CAS
CERN ACCELERATOR SCHOOL
APPLIED GEODESY FOR PARTICLE ACCELERATORS

CERN, Geneva, Switzerland
14-18 April, 1986

PROCEEDINGS
Editor: S. Turner
Propriété littéraire et scientifique réservée pour tous les pays du monde. Ce document ne peut être reproduit ou traduit en tout ou en partie sans l'autorisation écrite du Directeur général du CERN, titulaire du droit d'auteur. Dans les cas appropriés, et s'il s'agit d'utiliser le document à des fins non commerciales, cette autorisation sera volontiers accordée.

Le CERN ne revendique pas la propriété des inventions brevetables et dessins ou modèles susceptibles de dépôt qui pourraient être décrits dans le présent document; ceux-ci peuvent être librement utilisés par les instituts de recherche, les industriels et autres intéressés. Cependant, le CERN se réserve le droit de s'opposer à toute revendication qu'un usager pourrait faire de la propriété scientifique ou industrielle de toute invention et tout dessin ou modèle décrits dans le présent document.

© Copyright CERN, Genève, 1987

Literary and scientific copyrights reserved in all countries of the world. This report, or any part of it, may not be reprinted or translated without written permission of the copyright holder, the Director-General of CERN. However, permission will be freely granted for appropriate non-commercial use.

If any patentable invention or registrable design is described in the report, CERN makes no claim to property rights in it but offers it for the free use of research institutions, manufacturers and others. CERN, however, may oppose any attempt by a user to claim any proprietary or patent rights in such inventions or designs as may be described in the present document.
ABSTRACT

This specialized course addresses the many topics involved in the application of geodesy to large particle accelerators, though many of the techniques described are equally applicable to large construction projects and surveillance systems where the highest possible surveying accuracies are required. The course reflects the considerable experience gained over many years, not only at CERN but in projects all over the world. The methods described range from the latest approach using satellites to recent developments in conventional techniques. They include the global positioning system (GPS), its development, deployment and precision, the use of the Terrameter and the combination or comparison of its results with those of the GPS, the automation of instruments, the management of measurements and data, and the highly evolved treatment of the observations.
## Course on Applied Geodesy for Particle Accelerators, CERN, 14th-18th April, 1986

<table>
<thead>
<tr>
<th>Time</th>
<th>Monday 14th April</th>
<th>Tuesday 15th April</th>
<th>Wednesday 16th April</th>
<th>Thursday 17th April</th>
<th>Friday 18th April</th>
</tr>
</thead>
<tbody>
<tr>
<td>08.30</td>
<td><strong>GLOBAL POSITIONING SYSTEM (GPS)</strong></td>
<td><strong>HIGH PRECISION GEODESY FOR PARTICLE ACCELERATORS</strong></td>
<td>9.00</td>
<td>9.45</td>
<td>10.30</td>
</tr>
<tr>
<td>09.00</td>
<td>REGISTRATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09.15</td>
<td>Opening Address</td>
<td>The LEP Trilateration Network</td>
<td>High Precision Geodesy at CERN</td>
<td>Computer-aided Geodesy for LEP Installation (I)</td>
<td>On continuous theory for Geodetic Networks - K. Borre</td>
</tr>
<tr>
<td>10.00</td>
<td>A. Allan</td>
<td>J. Gervaise</td>
<td>J. Gervaise</td>
<td>W. Coosemans</td>
<td>New Trends in Mathematical Geodesy</td>
</tr>
<tr>
<td>10.30</td>
<td><strong>COFFEE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.30</td>
<td>GPS Development and Deployment</td>
<td>Deviation of the Vertical</td>
<td>Gyroscope Technology Status and Trends</td>
<td>Computer-aided Geodesy for LEP Installation (II)</td>
<td></td>
</tr>
<tr>
<td>11.30</td>
<td>S. Baker</td>
<td>B. Burki</td>
<td>W. Caspary</td>
<td>M. Mayoud</td>
<td>J.P. Queue \n</td>
</tr>
<tr>
<td>11.30</td>
<td>GPS Receiver Technology</td>
<td>Comparison between Terrameter and GPS Results</td>
<td>Underground Geodesy for LEP</td>
<td>Experiments in Metrology</td>
<td></td>
</tr>
<tr>
<td>12.30</td>
<td>C. Boucher</td>
<td>C. Beutler</td>
<td>M. Mayoud</td>
<td>C. Lasser</td>
<td></td>
</tr>
<tr>
<td>12.30</td>
<td><strong>LUNCH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.00</td>
<td>Precise Relative Positioning with GPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.00</td>
<td>C. Goad</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.30</td>
<td>GPS Orbit Corrections and Refinements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.00</td>
<td>C. Beutler</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.00</td>
<td><strong>TEA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.30</td>
<td>Accuracy Problems when Combining Satellite &amp; Terrestrial Observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.00</td>
<td>W. Welsch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.30</td>
<td><strong>WELCOME COCKTAIL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.30</td>
<td><strong>BANQUET</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CERN ACCELERATOR SCHOOL

APPLIED GEODESY
FOR PARTICLE ACCELERATORS

14–18 April 1986 at CERN
CH-1211 Geneva 23, Switzerland

Lectures:
Global Positioning System (GPS),
   Development and Deployment
GPS Receiver Technology
Precise Relative Positioning with GPS
GPS Orbit, Corrections and Refinements
Accuracy Problems when Combining Terrestrial
   and Satellite Observations
The LEP Trilateration Network Measured
   with a Terrameter
Deviation of the Vertical
Comparison between Terrameter and GPS Results
High-precision Geodesy at CERN
Gyroscope Technology, Status and Trends
Underground Geodesy for LEP
Computer-aided Geodesy for LEP Installation

Experiments Metrology
Tridimensional Adjustment in a Local Reference
   System
Deformation Analysis using Datum Invariant
   Geodetic Elements
Discontinuous Theory for Geodetic Networks
New Trends in Mathematical Geodesy

Seminars:
Very-long Base Interferometry
Future Accelerators
Challenge of the Future for Geodesy

Visit:
CERN Geodesy Installation

General information Participants should have a basic knowledge of geodesy. Persons wishing to attend this course can obtain further information and application forms from the CERN Accelerator School, Mrs. S. von Wartburg, LEP Division, CH 1211 Geneva 23. Application forms must be returned no later than 10 February 1986. A registration fee of SFr 200.— will be charged. All participants will receive a copy of the proceedings. The number of participants is limited to 120. The CERN Accelerator School Secretary, Mrs. S. von Wartburg will be pleased to help participants find accommodation in or near to CERN and a courtesy bus service will be provided to serve selected, nearby hotels.

Administration Head of School: P.J. Bryant; Organizing Committee: J. Gervaise, P.J. Bryant, M. Mayoud, S. Turner, S. von Wartburg, S. Beauquis.
FOREWORD

The CERN Accelerator School (CAS) was founded in 1983 to preserve and disseminate the knowledge accumulated at CERN and elsewhere on particle accelerators and storage rings of all kinds. Subsequently, a biennial programme of basic and advanced courses on general accelerator physics was set up. This is supplemented by specialized and topical courses, and workshops which are arranged as the needs arise. The present proceedings on "Applied Geodesy for Particle Accelerators" is the latest in the series of specialized courses.

Since the 1920's the positioning of particle accelerator components has evolved from a bench-top problem to one involving tunnels of tens of kilometres in extent while still maintaining sub-millimetric accuracies. Recent machines have required this accuracy to be transmitted from the surface into deep underground tunnels, while for future machines the trend towards particle beams of micron or sub-micron dimensions will impose even stricter alignment and stability conditions with overall dimensions still of the order of many kilometres. These proceedings describe the current methods for achieving the highest accuracies, notably with the Navstar Global Positioning System and the Terrameter, taking into account the ellipsoid and local geoid. Also described is the automation of such instruments as the gyroscope and Distinvar and the highly evolved treatment of all the observations necessary in order to accurately locate the many thousands of components in a modern accelerator.

The geodesy course and the publication of its proceedings would not have been possible without the generous support of many people. On behalf of the CAS we thank the CERN Directorate for their sustained help and encouragement. In particular, we thank J. Gervaise, the (now retired) Head of CERN's Applied Geodesy Group for proposing this course and for his tireless efforts to make it a success. The lecturers also deserve special appreciation for so freely giving their time and effort to prepare the lectures and their chapters in the proceedings. The contributions of the personnel in the different CERN services who assisted in the smooth running of the course and the publication of these proceedings are too numerous to mention individually but are gratefully acknowledged. Finally, we are indebted to the participants in the course whose presence made it worthwhile.

P.J. Bryant, Head of CAS
S. Turner, Editor
## CONTENTS

### Addresses and Speech

- **J. Gervaise**
  - Welcoming address
  
- **A.L. Allan**
  - Opening address - The People of Geodesy

- **E. Picasso**
  - After Dinner Speech

- **J. Gervaise**
  - Closing Address - Challenge of the Future for Geodesy

### Lectures

- **L.S. Baker**
  - **GPS Its development and deployment**
    - Development
    - Geodetic receivers
    - Deployment
    - Summary

- **C. Boucher**
  - **GPS receiver technology**
    - The GPS system concept
    - The NAVSTAR space segment
    - Measurement techniques
    - Error budget and measurement modelling
    - Review of existing receivers
    - Conclusion

- **C.C. Good**
  - **Precise positioning with the GPS**
    - Introduction
    - The phase observable
    - Data problems
    - Single site cycle fixing
    - Filtering techniques
    - Application to dynamic problems
    - Unmodelled errors

- **G. Beutler**
  - **GPS orbit determination using the double difference phase observable**
    - Introduction
    - Precision requirements
    - Principles of orbit determination
    - Applications
# Accuracy problems when combining terrestrial and satellite observations

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>66</td>
</tr>
<tr>
<td>The accuracy of satellite aided network observations</td>
<td>67</td>
</tr>
<tr>
<td>The accuracy of terrestrial network observation results</td>
<td>67</td>
</tr>
<tr>
<td>Comparison of satellite and terrestrial observation results</td>
<td>68</td>
</tr>
<tr>
<td>The combination of terrestrial and satellite aided network observations</td>
<td>69</td>
</tr>
<tr>
<td>- the functional model</td>
<td>69</td>
</tr>
<tr>
<td>- the stochastic model</td>
<td>71</td>
</tr>
<tr>
<td>- results</td>
<td>73</td>
</tr>
<tr>
<td>Changes in positions and heights due to the network combination</td>
<td>75</td>
</tr>
<tr>
<td>Conclusions</td>
<td>76</td>
</tr>
<tr>
<td>Appendix A - S-transformation</td>
<td>77</td>
</tr>
<tr>
<td>Appendix B - Variance component estimation</td>
<td>79</td>
</tr>
<tr>
<td>Appendix C - The Inn valley network</td>
<td>82</td>
</tr>
</tbody>
</table>

