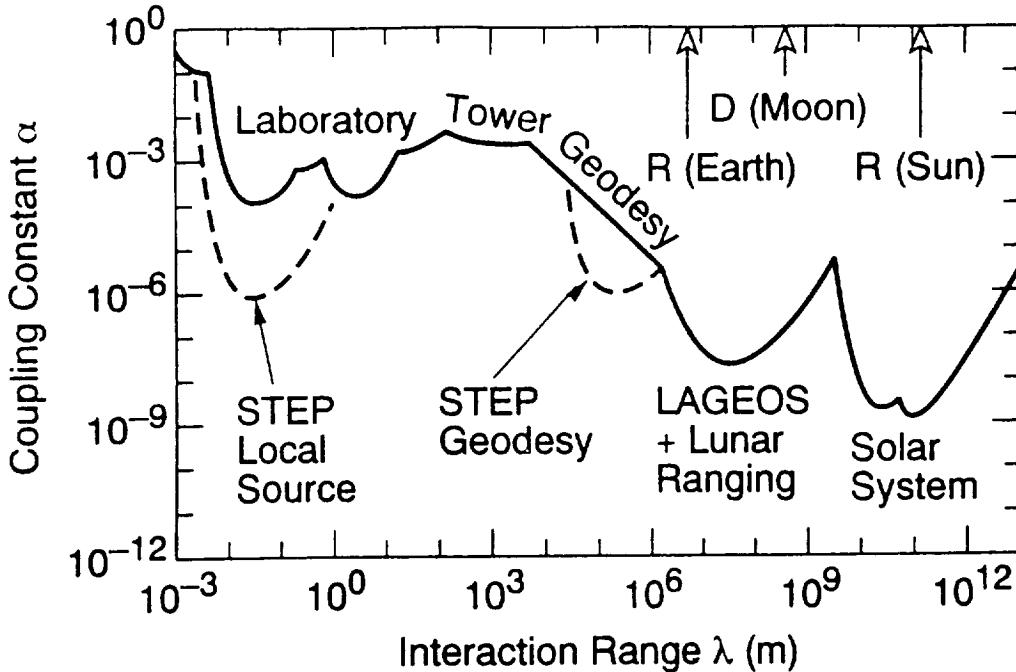


Fig. 5.4. Existing limits and resolutions expected from STEP for a violation of the inverse square law. The solid curve represents the existing limits in α (the parameter entering Eq. 1.1) as a function of λ , adapted from Adelberger et al. (1991). The dotted lines represent resolutions expected from STEP. The STEP G experiment is configured to improve α by two orders of magnitude in the cm range. As a by-product, the STEP geodesy experiment improves α by an order of magnitude around 300 km.

The G experiment provides a cross check for the composition dependence. In the absence of any composition dependence, the two G values obtained with the inner and outer gradiometers must agree within the experimental error, regardless of any violation of the ISL. On the other hand, a disagreement between the G values implies existence of a composition-dependent ISL violating force. The G and ISL results can be compared for consistency. The G experiment, therefore, constitutes a crucial cross check for the third scenario above. In return, the ISL test gives an essential systematic check for the G experiment. Clearly, the double gradiometer configuration is extremely useful.

5.4 Metrology and calibration requirements

Although our geometry reduces the errors in the relative positions of the masses to second order, the shapes and density homogeneities of the test and the source masses must be determined to 10^{-6} and 10^{-5} , respectively, for both the G and ISL experiments. This rather stringent metrology requirement comes from the proximity between the source and the test masses, which makes them couple strongly to higher moments of each other. By fabricating the test masses out of optically transparent single-crystal materials, the more stringent requirement is met. The source mass, being smaller, has a less stringent requirement, which could be satisfied with a standard material such as Pt-Ir.

The gradiometers must be calibrated *absolutely* to at least 1 part in 10^6 at the G signal level of approximately $5 \times 10^{-9} \text{ m s}^{-2}$, and their scale factors must be linear to 1 part in 10^4 at the ISL signal level of about $10^{-13} \text{ m s}^{-2}$. The geodesy signals cover the range between these levels and require a relative calibration to about 1 part in 10^2 over the acceleration range of five orders of magnitude. The STEP gradiometers are expected to satisfy the linearity requirement for the ISL experiment.

The relative calibration for geodesy can be performed in orbit by using the acceleration signals from the source masses measured over the entire range of motion. Thus the G and the geodesy experiments complement each other.

Various options have been investigated for the absolute calibration of the gradiometers. The best option appears to be the use of the Earth's gravity gradient as the calibration signal. The STEP spacecraft will be in a near circular polar orbit at a nominal altitude of 550 km, with residual eccentricity of about 10^{-3} . The sensitive axes of the gradiometers are aligned perpendicular to the orbit plane. The resulting variation in the radial position of the spacecraft, $\Delta R \sim \pm 7$ km, will generate an orbit frequency component of the Earth's gradient,

$$\Delta\Gamma_E = (GM_E/R^3)(3\Delta R/R) \sim \pm 4 \text{ E} \quad (5.1)$$

where $1 \text{ E} = 10^{-9} \text{ ms}^{-2}/\text{m}$ is the unit of gravity gradient. The Earth's quadrupole moment will contribute a second harmonic component of comparable amplitude. These acceleration levels coincide with the levels of the G signals and eliminate calibration errors that could result from nonlinearity of the scale factors. The calibration can be performed in less than 10^5 s.

This calibration requires knowledge of the gradiometer baselines to $0.5 \mu\text{m}$. This can be done by locating the test masses with respect to two reference surfaces in the quartz block whose distance is accurately determined during assembly by optical means.

Figure 5.5 shows the Fourier spectrum of the Earth's gravity signal that the STEP gradiometers would resolve in one day and the uncertainties expected from gravity model errors. The STEP orbit and the gradient signal have been simulated by using the gravity model GEM-T2. The signal is contaminated by the errors in the higher harmonic coefficients of the Earth's gravity. As the Earth rotates under the orbit, the spacecraft flies over different parts of the Earth from orbit to orbit, with the orbit and gradient signal being affected by the spatial dependence of the tesseral harmonics. The upper uncertainty curve in Fig. 5.5 represents a limit from the existing gravity models. Fortunately, the gravity model will be improved by STEP by more than two orders of magnitude. The lower uncertainty curve represents the limit expected from the STEP geodesy experiment. Calibration to 1 part in 10^6 appears feasible. This could be improved further by choosing an orbit with a slightly larger eccentricity.

5.5 Disturbance control requirements

Major disturbances that affect all STEP accelerometers have been summarised in Section 2.5. The requirement for magnetic shielding is discussed in Section 5.1. Here we discuss only the disturbances that could have more serious effects on the G/ISL experiments.

The linear and angular motions of the spacecraft couple to the gradiometers through misalignment of the sensitive axes. We set the alignment requirements for the parallelism and concentricity of the sensitive axes at 5×10^{-5} rad and 5×10^{-5} of the baseline length, the same as for the EP accelerometers. The spacecraft residual acceleration and jitter levels that produce errors equivalent to the intrinsic noise of the instrument are then $3 \times 10^{-10} \text{ m s}^{-2}/\sqrt{\text{Hz}}$ and $4 \times 10^{-10} \text{ rad s}^{-2}/\sqrt{\text{Hz}}$, respectively. Figure 2.10 shows that the actual residual acceleration levels of the spacecraft at 0.003 Hz are about $10^{-10} \text{ m s}^{-2}/\sqrt{\text{Hz}}$ and $2 \times 10^{-10} \text{ rad s}^{-2}/\sqrt{\text{Hz}}$, respectively. Thus the basic drag-free control satisfies the G/ISL requirements.

Angular motion of the spacecraft can produce significant errors through two additional mechanisms: modulation of the Earth's gravity gradient and centrifugal acceleration. Fortunately, both of these become second-order effects due to the orientation chosen for the STEP gradiometers, the inertial orientation with their sensitive axes horizontal. This greatly reduces the pointing and pointing stability requirements for the spacecraft. The absolute pointing requirement of the gradiometer

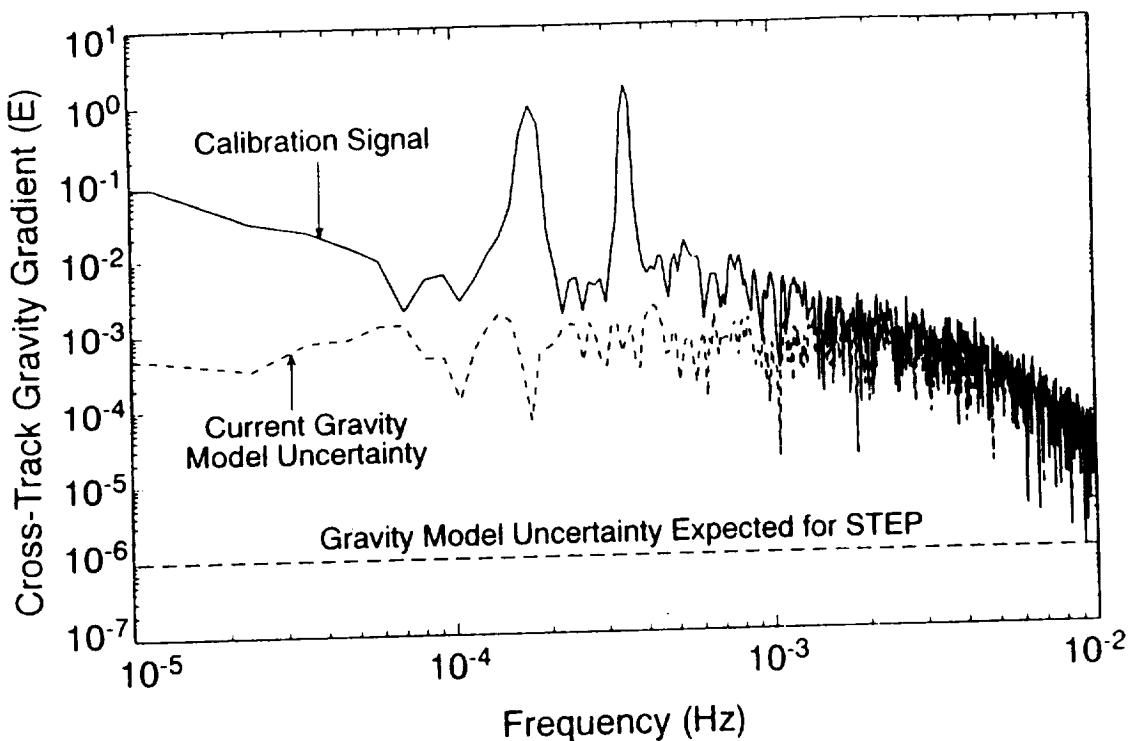


Fig. 5.5. Spectra of the calibration signal and uncertainties for the G experiment. The upper curve is the signal spectrum from one day of measurement simulated by using the gravity model GEM-T2 to predict the STEP orbit and the gradient signal. The signal was attenuated slightly by the Hanning window which was employed to accommodate the effect of finite data record length. The middle curve is the uncertainty from the existing gravity models, obtained by differencing the signals simulated using OSU91A and GEM-T2. The lower curve represents the true calibration limit. This curve was generated by covariance propagation using the predicted variances of the STEP GPS/gradiometer improvement to the gravity field.

axes, which arises from the need for calibration, now becomes 5×10^{-4} rad from the orbit normal. This can be satisfied easily by using the star trackers on board the spacecraft. The attitude and attitude rate stability requirements for the G/ISL experiments become 10^{-3} rad/ $\sqrt{\text{Hz}}$ and 10^{-6} rad $\text{s}^{-1}/\sqrt{\text{Hz}}$, respectively, at half the signal frequency. These are satisfied by the attitude control of the spacecraft with a margin of at least a factor of 100. The geodesy experiment also benefits from this ideal orientation of the gradiometers.

The energetic charged particles from the South Atlantic Anomaly, which is one of the most important error sources for the EP test, cause significantly less disturbance to the G/ISL experiments. This is because the test mass positioning requirement for the gradiometers is four orders of magnitude less stringent than for the EP accelerometers, 10^{-6} m instead of 10^{-10} m and the signal frequency for the G/ISL experiments can be tuned away from the orbit harmonics where the disturbance occurs. The calibration must be done at the orbit fundamental or second harmonic but the required acceleration resolution is 100 times less stringent than in the EP experiment. Charge control circuits will be provided in the gradiometers, but with much relaxed requirements.

The helium tide, however, is a potentially critical error source for the G/ISL experiments. This is especially true because the gradiometer test masses cannot be configured to reduce the higher-order coupling to local gravity field, as in the EP experiment. The main tidal bulge of helium, due to the gravity gradient of the Earth, rotates around the symmetry axis of the dewar and

does not couple to the gradiometers. One is concerned about a possible *horizontal* component of the helium tide at the orbit fundamental and second harmonic, which can contaminate the absolute calibration of the gradiometers. However, the gravity gradient alone cannot deform the helium at the fundamental frequency because the once-per-orbit gradient term arises from the altitude modulation of the monopole term of the Earth's gravity for which the vertical and horizontal components are always proportional to each other, by the ratio of 2:1. This is not exactly true for the twice-per-orbit term which arises from the oblateness of the Earth. The surface tension complicates the analysis. In the worst case, the helium tide could produce an error about one order of magnitude greater than allowed for the 10^{-6} calibration in the G experiment. The source masses also pull the helium around as they move. Fortunately, this effect is negligible due to the smallness of the masses and the symmetry in their motion. So the ISL experiment appears to be safe.

As described earlier, in order to solve the helium tide problem in the EP experiment, the helium tank will be divided into two compartments, which will keep the liquid/gas interface at least 40 cm away from the accelerometers (see Fig. 3.7). This alone may not be enough for the G experiment. However, there is a simple solution: conduct the calibration and the G experiment half way into the mission, at the time when the inner helium tank is full and the outer tank is empty. This condition can be well satisfied since the whole G experiment can be completed in less than a few days.

6. Geodesy

In August 1969, under the guidance of William M. Kaula, leading scientists and NASA management personnel met at Williamstown, Mass. to define solid-Earth and ocean physics application of space and astronomic techniques. The recommendations of this workshop became widely known in Earth sciences as the Williamstown Report. All main objectives of the programme defined then are met by now, with one important exception: the resolution of the spatial variations of the gravity field first to 250 km half wavelength, and ultimately to 100 km half wavelength by a dedicated low-flying gravity field mission, for the purpose of improving geotectonic analysis and for a determination of the general circulation of the oceans.

Over the years, oceanographers, solid-Earth physicists and geodesists joined efforts in Europe and the USA to establish such a mission. Many concepts were put forward, SLALOM and GRAVSAT in the seventies, GRM and ARISTOTELES in the eighties and early nineties, and recently BRIDGE and GAMES. All of these efforts focus on one and the same goal: the realisation of an improved gravity field, desperately needed in Earth sciences. Each of these concepts would serve this purpose, some slightly better than STEP, like ARISTOTELES or GRM, some on the same precision level, like GAMES. The problem with all of them is that none has been realised so far. STEP is a mission in the field of fundamental physics. Not a single important design element is determined by the geodesy experiment alone. Despite the fact that this mission design is driven by non-geodetic factors, STEP would produce an excellent high resolution gravity field model. STEP would be the first gravity gradiometer in space and it would serve a wide range of science objectives in geodesy, solid-Earth physics, oceanography and climate research.

6.1 STEP and Geodesy

STEP is equipped with a very accurate non-conservative force compensation system. Thus, once in orbit, the spacecraft itself can be considered a test mass in almost perfect free-fall, its trajectory determined solely by the Earth's gravity field. With an orbit rather low and almost polar and with the Earth revolving under the slowly precessing orbit plane a global map of the spatial structure of the Earth's gravity field would be derived from the orbit perturbations. A prerequisite is that the trajectory of the spacecraft can be determined accurately and over most of the mission period. For this purpose, a receiver would track in various combinations at least four of the 24 satellites of the Global Positioning System (GPS). Applying, in addition to pseudo-ranging, differential carrier phase measurements relative to a network of ground stations, orbit reconstitution at centimetre level is feasible. We denote this part the Geodesy free-fall (FF) experiment. Its principle is shown in Fig. 6.1.

The FF-experiment, because of the continuous and three-dimensional tracking of a low flying drag-free satellite, would mark the starting point of a new era of gravity field modeling. In principle, however, this concept could be applied to any satellite of this type. STEP has more to offer to geodesy.

Galileo's original Equivalence Principle experiment suffered from three shortcomings: air resistance, limited duration, and the effect of the unknown gradient of gravity between the two falling test masses. Putting the experiment into space and equipping the spacecraft with drag compensation eliminates the first two shortcomings. The gradient effect is removed by using two concentric test masses, hence each EP differential accelerometer can not be used to measure gradients. However, since there are six Equivalence Principle accelerometers at different locations, the acceleration difference of any two of them taken over the known baseline length yields a very accurate gravity

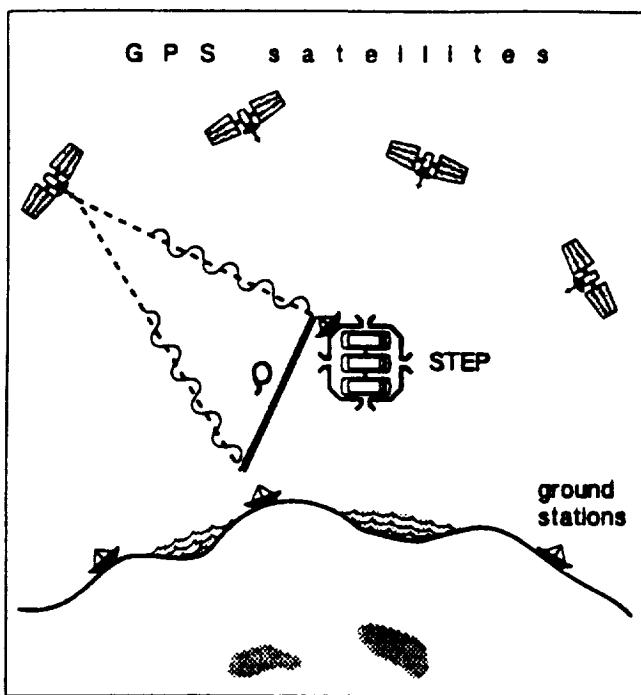


Fig. 6.1. Principle of the Geodesy free-fall experiment. GPS carrier phase measurements onboard STEP relative to a network of ground stations permit orbit determination with centimetre precision.

gradient reading. The pair of accelerometers of the G-experiment comprise an even better gradiometer because of the relatively long baseline, and because the gradient direction is perpendicular to the orbit plane and therefore fixed in space (see Fig. 2.3). We shall refer to this gradiometer as the Geodesy Gravity Gradient (GG) experiment.

The FF and GG experiments are highly complementary. While from the GPS measurements a large-scale map of the Earth's gravity field can be recovered, gravity gradiometry is particularly sensitive to more detailed field features.

6.2 Status of gravity modeling

Gravity is the combined effect of the Earth's gravitational attraction and the centrifugal acceleration. The latter is caused by the Earth's spin rate, its maximum value, at the equator, being only 0.5% of g. The gravity field is accessed in practice either by reference to its equipotential surfaces, in particular the one at mean sea level, the geoid (which corresponds to the ocean surface at rest), or in terms of gravity anomalies (the difference between the actual gravity and the corresponding normal gravity of a best-fitting ellipsoidal reference body). Gravity anomalies exhibit considerable correlation with terrain variations and shallow geophysical features with maximum values up to 10^{-3} of g. The deviations of the geoid from the reference ellipsoid on the other hand (geoid undulations) are rather smooth ranging from -100 m to +80 m.

Present day gravity models are based on three data sources:

- Terrestrial (and shipborne) gravity anomalies: They are determined from gravity (and height) surveys and represented, according to international standards, as mean values over certain equi-angular (e.g. $1^\circ \times 1^\circ$ or $30' \times 30'$) blocks. Even after more than 50 years of worldwide effort, coverage and accuracy are satisfactory only for certain parts of our planet (see Fig. 6.2,

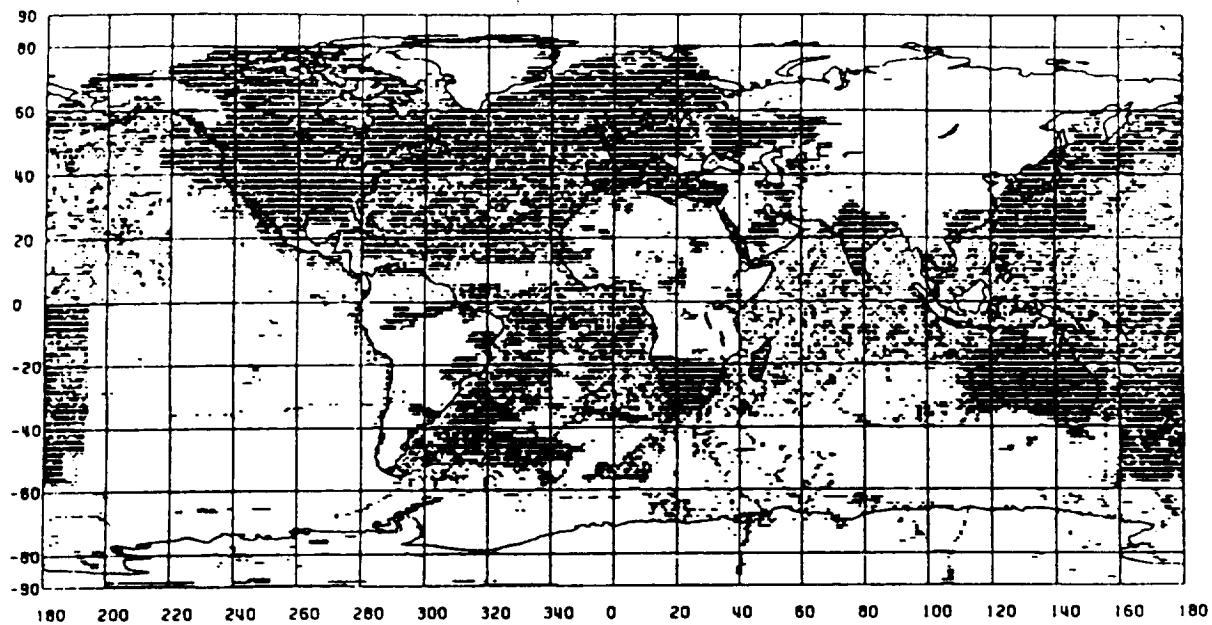


Fig. 6.2. Distribution of $1^\circ \times 1^\circ$ mean gravity anomalies with an accuracy below 10 mGal (1 mGal = 10^{-5} m s^{-2}). This map shows that the gravity data over large parts of Africa, South America, Asia and the polar regions are insufficient and that there is hardly any data over most of the oceans.

from Pavlis, 1988). This type of data contains medium and high frequency information.