# The LEP trilateration network

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>85</td>
</tr>
<tr>
<td>Principle of measurement with instruments using two different wavelengths in the visible spectrum</td>
<td>87</td>
</tr>
<tr>
<td>LDM 2 terrameter (Terratechnology Corporation)</td>
<td>90</td>
</tr>
<tr>
<td>Measurement of the LEP geodetic network with the terrameter</td>
<td>94</td>
</tr>
<tr>
<td>Conclusion</td>
<td>103</td>
</tr>
</tbody>
</table>

# Deviation of the vertical

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>105</td>
</tr>
<tr>
<td>The gravity field of the earth</td>
<td>105</td>
</tr>
<tr>
<td>Some examples</td>
<td>110</td>
</tr>
<tr>
<td>Determination of ( (\xi, \eta) ), ( \phi ), ( N )</td>
<td>112</td>
</tr>
<tr>
<td>Applications on the CERN-LEP control network</td>
<td>115</td>
</tr>
</tbody>
</table>

# Comparison between terrameter and GPS results - and how to get there

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measuring engineering type networks</td>
<td>119</td>
</tr>
<tr>
<td>The 1994 CERN macrometer campaign</td>
<td>122</td>
</tr>
<tr>
<td>The 1995 CERN SERCEL campaign</td>
<td>125</td>
</tr>
</tbody>
</table>

# High precision geodesy applied to CERN accelerators

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>128</td>
</tr>
<tr>
<td>Transverse coordinates</td>
<td>130</td>
</tr>
<tr>
<td>Bending magnets and magnetic rigidity</td>
<td>132</td>
</tr>
<tr>
<td>Quadrupole magnets and focusing</td>
<td>134</td>
</tr>
<tr>
<td>The gutter analogy of focusing</td>
<td>135</td>
</tr>
<tr>
<td>Alternating gradient focusing</td>
<td>137</td>
</tr>
<tr>
<td>The equation of motion</td>
<td>138</td>
</tr>
<tr>
<td>Floquet coordinates</td>
<td>139</td>
</tr>
<tr>
<td>Orbit distortion from one kick</td>
<td>140</td>
</tr>
<tr>
<td>Sources of closed orbit distortion</td>
<td>142</td>
</tr>
<tr>
<td>Fourier analysis of orbit distortion</td>
<td>143</td>
</tr>
<tr>
<td>Local orbit correction</td>
<td>145</td>
</tr>
<tr>
<td>CERN proton synchrotron (CPS) (1954-1959)</td>
<td>146</td>
</tr>
<tr>
<td>Intersecting storage ring (ISR) (1966-1971)</td>
<td>153</td>
</tr>
<tr>
<td>Super proton synchrotron (SPS) (1971-1976)</td>
<td>154</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>158</td>
</tr>
<tr>
<td>Length measurement devices</td>
<td>159</td>
</tr>
<tr>
<td>Offset measurement devices</td>
<td>161</td>
</tr>
<tr>
<td>Conclusion</td>
<td>163</td>
</tr>
</tbody>
</table>
W.F. Caparay

Gyroscope technology, status and trends
  Introduction 166
  Historical background 167
  The free gyroscope 168
  The suspended gyrocompass 169
  Equations of motion 171
  Observation methods 172
  Automated observation 174
  Calculation of the azimuth 176
  Accuracy considerations 177
  Non-conventional gyroscopes 179
  Conclusion 181

J.-C. Fiscoher, M. Hayotte, M. Mayoud, G. Trouche

Underground geodesy 183
  Introduction 183
  Link between surface and underground geodesy 185
  Guiding control with gyro traverse 187
  Transverse geometrical controls of underground works 198
  Conclusion 207

M. Hublin

Computer aided geodesy for LEP installation 209
  Part I: The installation procedure 209
    Introduction: some information from the survey point of view 209
    LEP, what does it imply for us? 209
    The installation procedure 210
    Conclusion 218

W. Coosemans

Computer aided geodesy for LEP installation 219
  Part II: Instrumentation 219
    Introduction 219
    Instruments 220
    Terminals and field computers 222
    The office computer 223
    CERN computer centre 224

M. Mayoud

Applied metrology for LEP 233
  Part I: Computing and analysis methods 233
    Introduction 233
    Theoretical positioning of LEP 234
    Adjustment of observations 240
    Expected results for LEP metrology 243

J.-P. Queenel

Applied metrology for LEP 248
  Part II: Data logging and management of geodetic measurements with a database 248
    Introduction 248
    Data architecture 248
    Using the database 251
    Conclusion 254
C. Lasseur

**Metrology for experiments**

- Introduction
- How experiments are carried out at CERN
- The surveyor's role
- Surveying procedure
- Means available in the field for the surveying of large physics experiments
- Computer processing
- Specific features for the LEP experiments
- Conclusion

P. Baldi and M. Unguendoli

**Geodetic networks for crustal movements studies**

- Introduction
- Some examples of ground deformations
- Geodetic measurements
- Conclusions

J. Iliffe

**Three-dimensional adjustments in a local reference system**

- Introduction
- The geodesy of LEP
- The observations
- The normal equations
- The input procedure
- Relative errors
- The results obtained

H. Pelzer

**Deformation analysis*)**

H.M. Dufour

**Inverse regional reference systems: a possible synthesis between three- and two-dimensional geodesy**

- The historical perspective of the problem
- Two-dimensional problems on the sphere
- Exploitation of the stereo representation of the sphere
- Two-dimensional problems on the ellipsoid
- Three-dimensional problems on the sphere
- General spatial problems
- Regional inverses of the ellipsoid (E)
- Three-dimensional corrections
- Possible uses
- Conclusion

K. Borre

**On continuous theory for geodetic networks**

- Introduction
- Systematic analysis of CERN-like networks
- Conclusion
- Appendix - Covariance matrices/functions for various 1-D geodetic problems

*) Contribution not received
New trends in mathematical geodesy

Handling large data sets 362
The power of symbolism 363
Hilbert space techniques 365
Spectral methods 368
Concluding remarks 373

Seminars

J. Campbell

Very long baseline interferometry 375
Introduction 375
Fundamentals of the VLBI-technique 376
Scientific interest; programs and results 396

K. Johnsen

Future accelerators 396
Introduction 396
Studies of future hadron colliders 398
Prospects for electron-positron colliders 402
Concluding remarks 409

List of participants
WELCOMING ADDRESS

J. Gervaise
CERN, Geneva, Switzerland

On behalf of CERN's Survey Group, it is with great pleasure that I welcome you to our School on Applied Geodesy for Particle Accelerators. I am very pleased to see how numerous you are. Among the faces I see many old friends, thank you for coming. I also see many familiar faces of friends and colleagues with whom I have had close and fruitful cooperation in the past, thank you for coming. There are also many new and younger faces who will, no doubt, be organising and lecturing in the schools of the future. I look forward to making their acquaintance over the next few days and thank them for coming.

As you all know, CERN is presently constructing LEP, the Large Electron-Positron Collider. There is neither the time nor the necessity to go into the details of this project at this moment but, at least, let me remind you that, in a tunnel some 27 km long and many tens of metres below the surface, we will be required to align several thousands of the accelerator's machine components to an accuracy of tenths of a millimetre. While CERN is well used to installing its equipment under the ground, LEP has presented us with the opportunity to go to the other extreme by using satellites hundreds of kilometres above the ground in order to construct the required survey network.

Thus the survey work here is very much state-of-the-art and has required close collaboration with many institutes, industry and individuals working in this field. It is for this reason that we felt it appropriate to bring all our experience together in the hope that it can be applied more generally and to our mutual benefit. I hope the course is well balanced so that you will find something new and interesting, something you desperately need, and something you already know. This will also be an opportunity for you to meet each other; a most important aspect of a meeting like this. And for this reason I hope that the lecturers will make a special effort to stay as long as possible and discuss with the participants.
OPENING ADDRESS

THE PEOPLE OF GEODESY

A.L. Allan
University College London

I think that the English poet, William Blake, most certainly did not have geodesists and CERN in mind when he wrote the words "Great things are done when men and mountains meet". The sentiment does reflect, however, the achievements of CERN, not only in physics itself, but also of the team of geodetic engineers which serves them, which has been associated with these men and those mountains for a period of over 30 years. It is also to be hoped, and expected, that the activities of the forthcoming week will be no exception.

It is customary for university teachers to talk easily on some quite specific topic; like a scientific principle, a new instrument, a system of analysis, or "how I sowed my wild oats in Botswana". To give the opening address at a happening, such as this, casts one in an entirely different role; because the subject matter has never been defined, the objectives are obscure, and the audience is itching to begin its real work. In truth there could be much merit in dispensing with the address altogether, thus permitting us all to sit here in silence; to think about our various problems, and to prepare to share prejudices with each other. However, as I have brought my harp to the party, you are going to hear it, like it or not. I have been asked to speak for about 40 minutes, which I shall do. You have to listen for that period (or think quietly to yourself). If you finish before I do, please let me know.

Conferences are like aeroplanes, difficult to get off the ground, but once in full flight, continue with consummate ease; so fasten your seat belts for take off.

To change the metaphor, I see this event as a promising feast; the menu has been well selected with foods in season, the cooks are in good form, the chef de cuisine is ecstatic, some excellent wine (unadulterated) has been laid up in the shape of a friendly and helpful administration; and it is my pleasant task to serve the hors d'oeuvre which you will recall performs the useful gastronomic function of stimulating the salivary juices, to initiate the whole digestive process.
Initially, I was under the misapprehension that we were gathering for a seminar in which geodetic engineers informed their clients, the physicists at CERN, about latest developments in their measurement science, and surreptitiously sang the praises of the research and development prowess of the CERN geodetic team. If such had been the case, my objectives would have been clearer, and a quick résumé of GPS, VLBI, distinvars, generalised inverses etc., would have followed.

However, it transpires that such is not the case. This event is described by the organisers as a "course" in applied geodesy; with objectives, inter alia, of bridging the gap between the mere dissemination of knowledge, as at conferences on the one hand, and the application of knowledge to large engineering problems, on the other. If the idea itself is good, the manner of its execution is even better. A considerable effort has been made to restrict the numbers and mutual interests of the participants, in order to achieve an effective outcome at a mutually constructive level. Bearing in mind, therefore, that we have now embarked upon a course as part of our continuing education, I do wonder what form of assessment is in store for each participant? We all know that, unless the subject is examined, the students show little interest! In this didactic spirit, therefore, I now pose a question. "Who is the best known writer in the world to have treated a geodetic subject?" Could it be Pythagoras, or Mercator, or even Amerigo Vespucci? The answer will be revealed later.

Having thus established the purpose of our meeting here it becomes clear that the opening address should make at least passing reference to the following topics.