- **Satellite altimetry:** After a series of altimeter missions for all (but the polar) ocean areas, precise sea surface heights are available from satellite radar altimetry. Disregarding the topography of the sea surface, i.e. the deviation of the actual ocean surface from the geoid (order of magnitude $\pm 1 \text{ m}$), this data can be taken as moderately accurate geoid information. This data is of considerable value for geophysical investigations but hardly of use to oceanography. Altimetry provides information in the medium range of frequencies.
- **Satellite orbit analysis:** For more than three decades, several institutions have regularly determined geopotential models. They are derived from the combined analysis of orbits of a large number of mostly non-geodetic satellites with different orbit elements and based on a variety of tracking data types. These models are presented as coefficients \bar{C}_{lm} and \bar{S}_{lm} of a spherical harmonic expansion of the field and they provide information on the long wavelength part of the spectrum. A representative example of the best currently available geopotential models, based purely on satellite orbit analysis (no altimetry, no terrestrial surface gravity), is the Joint Gravity Model JGM-1S, computed jointly by NASA Goddard Space Flight centre and the Space Research centre of the University of Texas (Lerch et al., 1992). . It is complete to degree $l_{\max} = 60$ and order $m = 60$. This corresponds to a spatial half wavelength of $L = \lambda/2 = 330 \text{ km}$. A stabilisation technique had to be employed to attain a solution for the complete set of coefficients. Actually, for some groups of coefficients of this model, the error estimates approach 100% of the expected size of the terms, particularly above degree 36 (or for half wavelengths $L \leq 560 \text{ km}$).

Resolution, in spectral terms, is expressed by the highest degree of the set l_{\max} and can be

translated into a corresponding spatial resolution on the Earth's surface by

$$L[\text{km}] = \frac{20000 \text{ [km]}}{l_{\max}} \quad (6.1)$$

The present state-of-the-art is best understood by comparing the signal spectrum of the Earth's gravity field with the uncertainty in the estimated coefficients. The spectrum of the signal is represented in terms of the root mean square (rms) or average value per degree and order:

$$c_{lm} = \sqrt{\frac{1}{2l+1} \sum_{m=0}^l \{ \bar{C}_{lm}^2 + \bar{S}_{lm}^2 \}} \quad (6.2)$$

It could be expressed by the coefficients of one of the so-called combined geopotential models (Rapp et al., 1991), see below. But it suffices to take the rule of thumb of Kaula (see the signal-line of Fig. 6.4). The corresponding error spectrum is derived with the same formula, by replacing the coefficients \bar{C}_{lm} and \bar{S}_{lm} by their corresponding error standard deviations. In Fig. 6.4 the error coefficients of JGM-1S are displayed, too. One sees that, between degree 30 and 50, the uncertainty level reaches 100%.

There also exist geopotential models combining the above three data sources. The latest and most advanced one is JGM-1 (Lerch et al., 1992). This model represents the optimal transformation of satellite orbit analysis, altimetry and terrestrial gravimetry into one set of spherical harmonic coefficients. However, even such a field is weak in areas with poor data coverage and biased in ocean areas, because the employed altimetry does not fully distinguish between actual ocean surface and geoid. For more explanations, see Balmino (1986).

6.3 Geodesy free-fall experiment

The FF experiment relies on the accurate, continuous, and uniform tracking of STEP from the GPS constellation. STEP's very accurate drag-free system ensures that its orbit is governed entirely (to a very high degree of approximation) by the gravitational field of the Earth (and Sun, Moon and planets). The low altitude of only 550 km makes STEP rather sensitive to the detailed structure of the Earth's field. Due to the almost polar orbit ($I = 97^\circ$) nearly complete coverage of the Earth with ground tracks will be attained within six months. A rough rule of the spatial resolution attainable from such an orbit is given by the empirical relationship between resolution (L) and orbital altitude (H): $L \approx (1/2)H$; for STEP, $L \approx 300 \text{ km}$.

6.3.1 GPS receivers

STEP will be equipped with a GPS receiver and two antennas placed in such a way that at least four Global Positioning System (GPS) satellites will be visible at any time. GPS is a military navigation system maintained by the US Department of Defense (DoD), comprising 24 (21 + 3 spare) satellites. The satellites are at an altitude of 20,240 km, the orbit period is 12 hours, and their inclination 55° . The satellites are evenly distributed over six orbital planes. They transmit three different kinds of pseudo-random noise (PRN) codes at two carrier frequencies ($L1 = 1.57542 \text{ GHz}$ and $L2 = 1.22760 \text{ GHz}$). These are the C/A code, the P-code and the Y-code. The PRN codes allow determination of the absolute position of the receiver antenna by pseudo-ranging. Pseudo-ranging means that the distance between GPS satellite and receiver can only be determined up to a constant, the unknown receiver clock bias. The latter is computed by observing more than the strictly necessary number of three simultaneous ranges. Range changes can be tracked with mm-precision from continuous carrier phase observations.

To ensure the integrity of the system, the DoD applies two levels of encryption to the information/signals transmitted: Selective Availability (SA) and Anti-Spoofing (AS). SA degrades the GPS-broadcast ephemeris (irrelevant for STEP) and dithers the frequency of the signal (accounted for in post-processing). AS is the mixing of the precise code with a secret code, not available to the general user, to produce the Y-code. For the AS encryption, the user has four choices: (1) use a standard type receiver, thereby incurring a significant loss of accuracy if or when AS is activated; (2) use a Y-code receiver, preserving the full accuracy but requiring classified hardware and software; (3) use a codeless receiver, which requires no classified information but still recovers dual-band ionospherically corrected high precision observables with a minor penalty in accuracy; (4) arrange for an agreement with the US Department of Defense to disable AS during the 6-month STEP mission. AS is a serious issue to consider in selecting the onboard receiver type. Option (1) is not practical for STEP, which requires high accuracy; option (2) may lower the cost for the flight receiver itself, although operation costs will be higher and it may be awkward for an international mission such as STEP; option (3) is ideal provided that the cost of such an instrument is not extremely high; and option (4) is also a possibility, since TOPEX/POSEIDON were able to obtain such an agreement.

The GPS flight antennas will each provide approximately a hemispherical field of view. They should be arranged as shown in Fig. 6.3, so that at no time does an antenna point directly at the Earth. Each antenna should be directed with maximum gain perpendicular to the line connecting the satellite with the Earth's centre. In Fig. 6.3 they are pointing in the cross-track direction (towards and away from the Sun).

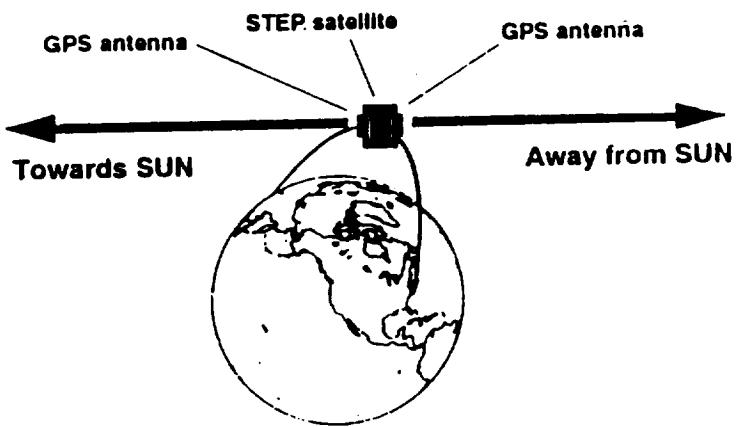


Fig. 6.3. STEP GPS antenna configuration with two oppositely directed antennas.

Simulations carried out by JPL show that carrier phase multipath can reach 5 mm with the patch and other antennas. Therefore, the multipath portion of the data error should be reduced by a factor of 10 by: (1) performing phase measurements and calibrations with the antennas on the satellite; and (2) creating software to perform detailed multipath calculations and modeling for the specific antenna-spacecraft configuration of STEP. (1) and (2) should be consistent and intercompared before launch. Other antenna designs should be considered in addition to the patch antenna. The geometric position of each antenna relative to a fixed reference point on STEP would be determined prior to launch. The centre of mass offset from that reference point will be monitored and determined from telemetry information throughout the mission. The orientation in space at any instant is derived from attitude information. This configuration will comprise either two GPS receiver units or a larger capacity single receiver. Nominally, each receiver unit will be $15 \times 12 \times 7 \text{ cm}^3$

or less, weighing less than 2 kg and requiring less than 7 W. It is expected that each will have 6 dual-band channels for each of the two antennas (the minimum required per antenna is 4 channels).

The collected data will be distributed as ionospherically corrected combinations of L1 and L2 carrier phase and pseudo-range. The measurement rate shall be 15 s. In the case of a code-tracking receiver, the ionospherically corrected pseudo-range rms error will be ≤ 15 cm and that of carrier phases ≤ 0.05 cm (receiver noise only, 15 s data rate), with systematic errors of ≤ 1 cm and ≤ 0.01 cm, respectively. Corresponding values for a codeless receiver are ≤ 100 cm for pseudo-range, ≤ 0.2 cm for carrier phase, with systematic errors ≤ 10 cm and ≤ 0.1 cm, respectively.

The leading candidate Y-code receiver at present is the Next Generation Monarch receiver from Motorola. For the codeless type, the leading candidate is a flight version of the TurboRogue receiver, a high-performance ground receiver. Apart from the basic difference between these two types, there are other factors. However, considering the fast pace of evolution in these technologies, they can be left for later consideration, closer to the time of taking a final decision.

6.3.2 Gravity field recovery

The global gravity field recovery by GPS is based primarily on the combined carrier phase measurements from STEP and a network of ground receivers. The principle of the FF experiment is shown in Fig. 6.1. For this purpose at least 12 ground sites are required, globally distributed and with the station coordinates known with cm-precision in one geocentric coordinate system. These sites should be equipped with high-performance dual-band GPS receivers with similar or superior capabilities to that on STEP (i.e. > 6 -channels). At present, the International Earth Rotation Service (IERS), in cooperation with several national agencies and institutions (including ESA and NASA), is in the process of establishing a global GPS tracking network with continuous automated operations. The International GPS Service (IGS) will operate high-performance dual-band receivers and will be responsible for collecting, archiving, and distributing data from this network. The data would be processed at one or more analysis centres and precise GPS orbits will be available within a few days from data collection. For the STEP GPS flight data, we estimate that the data rate (downlink) would be about 100 bps for carrier phase data every 15 seconds (Crow, 1992).

On the basis of the above described scenario a gravity field recovery covariance analysis has been performed. The gravity field is represented by a series of spherical harmonics. The method is described in Schrama (1992). The analysis consists of a least squares error propagation with 80^2 unknown spherical harmonic coefficients. We assumed an almost circular orbit at 550 km altitude with an inclination of 97° . The ground track pattern is repeated every 12 days, resulting in 15 repeat cycles in 6 months. Furthermore, it is assumed that ionospherically corrected pseudo-range and range rate measurements are available from GPS with a rate of 15 s resulting in a 3 cm-accurate (uncorrelated noise) STEP 3-D position. The noise value is on the conservative side so as to leave margin for systematic error sources, such as ground site coordinate inconsistencies, unmodeled multipath errors or residual ionospheric and tropospheric effects. The outcome was verified by independent studies such as the end-to-end simulations of the FF experiment described in Pavlis (1992) and in Muellerschoen et al. (1993).

The rms errors per coefficient derived from the simulation are included in Fig. 6.4. We see that, from STEP GPS alone, currently available geopotential models like JGM-1S can be improved both in terms of accuracy and resolution. Between spherical harmonic degrees 5 and 30 ($L = 4000$ km to 700 km), the improvement is about one order of magnitude, the resolution being around degree 40 ($L = 500$ km).

6.3.3 Alternative tracking concepts

Satellite Laser Ranging (SLR) is recommended as a supplementary tracking system for STEP. The space segment consists of a set of passive corner-cube reflectors. While SLR cannot achieve what continuous GPS tracking will do, the combination of GPS and SLR is very effective. Differential GPS can provide global, three-dimensional coverage with centimetre precision, while laser measurements can remove ambiguities and convert that precision into accuracy. Short-arc orbit solutions with a small cluster of SLR sites were recently done for TOPEX/POSEIDON and proved very useful. In addition, in case of a GPS failure, a precise backup tracking system exists, which is not qualified for an FF-type experiment, but sufficient to support the EP, G, and gradiometry experiments.

Besides GPS and SLR, the microwave tracking systems DORIS and PRARE(E) were considered. Both systems are attractive, but cannot offer what GPS does: uninterrupted tracking in three dimensions. The benefit drawn from the inclusion of either of these systems needs to be studied further. In Table 6.1 a very schematic list of pros and cons of the four considered tracking concepts is given.

Table 6.1. Comparison of GPS, DORIS, PRARE(E), and SLR.

GPS:

- + operational (TOPEX/POSEIDON), Explorer Platform (EP)
- + two frequencies (carrier phase) [1200/1600 MHz]
- + global, continuous, and three-dimensional
- military system
- multi-path

DORIS:

- + operational (SPOT-2, TOPEX/POSEIDON)
- + two frequencies [401/2036 MHz]
- coverage ($\approx 50\%$)
- one receiver at a time
- only one-dimensional
- coordination ground stations

PRARE(E):

- + range and range rate
- + two frequencies [2200/8500 MHz]
- + 4 receivers simultaneously
- coverage ($\approx 50\%$)
- coordination ground stations
- not yet space proven

SLR:

- + very precise
- + inexpensive and passive space segment
- + almost no refraction
- + very precise reference frame (link to LAGEOS)
- poor and uneven coverage ($\approx 10\%$)
- weather dependent

6.4 Geodesy gravity gradiometry experiment

Gradiometry is the measurement of the second derivatives of the gravitational potential (or gradients of gravity):

$$V_{ij} \equiv \frac{\partial^2 V}{\partial x^i \partial x^j} \quad (6.3)$$

The full gravity tensor V_{ij} consists of nine components. As the field is conservative ($\vec{\nabla} \times V = 0$) and harmonic ($\nabla^2 V = 0$ outside the Earth), only five are independent:

$$V_{ij} = \begin{pmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{xy} & V_{yy} & V_{yz} \\ V_{xz} & V_{yz} & V_{zz} = -(V_{xx} + V_{yy}) \end{pmatrix} \quad (6.4)$$

Gravity gradients are expressed in Eötvös units ($1 \text{ E} \equiv 10^{-9} \text{ s}^{-2}$). The largest component is the vertical gravity gradient, being about 3000 E (gravity changes by $0.3 \times 10^{-6} \text{ m s}^{-2}$ per metre). The horizontal diagonal terms are approximately half this size and negative, off-diagonal components are below 100 E. See Rummel (1980) for more about gradiometry.

Gravity gradients are highly sensitive to the local features of the field in the proximity of the measurement location. This is why in the past terrestrial gradiometry has been applied in exploration geophysics, using torsion balances. For the same reason a spaceborne gradiometer of a given accuracy will result in a more successful mission if it is in as low an orbital altitude as possible (gravity signal attenuates exponentially). Tests with airborne gradiometry were carried out some years ago with modest success (Jekeli, 1980). In space the high sensitivity of a gradiometer for spatial details of the Earth's gravity field should make it ideally suitable for gravity field refinement. No actual experiment has been carried out so far.

In the case of STEP gradient components could in principle be derived from the EP or the G experiment. A gravity gradiometer component is derived from differencing the readings of two accelerometers over the distance of their baseline. If the two accelerometer sensitive axes and the baseline axis are aligned, a diagonal component is derived. If the accelerometer axes are perpendicular to the baseline, an off-diagonal component is obtained. In the STEP situation, the EP accelerometers allow the formation of off-diagonal components only, which are highly sensitive to uncontrolled angular motion. A superior choice is the use of the two pairs of accelerometers of the G experiment (see Fig. 2.9). From them, a very accurate "out-of-plane" (cross-track) diagonal gravity gradient can be determined. It is ideal because (1) the baseline is long (75 cm), (2) two simultaneous independent observations can be carried out with the inner pair of accelerometers and with the outer pair, and (3) the out-of-plane component is affected by angular effects only to second order.

The G experiment with its two accelerometers benefits from the ideal orientation of the gradiometers, as explained in Section 5.4. The inertially fixed, horizontal orientation makes both the angular modulation of the Earth's gravity and the centrifugal acceleration second order errors, thus minimising the pointing requirements. This leaves the coupling to the angular acceleration through a misalignment of the sensitive axes as the dominant angular error source.

The intrinsic noise of the G gradiometer is better than $10^{-4} \text{ E}/\sqrt{\text{Hz}}$ for frequencies above $2 \times 10^{-4} \text{ Hz}$. Below this frequency, the $1/f$ power noise of the SQUID amplifier dominates. The drag-free and attitude control of the spacecraft, combined with the passive common-mode rejection ratio of 10^{-4} , keep the common-mode acceleration error below the intrinsic noise level of the instrument for frequencies below $5 \times 10^{-3} \text{ Hz}$. As the signal frequency increases, the drag-free control attenuation deteriorates (see Fig. 2.10). As a result, the error from the residual acceleration of the spacecraft can

reach a level 40 times the intrinsic at the highest geodesy signal frequency 0.02 Hz. The common-mode rejection can be improved to 10^{-6} by measuring the error coefficients and compensating for the linear and angular accelerations, as has been demonstrated with a laboratory superconducting gravity gradiometer (Paik et al., 1992). This active common-mode rejection will be applied to the STEP gradiometers to achieve a gradient sensitivity of $10^{-4} \text{ E}/\sqrt{\text{Hz}}$ over the entire geodesy frequency bandwidth from 2×10^{-4} to 2×10^{-2} Hz.

Other potential error sources, such as time-variable self-gravitation (helium tide and helium consumption) can be kept below this level. The gradiometer will be calibrated over its entire measurement range by means of the G experiment. This is discussed in Section 5.3.

The error propagation simulation can now be repeated, employing the same parameters for orbit and mission duration as for the FF experiment. The simulation has been carried out for the one-component out-of-plane gradiometer alone and for the combination of gradiometry and GPS. The assumed mission and instrument parameters were: mission duration 6 months, almost circular orbit at 550 km altitude with an inclination of 97°; a gradiometer observing the cross-track diagonal term with $10^{-4} \text{ E}/\sqrt{\text{Hz}}$ at a sample period of 1 s and a GPS range rate precision of 3 cm (uncorrelated noise) at 15 s sample interval.

The outcome of the simulation is summarised in Fig. 6.4 (upper panel). It shows the expected average size of dimensionless spherical harmonic coefficient of degree l of the Earth's gravity field and four types of error curves. The gravity signal spectrum is expressed by Kaula's rule of thumb. The error spectra are those of the JGM-1S geopotential model, the noise level reaching that of the signal between degrees 20 and 40 (= resolution), that of the STEP GPS measurements alone (resolution ≈ 40 to 60), of the combination of STEP gradiometry and GPS (resolution ≈ 150), and of ARISTOTELES (resolution ≈ 240). The JGM-1S error spectrum is representative for the current state-of-the-art of satellite gravity field modeling. STEP-GPS is representative for any low-flying drag-free satellite equipped with GPS; GP-B would be another example. The error curve of GAMES is very much comparable with the "STEP (GPS + Gradient)" curve. ARISTOTELES represents the ultimate goal of a dedicated gravity field experiment.

One can see that the "STEP (GPS + Gradient)" curve attains a resolution of almost degree 150 which corresponds to 130 km half wavelength on the Earth's surface. Up to degree and order 80, our current knowledge of the Earth's gravity field can be improved by at least one order of magnitude. The accumulated geoid error at $l_{\max} = 140$ shall be about 10 cm versus 1 m today, the gravity anomaly error $2 \times 10^{-5} \text{ m s}^{-2}$.

On theoretical grounds the gradiometry and GPS parts are expected to be complementary, the strength of the GPS part being the long wavelength range of the gravity spectrum, that of gradiometry being the medium wavelengths. The simulations confirmed this rule, although the STEP gradiometer is so extremely accurate that it helps to improve the long wavelength part of the spectrum. This is also the reason why it dominates ARISTOTELES at degrees below 40, despite the much lower altitude of 200 km of the latter and the fact that it is assumed to measure two gradient components. The ARISTOTELES gradiometer is assumed to operate at room temperature with $10^{-2} \text{ E}/\sqrt{\text{Hz}}$ only. Note that most of the geoid signal is contained below degree 30! The extremely high precision of the STEP gradiometer makes it also competitive with the gravity mission concept GAMES, which would be based upon laser satellite-to-satellite range rate tracking measurements (0.1 mm/s) between two low flying satellites (below 300 km), also equipped with GPS.

All GPS and gradiometry simulations were based on the principle of least squares error propagation. In order to stabilise the numerical solution, the expected gravity signal size was incorporated as prior knowledge, so as to allow the noise nowhere to exceed the gravity signal. This is a generally adopted procedure in gravity field modeling. It is the reason why all error spectra converge towards the gravity signal curve.

Also, variations in measurement precision (10^{-3} vs. $10^{-4} \text{ E}/\sqrt{\text{Hz}}$) and altitude (350 km, 450 km, 550 km) were investigated, with all other mission parameters left unchanged. The results are summarised in Fig. 6.4 (lower panel). The expected average size of the spherical harmonic coefficients of the Earth's gravity field is compared with error spectra. All error spectra refer to a STEP combined gradiometry and GPS geodesy experiment. The mission parameters are those of Fig. 6.4 (upper panel). Shown are the error curves at 550 km altitude for a $10^{-4} \text{ E}/\sqrt{\text{Hz}}$ and a $10^{-3} \text{ E}/\sqrt{\text{Hz}}$ gradiometer (plus GPS). For a $10^{-4} \text{ E}/\sqrt{\text{Hz}}$ gradiometer (plus GPS), the error spectra for altitudes 350 km, 450 km, and 550 km are displayed too. The inclusion of the expected average signal spectrum of sea surface topography should provide insight into the contribution of STEP GPS+gradiometry for ocean circulation determination.