(1) As this is an occasion of some historic importance in the life of the geodetic community at CERN, a generous dose of nostalgia is appropriate.
(2) The pedagogic aim of the course must be examined and some effort made to see that each specialist learns something of other specialisations, and that all non specialists, at very least, extend the horizons of their

![Image of NOSTALGIA, GEODESY, SPECIALISTS]
(3) The serious void which is opening up between the latest technology and the inability of educational establishments to purchase it, poses serious problems for the education of the new generation of geodetic engineers. I shall take the opportunity to demonstrate that much didactic advantage can accrue from an intelligent use of inexpensive technology to model such expensive systems.

(4) Because much attention will be paid to hardware and software during this week, some thought should be given to the pinkware – the people for whom, by whom, and with whom such systems are created.

(If you are thinking the answer to my early pose is Brahe or Kepler or Cassini or Newton, you may be right!)

In the current frenetic world, it does no harm at all to pause, now and then, to think or just to pause, and then to reflect on past events in all their glory, and in all their stupidity. Each generation faces the everlasting problem of dealing with permanent transitions from one accepted position to another. Some persons thrive on change, others hate it; others are pragmatic, often clinging stubbornly to a "tradition" they have only just acquired. The reluctance to jettison hard earned skills or understanding is surely pardonable. Contemporary students listen with unbelief to tales of heated arguments about the best way to find a square root on a hand calculating machine. As for logarithms! Yet, only recently a former student returned from the Antarctic with the news that, but for Briggs and his logarithms, he would not have calculated his work because of recalcitrant batteries.

(Are Napier or Leibnitz contenders for my prize?)
Not so long ago, C.A. Biddle, who had worked in India with Bomford in the 1930's said of his geodetic work, "It would not have mattered if we had not done it properly". Can this be true? Is it true of the early work here at CERN? Does it make any sense at all to chase after the unattainable, when to wait for the right technology is wise? Something of this thought lies, I think, behind the UK's failure to establish a numerical cadastral system. Today with modern technology, the situation is quite different, and the potential of a multipurpose cadastral is no illusion. It is therefore with some pleasure to hear of its possible inclusion in the work of the geodetic engineers at CERN.

(Surely the prize cannot go to Torrens or even Napoleon?)

I often think how sad it is that the new generation of "surveyors" is unable to meet the giants of my previous generation, Marussi, Meissl, Schermerhorn, Thompson and Hotine, to name some known to me personally. Not to have rejoiced in their company, nor laughed at their undoubted wit, nor marvelled at their intellect somehow impoverishes one's life. Is there no case for the use of video television to preserve some small record for posterity of today's giants?

I remember a remark made by the late Pietro Alaria, who surveyed the Italian side of the Mont Blanc Tunnel, when quizzed about his lack of least squares adjustment "there was no need, I just observed it properly." Surely a worthwhile thought?

A study of history of any kind, but also of geodesy, gives us a sense of belonging to a renowned line of men, and a few women, who largely posed all the right questions years ago. In the classical determination of the figure of the earth of the early 18th century, Bouger tackled the problems of heights from aneroids, terrestrial and celestial refraction, instrument calibration, and deviation from the vertical - all topics of interest this week. Galileo knew the principles of electromagnetic distance measurement but lacked the technology to put them into effect.

The study of history, however, has an even more practical purpose, in that it often reminds us of lost ideas and concepts. I recall being present at the measurement of a geodimeter line in the West Indies in 1956, when a large plane mirror reflector was being used with great difficulty in spite of the fact that the corner cube had been known for over 50 years. The late Keith Rylance also came across an illustration in an old French text of mathematics which stimulated him to develop a nylon wire bearing transfer device.

(Perhaps Bergstrand is the winner for the invention of the geodimeter, or Froome for the mekometer?)

Mention of such instruments leads us naturally to remember the old adage "today's physics is tomorrow's engineering", and in particular to give a little thought to those indigenous clients of geodesy, the physicists here at CERN. The mere existence of CERN is a lesson in international cooperation. In 1952 eleven nations set up an interim body called the Conseil Européen pour la Recherche Nucléaire, which since then, with a change of name, has made massive contributions to fundamental particle theory, a hallmark of which was the award of the Nobel Prize in Physics to two of its scientists in 1983. Their discovery ranks alongside that of James Clerk Maxwell's formulation of the dynamical theory of the electromagnetic field in 1864. I feel sure that all
geodetic engineers here present, will wish me to say publically how much we admire the work in "Big Science" at CERN, and to express the hope that it will continue successfully. I have no personal acquaintance with the physicists at CERN, other than those of my own university, with the exception of the late Sir John Adams, whose charm was most evident. Of past physicists, being a Glasgow graduate myself, I was nourished by tales of Kelvin's ballistic pendulums and the apocryphal tale that, when a student, Adam Smith sent his servant down to look at the examinations list to "see who was second". To remind us of Lord Kelvin's remark "Science is measurement", I wonder what you make of the following parody of a poem by Robert Burns (Burns was not a geodesist!) which I have anglicised a little for your benefit.

Rigid Body Sings
If a body meets a body
Flying through the air,
If a body hits a body,
Will it fly, and where?
Every impact has its measure,
Never a one have I,
Yet all the lads they measure me;
Or, at least, they try!

James Clerk Maxwell 1831-79

By way of illustration that geodesists too can write similar jingles, here is one by Hotine, as I remember it.

When the last observation is taken, and the theodolite stowed in its box,
And the surveyor is coffined up also, and his soul at St Peter's gate knocks,
Says he to St Peter, "do tell me, what spheroid you use and all that?"
Says he "you're a silly surveyor! Don't you know? This is heaven, it's flat."

I have already said that, at one time, I thought this seminar would be an occasion for communication between the geodetic engineer and his client, the physicist, and I have further said that we the geodetic community, admire the work of the CERN client. What do they think of us? Couched in general terms, this is an important question. What is the informed public's opinion of the geodetic engineer and surveyor? In a recent FIG Congress paper the results of a fact finding enquiry into this question were so alarming and generally so unfavourable as to engender a resolution for all geodetic engineers to examine local libraries and careers literature for correctness and to rectify any errors and omissions. The relationship between surveyor and client is a vital one, and especially the level and moment at which consultation takes place. Often this is too lowly and much too late. It does appear, fortunately, that such is not the case here at CERN, where geodesy is accorded a proper status and proper timing for its consultation. The reasons for this I think are obvious.
I now wish to touch upon the question of the education of the new generation of geodetic engineers. As many of you will know, I was fortunate to make a study (on behalf of the FIG) of the educational systems in Europe and to visit all EEC member states, with the exception of Greece. It is all too easy to sit in judgement from a position of bigotry, to forget that education reflects the social and cultural history of each nation, and that it has no optimal pattern. Everywhere there is need for improvement and no room for complacency. Those with well established systems must guard against atrophy and a tendency to seek the regulation of the changing world in which we live and which pays scant heed to established sacred cows unless they are of demonstrable value. Those countries in need of assistance should be given it, in the spirit that permeates this establishment, one of goodwill and good grace.

For years, we in the UK have found the general use abroad of the word "geodesy", to embrace all branches of surveying and "higher geodesy" proper, an irritating devaluation of the term "geodesy", and have doggedly stuck instead to the equally all-meaning word "surveyor", I think to our disadvantage. Although to my mind the distinction between "geodetic engineer" and "geodesist" is worth preserving there is merit in changing the title if the older name has unfavourable connotations in the public eye. The advent of trendy new terms like "remote sensing", "computer assisted design", and "information technology" seem to be one of the unhappy facets of modern society, that just have to be tolerated. We should accept that geodetic engineering is part of them all and they of it, if we are to share in the prizes on offer.

As geodesy sees a reduced need for people who delight in outdoor pursuits, and in the challenges for survival and for work in hostile environments, and now leans towards indoor types who delight equally in soldering irons, green screens, and software syllogisms, we have to ask if we are recruiting the right type of person for the future. Within the space of one hour, I overheard the following two contradictory remarks.

"Without computing, surveying is dull" and
"Computing has taken all the fun out of surveying".

Which, if either is correct? The late Professor Thompson (also a candidate for our competition) would often take delight in a story about Australian surveyors bemoaning the disappearance of axemenship skills among new surveying graduates. Has the axe been replaced by the soldering iron - or do we need both?
Throughout the world there is another contender clamouring for inclusion in the curriculum, disguised in the name of "management". Whilst it is clear that management skills are required in all professions, it does seem ironic that the very survival skills on which modern management courses rely, have almost been removed from the self-same courses in geodesy. There is nothing quite like the challenge of field work in a distant or hostile environment to breed management skills, the foundations of which were laid in early life by such organisations as the Boy Scouts! But individualism and resourcefulness are also required in academic subjects; the mental attitudes for the one are also vital for the other. What competence and skills will be required of the geodetic engineer during the next decade, and which will not be required? Since no one can predict successfully, our aim must be to create something of value which is permanent - the correct state of mind which will be capable of responding to change. In a recent paper, Dr. Gervaise made the remark that "the geodesist is still one of the few scientists who stays very close to the reality of the quantity being measured". Whilst this is true of professionals, it is increasingly impossible for students to do so, because of the very high costs of the capital equipment needed to provide this connexion. Instead, except for a favoured few, knowledge of new methods is purely theoretical, or gleaned from a brief sight of equipment during demonstrations; although I must say that manufacturers are more than generous with the loan of equipment, especially for project work at the Masters level, and by releasing data for theoretical treatment and computer analysis. Is there not a better way to solve this problem?

As a university teacher, I never ceased to be amazed by at least two attributes of my students (1) their enthusiastic work rate in response to challenges presented to them and (2) the added interest which derives from processing their own data. This latter spurs them to even greater efforts. Occasionally I forget the powerful motivation which derives from these personal observations, (with dire consequences). One's own observations are ten or even hundreds of times more interesting than those of another often gleaned from a book, or made by the teacher at some distant time. This psychological factor also affects professionals in large organisations like CERN. Hence the motivation to learn theory and principles is strongest if one has made the observations. How is this possible today?

Again, if one subscribes to the view that "education is what is left after one has forgotten all one has learned", as I do, how can the fundamental and permanent thought processes be inculcated and principles grasped if these are hidden away behind a confusing screen of new
jargon, complex knobs and hermetically sealed boxes, not to mention, user hostile software? In the course of my visits to educational establishments throughout Europe, I have noticed that the stresses created by the finance/technology gap are causing teaching to become more and more manifestly theoretical, whilst practical work with theodolites and EDM is reduced. A proven way out of this dilemma is to employ these necessary, though rather prosaic instruments, as a means of modelling their more expensive counterparts. It just so happens that the colossal growth of industrial surveying, at least in the UK and also at CERN, to augment and replace the traditional activities of the metrologist, comes to our aid. The teaching of industrial surveying can be used to model larger scale geodetic activities in which the student still (1) designs the observational programme, (2) makes useful observations, (3) captures the data manually and automatically, (4) produces and presents results, and (5) analyses errors of measurement and deformations in the work piece.

A few examples should be sufficient to illustrate these ideas. One needs to operate with a work space of at least 4m x 4m x 3m at the centre of which is a three axis measuring machine with a 1m x 1m x 1m capability. (The testing of such machines and of industrial robots is a current major activity in most industrialised nations).

Current practice is to set up a system of intersection from two theodolites, obtaining scale from a scale bar. The whole problem is analysed by the now traditional use of matrix transformations, including a solution in entirely geometrical terms, either neglecting gravity or using it as an input measurement.

The one-sidedness of the geometry clearly illustrates a similar geometrical feature of position fixing on the surface of the earth eg by GPS. Differential methods yield good results in the region of the work piece, which to give it topicality will be a not too spherical terrestrial globe. Thus the student can personally experience the merit of translocation and the like. So far I have worked solely with static work pieces, although dynamic use is not precluded.

A whole series of theoretical problems arises, ranging from the design problem of optimising the measurement positions (the GPS constellation problem) establishing lines of sight for photographic measurements by photo theodolite, the combination of differing measurement systems and so on.

For its intrinsic value, I have been experimenting with plane mirrors to fold directions and distances in industrial measurement, thus introducing range-sum measurements, and therefore ellipse geometry; and range differences, and therefore hyperbolic geometry; and the direction cosines of VLBI, as well as the more usual ranges and directions. Of course to be consistent with this overall spirit the theodolite is presented as an orthogonal goniometer measuring the rotations of the transformations involved.

A recent use of a spherical mirror gives an excellent model for the geoid/spheroid problem, since the mirror normals were plotted relative to an idealised sphere.

In case I may be misunderstood, I must point out that these are not the only practical exercises that my students carry out. For example we have also used microwave tellurometers to...
model VLBI, yielding directions, during foggy weather; we employ small-boat Doppler sets to derive translocations good to 15 metres; as well as some more conventional survey work, such as the use of sextants at sea for elementary hydrographic surveys.

Although I claim no monopoly for these ideas, and feel them to be necessary for the reasons mentioned, I make a further plea to those involved in education and research, to give a thought for the imaginative teaching in your whole establishment. Some that I have witnessed in Europe needs radical overhaul.

The application of such educational modelling can be applied to many subjects. It is no more needed than in the teaching of computing, where numerical principles, logical sequences and input-output rules are more often than not hidden in a soup of over complicated computer languages and operating systems, whose description defies comprehension, written in jibberish resembling the nonsense tongues we invented, as school children, to confuse our parents.

Having made a plea for imaginative, efficient teaching, let us re-assert the need for professional experience with real modern systems and equipment. In this connexion I must commend the *Stagiaire* system here at CERN. I plead for its extension in two ways (1) to include more countries if appropriate, and (2) to create here a real life laboratory for geodesy. It is my understanding that hitherto only French students, who need the experience to qualify as members of l'ordre du Géomètres Experts, and British, who need similar experience for the Royal Institution of Chartered Surveyors' membership, have been so employed. These professional bodies act as the clearing houses for likely candidates. This is a necessary function, because not all graduates have sufficient interest or intellectual make up to be happy at CERN. It surely must be equitable to extend this facility at least to other European countries contributing to the CERN budget.

It is an increasing tendency, at least in UK companies, for employers to seek surveyors "with experience" without offering to give newcomers an opportunity to gain this experience. The importance of the arrangements at CERN in helping to break this vicious circle are self evident.

As to the second suggestion, perhaps the CERN geodetic activities could be made available for educational visits of suitable students to see equipment and processing facilities. In addition, data could be made readily available for re-computing and assessment by serious research workers. From my own experience the CERN authorities have been only too willing to assist, but we must also realise that CERN is a geodetic production organisation whose main efforts must remain the provision of services to its clients.

Until now I have quite deliberately steered a pathway between geodetic engineers of the past and of the future, leaving some comment on contemporaries to the last. It is probably bad form to sing their praises too loudly and fashionable to expound upon their faults. I need hardly remind you that we are to be treated to presentations by a dazzling constellation of stars together with a selection of no less glittering jewels; the latter because there is just as much underground excellence here as on the surface. With one exception I shall not name names, other than to give the answer to my early connundrum. The most well known writer on a geodetic subject is of course, Jules Verne who wrote about the "Adventures of three Russians and three Englishmen in S. Africa", now known under the English title of "Measuring a meridian". In this, not very
memorable work, Verne depicts characters which are quite typical of people we all know, and the following extract seems typical of the precision engineering here at CERN.

"'Now' said Col Everest "we've got to find out the exact starting-point of our operation (the base terminal), by hanging a plumb-line from the end of the first rod".
"Yes" replied Matthew Strux "but only so long as we take into account the half-thickness of the line at the point contact!"
"I quite realise that!!" retorted Col Everest!!"

HELPFUL

BIG

VOICE!

MANUFACTURER!

Many professions have their popular writers; where are the James Heriots and Patrick Moores of geodesy?

Another fascinating feature of CERN is the cordial way in which so many scientists and institutions cooperate with each other. Is this because "the organisation shall have no concerns with work for military requirements and the results of its experimental and theoretical work shall be published"?

To those unfamiliar with industrial surveying, the tolerances required by the CERN engineers are quite mind bending, and the specialised equipment developed here is much admired by all concerned in this field (no pun intended). There can be no doubt that all this achievement has only been possible because of the good sense, intelligence, drive and personality of the geodesy team leader over its 30 year period. I refer to Dr.Gervaise who began his work here in December 1954, and as assistant leader of the Survey Group was responsible for the alignment of CERN's first accelerator, the Proton Synchrotron. Since then, after promotion to leader of this group in 1962, he has led and inspired the development of geodetic practice here, finding time to write over 50 important papers, presented world wide with gallic logic and enigmatic charm. From 1972 to 1984 he has been a professor of the Ecole Supérieure des Géomètres et Topographes in Paris, and is an officer of the Ordre des Palmes Académiques. In addition to these technical achievements, he also possesses another skill so essential to a professional man - the ability to deal with his clients at the appropriate design level; in other words, he is a supreme example of the Diplomatic Geodesist. It is fitting that this present company should thank him now on behalf of the global geodetic community.
- Ladies and Gentlemen.

To those who have not visited CERN before, I assure you that a geodetic treat is in store - and, returning to the gastronomic metaphor, the reality of the meal will more than match the eager anticipation with which you have perused the menu. I implore you not to bolt this food, but to savour it in the goodly company described by Sir Alec Merrison, President of the CERN Council in 1984, in the following words

"The vitality of the CERN laboratory stems from the abilities and enthusiasm of its staff. It has therefore been very satisfying to see the positive evolution of the relationships between the staff, the CERN management, and the representatives of the CERN Member States during the past few years. The channels for communication and discussion are now more solidly in place and I have every hope that, with a constructive attitude on all sides, the remarkable spirit, which has always prevailed at CERN, will continue to flourish."

At the beginning of this address, I described the objectives of this course in "Applied Geodesy for Particle Accelerators", as an effort to bridge the gap between the mere dissemination of knowledge, as at conferences, and its application to large engineering problems such as accelerators. It is to be hoped and expected, that all participants, by contributing in various ways - word, thought, attention and wit - will return to their homes enriched with new knowledge, fired by new enthusiasms, perhaps puzzled by new problems, but above all, with hearts gladdened by the warmth of new friendships made, and old ones re-kindled. There is no doubt in my mind that the geodetic community will have accumulated yet another debt due to Dr. Gervaise and his team for organising this occasion for our mental rejuvenation.
AFTER DINNER SPEECH

E. Picasso
CERN, Geneva, Switzerland

On the occasion of this conference on applied geodesy for particle accelerators, I have been asked to say a few words about the Large Electron-Positron collider (LEP) now in construction here at CERN.

I am very grateful to the organizers of the CERN Accelerator School for inviting me to give this after dinner speech.

* * * * *

Why build a big new collider?

I have no better answer to offer than a few general remarks.

Mankind’s curiosity about Nature has surely been the dominant force in the development of civilization. Man’s history has been shaped by the irresistible urge to find out how the world works and how we can use this knowledge to improve our way of living.

Among the most basic questions we have always asked are: What is the ultimate structure of matter? What are the forces through which matter interacts? How did the Universe begin? Will it ever end? as well as many other questions in the same vein. These questions fall within the domain of elementary particle physics, upon which our knowledge of other sciences ultimately rests. As biology and medicine are founded in chemistry, and chemistry in physics, so all the physics is founded in the study of elementary particles and forces that govern their behaviour.

Until very recently it could scarcely have been foreseen that some of the age-old mysteries of astronomy might be solved by looking down to elementary particles, where powerful forces give all matter its form. Particle physicists, together with nuclear and atomic physicists, astrophysicists and cosmologists are beginning to understand not only what matter is, but where it comes from, when and how. Questions such as these are important since everything else about which we would like to know more is intimately linked to them.

LEP is one of the powerful and precise scientific instruments upon which further advances in our knowledge of Nature critically depend.

LEP will be the largest electron-positron storage ring in the world. In this giant accelerator, matter (e⁻) and antimatter (e⁺) will be brought into head-on collision and in the resultant mutual annihilation the total energy will become available for the production of other particles. The magic process through which energy can turn into matter has been stated by A. Einstein in his well-known
equation \( E = mc^2 \), where \( E \) is the energy of a particle, \( m \) its mass and \( c \) the speed of light. This formula essentially means that matter and energy are two different aspects of the same entity which can change from one to the other.

Only one storage ring is needed for electrons and positrons. Because of their opposite electric charges they are bent in opposite directions by the same magnetic fields. Thus they can circulate in opposite directions in the same ring, resulting in head-on collisions.

In the ring, a series of bending and focusing magnets will ensure that the particles follow the desired paths and stay concentrated in a narrow beam. A very good vacuum in the tube in which they circulate will greatly reduce the number of particles lost due to collisions with air molecules.

The amount by which the beams are bent depends upon the magnetic field and upon the energy of the particles. As particles are accelerated to higher energies, by passing them through electric fields generated by a radio frequency system, it is necessary to increase the magnetic fields to hold the beams on the same orbit.

The electron and positron beams will each be injected into LEP in four bunches equally spaced around the ring. Each of these bunches will be about 2 cm long, 200 m wide and 20 m high at the crossing points. As the beams circulate in opposite directions, these bunches will collide at four points around the ring and it is around these positions that detection systems will be installed to observe the results of the collisions. These experiments will involve complex detector systems, which generally require large collaborations of physicists for the conception, construction and operation of the detectors and for the analysis of the results. The realization of such complex systems requires new advanced technologies and collaboration with specialized industries is essential.

The LEP Project includes the construction of two Linacs and an Accumulator Ring (EPA), the necessary modifications to the existing PS and SPS machine complex to enable acceleration of electrons and positrons up to 22 GeV energy, and above all the construction of the main ring of LEP (see Fig. 1).

The construction of the LEP machine necessitates the use of well-advanced technologies such as magnets, radio-frequency cavities, vacuum systems, extremely stable power converters, etc. The machine control systems, including positioning of the beam and communications, constitute a fascinating aspect of the realization and the operation of the collider, and its realization demands the use of the more recent technological developments. LEP construction is not only an intellectual adventure but also a technological realization ahead of its time.