Two conclusions can be drawn. A $10^{-4} \text{ E}/\sqrt{\text{Hz}}$ gradiometer (with 1 s sample rate) is highly desirable, the improvement above a 10^{-3} E instrument being very significant both in terms of resolution and accuracy. In the case of a 10^{-3} E instrument, the strength of GPS in the long wavelength part of the spectrum becomes visible. The geodesy experiment would benefit from any decrease in mission altitude. At 350 km altitude, STEP would make ARISTOTELES obsolete. At this point, some margin for variation in measurement precision and optimal sample rate must be left open for further study.

6.5 Applications in geodesy, solid-Earth physics, and oceanography

STEP would improve our current knowledge of the Earth's gravity field in the long and medium wavelength range by an order of magnitude. New, more detailed, geoid features would become visible and uncertainty about their size, location and gradients shall decrease significantly.

For geodesy the implications would be rather direct and practical:

1. much more accurate orbits for altimeter and geodynamic satellites
2. unification of worldwide height systems for sea level and crustal motion monitoring
3. from the STEP gravity model in combination with regional terrestrial gravity, local high precision geoids for "levelling by GPS"

Solid-Earth physics suffers intrinsically from a lack of direct measured evidence. Observations are confined to the surface of the Earth. Principal sources of information about the dynamics of the Earth's interior are the magnetic and gravity field and seismic wave propagation. The field of seismic tomography holds a lot of promise, but the translation of the computed velocity anomaly fields to density variations proves problematic. We know that the long and medium wavelengths of the gravity anomaly spectrum are generated by the dynamics of the Earth's interior, the core/mantle boundary, the convecting mantle, the upwelling ocean ridges and plumes, the subducting continental slabs together with their associated backarc volcanoes or mass readjustment as a consequence of postglacial uplift. Therefore, any significant gravity field improvement in this spectral range, in particular over the polar areas unsurveyed so far, is desperately needed. It is our conviction that the combination of seismic tomography, topographic modeling and the STEP gravity field model would lead to a much better understanding of solid-Earth dynamics. The range of models that can be brought into agreement with observed evidence would significantly decrease. A detailed assessment of solid-Earth applications is given in the proceedings of the Gravity Workshop (1987) and in Lambeck (1990).

Physical oceanography would most profoundly profit from STEP. Satellite altimetry proved very successful for determining the time variations of ocean surface topography, caused by meandering currents, eddy motion, tides, etc. The latest altimeter missions are ERS-1 and TOPEX/POSEIDON. Altimeters alone, however, cannot detect the quasi-stationary sea surface topography, i.e., ocean surface circulation itself. This would require precise knowledge of the ocean equilibrium surface,

the geoid. A precise geoid in combination with satellite altimetry, not only provides geostrophic surface circulation, globally and at any time, it also defines the necessary boundary condition for circulation computation at depth from hydrographic data, thus eliminating the vague concept of a "level-of-no-motion" assumption.

The ocean currents, together with radiation influx and atmospheric chemistry, determine the climate of our planet and how it changes with time. Or, as P. Morel, Director of the World Climate Research Programme (1990) expressed it: "In final analysis, the problem of estimating the heat intake of oceans, in the course of transient climate warming occurring over a period of several decades, calls for a detailed understanding of global ocean circulation". Any improvement in the ocean geoid at half wavelengths larger than about 50 to 100 km (the first baroclinic Rossby radius) is important to this aim (Mueller and Zerbini, 1989). ARISTOTELES was tailored to this objective. STEP, at 550 km altitude, would not be capable of quite reaching the same spatial resolution and accuracy. Nevertheless, in constraining circulation models at half wavelengths above about 230 km, it would significantly advance ocean circulation modeling. This is shown in Fig. 6.4 (lower panel).

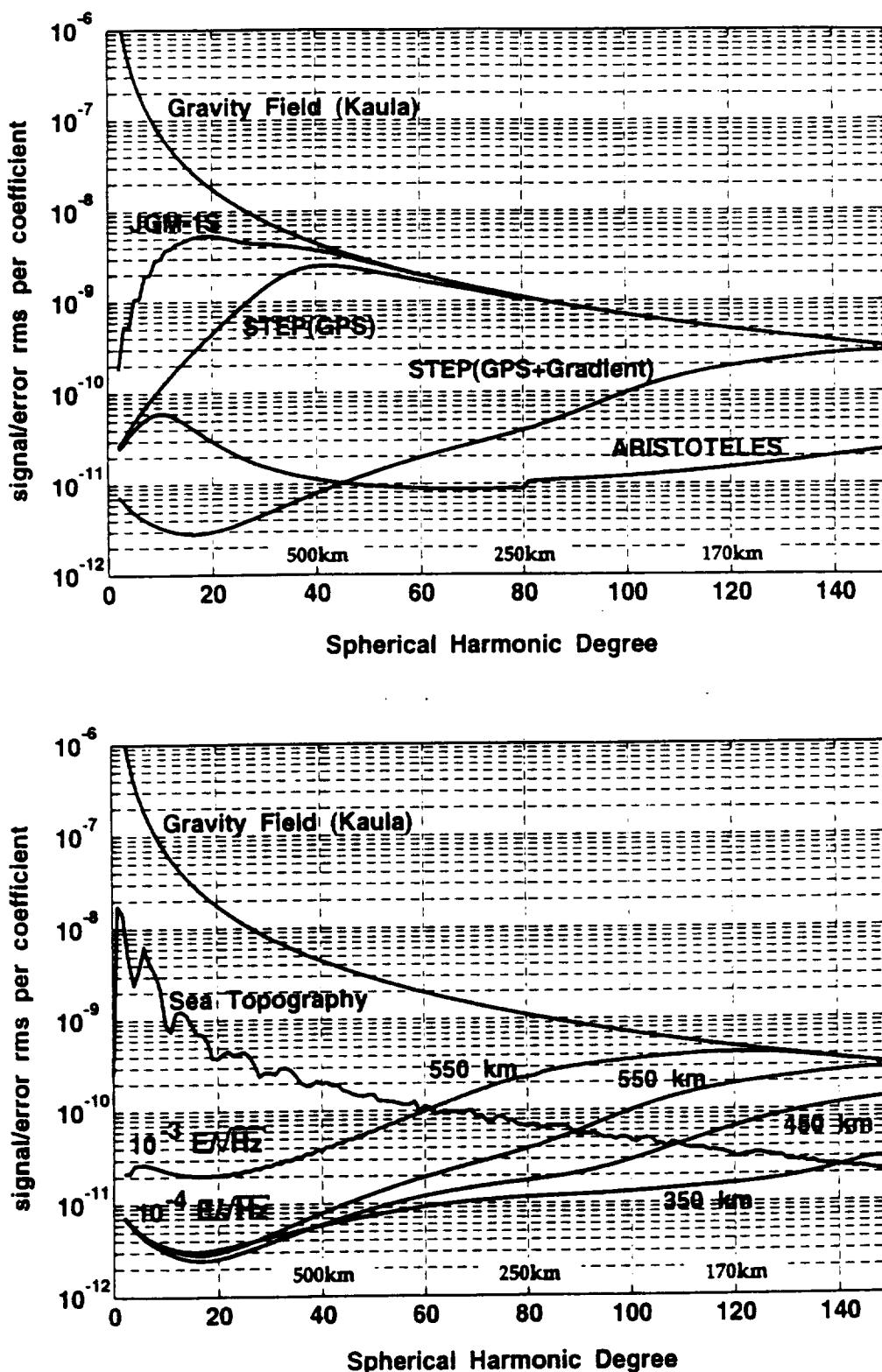


Fig. 6.4. STEP gravity field sensitivity in comparison with the gravity signal (Kaula), the current state-of-the-art (JGM-1S) and ARISTOTELES (upper panel), and for two different gradiometer precisions and different satellite orbital altitudes (lower panel). The lower panel includes the sea surface topography spectrum. The x-axis also indicates the spatial resolution in kilometres.

7. Aeronomy

The STEP satellite will orbit the Earth in an infrequently sampled region of the atmosphere, the base of the exosphere. Analysis of the thruster activity and accelerometer measurements of unprecedented accuracy and time resolution required to keep the STEP satellite in a drag-free orbit can potentially provide useful information about atmospheric density and winds at orbital altitudes. See Jafry (1992) for a detailed discussion on the subject of extracting aeronomics data from the drag-free controller activity.

Before the 1970's, the analysis of satellite drag was the principal method of determining atmospheric density near satellite perigee, but the vast majority of the satellite perigees were below the altitude considered for the STEP satellite. Accelerometers capable of detailed measurements of drag along the satellite orbit have been flown on several satellites, but have not had sufficient sensitivity to reach above about 400 km. Extensive in situ density and composition measurements have been obtained in the thermosphere with mass spectrometers, but again with satellite perigees generally well below the STEP altitude so that instrumental parameters were not optimum for higher altitudes and measurements were difficult to make or not obtained.

The orbit of STEP provides an opportunity to measure atmospheric conditions at the base of the exosphere, and thus provide a lower boundary condition for studies of an atmospheric region where particle mean free paths become long enough that a substantial population of the atmospheric atoms can orbit the Earth without collisions or even escape. This region is of considerable interest because escape is believed to have an important influence, over geologic time, on the abundance of water and isotopic ratios of certain atomic species on Earth.

The atmosphere at STEP altitudes will be principally composed of neutral atomic oxygen, helium, and some hydrogen in near diffusive equilibrium (density of heavier species decreases with altitude more rapidly than lighter species). As found at lower altitudes, it is expected that there will be significant seasonal, local time, and other variations in composition which stem from dynamic interactions between wind fields (global and local) and the background atmosphere. It is, unfortunately, this intricate interaction between density, temperature, wind, and composition, that limits the ability to test theories and interpret observations when only a single parameter, such as total density, is measured.

Aeronomy measurements on STEP would complement the planned TIMED (1991) mission for the lower thermosphere/mesosphere. If both satellites are launched as planned in the same time frame, there would be an unprecedented opportunity to study the coupling of atmospheric phenomena between the mesosphere and the exosphere.

7.1 Atmospheric science topics for STEP

Specific science topics that could be addressed by in situ measurements from the STEP satellite include: acoustic/gravity waves; composition and seasonal/latitude variations; magnetic storm variations; nature of the wind field; drag coefficient accuracy; and testing/updating of empirical models.

The determination of total density with sufficient accuracy from thruster activity and accelerometer measurements could:

- detect the presence or absence of density fluctuations from acoustic/gravity waves which are common at lower altitudes but may or may not penetrate to these altitudes
- quantitatively determine effects of magnetic storms at these altitudes

- test and constrain empirical models in a unique region; and measure cross track winds, the dawn/dusk STEP orbit being particularly well suited to determine winds from the solar wind driven polar vortex.

Drag information alone, however, is generally insufficient to further analyse the source of observed total density variations or discrepancies with models since they could arise from different physical processes which are only distinguishable through different composition and wind variations. Thus, to be most effective for aeronomy study, additional instrumentation would be highly desirable. The most important such instrument would be a mass spectrometer and then an energy analyser (separately or integrated with a mass spectrometer).

The combination of drag, mass spectrometer, and energy information could allow:

- detailed analysis of observed waves
- determination of seasonal/latitude composition variations and resolution of differences between historical models
- separation of oxygen and possibly hydrogen into hot and normal components
- escape flow properties of hydrogen; improve empirical models
- test accuracy of drag coefficient calculations
- confirm drag wind measurements and obtain along track component, although wind measurements are likely to be quite difficult.

Furthermore, the GPS receiver provides an opportunity for atmospheric measurements at locations remote from the satellite. Ionospheric density could be determined above the F-peak and would provide important input to ionospheric models. Density and temperature could be measured in the stratosphere and provide additional coverage or cross-check with other methods being used to study global change.

7.2 Mission issues for aeronomy

The relative accuracy of the drag force data needs to be about 2% or better, and absolute accuracy 5% to provide useful atmospheric densities. It appears that the planned continuous release of He, such that the net atmospheric drag is only a small part of the total thruster force, could prevent reaching the required accuracies unless the He release not needed to counter drag can be cycled with a period of a few seconds to minutes. Duty cycling would be particularly important for the cross-track wind determination because net forces are two orders of magnitude smaller than the in-track drag force.

The inclusion of a mass spectrometer and energy analyser would greatly increase the aeronomy topics that could usefully be addressed as discussed above. While the nominal one spin per orbit mission profile is not ideal for an instrument that needs to look within 90 degrees of ram to obtain atmospheric data, it is anticipated that normal operations will result in sampling over a wide range of latitudes. Such instrumentation would likely weigh in the range 3 to 18 kg, require power in the range 5 to 18 W, and telemetry of 250 to 500 bits/sec.

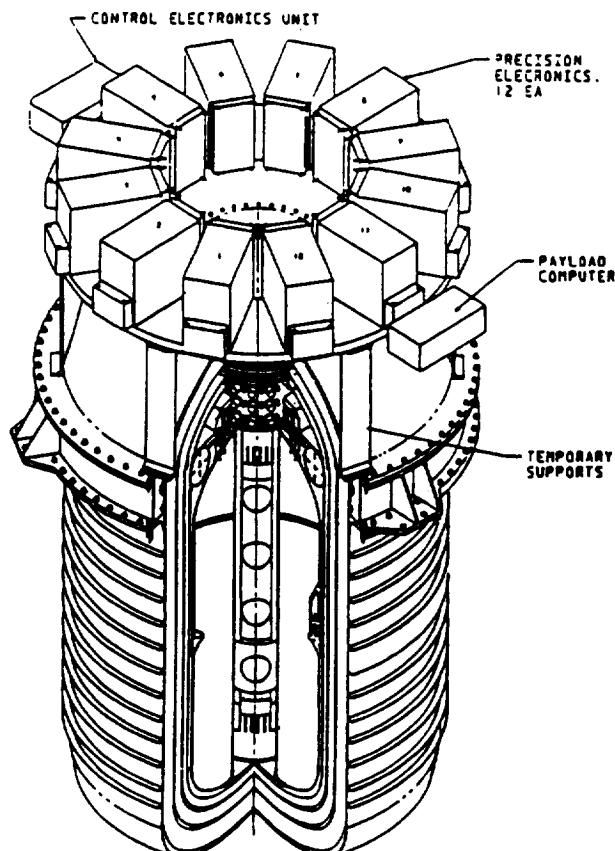
If additional instrumentation, such as a mass spectrometer, were included on the mission it is highly recommended that low cost means (such as gravity gradient or magnetic torque) be considered to stabilise (or spin) the satellite for an extended mission lifetime without the helium thrusters. This would be advantageous because most atmospheric phenomena have seasonal and annual variations which could be better examined with a longer mission.

8. Cryogenic Payload System

The STEP Cryogenic Payload System (Fig. 8.1) consists of :

- the instruments for all scientific experiments, including:
 - the EP experiment (3 Stanford + 3 European differential accelerometers)
 - the G/ISL experiment (2 gradiometers)
 - the SC experiment (1 differential accelerometer)
- the support electronics for all scientific experiments
- the probe assembly
- The cryogenic dewar

Figure 8.2 illustrates the breakdown of the STEP Cryogenic Payload System into major sub-elements, subsystems, and components. The Cryogenic Payload System does not include the STEP spacecraft or any of its subsystems. The experiment support functions provided by the Payload System include cryogenic environment and thermal control, sensing and data collection, data management, and helium tide control.



ORIGINAL PAGE IS
OF POOR QUALITY

Fig. 8.1. Cryogenic Payload System.

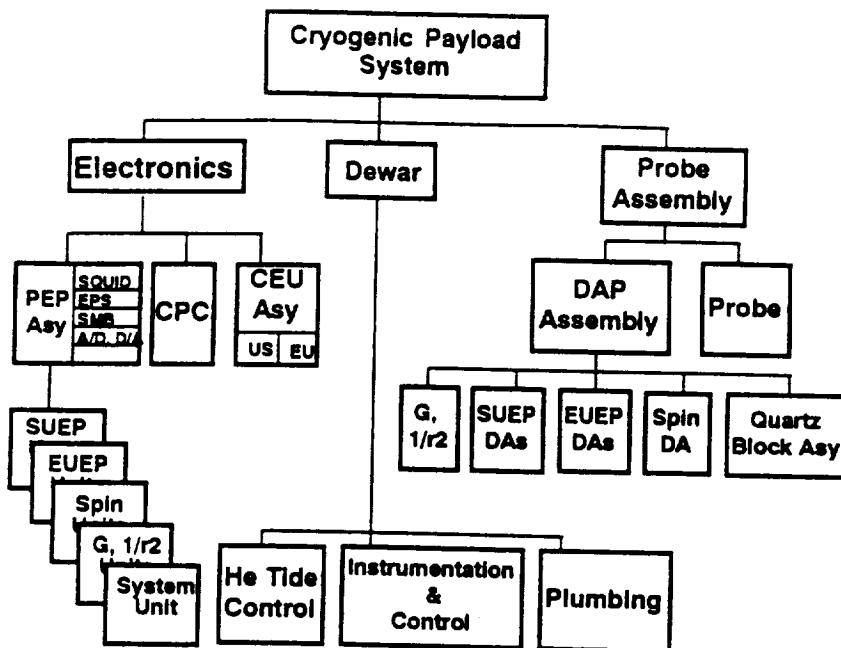


Fig. 8.2. Cryogenic Payload System block diagram.

8.1 The probe assembly

The probe assembly consists of the probe, the quartz block, and the differential accelerometer package (DAP). This is a mid level assembly to be integrated with the dewar and the electronics packages to form the cryogenic payload. The differential accelerometer package will be integrated as an independent subsystem.

8.1.1 Differential accelerometer package (DAP)

The DAP is the cryogenic assembly containing the differential accelerometers. Figure 2.6 shows the configuration for the scientific sensors in the quartz block.

8.1.2 Probe

The probe configuration was shown in Fig. 8.3. The probe provides:

- an ultra-high vacuum container for the accelerometers (10^{-11} torr)
- superconducting magnetic shielding
- mechanical support for the quartz block assembly
- electrical connections to the ambient temperature electronics
- thermal isolation from the heat load coming from the instrumentation wires.
- shielding from thermal radiation

A leak tight chamber inside the probe will contain the ultra high vacuum and the DAP. The integrity of this vacuum is protected by a surrounding guard vacuum. This is the vacuum in the dewar which is opened to space.

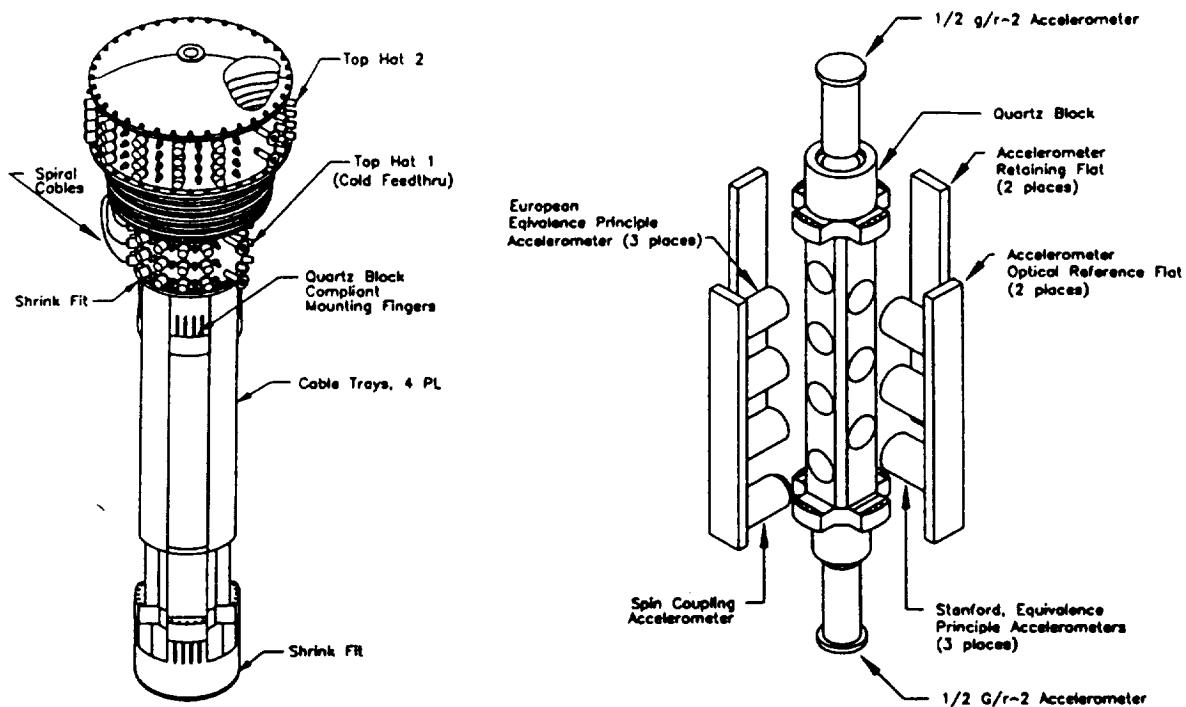


Fig. 8.3. Probe assembly diagram.

A lead shield, fixed to the probe's vacuum chamber, will be cooled through its superconducting transition temperature in a low magnetic field facility prior to final integration. This shield traps the low field required for the experiments.

The probe contains 200 cables and 1020 wires required for instrumentation and control of the DAP. These cables are thermally grounded to a series of vapor cooled shields located in the probe. Spiralling the cables increases their length, further reducing the heat flow into the DAP and the superfluid helium tank.

8.2 Dewar

The dewar provides a cryogenic environment for the STEP instruments and provides propellant gas for the proportional thrusters. The dewar (Fig. 8.1) holds approximately 200 liters of superfluid liquid helium. The helium keeps the temperature of the apparatus at about 2.0 K. Operation below the lambda point (2.17 K) greatly increases the thermal conductivity of the helium, facilitating temperature uniformity within the dewar. A "superfluid plug" is used as a phase separator for the helium liquid and gas. The boiloff gas cools a series of thermal shields as it escapes, improving the thermal efficiency of the dewar.

The boiloff gas is used as reaction mass for the drag-free and attitude control system. The warmed gas is throttled through a set of proportional thrusters. The total flow regulates the temperature of the instrument, and the differential flow, through thrusters pointing in opposite directions, determines the net thrust and torque on the satellite.

The precise operating temperature is not critical, but the variation in temperature at orbit frequency should be no more than 1 mK across the volume of the experiment, and no more than 1 mK

per orbit. The rate limit is ten times smaller in specific parts of the apparatus. The temperature requirement limits disturbances from gas pressure and supercurrent penetration depth variation.

8.2.1 Requirements

The predicted level of performance for the dewar and probe system meet the requirements. All of the 14 dewar advanced hardware technology components have been demonstrated on prior cryogenic hardware programs (ID dewar, GP-B, CLAES, SPIRIT III, TEAL RUBY, LLC, ELC). The critical probe technologies have been demonstrated during the development of the GP-B flight program. This high degree of demonstrated hardware technology greatly reduces the technical, schedule and cost risk associated with the program.

8.2.2 The helium constraint system

The purpose and principles of the helium constraint system was described in Section 3.4.2. The free surface of the helium will always be in the outer chamber, and will be constrained electrostatically. The electrostatic field will be produced by a single cylindrical electrode fed with high voltage of 1 to 2 kV. The high voltage will be generated by a supply with foldback current limiting protection to guard against breakdown in the helium. Additionally, there will be a monitoring assembly, consisting of temperature and liquid level sensors and superconducting or capacitance-based helium motion sensors. The helium constraint system will interface with the payload computer to enable flight and ground monitoring.

8.2.3 Lifetime

The dewar with all system cables has been studied and the predicted lifetime exceeds the mission requirements of 6 months. Lifetimes and equilibrium vent rates for a range of dewar vacuum shell temperatures are shown in Fig. 8.4.

8.2.4 Thermal and structural performance analyses

The experiment must be maintained at ≈ 2 K with a stability of 1 mK. Much of this thermal requirement is met passively by the dewar and guard tank. The payload electronics outside the dewar may require small heaters to maintain acceptable operating temperature ranges.

The thermal performance of the dewar is summarised in Fig. 8.5. A guard tank, filled with normal liquid helium, maintains a small heat flow to the superfluid before launch. It is vented just prior to launch.

Transient thermo-mechanical effects

Detailed spacecraft transient thermal analyses of the dewar vacuum shell temperature covering a 6 month period shows the maximum change in average temperature at any point, per orbit, is 0.4 K. The maximum temperature gradient across the shell per orbit is 0.7 K. These transient temperatures cause mechanical distortions in the vacuum shell, as large as 0.0005 cm. These distortions, in turn, change the gravitational coupling to the accelerometers. These periodic disturbances could mask or be mistaken for an Equivalence Principle signal, since they always come at signal frequency. The transient mechanical displacements per orbit were calculated using a structural dynamic finite element model. The maximum change in gravitational coupling was calculated using a point mass at the radius of the dewar shell as a model. The point mass was assumed to have half the mass of half of the entire shell. For the baseline EP test masses (belted outer cylinder, straight inner cylinder), and assuming the maximum distortion, the change in differential acceleration would be less than $4 \times 10^{-18} \text{ m s}^{-2}$.

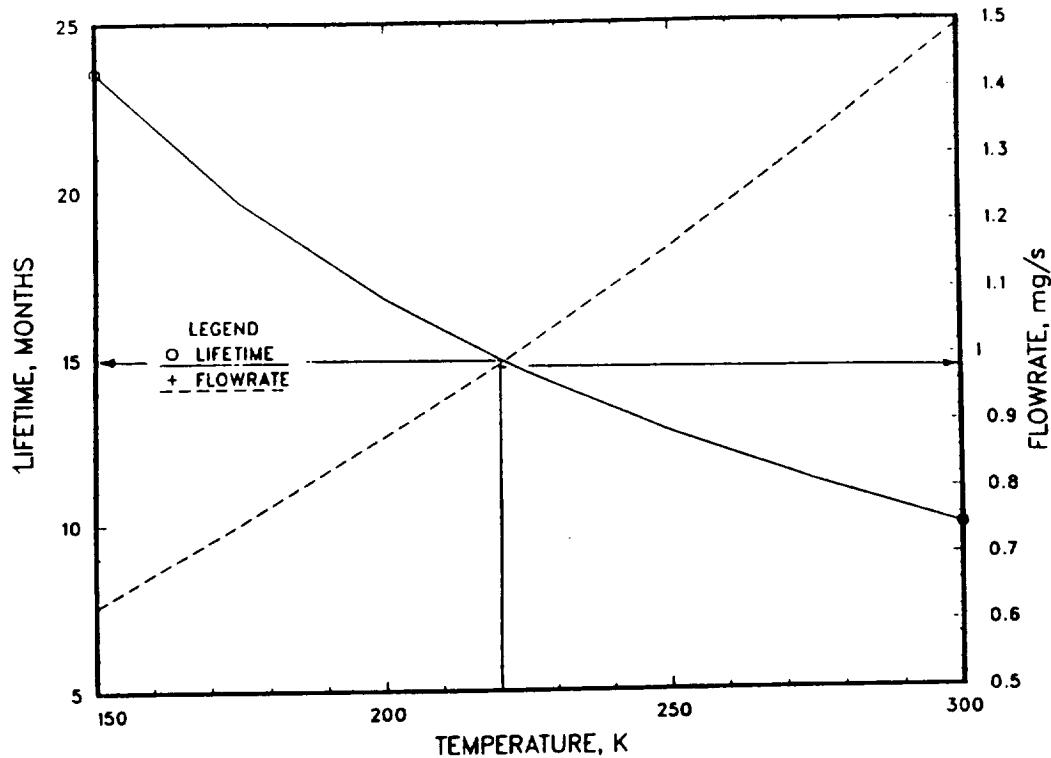


Fig. 8.4. Dewar lifetime plot.

8.3 Electronics

Precision electronics (PE)

The precision electronics are contained in twelve boxes located above the dewar and close to the probe top hat to minimise the low level signal path from the accelerometers (Fig. 8.1). They are mounted on a pallet that will be integrated with the spacecraft without disconnecting the cables. The complete payload integration can be performed prior to integration to the spacecraft.

The precision electronics consists of all non-cryogenic electronics necessary to perform measurements with the accelerometers. The twelve boxes are allocated to the four experiments and the common system as follows:

Box	Experiment
1-4	Stanford EP experiment
5-8	European EP experiment
9-10	G/ISL experiment
11	SC experiment
12	common system box

The precision electronics will measure the outputs of the science instruments and several cryogenic instruments used to control the experiments. They will also send control signals to various cryogenic components through (nominally 12-bit) D/A converters. The D/A converters cannot be multiplexed. The control functions include: magnetic bearing control, SQUID control, and EPS control. Although

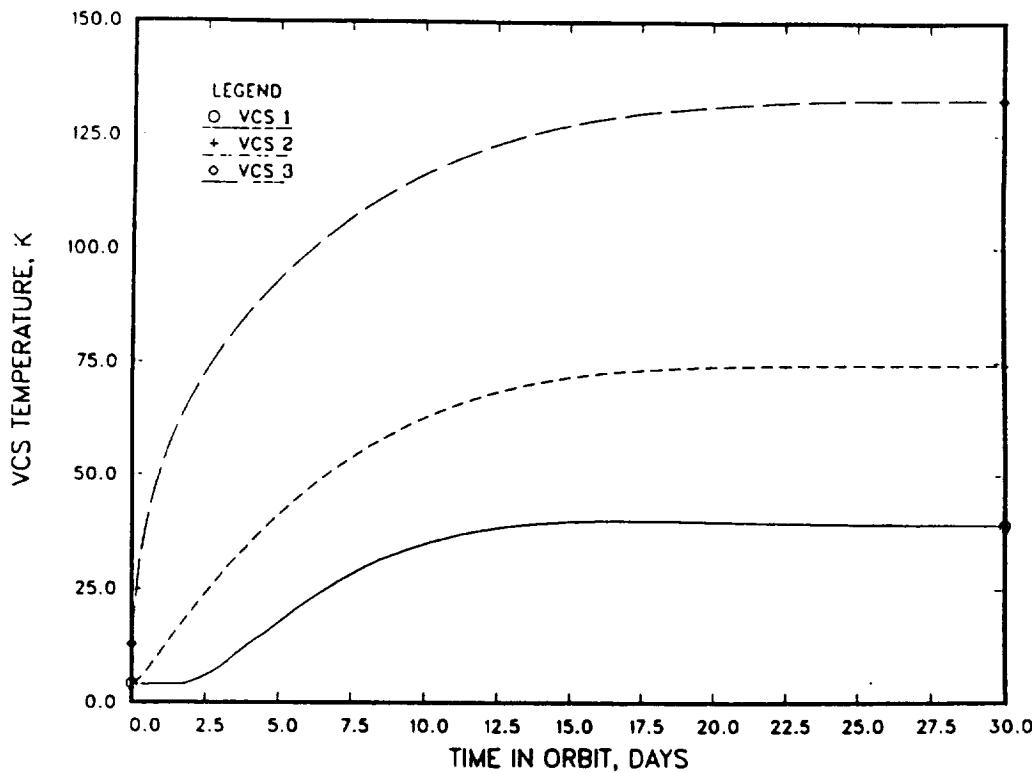


Fig. 8.5. Dewar thermal transient after launch. Plot shows the temperature histories for the three vapour cooled shields (VCS).

the EPS control signal circuit implementation is yet to be decided, at least 15 bipolar D/As will be required (possibly as many as 30).

Payload computer

The payload computer is located on the spacecraft and will have mostly digital lines and non-critical analog connections to the cryogenic payload.

The payload computer will perform the following high level functions:

- experiment management,
- payload data management,
- remote bus interface (RBI) with the spacecraft computer,
- helium constraint system monitoring and control,
- digital signal processing.

The payload computer will contain separate software modules for data management and control of the Stanford EP, European EP, SC, and G/ISL accelerometers.

Since the spacecraft mass memory unit (MMU) will store experiment data to be burst-transmitted or “dumped” to the ground periodically, the memory requirements for the payload computer will be small. Memory size is driven by the MMU interface and ground upload functions. As with any

flight computer hardware, the requirements for space environment qualification (radiation and thermal) and reliability, limit the miniaturisation, efficiency and technological novelty of the computer hardware.

Control electronics unit (CEU)

The Control Electronic Unit will be located on the spacecraft and will have digital and noncritical analog connections to the cryogenic payload and include power supplies and electrical drivers for the noncritical electronic functions, such as stepper motor drives and heaters.

8.4 Integration and test

The integration tasks are shared between Europe and the United States. The DAP will consist of three major subsystems that include an assembly of the Stanford EP accelerometers, an assembly of the European EP accelerometers plus the SC accelerometer, and an assembly of the G/ISL accelerometers with the quartz block. These subassemblies can be integrated and tested on the subsystem level so, at final integration, the risk of failure will be reduced. Contingency planning could be made for schedule slippage in the development of the different systems.

Integration of the probe and DAP into the dewar will be performed at the Cryogenic Payload Integration Facilities in the United States (either at university or contractor sites). The integrated Probe, DAP, and dewar, together with the payload electronics, constitute the fully integrated cryogenic payload. After the probe and DAP, including all instruments, cryogenic support equipment and cabling, have been integrated into the dewar and tested, the payload, still cryogenically cooled, will be shipped to ESTEC for integration with the STEP spacecraft. At this time, the precision electronics pallet is mounted on the dewar top hat, and the cryogenic payload computer is mounted on a selected spacecraft panel. Final verification of the payload will be carried out during the spacecraft functional test programme.

9. Spacecraft

9.1 Spacecraft requirements

The spacecraft is the platform on which the experiments will be flown. As well as providing all the normal resources usually supplied to any space experiment, such as power, command capability and telemetry, the STEP payload defines some unique requirements in isolating the payload from the external environment. In addition to the requirement to maintain accurate three axis attitude control, the payload must be shielded from external forces such as air drag and solar pressure. This is achieved by operating a combination of proportional thrusters to allow the spacecraft to fly *drag-free*.

In order to resolve a periodic Equivalence Principle signal of 10^{-17} g we need to have a very quiet environment of $5 \times 10^{-11} \text{ m s}^{-2}/\sqrt{\text{Hz}}$ in all directions in a narrow band of 10^{-5} Hz around the signal frequency. The signal frequency varies from 1 to 3 times orbital rate depending on the operational mode. Furthermore, the common-mode motion of the accelerometers must be limited to avoid coupling with spacecraft gravity gradients.

Since the signal frequency is below the bandwidth of the accelerometers, the drag-free and common-mode motion requirements can be equivalently defined as limits in the relative common-mode displacement of each differential accelerometer in X, Y and Z directions. These are listed below.

Table 9.1. Drag-free requirements.

	Signal Frequency $\pm 0.5 \times 10^{-5}$ Hz	Other frequencies
RMS common-mode displacement along:		
X, Y axes	10^{-8} m	10^{-6} m
Z axis	3×10^{-9} m	10^{-6} m

These requirements have to be taken into account by the attitude & orbit control system. They also dictate the level of disturbance coming from the spacecraft. This necessitates that mechanisms are reduced to a minimum so no gyroscopes, reaction wheels or tape recorders are used. The only mechanisms to operate during drag-free operation will be those inside the thruster valves. Another fundamental requirement is that the gravitational field produced by the spacecraft must not cause differential accelerations of the test masses greater than $10^{-17} \text{ m s}^{-2}$ in a bandwidth of 10^{-5} Hz around the signal frequency. This dictates that the spacecraft structure must be stable and the thermal environment changes minimised. This requirement is more easily met by choosing a Sun-synchronous orbit such that the satellite is never subject to eclipses.

9.2 Spacecraft configuration

The spacecraft in-orbit and launch configurations are shown in Fig. 9.1. The spacecraft consists of two service modules placed below and above the dewar. Each service module has the shape of a regular eight-sided polygon inscribed in a cylinder of diameter 2 m. The heights of the lower service module and upper service module are 0.6 m and 0.46 m respectively. The separation between the two modules is 1.74 m. Eight tubular struts connect the upper module to the lower one. The spacecraft will have a mass of about 1100 kg of which 500 kg will be the dewar and instruments.

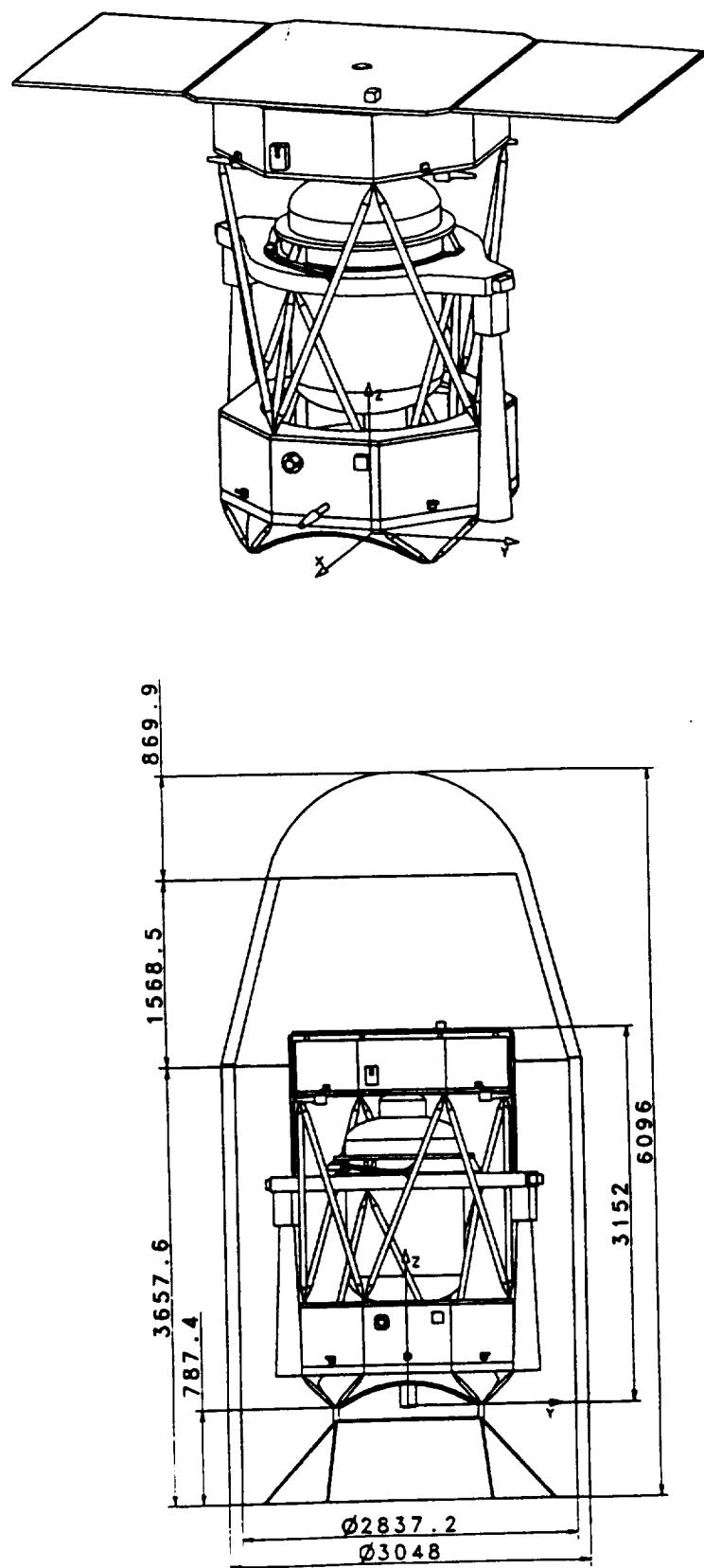


Fig. 9.1 STEP spacecraft: (a) in-orbit configuration; (b) launch configuration

In normal flight operation the upper face of the spacecraft lies in the orbit plane so that the solar array is Sun-pointing.

The spacecraft is open in the middle so that the dewar, which is in the shadow of the solar array, can radiate freely to all sides. This keeps the outside of the dewar as cold as possible to minimise heat leaks. The dewar is quasi-isostatically attached to another platform, the intermediate platform, at three equally spaced points. The intermediate platform, in turn, is connected directly with the floor of the lower service module at three points in a circle. A set of six tubular struts provide this structural support.

A conical transition structure provides the connection between the lower service module interface circle with the adapter interface ring. The height of the transition structure is 0.275 m and the slope of the cone is 60 degrees.

A rectangular solar array using conventional silicon cells is placed on top of the upper service module. The array is made up of three sections, one fixed and two deployable. The deployable wings are folded at launch against the $\pm Y$ sides. The total surface area of the array is 7.5 m^2 .

The intermediate platform supports two star trackers viewing away from the Sun in the $-Z$ direction. This location is chosen to provide a common structural fixation with good mechanical stability to that of the accelerometers. To avoid stray light, the star trackers are fitted with long baffles reaching all the way to the satellite floor. Two medium-accuracy magnetometers are also mounted on the intermediate platform.

The telecommunication S-band antennas are located on the $\pm X$ sides of the lower service module, providing full sky-coverage at all times in the mission. Two GPS patch antennas are placed on top of the fixed solar panel and on a dedicated structural support protruding below the spacecraft adapter interface plane. The GPS antennas together provide 4π steradians coverage with the axis of their field-of-view towards the orbit poles throughout the mission. Two hemispherical laser retroreflectors are placed on opposite sides of the lower service module, providing unocculted coverage over 4π steradians. A functional block diagram of STEP is given in Fig. 9.2.

9.3 Drag-free control and AOCMS

The Attitude & Orbit Control & Measurement System (AOCMS) main function is to control the spacecraft attitude and position to provide the drag free environment for the payload. The AOCMS must also support those phases of the mission prior to entering drag-free control.

9.3.1 Sensors and thrusters

In the nominal orientation, the X - Y axes are in the orbit plane and the Z axis is normal to the orbit plane. There are three EP accelerometers in the X direction, one SC accelerometer in the X direction, three EP accelerometers in the Y directions, and two G/ISL accelerometers in the Z direction. The common-mode displacements will be measured by SQUID displacement sensors.

Star trackers will be used for initial setup of the orientation system, for this they will have a resolution of about two arc seconds. Two star trackers are needed for redundancy. The star trackers will be used for systematic checks after setup phase since they will be sensitive to thermal distortion of the spacecraft. The major requirement during setup is that the star trackers be able to resolve less than 0.1 orbital rate to ensure that the spacecraft is not rotating so fast as to prevent the orientation system from working. In the operational mode, the star tracker will provide external orientation reference for the attitude estimator, which will be described later.

The helium gas boil-off will be used as the propellant for the set of drag-free and attitude control thrusters. By directing the escaping gas in specific directions, forces and moments can be generated for both translation (drag-free) and attitude control. The specific impulse of the helium thrusters is about 130 seconds. About 2 mg/s of helium gas is available as boil-off from the dewar.

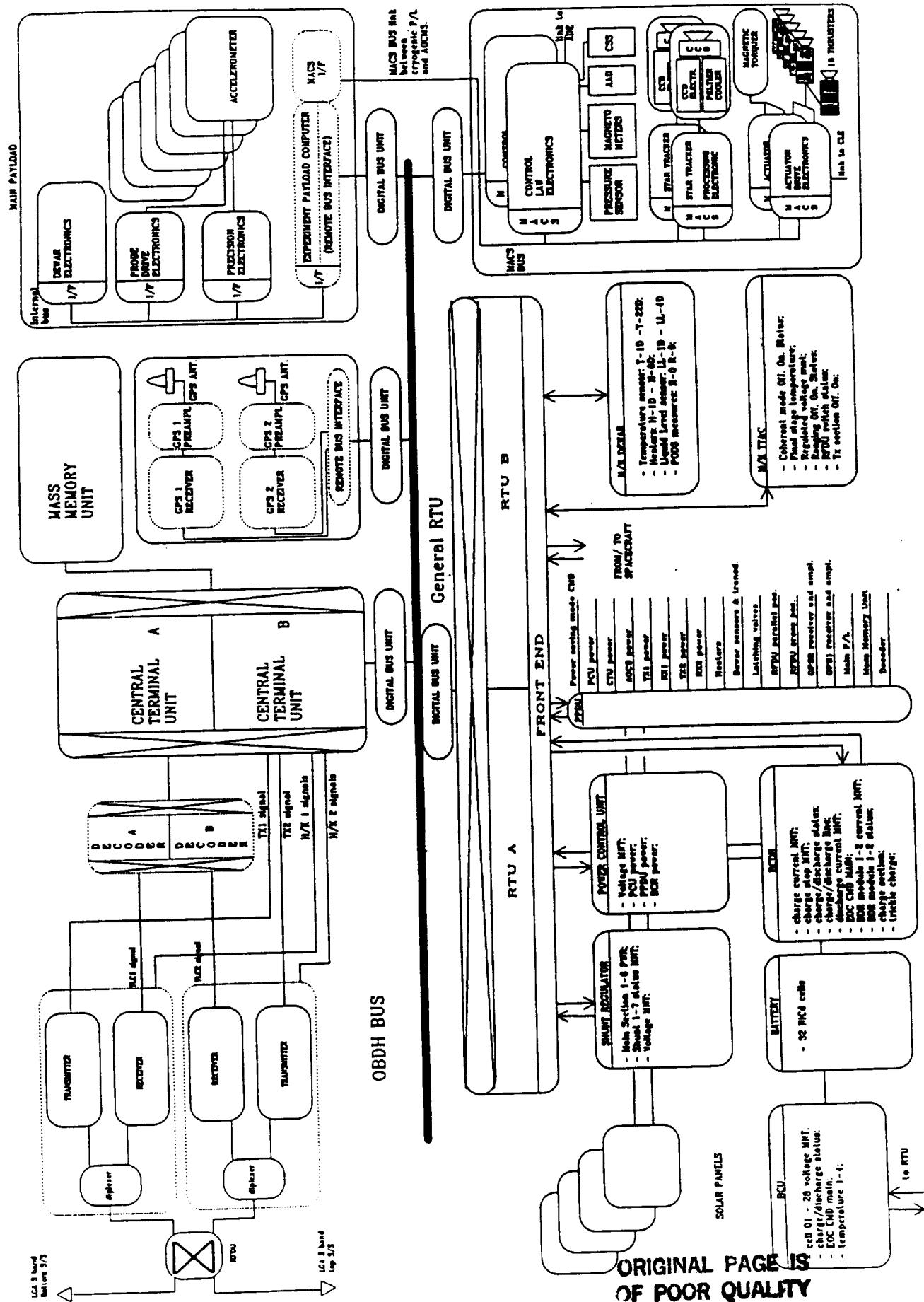


Fig. 9.2 *STEP spacecraft functional block diagram.*

This can produce a total thrust of 2.5×10^{-3} N. This has been shown to be enough control authority to counter the disturbance environment in the baseline orbit.