The main tunnel consists of eight arcs and eight straight sections, the latter containing the beam collision points. The underground system will consist of a tunnel approximately 27 km in length with an internal cross-section diameter of 3.76 m, after concreting. This will lie for approximately 10% of its length in the limestone strata, and for 90% in the molasse. Along this tunnel at a location corresponding to four of the eight beam collision points, will be constructed the experimental zones at P 2, P 4, P 6 and P 8 as seen on the map in Fig. 1. Additional galleries are built in parallel to the main tunnel at P 2 and P 6 to house the klystrons and the other radio-frequency
equipment. Let me point out that the underground excavations are about one million m$^3$, and each experimental hall is about 25,000 m$^3$, there are 19 pit shafts, the average depth of which is approximately 100 m, the diameter varying from 9 m to 24 m. The construction of LEP is an impressive task carried out by a team of scientists, technicians and skilled craftsmen. It is for me a great honour to be the conductor of such a wonderful orchestra.

So much for LEP. Let me say a few words about the work that the applied surveyors such as J. Gervaise and M. Mayoud and their group have done for LEP construction.

I am one of the scientists that has been able to measure quantities of quantum electrodynamics with extremely high precision. When I mention accuracy of one part per million or more, people undoubtedly express admiration for the accuracies that have been measured in such experiments. Let me express, in turn, my admiration for what you have done in your field. J. Gervaise and his collaborators have been able to establish a network of points of reference on the surface and deep underground so that it has been possible to drive the three huge full-face boring machines with great accuracy. It is for me incredible that man is able to guide these huge tunnelling machines, the heads of which weigh about 170 tons, with errors no greater than one or two centimeters! It is a fantastic achievement if one thinks that everything happens at about 100 m underground and the only communication with the surface reference points is via shafts around 3.5 km apart. The use of sophisticated devices such as the terrameter, satellite measurements, and the use of more conventional devices such as the gyroscope and similar ones have allowed us to do a wonderful job.

LEP is planned to operate at the end of 1988 at 50 GeV per beam and later, around 1993/1994 at an energy of about 100 GeV per beam.

It is clear that LEP is a joint adventure between CERN scientists, European industrial competence and collaboration with scientists of other disciplines, of which applied geodesy. If LEP operates according to our design a good share of the thanks goes to you all.

Fig. 1 General layout of the underground areas
CLOSING ADDRESS : CHALLENGE OF THE FUTURE FOR GEODESY

J. Gervaise
CERN, Geneva, Switzerland

ABSTRACT
According to our present understanding, High Energy Physics in the 21st century will require colliding-beam luminosities proportional to the square of the energy of the centre of mass. To obtain such performance with current estimations of cost, power consumption and size, new concepts and methods of acceleration will be needed. Amongst the many ideas under discussion at the present time, the author tries to find a new understanding in the survey field and a way to meet the future challenge. Applied Geodesy will face for very large accelerators.

On Monday the 19th of January 1983, Carlo Rubbia called a meeting in the conference room of the PS to announce the discovery of the intermediate bosons \( W \) and \( Z^0 \). This announcement was made to 50 of the most important people at CERN and a handful of others. I had the honour to be there. During our work in the alignment and construction of the experiment UA1, we had realized that the complexity, inspiration and ingenuity of this experiment significantly pushed back the frontiers of what we had thought to be possible.

On returning to my office, which was next to that of Sir John Adams, we discussed the results that had just been revealed to us. John summarized the immensity of this discovery in one phrase "Jean" he said "this is not even the discovery of a needle in a haystack, it is the discovery of a particular fragment of hay in a haystack".

Looking back on such feats, how can we say something is impossible, how can we fix the limits of the human imagination and ingenuity in front of seemingly unsolvable problems.

This morning, after the figures presented by Kjell Johnsen, I would have liked to say that the projected tolerances are beyond applied geodesy. They are beyond the limits of the stability of the earth on which the new accelerators or colliders will be constructed. There are too many unknowns, for example the vibrations of the supports of the quadrupoles and dipoles which have never been seriously studied up to now, having never hindered the smooth functioning of the machines.

The Earth is forever moving. It is subject to earth and marine tides, earthquakes, movements of faults both active and so called dormant since we cannot measure the movements, volcanoes and of course the irresistible movement of the crustal plates themselves. The latter we have only recently started to measure following the launching
of the Navstar satellites of the Global Positioning System and by studies with Very Long Base Interferometry. I could leaf through any book of Geophysics and give many more examples. At this level of precision everything acts on everything else. For the orbit of the particles in an accelerator, the most important would be differential movements over relatively small distances of the order of several hundreds of metres. Wouldn't these differential movements, say of the order of 0.1 μm for a period of 0.1 ns, induce oscillations which would enlarge the particle beam beyond acceptable limits? As measurements of this precision have not yet been achieved, no one can say.

This talk takes me back 31 years, to when I arrived at the PS, and the Parameters Committee asked me to quote the precision I could expect to attain in the alignment of the accelerator's 100 magnets. In a CERN report dating from 1958, the tolerances required for the alignment of the PS magnets were as follows:

- 0.3 to 0.6 mm in the radial plane,
- 0.6 to 1.2 mm in the vertical plane.

In 1959, before the accelerator was put into service, the standard deviations achieved were as follows.

In the radial plane:

- 0.128 mm for the magnets situated in the middle of the octants (the worst determined),
- 0.018 mm for those situated opposite reference pillars.

In the vertical plane:

- 0.1 mm.

When the measurements were retaken in 1960, standard errors were greater than 1.0 mm radially and these errors were distributed in a periodic way throughout all the octants. These movements were due to temperature change acting on the concrete beam which supports the magnet. In the vertical plane the standard errors were also greater than 1 mm due to the movement of the earth over the machine tunnel and the construction of the East experimental zone.

Figure 1 shows the movement of the central pillar of the PS during the period 1965-68. We note two areas of quasi-stability separated by 0.11 mm. The pillar of diameter 1.5 m was sunk into the molasse and the exposed part was inside a building and protected by an insulated hut to avoid thermal change. These movements cannot be explained by instability of the geodetic network. They were detected by using horizontal pendulums. There had been no construction in the immediate vicinity of the pillar during the period of measurement, nor movement of the water table.
Fig. 1 Movement of the PS central pillar from 1965-68

Results from the SPS show that after the smoothing of the quadrupoles, immediately before putting the machine into service, the radial standard deviation in relative position was 0.08 mm over the 7 km circumference of the machine. The standard error in the vertical plane was of the same order.
Thus from experience gained in the installation of the PS and SPS, both of which will serve as part of the long chain of pre-injectors for LEP, the precisions expected for the alignment of functional components of accelerators are:

- Radial standard deviation (1 σ) \( dr = 0.08 \) mm
- Vertical standard deviation (1 σ) \( dz = 0.10 \) mm
- Tilt standard deviation (1 σ) \( dt = 0.02 \) mrad

As LEP is a collider, it is important to take care of the length of the bunches of particles (about 30 cm) to ensure that they collide at the correct place and time. This requires the knowledge of an additional parameter, the length of the circumference, to a relative precision of \( 1.10^{-6} \). This is attainable using the techniques available at the present time.

The values quoted above can be slightly improved without resorting to new geodetic methods. The forced centring sockets on the magnet which take the target used for the alignment could be machined to a greater precision, but this would obviously cost more. The relationship between the magnetic axis and these targets could also be measured to a greater precision.

To illustrate this, for the alignment of the SPS the tolerances put forward by the Applied Geodesy Group took into account the uncertainties of the relation between the reference socket and the magnetic axis. We estimated:

- standard deviation of the position of the reference : 0.15 mm,
- standard deviation of the alignment in radial direction : 0.15 mm,

and in adding these quadratically, we find:

\[
\epsilon_t (dr) = \sqrt{(0.15)^2 + (0.15)^2} = 0.21 \text{ mm}.
\]

The above figures show the margin of error allowed for in these estimations. The actual value, just before commissioning the accelerator was 0.08 mm. The ease and precision of adjustment of the magnet jacks, and general improvements of the instruments used, were responsible for the ameliorations. Obviously since then, the precision of this alignment has gradually deteriorated. Thanks to the understanding developed between the geodesists and applied physicists responsible for the beam orbit, certain quadrupoles have been moved to reduce the beam oscillations induced by this degradation. The first general realignment of the SPS took place in January 1986, 10 years after the accelerator was brought into service.

There is one point in which experience is of little help, that is the uncertainty of the installation schedule. We have never performed a laboratory or workshop type of metrology, even if the precisions achieved are close to those of the latter. What we have done, is the best possible metrology bearing in mind the real working conditions. The alignment of the magnets, one of the last operations, must cope with the delays already accumulated. This means working in conditions which would never have been accepted at the
design stage of the project. The LEP is no exception. We have had to design a metrology based on fixed tripods, before the installation of the quadrupoles, an eventuality which was not foreseen in the planning of the project but which was imposed by circumstance.

Of course, we can see the microns. The results of the calculations on the calibration base at CERN produce tenths of microns. We must however ask if they are real. 0.1 μm is 10^{-7} m or 1000 Å. The Consultative Committee for the Definition of the Metre recommended that the metre should be defined as the length of the path travelled by light in vacuum in an interval of 1/299792458 of a second. It thus depends on the measurement of time or frequency. At the present time the second and the hertz are defined by a caesium standard to a precision of 10^{14}. Thus the definition of the second is known to several orders of magnitude better than the metre and therefore this definition of the metre should last for a long time.

The accuracy of length measurement is in fact limited by other phenomena, in particular the effect of diffraction. Plane waves theoretically have an infinite transversal dimension, but when they cross our instruments they are collimated and subject to diverse optical aberrations. Their mean phase velocity is no longer exactly equal to c, even in vacuum.

The magnitude of this effect can be calculated on a collimated beam, at the crest of the wave front, where it is practically flat. For a Gaussian distributed beam of wave length λ = 633 nm of which 98% passes within a circle of 4 cm diameter, the relative error between the mean wave length and that of ideal plane waves is of the order of 2.10^{-10}. This error is quadrupled if the diameter of the circle is halved. The effect of collimation and other imperfections of the wave front on the wave length must be studied both experimentally and theoretically if we wish to reach and exceed an accuracy of the order of 10^{-10} for the definition of the metre.

Precisions of this order are obtained in a vacuum. We work in air, whose composition varies appreciably. Measurements for the geodetic network of LEP using the Terrameter, which eliminates the effects of variation of temperature and pressure, reveals a relative precision of 0.9 . 10^{-7} and exceptionally 3 . 10^{-9} over a distance of 7.5 km. Compare this with the 10^{-10} accuracy of the standard metre.

Unfortunately the Terrameter cannot be used for distances of less than 500 m. Even if this was possible, the time necessary to set up the instrument is much longer than that available to the geodesist involved in the alignment of an accelerator. However smaller instruments using the same principles adapted to short distances could be one possibility for the alignment of future accelerators.