9.3.2 Control in translation

In the X and Y directions, the drag-free feedback signal will be derived from one accelerometer or a weighted average of all three accelerometers. In the Z- direction, common-mode output from the G accelerometers will be the drag-free control signal. The drag-free controller will notch the disturbances at the EP signal frequency to make sure the performance at that frequency is satisfied. This can be realised by either designing a notch filter at the signal frequency or by designing an estimator to estimate the disturbance component at the signal frequency and feed it forward to cancel the disturbance. Control bandwidth of about 0.3 Hz is enough to satisfy the requirements at other frequencies. The translation control loop is shown in Fig. 9.3 and the closed loop frequency response in Fig. 9.4.

9.3.3 Control in rotation

For the rotation control we have a difference between control around the X and Y axes and that of control about the Z axis. The rotation control of X and Y axes is essential, since any rotation around these axes induces a relative displacement of the proof masses. The dynamics of these axes, is driven by the relative acceleration and gravity gradient acceleration between two proof masses. This means that there is strong coupling between the X and Y axis which is time-dependent. The solution to this control problem is to use the signal from the accelerometers together with the X and Y angles measured by the star tracker. The estimator gains within the control loop are truncated Fourier series (two terms are sufficient) which are predefined from the ground. The rotation control loop is shown in Fig. 9.5 and the closed loop frequency response in Fig. 9.6. The Z axis rotational control differs from the X and Y axis control because it is naturally decoupled from the other axes. Moreover, it is the only control that does not aim primarily at keeping the masses drag-free, but only ensuring a given rotational rate with a given accuracy. The control envisaged is a simple proportional-integral-derivative scheme with inputs from the star tracker tracking off-axis stars together with angular measurements derived from the accelerometer outputs.

9.3.4 Modes

The AOCMS has a number of modes shown in Fig. 9.7. These modes may be summarised as follows

Coarse acquisition phase: this phase begins at launch separation and ends after the solar array is stabilised to the Sun direction. Control in X and Y axis is sensed by two coarse Sun sensors while control in Z axis is sensed by a magnetometer. Actuation is by operation of the thrusters working in bang-bang control. One orbit will be required to recover from the launch residual rates and acquire the Sun.

Fine acquisition phase: after coarse acquisition phase the spacecraft carries out star recognition and transfers control to the star trackers. The thrusters are used in their linear range and the magnetorquer is used to trim out the spacecraft magnetic moment.

Medium accuracy pointing mode: this mode is under the control of the star trackers. Drag-free initialization takes place by positioning the test masses electrostatically to a point where the SQUIDs can take over sensing to allow switch over to drag-free control.

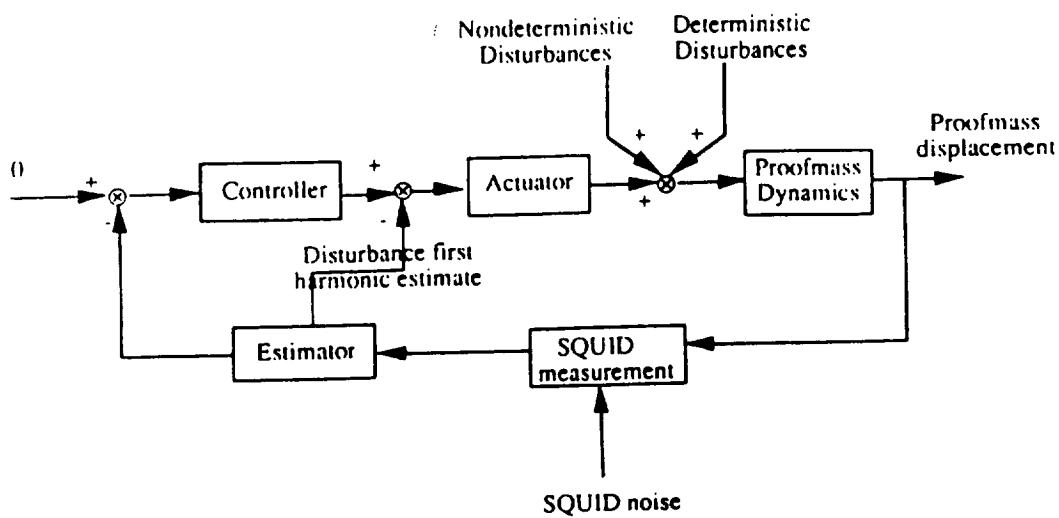


Fig. 9.3 Block diagram for drag-free control in translation.

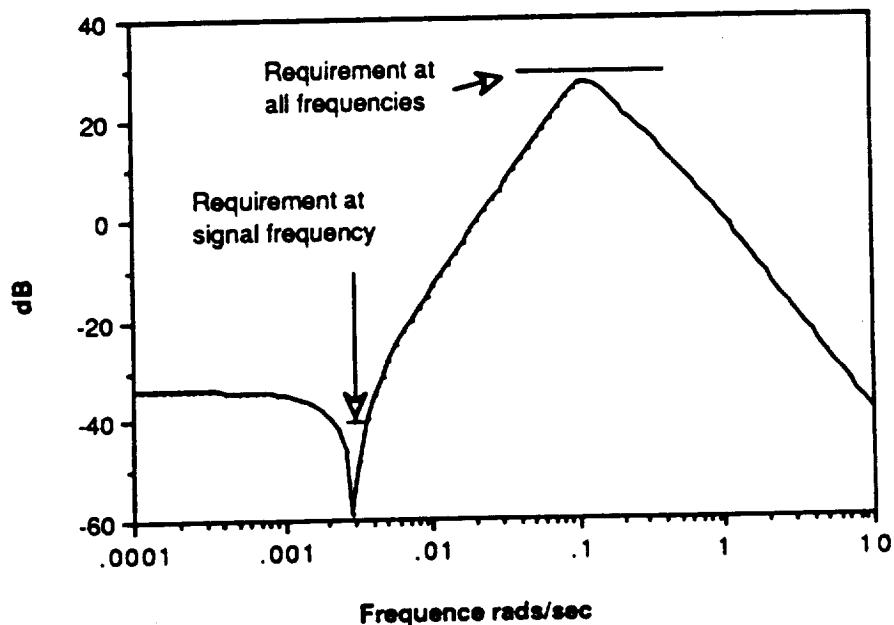


Fig. 9.4 Disturbance rejection ratio for drag-free control in translation.

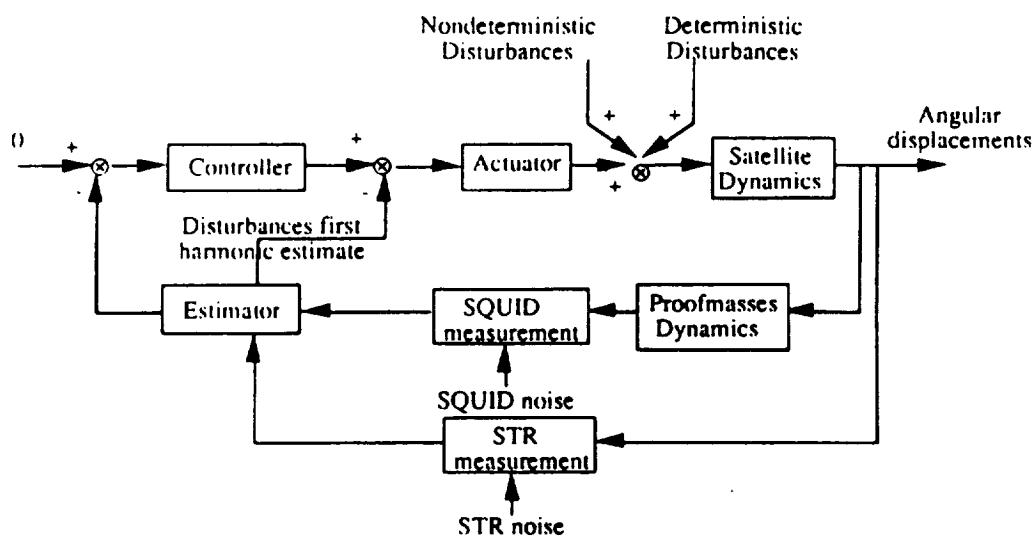


Fig. 9.5 Block diagram for drag-free control in rotation (X-Y axes).

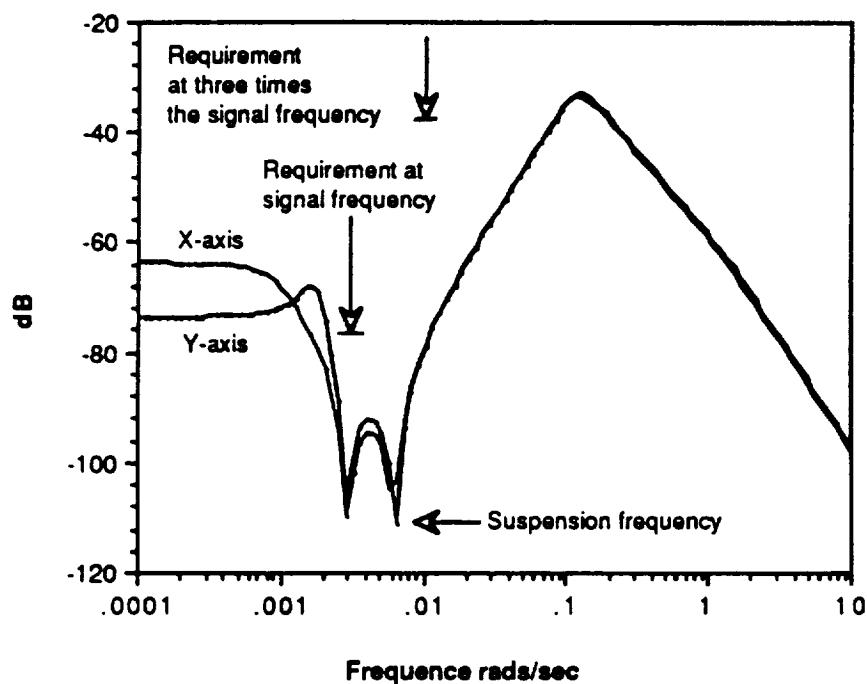


Fig. 9.6 Disturbance rejection ratio for drag-free control in translation (X-Y axes).

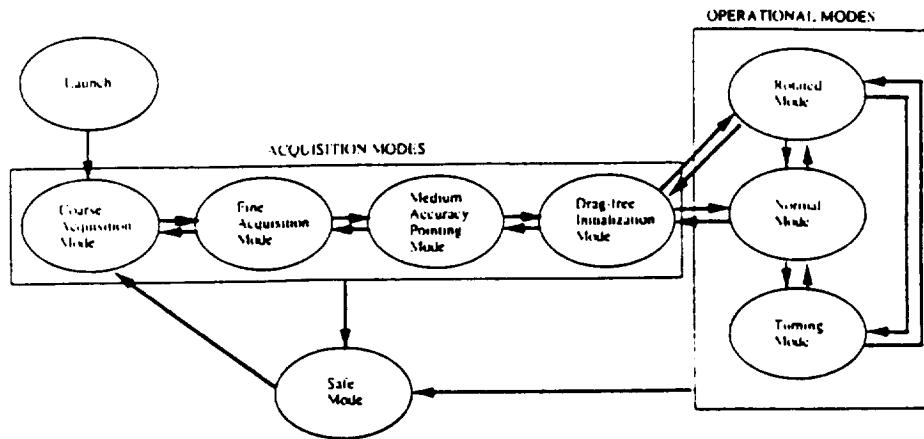


Fig. 9.7. Sequence of AOCMS modes.

Normal operational modes: these are the *Normal*, *Rotated* and *Turning* modes (N, R & T modes) when science data will be taken. In N mode the spacecraft is fixed in inertial space so the Earth rotates about the spacecraft at orbital frequency. The R mode is similar to the N mode but the spacecraft is rotated a fixed angle about the Z axis, from that in the N mode. In the T mode the spacecraft is rotated at constant rate about the Z axis, at rates between zero and three times orbital rate in either direction.

Safe mode: a failure causing an attitude drift of the satellite Z axis from the Sun will be sensed by a simple analogue sensor, the Attitude Anomaly Detector. Such a failure will place the spacecraft in the safe mode. This mode is similar to that of the coarse acquisition mode.

9.4 Structure

To be compatible with the Titan II launch vehicle, the structure has to have a fundamental frequency in the lateral axis above 10 Hz and in the vertical axis, above 25 Hz. To meet the Delta II back up launch vehicle requirements the structure has been designed to meet 15 Hz laterally and 35 Hz vertically. Additionally the structural strength must be adequate to meet, with appropriate safety factors, longitudinal accelerations of -10 g simultaneously with lateral accelerations of 2.5 g. It has been shown by structural analysis that these requirements have been met.

The upper and lower service module are formed by two horizontal platforms and eight lateral panels. Some of these panels are designed for easy dismounting to allow access to the spacecraft units. The panels are made from aluminium alloy sheets with honeycomb core.

The struts connecting the upper and lower service modules and supporting the dewar are hollow tubes in CFRP. This material has been selected because of its high strength and low thermal coefficient of expansion. The adapter cone between the spacecraft and the interface plane is an aluminium alloy machined part.

9.5 Thermal control

Although the orbit selected for STEP is a particularly benign environment from a thermal control stand-point, because of the absence of eclipses, the requirements are still demanding. The main requirements are as follows:

1. Maintain the average temperature of the dewar outer shell lower than or equal to 220 K.

2. Minimise the temperature variation of the dewar outer shell and the spacecraft structure along the orbit so that the gravity field requirements are met.
3. Provide the spacecraft units with an ambient temperature within their design limits.

The first requirement has dictated the overall configuration of the spacecraft to allow the dewar to radiate freely to space. The other requirements are met by a classical design in that appropriate surface finishes are applied to the spacecraft surfaces to achieve the required temperatures and temperature stability. The main surface finishes are listed as follows:

- Dewar covered in white MLI
- Upper and lower service module enclosures black painted
- Upper and lower service module lateral panels covered with MLI/OSR
- Black MLI on the outside of upper service module lower platform and lower service module upper platform
- Kapton MLI on rear + Z solar panels
- White MLI on struts

Thermal analysis has been performed under *hot case* and *cold case* conditions. Both cases occur for launch in the spring window. Steady state and transient cases were analysed, the transient cases took into account Earth albedo variations around the orbit. Acceptable temperature ranges were found for all units. The results of the transient analysis were used in the thermo-elastic analysis.

9.5.1 Thermo-elastic analysis

Of fundamental importance to the success of STEP is the control of the spacecraft gravitational field, as changes at orbital frequency can couple into the accelerometers and mimic an Equivalence Principle violation. Variations of the spacecraft gravity field are caused by changes in the mass distribution due to thermal deformation of the spacecraft structure. Since the STEP orbit is not affected by eclipses, the thermal deformation at signal frequency is produced mainly by the successive exposure of different portions of the spacecraft external surfaces to the thermal flux coming from the Earth.

To compute the differential acceleration induced on the accelerometers, the spacecraft has been schematised as a set of point masses, distributed on the base platforms of the upper service module and lower service module, on the intermediate platform and on the 14 struts connecting the platforms. The overall mass of the spacecraft dewar and solar panels excluded, have been concentrated on these points. The final model used for this analysis contained 184 mass points as shown in the figure below.

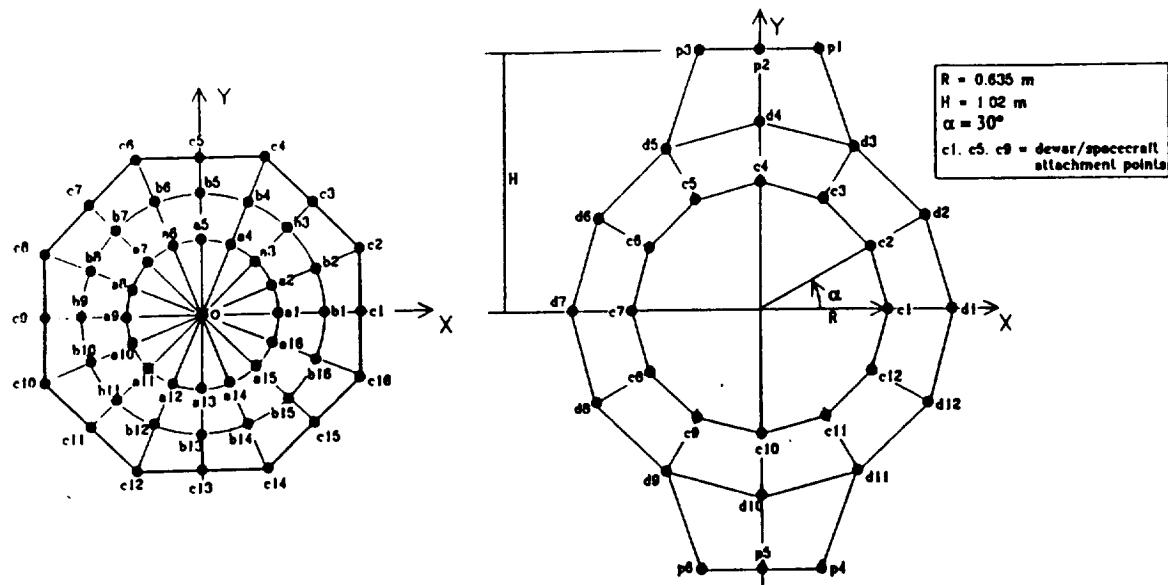


Fig. 9.8 Mass points distribution in the spacecraft model for the self-gravity stability verification.

From the thermal transient analysis of the spacecraft, the temperatures at 11 points around the orbit have been predicted. These values have been fed into a structural model of the spacecraft having the nodes coincidental with the point masses of the self-gravity model shown below.

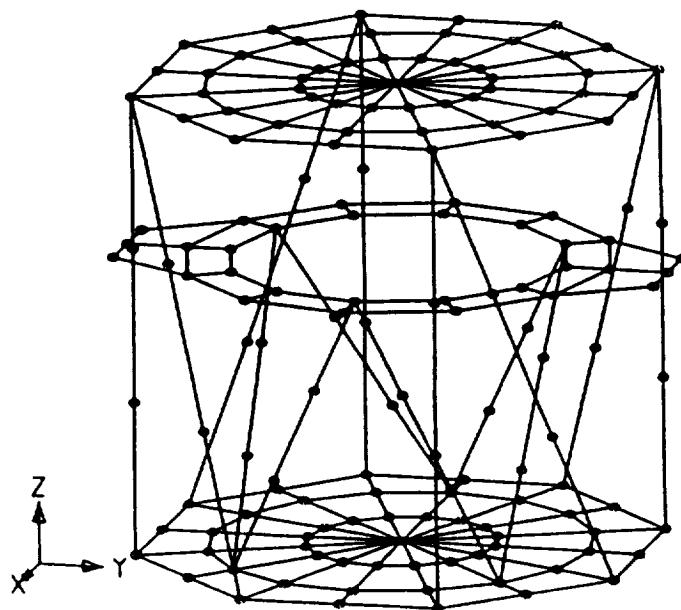


Fig. 9.9 Structural model for thermo-elastic analysis.

From the position of the point masses obtained as a result of the thermal structural analysis, the differential acceleration seen by the accelerometers has been calculated. The Fourier transformation of the time history of the differential acceleration has been calculated to find the amplitude of this variation at orbital frequency. These calculations have been performed for six positions within the dewar probe. The analysis has shown that the performance requirements can be bettered by a factor of two.

9.6 Tracking, Telemetry & Command (TT&C)

For a satellite in low Earth orbit like STEP communications with the ground will be infrequent. Either communication via stations of the NASA DSN or through the European DRS is possible. Both options are equally viable but for the STEP phase A study the former solution has been selected. The TT&C transmits and receives on S band with uplink frequency 2109 MHz and downlink frequency 2290 MHz. Two S-Band antennas provide a continuous omnidirectional coverage for any spacecraft attitude. Data will be transmitted and telecommands received during passes over ground stations situated at Goldstone, Madrid and Camberra. The telecommand bit rate will be 2 kbps, while the telemetry bit rate will be 732 kbps. The telecommunication ground systems are equipped with 26 m diameter antennas. This results in high gain margins (≈ 20 db) in the link budgets.

9.7 Onboard data handling

The STEP OnBoard Data Handling (OBDH) provides the spacecraft with all the necessary functions for the management of commands and data for all onboard subsystems and payloads, in accordance with ESA class 2 OBDH standards.

The onboard data rate is made up of the science and housekeeping data plus the AOCMS data rate. The AOCMS data rate is 0.6 kbps during science operations and 1 kbps during thruster calibration periods. The higher value plus a margin of 20% has been taken to establish the mass memory size and telemetry rate. The mass memory size depends on the time gap between two successful station passes. The nominal time is 16 hours and the worst case is 19 hours. This equates to a mass memory requirement of 166 Mbit and a downlink data rate including overhead of 732 kbps. A mass memory of 192 Mbits is proposed based on the 4 Mbit DRAM as used in the Mars 94 mission. Because of the low size requirements it is proposed to integrate the mass memory into the same box as the CDMU.