An other instance where the geodesist has seen values of the order of 0.1 μm was in the calibration of an H invar rule of 1 m, used for the standardization of invar wires. This rule was manufactured by SIP and was calibrated before delivery. At that time, the distance separating the two lines which define the length of the rule, at 20°C, was
1 mm + 0 \mu m. This rule was subsequently recalibrated several times at BIPM and the results are shown in Table 1.

**Table 1**

Calibration of a 1 m H invar rule

<table>
<thead>
<tr>
<th>Date</th>
<th>Calibration value at 20°C</th>
<th>Standard deviation of this value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02.04.1962</td>
<td>1 mm + 0</td>
<td>0.026 \mu m</td>
</tr>
<tr>
<td>31.07.1968</td>
<td>1 mm + 1.17 \mu m</td>
<td>0.018 \mu m</td>
</tr>
<tr>
<td>11.04.1972</td>
<td>1 mm + 1.11 \mu m</td>
<td>0.020 \mu m</td>
</tr>
<tr>
<td>29.05.1975</td>
<td>1 mm + 1.15 \mu m</td>
<td>0.020 \mu m</td>
</tr>
<tr>
<td>05.10.1979</td>
<td>1 mm + 1.08 \mu m</td>
<td>0.018 \mu m</td>
</tr>
<tr>
<td>27.03.1985</td>
<td>1 mm + 0.94 \mu m</td>
<td></td>
</tr>
</tbody>
</table>

Assuming a linear relationship between time and length of the rule, it reduced in length by 0.007 \mu m each year, with the standard deviation of the residuals never being greater than 0.03 \mu m. Thus after two years a reduction in length of the order of 0.014 \mu m could be expected, but this cannot be proved due to the uncertainties of the measurement. An interval of 4 or 5 years between measurements seems essential, except in the case of thermal or mechanical accident. The result of 27.3.1985 confirms this trend at around 0.013 \mu m each year. These results were obtained in a laboratory in optimum conditions which cannot be achieved in the tunnel of an accelerator.

Microns can also be seen using a laser interferometer, which replaced the 4 m rule as the standard for invar wire calibration at our base. The conditions of use are not as favourable as those of the BIPM distance measurement laboratory but still very good for standardization up to 50 m.

A measurement of distance between two forced centring survey sockets using a laser interferometer takes 15 minutes. Nothing has ever led us to suspect movement of the wall on which the microscopes are fixed during a measurement. The standard deviation of the thousands of calibrations made either with the 4 m rule or the laser interferometer is 2 \mu m.

The precision of calibration of a distinvar between the two sockets is 30 \mu m, which allows us to achieve measurements in the tunnel to an accuracy of less than 100 \mu m. We naturally take into account the temperature difference between the geodetic base and the tunnel where the magnets are located.

The development of the laser interferometer with servo-reflector which will be used intensively in the measurement of the LEP metrological network (consisting of the references on the quadrupoles) has permitted an amelioration of the precision.
A test was made on a SPS sextant over a distance of 900 m to simulate the method to be used for the alignment of the quadrupoles of LEP. The size of LEP makes it impossible to calibrate invar wires at the base, as has been the case up to the present. The distances (32 m) between the references of the quadrupoles in the SPS was measured using an automatic distinvar. The standard deviation of these measurements was 31 μm. Three distances of 32 m were measured, at the beginning, mid-point and end of the sextant, using the interferometer. The introduction of these values in the least-squares adjustment reduced the standard deviation of the distinvar measurements from 31 μm to 16.6 μm. Thus the intermediate calibrations performed in situ, improved the precision of the measurements by a factor of two.

Unfortunately, this improvement in the precision of measurement of distances, which is an order of magnitude greater, is such that the tolerances foreseen for the alignment of components of future accelerators would not be sufficient to conserve a bunch radius within the limit of 0.1 μm.

All other measurements, except those of tilt, are influenced by the atmosphere in the tunnel, temperature gradients, turbulence, air current, variation of pressure and humidity. Once again the geodesist is face to face with his old enemy the atmosphere, and what is worse, the atmosphere of a tunnel. At the moment of the measurement at time t, we can hope to position the elements of the machine to a standard deviation of 50 μm, providing the time is available to achieve the measurements and we know the meteorological variables to introduce into the calculations. At time t₁, after the smoothing of the quadrupole positions before commissioning the accelerator, there will be at least local deteriorations due to all that can influence the mechanical stability of the quadrupoles. At time t₂, those deteriorations will continue, without the possibility to remedy the situation by geodetic means.

All this is not very encouraging, because it is difficult to master the atmospheric problem as well as the instabilities and rapid movements of the earth, which can induce long or short term differential deformations of the beam. On the other hand, the use of optical measurements, in particular the laser, is also limited. Will new instruments be invented to satisfy the constructors of accelerators? This is possible, but unlikely, since at these precisions, even if the electronic dials give us as many insignificant figures as we want, in the end we always have the tendency to construct thermometers and barometers.

In suggesting 50 μm as a standard deviation for the position of the references in the radial, longitudinal and vertical planes, and for tilt, for a limited and unknown period, I am taking some risks. In any case this lower limit will not be possible for LEP, the precisions achieved will be of the same order as those obtained on other accelerators at CERN. To achieve the desired tolerances, it is going to be necessary to establish a permanent dialogue between the applied physicists responsible for the orbit of the particles and the geodesists. A dialogue based on a deep understanding of the requirements of the one and the service provided by the other, including a unified use of
the computer and, of course, a mutual understanding of the calculations. This is already practically the case for the SPS. But the dialogue must go even further to achieve a total control of the dimensions and behaviour of the beam.

The work of the geodesist is firstly to ensure one complete orbit, or preferably several complete orbits, of the accelerated beam at the energy where it is most sensitive to alignment errors. From this moment, analysis of the oscillations must reveal, within the practical limits of the geodesy, which quadrupoles have to be moved. For this it is necessary that the pick-up electrodes used to detect the position of the beam form, are part of the quadrupoles and that their relative position is known to a precision of 0.1 μm, no laughing matter.

It would also be necessary to foresee new types of detectors, which check the beam and by servo-command modify the position of the quadrupoles, or trigger correction elements be they steering magnets, sextupoles, octupoles or dodecapoles. Also, a certain degree of innovation is required, the geodetic measurements being used as a limit within which all the resources of electronic, automation and computing are brought into play. The quality of beam analysis could allow the geodesists to make local interventions in order to restore the beam to within the achievable limits. Thus by means of successive iterations, it would be possible to satisfy the demand of machine physicists, whose imagination would be equal to the challenge. I quote the example of the method invented by Simon Van der Meer to store antiprotons of variable energies and direction of flight and to homogenize them in energy and direction by "stochastic cooling".

The imagination and the will to succeed must allow us to solve even the most testing technical problems. On the other hand, the great unknown remains the behaviour of the earth and thus the foundations of the accelerator. If particle physics has known notable success since the beginning of the century, unfortunately the same cannot be said for the Earth Sciences, which despite considerable progress are still at a relatively early stage of development. Knowledge is largely empirical, due to the complexity of the studied phenomena, the problem of prediction of when and where they will arise, the inadaptation of the measuring instruments to the scale of the effect, the difficulty of data treatment and analysis even with computers and, as astonishing as it may seem, the lack of financial means and of global knowledge in the sense of the Earth treated as a planet.

The arrival of satellites must produce profound changes in all of this. The ten-year, "Global Change" studies now starting on the definition of the geoid, sea level, "hollows" in the surface of the ocean, and many other phenomena, will not permit us to foresee changes of the order of the nanosecond in the beams of accelerators, but perhaps the intervals within which the geodesists must intervene to put the particles back on the right tracks. That also is the challenge of the year 2000.
GPS ITS DEVELOPMENT AND DEPLOYMENT
Leonard S. Baker
Retired, Director, National Geodetic Survey

ABSTRACT
The Navstar Global Positioning System (GPS) is an all-weather, space-based navigation system under development by the US Department of Defense. The military needs and requirements are well documented, but the combined utilization, savings, and deployment by the civilian users throughout the world will surpass the benefits to be reaped by DoD.

1. DEVELOPMENT

Scientists have always reached for goals that at the time seemed unobtainable. When those goals were near, new goals were envisioned to stretch our imagination, deepen our thoughts, and yes, day dream a little - "WHAT IF?".

This simplistic view is so very true of the Physical Sciences. When the first breakthrough was made concerning matter, it only opened one door along the way, and you immediately planned the attack on the next door. The installations, at each of the laboratories around the world, have monuments to this path of discoveries. Here at CERN there are the Proton Synchrotron, the Intersecting Storage Rings, the Super Proton Synchrotron; and now, the Large Electron Positron Collider, LEP, is under construction.

In the United States, there is renewed interest in the Super Colliding Accelerator, that may have a diameter two or three times the size of LEP. As a geodesist, who has worked on an accelerator ring, I have gained a greater appreciation of the extremely high surveying accuracies required. Each new plan for the next generation of accelerators makes me stop and wonder, "will they be able to meet the surveying accuracy requirements?". GPS will play a major role in meeting those stringent demands.

Humans have never been content to accept the norm for very long, and in 1957, when the Russians launched SPUTNIK, the gestation period of GPS began. GPS has been born and has grown quite rapidly in its short life. However, we have been treated to only a very limited preview of what GPS will be. The real GPS will not occur until the full constellation of GPS satellites are in orbit, in 1988, or 1989.

The world has changed dramatically since SPUTNIK. Everyday, each of us enjoys the benefits of the "Space Age" technology, in our work, in our play, and in our normal living. The world is moving towards being a better place to live in, not only for us, but for all humans on this planet we call earth, and GPS will contribute in so many and various ways.
GPS is an all weather, space based, navigation system, under development by the United States Department of Defence (DoD), to satisfy the requirements of the military forces. Requirements include, accurately determined positions, velocity, and time in a common reference system, anywhere on or near the Earth, on a continuous basis. The Air Force Systems Command's Space Division acts as the executive agent for the DoD in managing the GPS programme. All branches of the United States military are represented, as well as the Department of Transportation, and NATO.

The Navstar GPS navigation and time-transfer system operates on two L-band frequencies, L1 (1575.4 MHz) and L2 (1227.6 MHz). The system consists of three major segments; a space segment, satellites that transmit radio signals; a control segment, ground-based equipment to monitor the satellites and update their signals; and a user segment, equipment which passively receives and converts satellite signals into positioning and navigation information.

The GPS system has already completed Phase I, concept validation. Phase II, engineering development, is nearing completion, and a constellation of six development satellites are being maintained to support testing of both the military and civilian applications. The contract for the development of the operational satellites has been awarded and, barring any additional delays such as the shuttle disaster, operational satellites are scheduled to be placed in orbit starting this year, with a planned completion of the constellation in 1988.