A single Remote Terminal Unit (RTU) which interfaces the OBDH bus with the onboard systems. Four Data Bus Units (DBU's) which make the physical connection between the OBDH buses and the users and the CDMU. The interface to the users is further standardised by fitting Digital Bus Interface units (DBI) into each subsystem.

9.8 Electrical power subsystem

The electrical power subsystem consists of

- a solar array with two deployable panels
- a shunt regulator
- a 21 a/h Ni-Cd battery and associated charge and discharge regulators
- a power protection and distribution unit

During the Phase A study, two configuration options were considered for the array, one with silicon solar cells and one with gallium-arsenide. The array with GaAs cells allowed mounting the array on the satellite body, without adding any deployable wings. It was however decided for the

Table 9.2. Spacecraft power budget.

Subsystem	Power Consumption [W]			
	LEOP	average	maximum	safemode
Payload	0.0	167.6	205.1	36.0
OBDH	28.8	32.2	40.3	31.6
TT&C	13.3	13.3	37.5	13.3
EPSS	36.9	42.3	60.1	25.0
AOCMS + RCS	42.0	53.8	81.4	30.0
Thermal Control	38.0	43.7	43.7	43.7
Harness Loss (2%)	3.2	7.1	9.4	3.6
Sub-Total	162.2	360.0	477.5	183.2
System Margin(15%)	24.3	54.0	71.6	27.5
Grand Total	186.5	414.0	549.1	210.7

Phase A study to baseline the conventional Si cell array, while GaAs could be reconsidered in later phases when the maturity of the technology would be demonstrated.

In orbit the required spacecraft power is 550 W. This will normally be provided by the 7.5 m^2 of solar array after conditioning by the shunt regulator to 28 V DC ($\pm 2\%$). During the initial deployment and in cases of safemode recovery, power will be provided by the battery through the discharge regulator.

The power protection and distribution unit provides the main bus distribution to the payload and spacecraft subsystems. Each line is protected against over-current by electronic circuit limiters.

9.9 Budgets

9.9.1 Power budget

The satellite power budget is given in Table 9.2. Four cases have been defined in the compilation of this budget. They are

- ascent phase
- normal operation phase
- operation, maximum power demand
- safemode

Power during ascent phase and safemode recovery is provided by the battery which can support this load for 2 orbits (=3.2 hours) with a DOD of 90%. Power during normal operation and maximum power demand is provided by the solar array.

9.9.2 Mass budget

The overall spacecraft mass budget is given in Table 9.3. On all payload elements and spacecraft subsystems a mass contingency is added based on the level of maturity of the design. When existing hardware design is proposed, a margin of 5% is taken but for new designs a 20% margin is budgeted. In addition, a system margin of 20% of the spacecraft mass is added. To this must be added the launch vehicle adapter which leads to a total mass of 1165 kg. For the reference 550 km circular orbit, this leaves a 55 kg margin.

Table 9.3. Spacecraft mass budget.

Subsystem	S/S mass [kg] (no margins)	Contingency margin [%]	Contingency mass [kg]	Overall S/S mass [kg]
Cryo P/L (dry)	416.8	12.1	50.6	467.4
GPS	5.5	16.4	0.9	6.4
Laser S/S	2.0	20.0	0.4	2.4
Payload (dry)	424.3	12.2	51.9	476.2
OBDH	24.7	10.0	2.5	27.2
TT&C	9.9	5.0	0.5	10.4
EPSS	66.3	11.1	7.3	73.6
Solar Array	28.5	17.2	4.9	33.4
AOCMS	41.2	12.3	5.1	46.3
RCS	46.8	20.0	9.4	56.2
Structure	143.0	20.0	28.6	171.6
Thermal Control Harness	17.8	20.0	3.5	21.3
	30.0	20.0	6.0	36.0
Spacecraft (dry)	408.2	16.6	67.8	476.0
System margin (20 %)	---	----	95.2	95.2
Total S/C (dry)	408.2	39.9	163.0	571.2
Total sat (dry)	832.5	25.8	214.9	1047.4
SFHe	38.8	5.0	1.9	40.7
Satellite (wet)	871.3	24.9	216.8	1088.1
Launch V. Adapter	77.0	0.0	0.0	77.0
Total Launch Mass	948.3	22.9	216.8	1165.1
Launch Vehicle Performance	-----	-----	-----	1220.0
Margin w.r.t. LV performance	-----	-----	-----	54.9

10. Mission analysis

The experiment requirements, the spacecraft design constraints and demands, the profile of non-gravitational forces and the control authority of the helium thruster system, require to fly the STEP spacecraft in a circular Sun-synchronous orbit in heights between 375 and 600 km above the Earth. Table 10.1 contains a summary of the facts that lead to this "compromise orbit". The design driving facts are underlined. For given launcher performances and dispersions and because of geometrical and environmental conditions, feasible orbits can only be expected in certain launch windows for a given altitude. This will be outlined in what follows.

10.1 Sun-synchronous orbits, eclipses, and launch windows

Near Earth orbits outside the Earth shadow should be almost polar with ascending nodes $\pm 90^\circ$ away from the Sun longitude λ_\odot . Since the Sun's mean motion in longitude is $360/365.15^\circ/\text{day}$, their orbital nodes also move by that amount. Such orbits are called Sun-synchronous. Making use of the J_2 -effect of the Earth gravity one can obtain such a node rotation by a proper combination of the semi-major axis (a) and the inclination (i) of the orbit. The inclinations of these orbits are close to 97° , i.e. almost polar. One can furthermore see that launches in spring require orbital nodes at $\lambda_\odot - 90^\circ$, and launches in autumn require nodes at $\lambda_\odot + 90^\circ$.

Though the drag is compensated in the motion of STEP neither the node nor the Sun will move uniformly. The nodal motion is influenced by the gravity of the Sun, the Moon, and higher zonal

Table 10.1. Summary of orbit selection requirements.

Origin	Requirement	Orbit
STEP and gravity G experiment	<ul style="list-style-type: none"> - no out of plane forces - no orbital node drift - <u>S/C below radiation belt and outside South Atlantic Anomaly (charge accumulation !)</u> - <u>providing sufficient gravity</u> - S/C high enough for tolerable Earth gravity gradient and Earth infrared radiation - <u>mission duration ≥ 6 months</u> - S/C at constant height 	<p>polar in heights (h) below 800 km</p> <p>but above 400 km</p> <p>$\epsilon \approx 0$</p>
Geodesy co-exp.	<ul style="list-style-type: none"> -orbit polar and as low as possible -long mission duration 	$h = '0' \text{ km}$
Thruster control authority	<ul style="list-style-type: none"> - <u>maximum drag force ≤ 120 dyne</u> for cross sectional area $A \approx 4.2 \text{ m}^2$ and drag coefficient $C_D \approx 2.4$ 	<p>375 km $\leq h \leq$ 600 km</p>
S/C design -thermal -AOCS	<ul style="list-style-type: none"> - <u>Avoid Earth shadow</u> - avoid transition of density bulge 	<p>Sun-synchronous with node $\pm 90^\circ$ from mean Sun longitude</p>

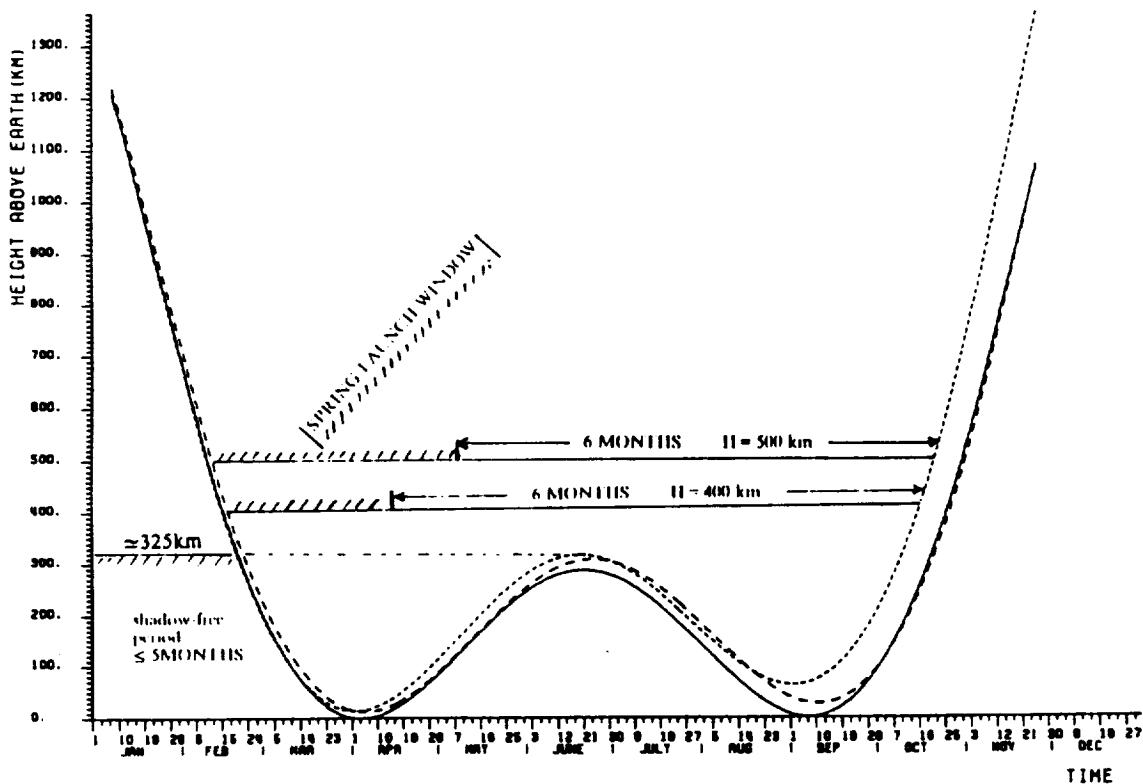


Fig. 10.1. Shadow heights for circular Sun-synchronous orbits in spring.

and tesseral terms, and the Sun longitude proper oscillates because of the eccentricity of the Earth orbit and its inclination with respect to the equatorial plane (the equation of time). The launcher will not be perfect. Injection errors in a and i will lead to further node drift anomalies.

A launch from a rotating Earth into a perfect node in inertial space can only be achieved at one single time of each day. Finite launch windows necessarily lead to node windows. A launch window of 30 minutes is equivalent to node biases of maximum $\pm 3.75^\circ$.

Even Sun-synchronous orbits with optimum nodes will enter the Earth shadow if their height drops below a certain boundary value (the shadow height). This height varies with the position of the Sun, though for a given initial node it becomes a well-defined function of time. Figure 10.1 shows the shadow heights for the nominal orbit after a spring launch. Also shown are the curves for two extreme bias cases with maximum node drifts of $\pm 0.02^\circ/\text{day}$ for injection errors and ± 0.25 hours of launch shift from nominal. A similar figure can be obtained for the autumn launches. One can see from such figures that, for a daily launch window of 1/2 hour, the eclipse-free period will suddenly drop below 5 months for perigee heights below 325/350 km for spring/autumn launches. For orbital heights at and above 350 km one obtains launch windows as shown in Table 10.2. For a launch from the Western Test Range (latitude=34.4°, longitude=120.6° West) the midpoint of the daily launch window is at about 2 h ($30 + 3.94d$) min where d is the number of days from spring/autumn equinox.

10.2 The drag-free condition and the minimum orbital height

The proportional helium thrusters on STEP must be able to compensate all non-gravitational forces and torques. One exception is the large torque due to the interaction of the mu-metal shield with the Earth magnetic field. This will be compensated using magnetorquers. Hence, most of the available helium thrust is needed for cancelling the air drag. The available omni-directional thrust

Table 10.2. Yearly launch windows for STEP orbits.

Height (km)	Spring (from)	Window (to)	open for (days)	Autumn (from)	Window (to)	open for (days)
350	Feb. 21-Apr. 13		51	Aug. 26-Oct. 12		47
400	Feb. 18-Apr. 16		57	Aug. 23-Oct. 16		54
450	Feb. 15-Apr. 19		63	Aug. 20-Oct. 19		60
500	Feb. 12-Apr. 22		69	Aug. 17-Oct. 22		66
550	Feb. 10-Apr. 27		79	Aug. 12-Oct. 28		77

authority in the orbit plane will be about 1.2 mN (120 dyne). The height of the STEP orbit must therefore be chosen such that the air drag does not exceed this value for extensive periods of time, though it is acceptable to lose drag free control for short periods corresponding to extreme solar activity.

The drag is a function of air density, spacecraft velocity relative to the air, and of satellite properties such as the drag-coefficient, C_D , and cross-sectional area, A . Spacecraft position and velocity are accurately known, and the surface interaction properties can be modelled fairly well. The following calculations are for $C_D = 2.4$ and $A = 4.2 \text{ m}^2$.

The density is, however, a complex function of spacecraft position and time. A good survey on density models can be found in Rees, 1989. The model used in the following calculations is 'MSIS' (Hedin et al., 1977 a, b). In the model, the solar flux, F10.7, and the geomagnetic activity index, A_p , are to be specified. These are the parameters which are strongly correlated with density variations. In order to predict the maximum drag force on the spacecraft, one needs predictions of F10.7 and A_p for the nominal orbit.

The solar flux shows a well known but not well predictable periodic variation over the 11 year solar activity cycle. The prediction of the geomagnetic activity is also prone to remarkable uncertainties. Figure 10.2 shows recorded and predicted data (Kerridge et al., 1989) for F10.7 and A_p for a time interval in which the STEP mission could take place. The predicted data are at a 95% probability level.

Combining the density model with the predicted solar and geomagnetic activity, one can predict the maximum drag over a mission period. The results can then be converted into minimum orbit altitudes pertaining to the available launch windows and thrust authority. Figure 10.3 shows the minimum heights below which the drag force will exceed the (1.2 mN) for the spring and autumn launch windows in 1998 - 2008. One can see from the figure that orbits below 450 km could be selected for launches in 1998 and in the years after 2004. For the years 1998 and 2004, this statement is not completely reliable because the prediction of the actual start or end of the period of highest solar activity after the year 1999 is not very reliable. On the other hand, experience with previous solar cycles suggests that the onset of the period of low activity can be predicted quite reliably, and hence we can reliably expect low activity and thus low densities for the few years following 2005.

The nominal STEP launch is scheduled for spring 2002. This dictates a nominal orbit above 550 km if one does not wish to risk losing the drag-free control for more than 5% of the mission duration. Final choice of the orbit height will be made closer to the launch date when the strength of the next solar cycle will be known. Any delay in the schedule will allow a lower orbit altitude.

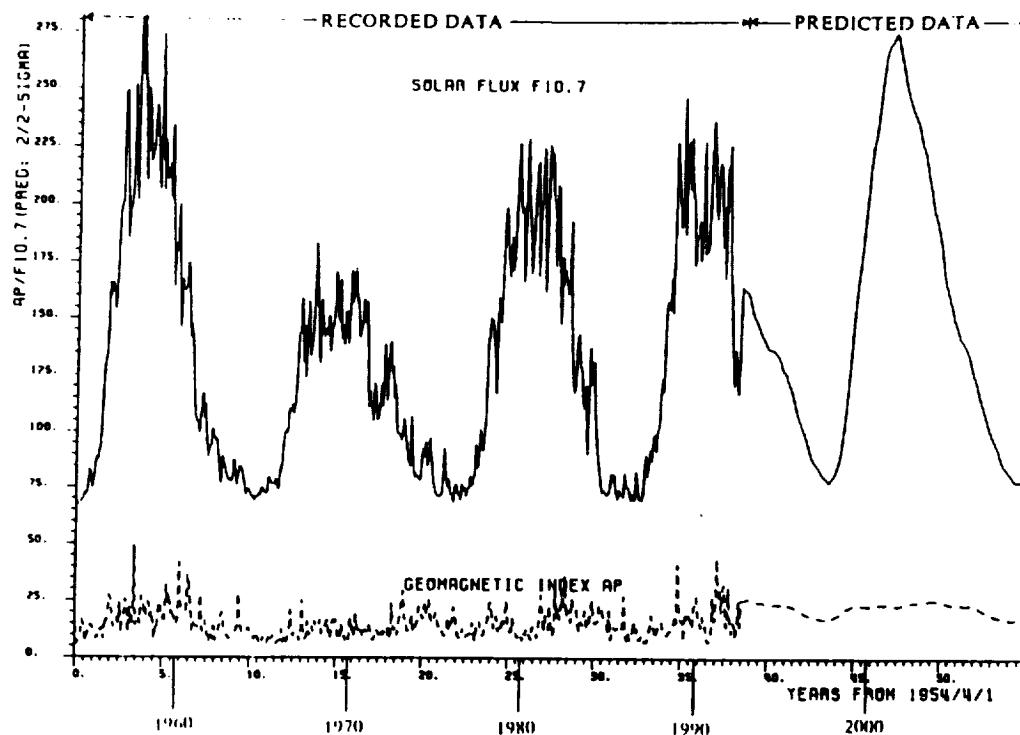


Fig. 10.2 History and prediction (95% probability level) for solar flux ($F_{10.7}$) and geomagnetic activity index (A_p).

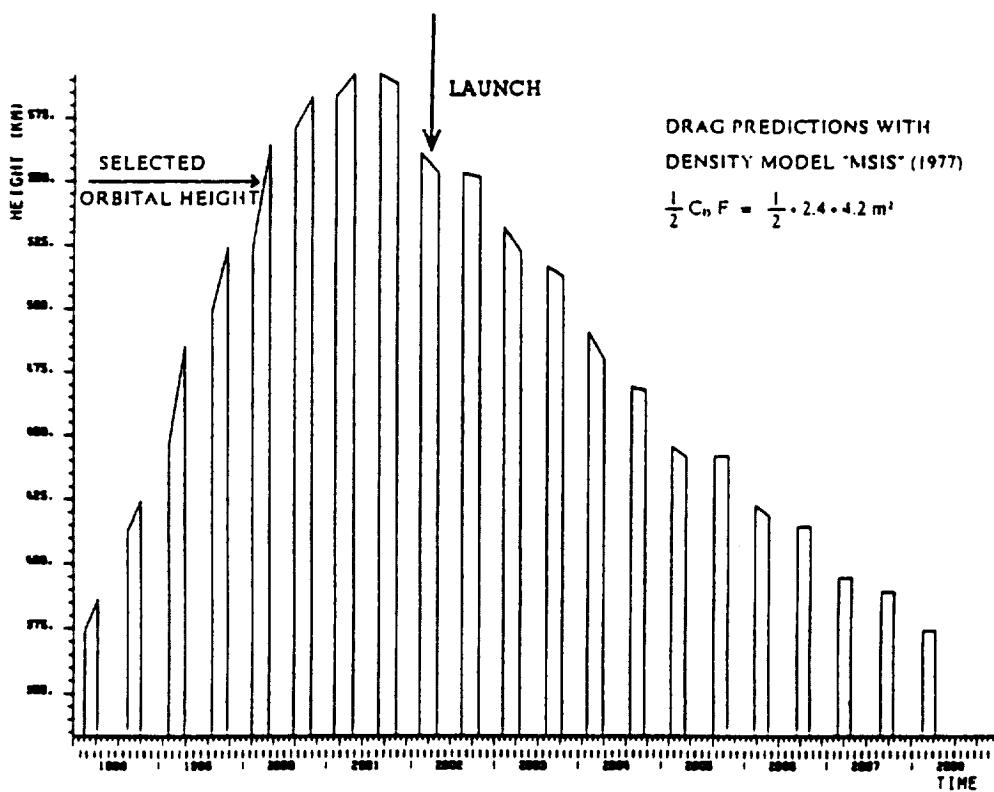


Fig. 10.3 Orbital heights below which the air-drag will exceed 1.2 mN.

10.3 Nominal STEP orbits

Figure 10.4 and Table 10.3 define the nominal orbits in the spring and autumn windows.

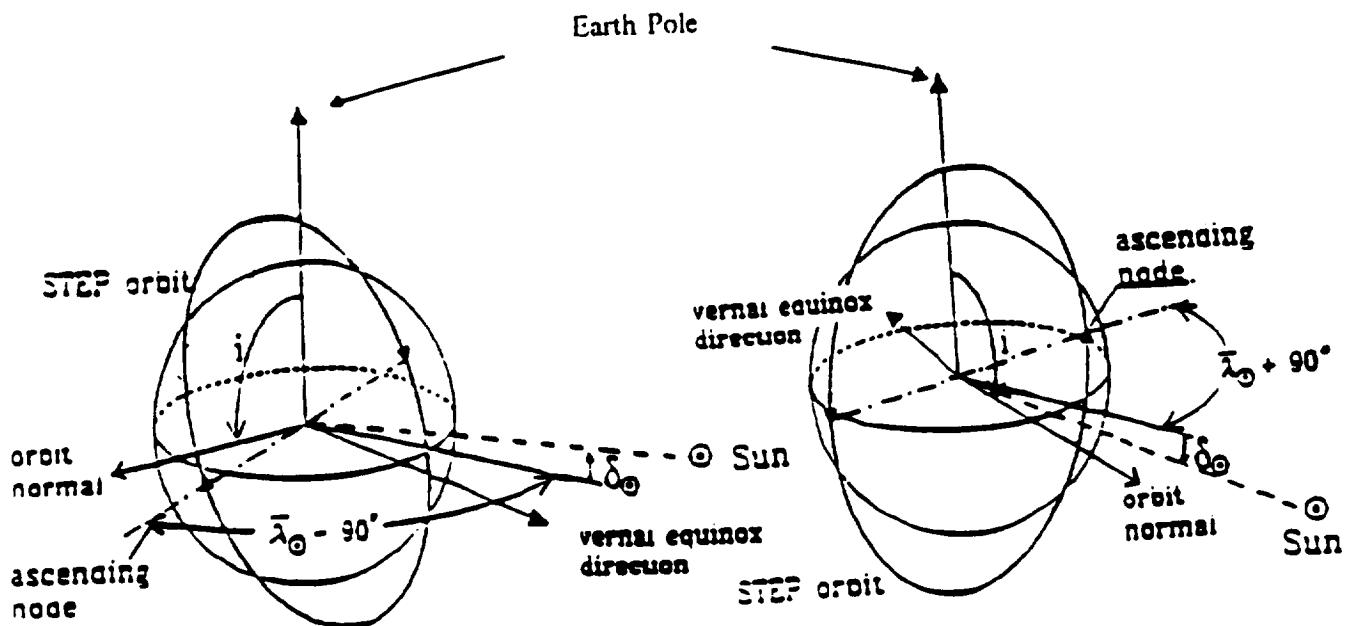


Fig. 10.4 Nominal STEP orbit geometry.