The full constellation will contain 18 satellites, in six orbital planes, 60 degrees apart, and each plane will have three satellites. The satellites will operate in circular 20,200 km orbits, at an inclination of 55 degrees, with 12-hour periods. Spacing is planned to ensure that at least four satellites will be in view to the user at any time on a worldwide basis.

This is a DoD programme that could serve as a model for other programmes, anywhere. It is well planned, funded, tested, and is now being put into place. The over-riding questions that have yet to be answered are - what level of accuracy, in real time, will users other than DoD, be allotted, and - will the signals be distorted?

2. GEODETIC RECEIVERS

A number of GPS geodetic receivers are already being used to gather data, while others are undergoing tests in preparation for their employment. At the present time these include:

- MACROMETER V-1000, the first GPS geodetic receiver developed for precise surveying by Macrometrics Inc. The Aero Service Division of Western Geophysical Company of America, Houston, Texas, acquired the Macrometrics Company, in March 1984.
- TEXAS INSTRUMENT TI 4100, developed by Texas Instruments, with funding from U.S. Government Agencies. This was the first dual frequency geodetic receiver to be produced utilizing both the P- and the S-code.

- GPS LAND SURVEYOR, MODEL 2002, developed by ISTAC Inc., Pasadena, California. It is an outgrowth of the SERIES technology, originally conceived and developed by MacDoran and his co-workers at the Jet Propulsion Laboratory, Pasadena, California.

- TRIMBLE 4000S, a second generation GPS instrument developed by Trimble Navigation, Ltd., Sunnyvale, California. The first GPS instrument developed by Trimble was the 4000A, a navigation instrument, developed primarily for the shipping industry.

- WM 101, a joint venture between the Wild Co., Heerbrugg, Switzerland, and Magnavox, Torrance, California. The instrument is presently undergoing testing, prior to being put into service.

- TR55, developed by Sercel, France. These units have already collected data, but no additional information is available.

There will undoubtedly be numerous new geodetic receivers available in the near future. Many of these receivers will be second generation receivers. They will shrink in size and decrease in price as production increases to meet the demand of the multitude of users.

3. DEPLOYMENT

No one can dispute the needs, requirements, and planned use DoD has for GPS. The accomplishments will be too numerous to tabulate. However, the combined utilization, savings, and deployment by the civilian users, throughout the world, will surpass the benefits to be reaped by DoD, perhaps not in the near term, but within a very short span after the system is fully operational.

Even in the development phase, GPS is being utilized to monitor seismic actions, both horizontal and vertical. It eliminates the time delay factor, and because the satellites are in space, ground truths can be located in regions completely outside the areas under study.

One of the major problems the scientist working at sea has always been confronted with, is accurate location while underway. Geophysical profiling is now, with the aid of computers, three-dimensional. When the profiles show a promising structure, the scientists can be confident that the coordinates are accurate, thereby ensuring that the drilling crew will be able to pierce the structure. Utilizing the GPS translocation mode, ships can now be assured of one metre accuracies, within a range of about 300 kilometres from the GPS receiver over a known location.
The most excitement GPS has generated, has been in the geodesy community. Using methods that will be discussed later by other speakers, geodesists have been able to achieve extremely high accuracies between two points. These are greater than the normal accuracies obtainable with traditional surveying equipment and approach, or sometimes exceed, the accuracies of special surveys at installations such as CERN.

When the electronic distance measuring equipment became available to the surveying and geodetic communities, it broke the bond the steel tape had held on them. Formerly, distances could only be measured if a clear line of sight was available and the angles had to be determined in order to permit the computations to be made. All that has changed. No line of sight is required, no angles are needed, only a window to the sky, a piece of space-age technology, an innovative computer program, and the radio signals from a military programme satellite.

3.1 Mapping

With the advent of satellites, mapping of the earth's surface and resources, has become a new science, a science that was only dreamed of 30 years ago. Remote sensing has been responsible for many new discoveries and has allowed new maps of the world to be produced, at least in regions where aerial photography is allowed. But the science of remote sensing has not reached the level of accuracy, or scale, required for both developed, and the developing nations. With GPS, employed in the translocation mode, and the other antenna mounted in an airplane, positions can be computed for the airplane, either in flight, or post-processed, to an accuracy of less than one metre within a range of 300 kilometres. With very little additional equipment, other than an aerial camera, the pictures taken can be used to produce maps to a scale of 1/40,000, with little or no ground control. Mapping on this scale is more than adequate for all developing nations who need to plan and establish the basic road systems, access the natural visible resources, manage the water resources, and institute major land reforms.

Most of the developing nations do not enjoy an adequate network of coordinated control, the existing system being established by methods now labelled as inadequate. The GPS instruments allow control to be established faster and less expensively than with other means. They can also be used to up-grade the existing facilities and to tie in separated areas to provide a single coordinated framework.

3.2 Air navigation

If the Precise Positioning Service (PPS) would be available to the civilian users, an accuracy of 15 metres in x, y, and z would be available for positioning aircraft in flight. That accuracy represents a distance that is less than the wing span of almost every aircraft in existence.

The present method for flight plans for commercial aircraft, is to assign a flight altitude in a designated corridor, with check points along the way. The aircraft is
tracked by ground control systems that relay this information to air traffic controllers, assigned to monitor that particular flight. None of the in-flight and the ground tracking instrumentation can provide the pilot with real time locations as accurately as the PPS GPS system. Barring the problems all systems experience, a pilot, flying an airplane equipped with GPS, could take off in zero visibility, navigate to the destination, and land in near zero visibility without any outside assistance. When the coordinates, x, y, and z are established by GPS at the end of every runway, and PPS GPS is available, each of them becomes an emergency landing site.

The accuracy of the Standard Positioning Service (SPS), to be available to the worldwide community of civil users, will be 100 metres, in x, y, and z with a 95% confidence level. This is still as much accuracy as a pilot, travelling at 600 knots/hour, can use in flight, and it is only the z that is so useful to them in landing with reduced visibility.

One can speculate on the adverse side of GPS in aviation. Will inexperienced pilots, not proficient in flying by instrument, substitute the advantages of GPS for experience and attempt to fly in adverse weather conditions or after dark? Even cautious pilots become lost, perhaps due to weather fronts moving into the flight path, and fail to reach their destination even during daylight hours. GPS can add a measure of safety.

We are so prone to think only of the flying conditions of our respective countries where we have excellent facilities. Why should less fortunate people, living in distant lands without adequate air traffic control or improved airports, be subject to more hazardous flights when flying is often their only link to the outside? With GPS they will be able to take a quantum leap forward in air navigation and safety, even with the SPS, 100 metre accuracy.

3.3 Sea navigation

Shipping has used Transit satellite generated positions for many years. Transit has a 90 minute interval between fixes, therefore is not an adequate system for navigating in close quarters. When GPS is available world wide, it may spell the end of "freedom of the seas". No ship has the right to endanger the coastline of another nation, by polluting its shores with an oil spill. GPS will allow the nations of the world to assign dangerous cargo vessels to certain shipping lanes that will not cross, except at designated points. Very few ships collide when going in the same direction; the majority of collisions occur in passing or crossing situations. GPS can not alleviate the crossing situation, but can eliminate the chance of a passing collision, if shipping lanes are established, and GPS is a requirement.

3.4 Multipurpose cadastre

There is a very critical requirement for a land information system in the United States. The system is needed to improve land-conveyance and to provide for equitable
taxation, information for resource management and environmental planning. We need a MULTIPURPOSE CADA斯特. We have needed it for more years than we care to admit.

The concept of the multipurpose cadastre, is a framework that supports continuous, readily available and comprehensive land-related information at the individual parcel level. The first component is a reference framework, consisting of a geodetic network. The geodetic control must be an integrated network of points on the ground, over the entire area of the parcels that are related to each other. For a true cadastre, the continental United States must have a single, densely spaced, related geodetic network. Most of the populated European countries have a densely spaced control network, that supports their cadastre. These nations have utilized, and benefited from the use of their multipurpose cadastre system for many, many years.

GPS presents to the United States, the first chance to have a truly continental integrated network of control, at an acceptable price, and within a reasonable time frame, a requirement for our cadastre. The continental network will be a part of the world network, to respond to the intercontinental needs.

3.5 Commercial application

One United States automobile manufacturer is planning to offer GPS in its 1989 models. A GPS receiver will be connected to a computer and digitized map system and will display on a small screen a continuous, detailed, map of the location of the automobile. This will be a very useful tool for travellers to a new location, and to those individuals who have difficulty finding their way. The system will allow a delivery firm to prepare a route map for the driver, increasing the efficiency of the driver and vehicle, and eliminating the, "I got lost", excuse. (How many more distractions can the human driver master?)

3.6 Rescue

GPS will become a very important element for air and sea rescue. When the search aircraft or vessel knows its exact location within a few metres, there is no doubt that the suspected area has been covered. If the aircraft, vessel, or person knows the exact location, and can relay this information, finding them becomes almost as easy as following the vehicle on the screen in the car.

3.7 Boundary locations

There are some countries that continue to have boundary disputes with their neighbours. Surprisingly enough, many of the disputes are not over the physical location, they are over the coordinates of the common turning points of the boundaries, such as the mountain peak. GPS may eliminate many of these problems by placing each country on an identical coordinate system, thereby establishing common coordinates for boundary points.
3.8 Daydreaming

If we are allowed to day dream, there are some excellent humanitarian projects where GPS could play a major role. What if we were able to build a GPS unit including a digitized map of Geneva on a small micro disk and small enough to fit into a back-pack. This could acquire the satellite signals, compute the position, and feed this information to the computer which would search the digitized map and then, verbally inform the person of his exact location.

I do not believe this is a "what if", but rather a "when" idea. When will it be available to all those people who need it now? Such a device would give them the freedom to move about with ease, with confidence, and with the assurance that they could go it alone and not be dependent upon others for guidance.

4. Summary

The original GPS scientists may never hear the many thanks that surely will be voiced on behalf of their invention. They may never receive the recognition rightfully due to them, but they can take a great deal of pride in knowing that their efforts have benefited and will continue to benefit mankind, all over the earth, in so many ways, for so many years to come.

THE USE OF GPS IS ONLY LIMITED TO OUR IMAGINATION AND INGENUITY.

* * *

References

R.W. King, E.G. Masters, C. Rizos, A. Stoltz, J. Collins; Surveying with GPS; Monograph No. 9, School of Surveying, The University of New South Wales, Kensington NSW, Australia 2033.

Need for a Multipurpose Cadastre; Panel on a Multipurpose Cadastre, Committee on Geodesy, Assembly of Mathematical and Physical Sciences, National Research Council; National Academy Press, Washington, D.C. 1980.

GPS RECEIVER TECHNOLOGY

C. Boucher
Institut Géographique National, Saint-Mandé, France

ABSTRACT
The NAVSTAR/GPS system is already widely used for geodetic positioning, although the full capability is not yet available. Various types of receivers, using the C/A and/or P codes or codeless, are available as prototypes or commercially. This paper reviews these receivers.