Table 10.3 Nominal STEP orbit elements.

	SPRING WINDOW	AUTUMN WINDOW
Epoch	TBD from launcher specific calcs. e.g. for a launch from WTR: $T = \text{March } 22 \text{ 2h } 22\text{m } 49.3\text{s}$	TBD from launcher specific calcs. e.g. for a launch from WTR: $T = \text{Sep. } 21 \text{ 14h } 20\text{m } 51.4\text{s}$
semi major axis a	6928.144 km	6928.144 km
eccentricity	0.000	0.000
orbit period	95.650 min	95.650 min
perigee height	550 km	550 km
apogee height	550 km	550 km
inclination i	97.565°	97.565°
ascending node Ω	$\bar{\lambda}_0(T) - 90^\circ$ -90.759° for above T	$\bar{\lambda}_0(T) + 90^\circ$ -90.386° for above T
node drift $\dot{\Omega}$	0.98563 deg/day	0.98563 deg/day
arg. of perigee ω	TBD from launcher specific calcs. 325.255° for above T	TBD from launcher specific calcs. 325.255° for above T
perigee drift rate $\dot{\omega}$	-3.419 deg/day	-3.419 deg/day
true anomaly ν	TBD from launcher specific calcs. -180° for above T	TBD from launcher specific calcs. -180° for above T

11. Launch vehicle, ground segment, mission operations

11.1 Launch vehicle

11.1.1 Baseline launch vehicle

STEP was always considered to be a shared mission with NASA, who, from the outset, offered to provide an appropriate launcher, operations, and a share of the payload module. The possibility of launching with Ariane 5, although in principle technically feasible, was not pursued because:

- the required STEP orbit (polar, Sun-synchronous) could not be achieved efficiently with Ariane 5, bearing in mind the low mass of the STEP spacecraft (about 1000 kg), which would mean finding a major partner to share the launcher;
- the cost of an Ariane 5 was not affordable within the budget allocated to the STEP mission.

The STEP spacecraft is baselined for launch on a Titan IIG from the Western Test Range into a near circular, Sun-synchronous, eclipse-free polar orbit with a nominal altitude of 550 km.

The Delta II is a backup to the Titan, and may be preferred by NASA. The Delta II has so much excess performance capability, that a shared launch with another payload compatible with STEP would be considered. The Delta II has higher modal frequency requirements than Titan IIG, but these have been taken into account in the structural design of STEP.

11.1.2 Titan II performance

As direct injection into a 550 km orbit would not allow sufficient payload mass, the proposed method is to inject into a parking orbit with apoapsis at the STEP orbit altitude and to use the Titan Attitude Control System (ACS) propulsion to perform a circularisation burn. The ACS Kit, with extra tanks can hold enough monopropellant to achieve the required ΔV .

For the baseline 550 km circular orbit, the payload capability is 1220 kg. This orbit is in the range of 200 to 600 km where a 100 km increase in altitude results in a 100 kg decrease in performance capability. For higher orbits the differential is 140 kg for every 100 km.

11.1.3 Satellite interface

The Titan payload adapter has a four-point attachment for the satellite on a 1742 mm diameter circle and incorporates a separation system. Figure 9.1(b) shows the satellite within the launcher shroud.

11.2 Tracking

11.2.1 Deep Space Tracking Network

The Deep Space Network (DSN) would provide tracking and data acquisition for the STEP project, using the 26 m network. Data is taken at 3.0 kbps, recorded on a 200 Mbit solid state recorder, and burst back on a S-band downlink to the 26 m network. The downlink data is transmitted at 865 kbps for 4 minutes every eight hours. Since the solid state recorder can store up to 19 hours of data, a DSN single pass can be missed without losing any STEP data.

The present spacecraft design uses a 1/3 watt transmitter that provides a carrier margin of 36 db and a data margin of 20 db.

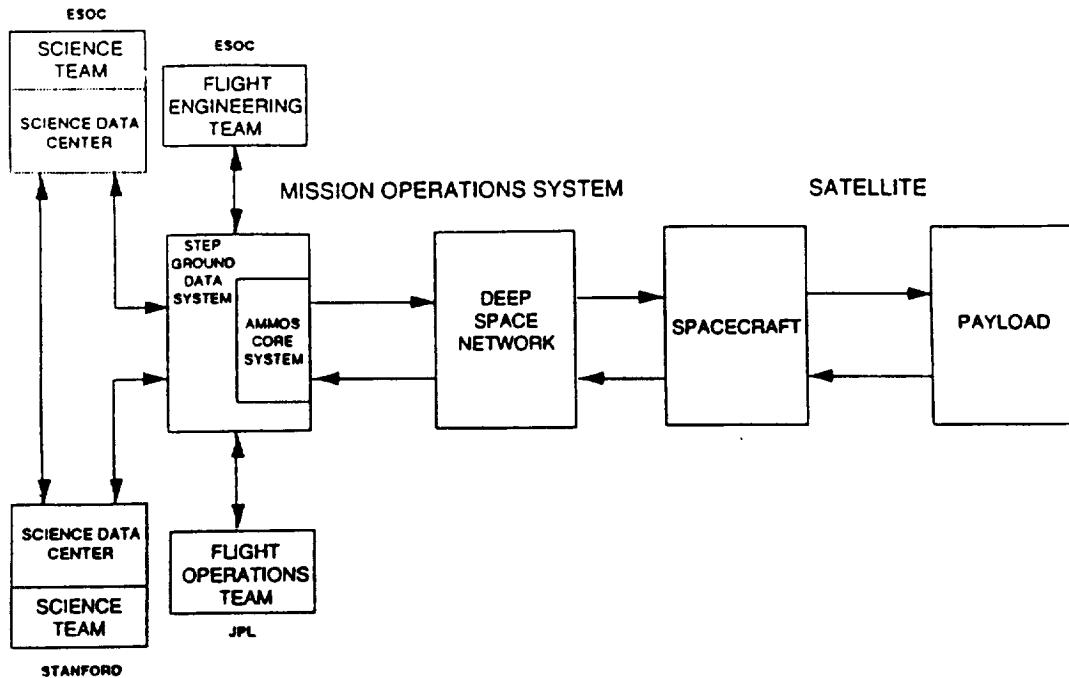


Fig. 11.1. Mission Operations System overview.

11.2.2 Global Positioning System (GPS) tracking

The DSN provides a 6-station GPS ground network used to improve knowledge of the GPS constellation's orbits. This improved knowledge of the locations of the GPS satellites would allow the onboard GPS receiver to accurately determine the position of the STEP spacecraft to a few centimetres. Currently, a 12-station GPS ground network is being considered in order to improve accuracy and network reliability. The locations of the present six GPS ground stations are: Goldstone, California; Madrid, Spain; Canberra, Australia; Santiago, Chili; Hartebeesthoek, South Africa and Usuda, Japan.

11.3 Mission Operations System

This section describes the option for operating the STEP spacecraft, as studied during the Phase A. Alternative operations concepts, using more European funding with a greater involvement of ESOC, may be considered for the eventual project, as the current estimates of the ESA costs show that more funds may be available to support the European contribution to the operations.

11.3.1 Characteristics of STEP Mission Operations System

The STEP Mission Operations System (MOS) is a set of functions, distributed on the ground, that operate cooperatively to control and monitor the operations of the STEP satellite and experiments, and to collect, transport, process, store and analyse the data and information of the mission. Figure 11.1 describes the top-level MOS architecture of STEP in the context of the End-to-End Information System (EEIS). Functionally, the STEP MOS can be divided into two processes: an uplink process and a downlink process. Figures 11.2 and 11.3 provide an overview of these two processes. The STEP MOS has the following key characteristics:

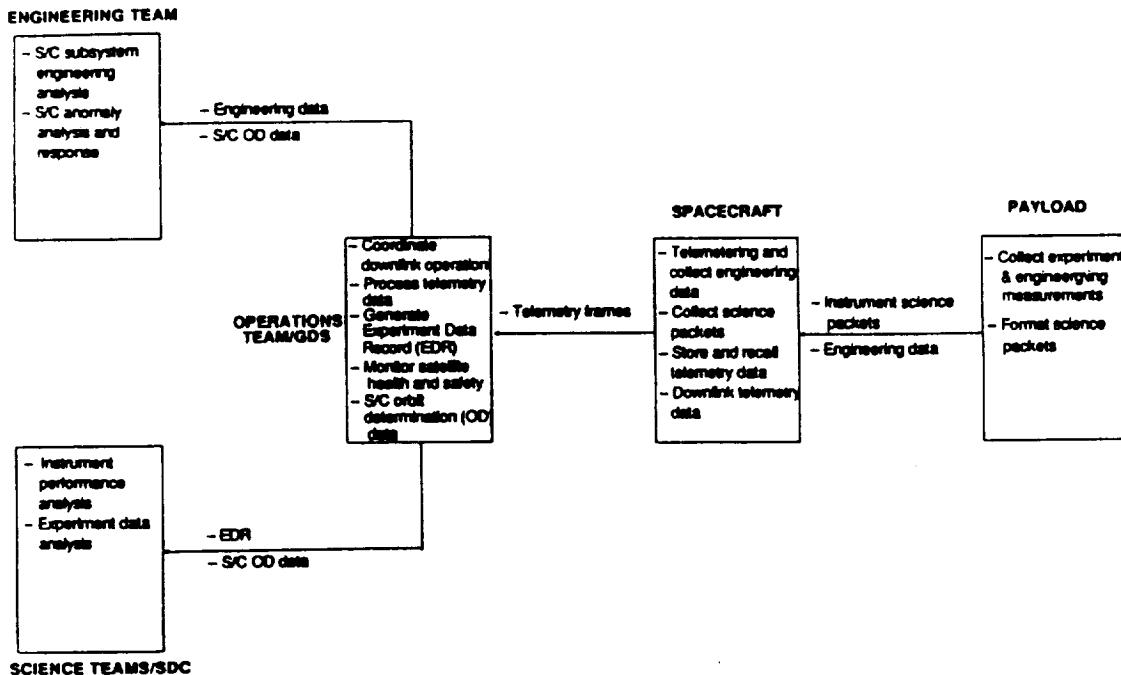


Fig. 11.2. STEP EEIS/MOS architecture uplink.

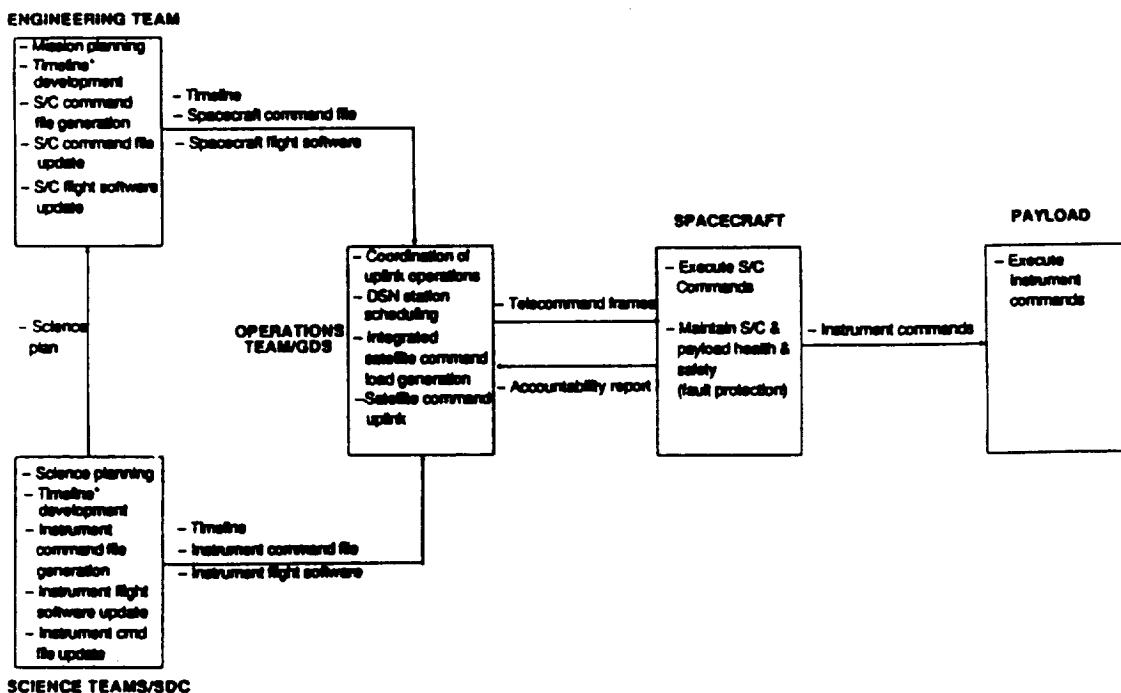


Fig. 11.3. STEP EEIS/MOS architecture downlink.

1. *Use of multi-mission capabilities.* The STEP MOS would use the capabilities provided by the JPL multi-mission systems. This includes the tracking and data acquisition services of the

DSN and mission control and data processing services available from AMMOS. The utilisation of the multi-mission systems by STEP has certain significant implications. For example, the STEP Ground Data System, as a project-specific element of the MOS, would be developed by integrating and adapting the multi-mission toolsets provided by AMMOS.

2. *Use of CCSDS interface protocol standards for uplink and downlink operations.* To ensure the highest degree of flight/ground compatibility and minimum development effort for the end-to-end data transport, the STEP MOS would use the channel services, data routing services, and packetization services as defined in CCSDS telecommand standards for uplink operations, and would also provide telemetry channel coding and packet telemetry services according to CCSDS telemetry standards for downlink operations.
3. *Distributed architecture.* The STEP MOS design is based on a geographically distributed architecture. In addition to the distributed DSN ground stations, the spacecraft operations would be conducted by the flight engineering team provided by ESOC and the payload operations by the science teams at their home facilities, while the overall mission operations coordination, on-line satellite monitoring and control, and data services would be performed by the flight operations team at JPL. In the baseline option, the ESOC Flight Engineering Team would be headed by a Manager who would be the ESA Representative in all matters relating to the mission operations.

12. Pre-launch activities, early in-orbit setup and calibration, science operations

12.1 Payload calibration and mission timeline

This section provides the preliminary schedule of events for the STEP mission. It covers all phases from instrument integration and calibration, through to final data reduction.

Instrument integration and calibration phase

The Equivalence Principle measurement programme begins prior to instrument assembly with a series of calibrations. Qualification of the test masses, which includes documentation of their mass, shape and density distribution, must be done before assembly. The masses should have their surfaces checked for contamination and electrical patch effect, to the extent possible on the ground. Other calibrations can only be performed after the instrument is partly or wholly assembled. These include calibration of the capacitance pickoff and electrostatic positioner for each test mass. These calibrations are used as standards for in-orbit calibration of the SQUID position detector. Also important is the calibration of the SQUID and magnetic bearing setup circuits. This includes measurement of the scale factors for sensitivity and spring constant as a function of current, as well as the scale factor for displacement of the equilibrium position. This information will be essential for quick in-orbit setup of the system to precalculated values. Calibration of the accelerometer thermal properties, that is, the heat capacity and thermal relaxation times, should also be done after assembly.

After final assembly and testing, the cryogenic instrument should be "heat flushed" by raising its temperature slightly above the superconducting transition temperature, in a small and uniform external field. This removes any accidentally trapped flux introduced into the instrument and superconducting shields during the early testing. When the temperature is carefully lowered again, the external low field will be trapped, giving a fresh environment. The cryogenic instrument should then be maintained in a cold (≈ 2 K) condition until launch, with no inputs except the minimum needed for testing.

Pre-launch phase

The prelaunch phase is defined as those activities needed to prepare the payload and spacecraft for launch and to verify their operation. It includes any maintenance needed prior to launch. A total of 31 days of pre-launch preparations are assumed.

Launch phase

The very brief launch phase is the period from lift-off until spacecraft release. One important operation to be performed during this phase is the opening of the cryostat vent valve.

Commissioning phase

Once in orbit the spacecraft is commissioned and the experiment enters an initialisation phase. This phase should last 1–2 weeks and is performed with close monitoring from the ground. After attitude stabilisation and after uncaging the test masses, the drag-free system may be turned on in a coarse mode using the capacitance pickoff for measurements. The masses must be moved into position with the electrostatic positioner before setup. This is because the flux trapped in each circuit for setup depends on its inductance, which depends strongly on mass position. The magnetic

bearings should be set up first, then the SQUIDs, using precalculated setup currents.

With the SQUIDs operating, the drag-free control can be switched to high accuracy mode, and calibration of the accelerometers and thrusters can begin. This is done by a combination of attitude and translation manoeuvres. First the outermost accelerometers are calibrated, by rotating the spacecraft around the proof mass, measuring the angular acceleration, and using the displacement of the outer accelerometers from the centre to calculate their absolute acceleration. This calibration is transferred to all accelerometers by a uniform translation. Bearing misalignment and intermode cross-coupling can also be measured during this phase. Many of these calibrations can be performed concurrently, either because they do not interfere with each other or because they use the same operations. Next the SQUID sensors are set up for accurate common-mode matching which is checked by dithering the spacecraft around them. The amplitude and phase of the residual differential acceleration is proportional to the correction which must be applied to the sensor. When there is no significant differential response to a common-mode acceleration, the sensor is balanced. The test mass centring procedure and gravity gradient attitude control reference should be done after the common-mode match. There may be a need to iterate these procedures if the initial setup is very different from the final. Some pre-measurement systematic checks, such as a background force measurement, are performed during the initialisation phase, and models of the apparatus are updated to reflect the real situation. This confirms that requirements on the apparatus and background forces have been met.

Measurement phase

The second and longest part of the science mission is the measurement phase. This is divided up into a number of measurements each lasting about 100 orbits.

The measurement cycle begins with a baseline measurement in normal orientation. This takes about a week, equivalent to 100 orbits plus setup time. Measurements are made in the other modes as described above, followed by a series of systematic checks. Each set of measurements might take 1–4 weeks and the associated checks a similar length of time. During this phase, modelling and analysis will indicate the probable level of Equivalence Principle violation and suggest changes to the setup and measurement procedure to confirm or deny the violation. The cycle begins again with re-initialisation as appropriate. The entire measurement cycle takes 1–3 months, and repeats until the helium refrigerant is exhausted.

It is not possible to completely specify in advance the series of measurements and checks that must be made. The disturbances that will be seen are either part of the error analysis, in which case we can design tests for them, or not. In the former case, the presence or absence of a disturbance at the predicted level determines the aim of the next set of measurements. If the disturbances have an unknown cause the schedule must be changed to study them.

Data-reduction phase

Finally, we estimate that about half a year will be necessary to completely analyse the data. This will be performed by four teams of scientists of the four fundamental physics experiments in Europe and the USA.

12.2 Science operations

A science operations plan will be established and agreed between the four experiment teams and the mission operations experts well before the launch.

This plan will be the improved and much more detailed version of the mission timeline which was worked out in detail during Phase A (not shown in this Report).

For the STEP mission lifetime of six months, a Flight Operations Team (FOT) will be established. The FOT will be accommodated in the Science Operations Centre (SOC), in close proximity to the Mission Operations Centre. The FOT will consist of representatives from each of the four fundamental physics experiment teams. At least one representative from each of the four fundamental physics experiment teams will be present in the SOC at all times to monitor the health of the experiment based on spacecraft housekeeping data, to analyse quick-look science data, and, if necessary, to request changes to the science operations plan. The experiment housekeeping and quick-look data will be available to the experimenters in the SOC in near-real time ($\leq 1/2$ h after each pass).

The full science data set will be transmitted within a few weeks to the home institutes (or other dedicated institutes) of the four PIs for detailed analysis.

13. Science management, programmatic and schedule

13.1 ESA/NASA collaboration

STEP is envisaged as a cooperative ESA/NASA mission. In the baseline scenario which was studied in the Phase A, NASA would provide

- the launch vehicle
- integration of the satellite into the launch vehicle
- part of the mission operations, including the Deep Space Network (DSN) and the mission operations centre
- three differential accelerometers to test the Equivalence Principle
- the quartz block assembly, the superfluid helium dewar, integration and testing of the complete payload at Stanford
- the proportional thrusters
- GPS receivers and antennas

ESA would provide

- the STEP spacecraft including integration and testing
- part of the mission operations, in particular, the spacecraft engineering team

In addition, European science institutes funded nationally would provide

- three differential accelerometers of a different design than the Stanford EP accelerometers and with different test mass materials
- two accelerometers to measure the constant of gravity G and to test the inverse square law
- one accelerometer to determine an upper limit for the coupling force between normal and spin-polarised matter
- integration and testing of the European payload elements before shipment to Stanford

Options on mission operations, involving more ESA funding, may be considered in Phase B.

13.2 Science management

The science management scheme outlined below is tailored to the special needs of the STEP mission. To see whether the structure envisaged for the post-approval phases of the project is practical, it was tried out as much as possible already during Phase A. The Science Study Team, chaired by the two Study Scientists formed three subgroups of experts in the fields of theory, hardware and geodesy (Fig. 13.1). This allowed to use the time of the experts most efficiently. The chairmen of the subgroups were automatically members of the Study Team. The Hardware Group had two chairmen, one from the US, one from Europe. The Hardware Team formed four "sub-subgroups" to study the four different experiments of the STEP payload in detail. The leaders of these four groups had the function of future Principal Investigators. Guidance on selection and definition of the payload was provided by the Theory Group, consisting of eminent European and US theorists. Each group met several times and reported to the Study Team at each of their meetings.

A Joint Planning Group, composed of the two Study Managers, the two Study Scientists and the ESA and NASA Headquarters Representatives, with two senior scientists from the Study Team as observers, oversaw the whole science study activity and made decisions regarding the overall

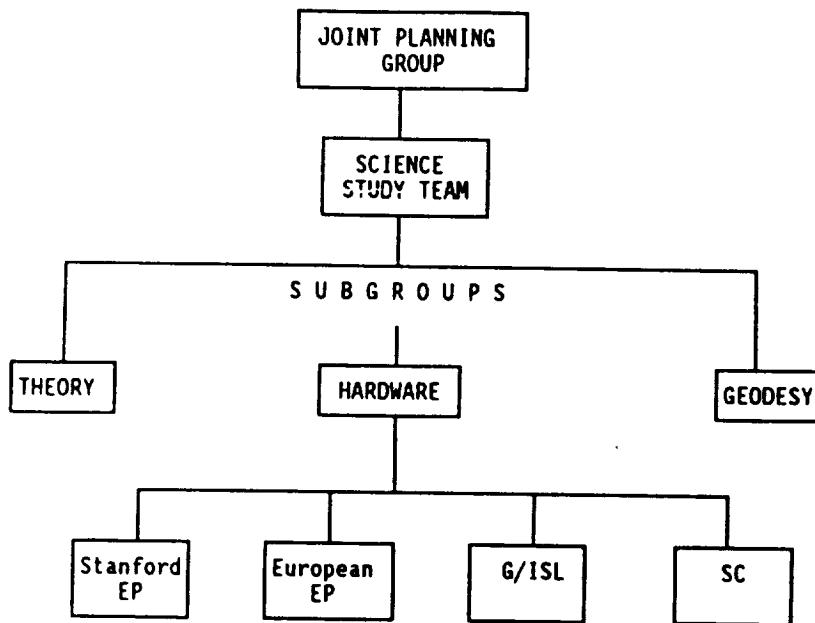


Fig. 13.1. STEP science management structure during Phase A.

direction of the study. This structure worked very smoothly and efficiently. As a result, the science payload could be dramatically improved during the Phase A study with a minor cost increase on the NASA side, no cost increase on the ESA side, and a significantly higher sharing of the payload by European science institutes. It is recommended to adopt this science management structure with some changes for the post-approval phases.

During the post-approval phases, the Joint Planning Group would be retained, possibly under a different name to reflect its advisory nature during Phase B rather than the management function that it had in Phase A. The previous Study Team and the Hardware Group would be merged into the Science Working Team (SWT). Its membership would comprise

- the ESA and the NASA Project Scientists and Project Managers
- the four PIs as the leaders of the four experiment teams and the four experiment managers
- the two engineers who are in charge of the pre-integration and testing of the European hardware and of the final integration and testing of the complete payload
- the two chairmen of the Theory Group and the Geodesy Group

This group would provide advice to the STEP Project Management. The STEP SWT would meet on average four times per year, with the venue alternating between Europe (ESTEC) and the USA (JPL). The meetings at ESTEC would be chaired by the ESA Project Scientist, the meetings at JPL by the NASA Project Scientist.

It is the task of the four experiment teams to develop, build, test and deliver the experiments according to the agreed schedule and to do the data analysis. Because of the highly integrated nature of the payload, the four teams would have to work closely together.

The four fundamental physics experiments would be selected competitively by a joint ESA/NASA selection committee. The selection would be based on proposals that would be submitted to ESA and NASA in response to a single joint ESA/NASA Announcement of Opportunity (AO).

While the four fundamental physics experiments would be nationally funded "PI experiments", the geodesy experiment would be a "facility". The hardware for this facility would be provided by JPL (GPS receivers and antennas) and by the G/ISL experiment team. A Geodesy Science Team composed of European and US scientists would be formed to ensure that the requirements for high quality and complete geodesy data are met, and that gravity field data analysis software is prepared. The selection of the members of the Geodesy Science Team would be based on proposals that would be submitted in response to the joint ESA/NASA AO described above.

At a later stage, about four years before launch, a second ESA/NASA AO would be issued, calling for proposals for two geodesy science centres whose task would be to perform the geodesy science data analysis in parallel.

The members of the Theory Advisory Group, which is composed of European and US scientists, would be appointed by ESA and NASA. The Theory Advisory Group would not meet at regular intervals, it would be convened ad hoc, if the need arises, by the two Project Scientists to clarify matters related to the theory of the four experiments and to provide theory support to the STEP payload as a whole, bearing in mind that the four fundamental physics experiments on STEP all address problems in the discipline area gravitation and new macroscopic interactions.

13.3 Archiving

According to the ESA policy on data rights, for the first six months after the end of the mission, the experimenter teams will have exclusive rights over their data. Thereafter, the data will have to be submitted to two science data centres where they can be accessed by the wide scientific community. The complete STEP data set will be available on CDROM at two STEP archive data centres, one in Europe (SSD/ESTEC), the other in the USA.

The complete STEP data set would comprise seven different data sets

- data from the European EP experiment
- data from the Stanford EP experiment
- data from the G/ISL experiment
- data from the SC experiment
- geodesy data
- aeronomy data
- charged particle data

The four fundamental physics data sets would be provided by the four experiment teams, the geodesy data by the two geodesy science centres, and the aeronomy data (essentially the calibrated drag-free controller data) and the data from the charged particle monitor by the Project Team with support from SSD.

The four fundamental physics data sets each include

- the respective science data set (raw and calibrated data)
- charged particle data
- selected spacecraft housekeeping data (this includes magnetometer data)
- software for the data analysis

The teams providing the various data sets have the following tasks

- performing a thorough end-to-end error analysis
- calibration of the science data
- development of appropriate software for data analysis
- production of an explanatory supplement
- timely (i.e. 6 months after mission end) delivery of the items above to the European and US STEP science data centres

The two STEP science data centres have the following tasks

- ensuring timely delivery of the items above
- verification of the contents of the CDROMs
- production of an appropriate number of copies of CDROMs and supplements
- responding to requests from the user community and sending out CDROMs and supplements as requested.

13.4 Management

After approval of the STEP project, a Memorandum of Understanding (MoU) and a Project Implementation Plan (PIP) will be prepared jointly by ESA and by NASA.

The MoU will define the agreement made between ESA and NASA for the implementation of the STEP project. It will contain, *inter alia*, agreements, responsibilities and deliverable items.

The PIP is a lower level document which defines the management structure of the project.

It is envisaged that the STEP project will be managed along similar principles to those applied to other cooperative projects. ESA will be responsible for overall project management and within this context NASA will be responsible for payload management. Special arrangements will be made to account for the intimate relationship between the spacecraft systems and the payload. These arrangements will include a joint working group with appropriate levels of management representation from ESA and NASA to resolve issues which stand to significantly impact either partner.

13.5 Schedule

The overall STEP schedule is given in Fig. 13.2. In this schedule the Phase B start is in late 1995. The Phase B duration is 15 months and the Phase C/L 50 months. An earlier start of the NASA activities (dewar and US payload) is possible, as soon as the instrument selection has been made. This is compatible with the required payload module deliveries. A payload QM is required in November 1998 which is the dewar qualification model plus structural-thermal simulators of the electronic units. This model needs to be upgraded to a model suitable for the engineering model (EM) test programme prior to September 1999 by replacing the electronics dummies with EM units. The flight payload needs delivering for integration and testing with the service module by June 2000. Corresponding delivery dates for the instrument are as follows:

- | | |
|------------------|----------------|
| - Instrument STM | December 1997 |
| - Instrument EQM | June 1998 |
| - Flight Unit | September 2000 |

Delivery of the satellite to the launch site must take place some months ahead of one of the two yearly launch windows (spring launch window: February 13 to April 17; autumn launch window: August 19 to October 19). In this planning, the satellite is delivered four months before the opening of the 2002 launch window. The launch is planned for the spring of 2002.

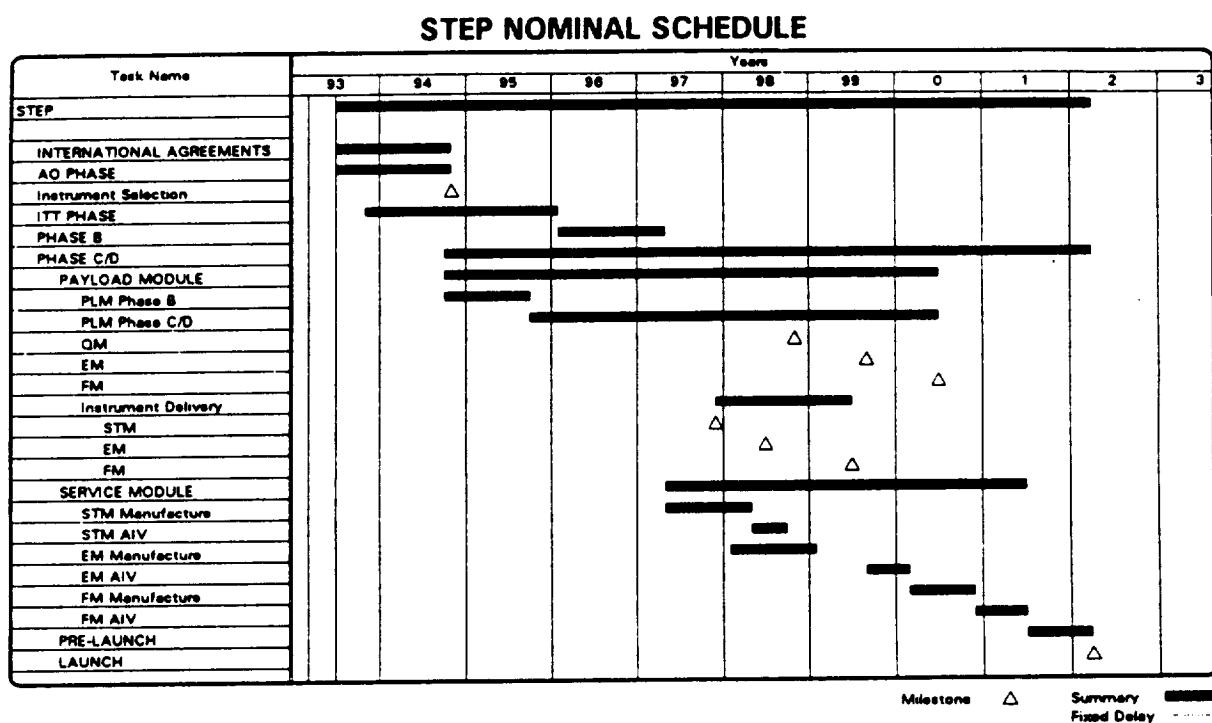


Fig. 13.2. STEP programme schedule.

References

- Adelberger, E. G. et al. (1990). Testing the equivalence principle in the field of the Earth: Particle physics at masses below 1 μeV ? *Phys. Rev. D*, **42**, 3267–3292.
- Adelberger, E. G., Heckel, B. R., Stubbs, C. W., and Rogers, W. F. (1991). Searches for new macroscopic forces. *Ann. Rev. Nucl. Part. Sci.*, **41**, 269–320.
- Balmino, G. (1986). Present status and future improvements in measuring the gravity field of the earth and planets. In *Space Geodesy and Geodynamics*, A. J. Anderson and A. Cazenave (Eds.), (pp. 19–54). Academic Press, London.
- Braginsky, V. B. and Panov, V. L. (1972). Verification of the equivalence of inertial and gravitational mass. *Soviet Physics JETP*, **34**, 463–466.
- Brun, R., Bruyant, F., Maire, M., McPherson, A., and Zanarini, P. (1984). GEANT3 User's Guide. DD/EE/84-1, CERN, 1984.
- Cavendish, H. (1798). Experiments to determine the density of the Earth. *Phil. Trans. Roy. Soc.*, **88**, 469.
- Chan, H. A. and Paik, H. J. (1987). Superconducting gravity gradiometer for sensitive gravity measurements. I, theory. *Phys. Rev. D*, **35**, 3551.
- Crow, B. (1992). Preliminary analysis of the Deep Space Network (DSN), support of STEP. NASA 2-009 Ref. WBS 2.3.6.
- Damour, T. and Nordtvedt, K. (1992). General Relativity as a cosmological attractor of tensor-scalar theories. Preprint IHES/P/92/94, submitted to *Phys. rev. Lett.*
- Damour, T. and Polyakov, A. M. (1993), *in preparation*.
- Damour, T. and Taylor, J. H. (1992). Strong-field tests of relativistic gravity and binary pulsars. *Phys. Rev. D*, **45**, 1840–1868.
- Fayet, P. (1986). A new long-range force? *Phys. Lett. B*, **171**, 261–266.
- Fayet, P. (1990). Extra U(1)'s and new forces. *Nucl. Phys. B*, **347**, 743–768.
- Feynman, J., Armstrong, T. P., Dao-Gibner, L., and Silverman, S. (1990). New interplanetary proton fluence model. *J. Spacecraft*, **27**, 403.
- Gravity Workshop (1987) :Geophysical and geodetic requirements for global gravity field measurements 1987-2000. Geodynamic Branch, NASA, Boulder, Colorado.
- Hedin, A. E. et al. (1977a). A global thermospheric model based on mass spectrometer and incoherent scatter data, MSIS 1, N₂ density and temperature. *Journal of Geophysical Research*, **82**, 2139–2147.
- Hedin, A. E. et al. (1977b). A global thermospheric model based on mass spectrometer and incoherent scatter data, MSIS 2, Composition. *Journal of Geophysical Research*, **82**, 2148–2156.
- Jafry, Y. R. (1992). *Aeronomy coexperiments on drag-free satellites with proportional thrusters: GP-B and STEP*. PhD thesis, Stanford University, Stanford, California.
- Jekeli, C. (1988). The gravity gradiometer survey system (GGSS). *EOS Transactions*, **69**(8), 105–117.
- Kerridge, D., Carlaw, V., and Beamish, D. (1989). Development and testing of computer algorithms for solar and geomagnetic activity forecasting. British Geological Survey T.R. WM/89/22C, Edinburgh.

- Kolb, E. W. and Turner, M. S. (1990). *The Early Universe*. Addison and Wesley.
- Lambeck, K. (1990). ARISTOTELES: An ESA mission to study the earth's gravity field. *ESA-Journal*, 14, 1-21.
- Lerch, F. J. et al. (1992). Gravitational modelling improvement for TOPEX/POSEIDON (abstract). *EOS Transactions, Fall Meeting Supplement*, 73(43), 125.
- Luther, G. G. and Towler, W. H. (1982). Redetermination of the Newtonian Gravitational Constant. *G. Phys. Rev. Lett.*, 48, 121-123.
- Moody, J. E. and Wilczek, F. (1984). New macroscopic forces ? *Phys. Rev. D*, 30, 130-138.
- Morel, P. (1990). The World Climate Research Programme, Science International. ICSU, no. 42.
- Mueller, I. and Zerbini, S. (1986). *The interdisciplinary role of space geodesy, lecture notes on Earth Sciences*, Volume 22. Springer, Heidelberg.
- Muellerschoen, R. J., Bertiger, W. I., and Sien C. Wu (1993). Gravity recovery analysis using GPS for STEP and a low-satellite mission. AAS/AIAA Spaceflight Mechanics Meeting, Pasadena, CA.
- Niebauer, T. M., McHugh, M. P., and Faller, J. E. (1987). Galilean test for the fifth force. *Phys. Rev. Lett.*, 59, 609-612.
- Paik, H. J., Canavan, E. R., Kong, Q., and Moody, M. V. (1992). Laboratory demonstrations of superconducting gravity and inertial sensors for space and airborne gravity measurements. In *From Mars to Greenland: Charting Gravity with Space and Airborne Instruments - Fields, Tides, Methods, Results*, O. S. Colombo (Ed.), Verlag, New York.
- Pavlis, E. C. (1992). Gravity field improvement prospects from the STEP mission. Presented at the 7th Intern. Symp. Geodesy and Physics of the Earth, Potsdam.
- Pavlis, N. K. (1988). Modeling and estimation of a low degree geopotential model from terrestrial gravity data. Dept. Geodetic Science and Surveying, 386, Ohio State University, Columbus, Ohio.
- Peccei, R. D., Solà, J., and Wetterich, C. (1987). Adjusting the cosmological constant dynamically: Cosmoms and a new force weaker than gravity. *Phys. Lett. B*, 195, 183-190.
- Pendlebury, J. et al. (1984). Search for a neutron electric dipole moment. *Phys. Lett.*, 136 B, 327.
- Rapp, R. H., Wang, Y. M., and Pavlis, K. K. (1991). The Ohio State 1991 Geopotential and Sea Surface Topography Harmonic Coefficient Models. Dept. Geodetic Sci. and Surveying, 410, The Ohio State University, Columbus, Ohio.
- Rees, D. (1989). Cospar Int. Reference Atmosphere: 1986, Part I: thermosphere models. *Advances in Space Research*, 8.
- Ritter, R., Winkler, L., and Gillies, G. (1993). Search for anomalous spin-dependent forces with a polarized-mass torsion pendulum. *Phys. Rev. Letts.*, 70, 701-704.
- Roll, P. G., Krotkov, R. V., and Dicke, R. H. (1964). The equivalence of inertial and passive gravitational mass. *Ann. Phys (NY)*, 26, 442-517.
- Rummel, R. (1980). Satellite Gradiometry, lecture notes in Earth Sciences, Vol. 7. In *Mathematical and Numerical Techniques in Physical Geodesy*, Sünkel, H. (Ed.), (pp. 318-363), Springer, Heidelberg.
- Scherk, J. (1979). Antigravity: a crazy idea ? *Phys. Lett.*, 88 B, 265-267.
- Schrama, E. J. O. (1992). STEP geodesy co-experiment: New results for an orbit at 450 km height. Internal memorandum, faculty of geodetic engineering, Delft.
- Spero, R., Hoskins, J. K., Newman, R., Pellam, J., and Schutz, J. (1980). Test of the Gravitational

References

- Inverse-Square Law at laboratory distances. *Phys. Rev. Lett.*, **44**, 1645–1648.
- Sumner, T. (1987). A calculation of the effect of a coaxial superconducting shield on the magnetic field distribution of an enclosed coaxial solenoid.
- Taylor, T. R. and Veneziano, G. (1988). Dilaton couplings at large distances. *Phys. Lett. B*, **213**, 450–454.
- TIMED (1991). Report of the Science Definition Team: Thermosphere-Ionosphere-Mesosphere Energetics and Dynamics (TIMED). Technical Report Volumes 1 and 2, NASA.
- Venema, B. J. et al. (1992). Search for a coupling of the Earth's gravitational field to nuclear spins in atomic mercury. *Phys. Rev. Lett.*, **68**, 135.
- Vitale, S., Bonaldi, M., Falferi, P., Prodi, G. A., and Cerdonio, M. (1989). Magnetization by rotation and gyromagnetic gyroscopes. *Phys. Rev. B*, **39**, 11993–12002.
- Weinberg, S. (1989). The cosmological constant problem. *Rev. Mod. Phys.*, **61**, 1–23.
- Will, C. M. (1992). The confrontation between General Relativity and experiment: a 1992 update. *Int. J. Mod. Phys. D*, **1**, 13–68.
- Worden, P. W. (1974). The Gyroscope Experiment. III tests of the Equivalence of Gravitational and Inertial mass based on cryogenic techniques. In *Experimental Gravitation*, B. Bertotti (Ed.), Academic Press, New York.
- Worden, P. W. (Jr.) (1976). *A cryogenic test of the Equivalence Principle*. PhD thesis, Stanford University, Stanford, California.

List of Acronyms

ACS	Attitude Control System
AIV	Assembly, Integration and Verification
AMMOS	Advanced Multi-Mission Operations System
AO	Announcement of Opportunity
AOCMS	Attitude and Orbit Control and Measurement System
AOCS	Attitude and Orbit Control System
ARISTOTELES	Applications and Research Involving Space Techniques Observing The Earth field from low-Earth orbiting Satellite
BIPM	Bureau International des Poids et Mésures
BRIDGE	not an acronym
CCSDS	Consultative Committee for Space Data Systems
CEU	Control Electronics Unit
CDMU	Central Data Management Unit
CDROM	Compact Disc Read Only Memory
CFRP	Carbon Fibre Reinforced Plastic
CNR	Consiglio Nazionale delle Ricerche
CP	Charge conjugation Parity
D/A	Digital to Analog
DAP	Differential Accelerometer Package
DBI	Digital Bus Interface unit
DBU	Data Bus Unit
DoD	Department of Defence
DoD	Depth of Discharge
DORIS	Doppler Orbitography and Radio Positioning Integrated by Satellite
DRAM	Digital Random Access Memory
DRS	Data Relay Satellite
DSN	Deep Space Network
EEIS	End-to-End Information System
EM	Engineering Model
EP	Equivalence Principle
EPS	Electrostatic Positioning System
EQM	Engineering Qualification Model
ERS	European Remote Sensing (satellite)
ESA	European Space Agency
ESOC	European Space Operations Centre
ESTEC	European Space Research and Technology Centre
FF	Free Fall
FM	Flight Model
FOT	Flight Operations Team
GAMES	Gravity and Magnetic Earthprobe Studies
GEM-T2	Goddard Environment Model T2
GG	Gravity Gradient
G/ISL	Constant of gravity G and Inverse Square Law
GP-B	Gravity Probe B
GPS	Global Positioning System
GRM	Geopotential Research Mission
GSFC	Goddard Space Flight Center
IAS	Institute for Advanced Study (in Princeton)

ID dewar	Internal Development dewar (Lockheed)
IHES	Institut des Hautes Etudes Scientifiques
ISL	Inverse Square Law
ITT	Invitation To Tender
JGM-1S	Joint Gravity Model 1S
JPL	Jet Propulsion Laboratory
LAEFF	Laboratorio de Astrofisica Espacial y Fisica Fundamental
MLI	Multi-Layer Insulation
MMU	Memory Mass Unit
MOS	Mission Operations System
MoU	Memorandum of Understanding
MSSL	Mullard Space Science Laboratory
N	Normal (mode)
NASA	National Aeronautics and Space Administration
NMR	Nuclear Magnetic Resonance
OBDH	Onboard Data Handling
OSR	Optical Surface Reflectors
PE	Precision Electronics
PI	Principal Investigator
PIP	Project Implementation Plan
PPN	Parametrised Post Newtonian formalism
PRARE(E)	Precise Range And Range-rate Equipment (Extended)
PRN	Pseudo-Random Noise
QM	Qualification Model
R	Rotated (mode)
RAL	Rutherford Appleton Laboratory
RBI	Remote Bus Interface
RTU	Remote Terminal Unit
SC	Spin Coupling
SLALOM	Space Laser Low-Orbit Mission
SLR	Satellite Laser Ranging
SOC	Science Operations Centre
SQUID	Superconducting Quantum Interference Device
SSD	Space Science Department (at ESTEC)
STEP	Satellite Test of the Equivalence Principle
STM	Structural Thermal Model
SWT	Science Working Team
T	Turning (mode)
TIMED	Thermosphere-Ionosphere-Mesosphere Energetics and Dynamics
TOPEX/	Ocean Topography Explorer
POSEIDON	not an acronym
TRIAD	Tracking, Telemetry and Command
TT&C	Technical University
TU	Vapour Cooled Shields
VCS	