1. THE GPS SYSTEM CONCEPT

The NAVSTAR/Global Positioning System (GPS) is a satellite-based system developed by the US DOD for worldwide navigation of any system moving in the Earth's environment (ground, maritime or aerospace vehicle). The capability of this system for geodesy has been illustrated many times. Although it will only be fully operational in the late eighties, a provisional constellation of satellites, available for a few hours per day, enables the geodesists not only to prove the efficiency of this system for precise positioning but also to use it for operational purposes, replacing efficiently many of the conventional procedures.

We shall review here the various receivers available for geodetic use, either prototypes, commercially available or under development. Similar surveys have already been made by J.D. Bossler and C.W. Challstrom\(^1\), J. Hannah\(^2\), P. Hartl et al.\(^3\) and B. Remondi\(^4\).

2. THE NAVSTAR SPACE SEGMENT (Fig. 1)

The GPS system uses an L-Band down-link from a cluster of several satellites in the common view of a receiver which can perform several types of measurements as described below.

![GPS satellite signals diagram](image-url)
Each satellite transmits a pair of carriers at $L_1 = 1575.42$ MHz and $L_2 = 1227.6$ MHz. These are modulated by three levels of codes; namely C/A (also called S in the final stage of the system), P (called Y in the final stage) and message. Each code is a flow of bits, 0 or 1, represented by a $\pm 180$ degree phase shift. The P code has a bit rate of 10.23 Mb/s, the C/A code of 1.023 Mb/s and the message code 50 b/s. The message and the P Code exist on both L1 and L2. C/A code exists only on L1 in quadrature with the P Code. For more details see Ref. 3.

3. MEASUREMENT TECHNIQUES

Three types of techniques can be conceived to perform measurements of geodetic interest:

3.1 Code correlation (Fig. 2)

This technique was adopted for normal use. The receiver produces a replica of the code (C/A or P) from a local oscillator. The code stream can be shifted in time and it is mixed with a beat signal from the received GPS signal and a pure carrier locally generated. The code is then shifted in order to obtain a maximum correlation between the received and the generated codes. The shift expresses directly the sum of the offset between satellite and receiver clocks, and the group propagation delay from the satellite to the receiver. This quantity, expressed in length through the velocity of light, is named pseudo-range.

The continuous alignment of code cleans the received carrier on which phase or Doppler measurement now gives useful information.

![Diagram](image)

**Fig. 2** GPS correlation channel

3.2 Signal squaring (Fig. 3)

The principal is to multiply the received signal by itself. In the case of the macrometer, this is done directly. As the code is in phase opposition, this technique removes it and produces a pure carrier with double frequency on which phase measurements can be performed.
In the SERIES technique developed at the Jet Propulsion Laboratory, one signal is shifted in time by half the period of the code before being mixed with the other. It can be shown that epochs of code transition can be obtained precisely (see Ref. 5).

3.3 Interferometry

This last technique is identical to the one used in VLBI: each received signal is down converted and recorded. Later, signals are correlated among two or more simultaneous trackings.

To my knowledge, this has been used only for test purposes at the Haystack Observatory using a Macrometer prototype.

4. ERROR BUDGET AND MEASUREMENT MODELLING

A typical error for the pseudorange measurement is at the meter level while, for the phase, it is assumed to be millimetric (5 × 10⁻³ cycle). At present, the limiting factor of the GPS technique (as far as the 1 ppm application is concerned) is not the measurement in itself. The error due to the orbits, to the insufficiently accurate propagation models and to the internal clocks are larger. If the 0.1 ppm application is required all these error factors must be removed at the same time, then the problem of the measurement uncertainties becomes more crucial.

5. REVIEW OF EXISTING RECEIVERS

Several receivers are already available, either as prototypes or commercially. Table 1 lists only instruments which can be used for sub-meter geodetic positioning and which had been announced at the time of writing.
Table 1

Presently available receivers and their characteristics

<table>
<thead>
<tr>
<th>Type</th>
<th>Manufacturer</th>
<th>Model</th>
<th>No. Chan</th>
<th>Frequency</th>
<th>Mode</th>
<th>No. SAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codeless receivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAC</td>
<td>Macrometrics</td>
<td>MACROMETER 1000</td>
<td>6</td>
<td>L1</td>
<td>SQ and CT</td>
<td>6</td>
</tr>
<tr>
<td>MAC2</td>
<td>Western Geophys.Cy Amer. Aero</td>
<td>MACROMETER II</td>
<td>12</td>
<td>L1/L2</td>
<td>SQ and CT</td>
<td>6</td>
</tr>
<tr>
<td>SERIES</td>
<td>Jet Propulsion Lab.</td>
<td>SERIES</td>
<td>3</td>
<td>L1/L2</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>SERX</td>
<td>Jet Propulsion Lab.</td>
<td>SERIES-X</td>
<td>3</td>
<td>L1/L2</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>IST2002</td>
<td>ISTAC, Inc.</td>
<td>Model 2002</td>
<td>1(3)</td>
<td>L1/(L2)</td>
<td>SD</td>
<td>35</td>
</tr>
<tr>
<td>GLS1991</td>
<td>GEO/HYDRO</td>
<td>GPS Land Surveyor</td>
<td>3</td>
<td>L1/L2</td>
<td>SD</td>
<td>n</td>
</tr>
</tbody>
</table>

| C/A and P code     |                                   |                        |          |           |            |         |
| ST15010            | Stanford Telecom. Inc.            | ST1 5010               | 2        | L1/L2     | CC and LS  | 1       |
| TI4100             | Texas Instruments                 | TI 4100                | 1        | L1/L2     | CC and MX  | 4       |

| C/A code           |                                   |                        |          |           |            |         |
| 4000A              | Trimble                           | Model 4000 A           | 1        | L1        | CC and FS  | 4       |
| 4000S              | Trimble                           | Model 4000 S           |          | L1        | CC         | 4       |
| WM101              | Wild-Magnavox                     | WM 101                 | 4        | L1/(L2)   | CC and FS  | 9       |
| TRSS               | SERCEL                            | TRSS                   | 5        | L1        | CC and CT  | 5       |
| LGSS               | Litton Aero Products              | Litton GPS Survey      |          | L1        | CC         | 8       |
| SEL                | Standard Electrik Lorenz          | "low cost"             |          |           |            |         |

5.1 Codeless receivers

The measurement technique is basically the signal squaring technique mentioned in Section 3.2. Several receivers are available on the market and as knowledge of the codes is not needed, they are usually double-frequency devices. On the other hand, they need to be synchronized before every experiment and the orbits need to be known from an external source.

5.2 C/A and P code receivers

The principle of these receivers is based on the code correlation. Both codes (C/A and P) must be known.

5.3 C/A code receivers

The technique is the same but only the C/A code (non-restrictive) needs to be known. As only one frequency is available the ionospheric delay must be removed by the use of models.
6. CONCLUSION

Even though the GPS system is not yet fully operational, it is widely used by the geodetic community. Many different receivers are actually available on the market. Their capabilities and measurement techniques are very different. At present, the precision of these instruments is sufficient for the 1 ppm applications. In the future, if significant improvements are realized in propagation models and in orbit determinations, this precision may be improved.

* * *

REFERENCES


PRECISE POSITIONING WITH THE GPS

C. C. Goad
National Geodetic Survey
Charting and Geodetic Services
National Ocean Service, NOAA
Rockville, Maryland 20852
U.S.A.

ABSTRACT
A review of current measurement reduction techniques for precise positioning using carrier phase observables from the Global Positioning System satellites is presented. The qualities of each are discussed. The main cause of data problems, cycle slips, is also given along with an example, and then a possible algorithm to correct slips over short time gaps when data from a single station are given. It is then argued that this algorithm will be useful for cycle slip detection in precise positioning of a moving platform as well as in a static relative positioning mode. Optimal estimation in the presence of unmodeled errors is discussed.

1. INTRODUCTION

The use of the Global Positioning System for precise surveys is now a reality. Several government, university, and commercial agencies are using the various carrier phase measuring receivers to obtain data which are processed interferometrically to provide base line vector estimates which far exceed minimum first-order standards.

The processing of these data can be separated into two generic classes depending on the handling of cycle slips which can be caused by problems inside the tracking hardware and by environmental causes. I call these two categories batch and interactive techniques. The batch techniques do not require decisions to be made by an analyst during the actual data-reduction. They provide fast and accurate results but do not take advantage of the integer nature of the ambiguities (more about ambiguities later). These batch techniques do not yield the most precise estimates of base line vectors. Interactive techniques do try to take advantage of the integer nature of the ambiguities, but because of the human intervention these solutions require far more time and thus more expense. This situation may change, however, as processing schemes and algorithms are improved. Obviously, more automation is desired from a cost per base line point of view. Also, such improvements in the static relative positioning processing might impact moving-platform applications as well.
The basic phase observable will be presented in Section 1. Section 2 will discuss the problem of cycle slips in phase data and how to fix them while Section 3 will address the cycle slip problem using data only from a single station. In Section 4, one technique to filter single station data will be presented with possible application to moving platforms discussed in Section 5. In Section 6, the problem of optimal estimation in the presence of unmodeled systematic errors is discussed.

2. THE PHASE OBSERVABLE

The basic carrier phase observable is the phase difference between the carrier signal transmitted at the satellite and a local receiver’s oscillator. Mathematically this is expressed as

\[ \phi_j^i(t_R) = \phi_j^i(t_t) - \phi_j^i(t_R) \]  

where superscripts refer to a particular satellite; subscripts refer to a particular ground receiver; and the argument is either the transmit time, \( t_t \), or the received time, \( t_R \).

Equation (1) is now rewritten in a more usable form. The right side is written so that the times are expressed in terms of the received time, \( t_R \). The distance from the satellite to receiver is given by \( \rho_j^i(t_t) \). Therefore the travel time in a vacuum is given by \( \frac{\rho_j^i(t_t)}{c} \) where \( c \) is the speed of light. Assuming that the oscillators run at a constant frequency, \( f \), Eq. (1) is now rewritten as:

\[ \phi_j^i(t_R) = \phi_j^i(t_R - \frac{\rho_j^i(t_t)}{c}) - \phi_j^i(t_R) \]
\[ = \phi_j^i(t_t) - f/c \rho_j^i(t_t) - \phi_j^i(t_R) + N_j^i. \]  

(2)

Notice that an integer \( N_j^i \) has been inserted in Eq. (2) to acknowledge that \( \phi_j^i(t_t) \), the phase observable at the first measurement time, is determined modulo one cycle. If lock is maintained thereafter, the same \( N_j^i \) will appear in every phase observable between receiver \( j \) and satellite \( i \). Some small terms have been ignored in the derivation of Eq. (2) such as the effects of relativistic, ionospheric and tropospheric refraction and time tag errors. Some of these will be addressed later. Since all time arguments in Eq. (2) are implicitly or explicitly a function of the received time tag, \( t_R \), the \( R \) designation will be dropped with the understanding that hereafter the time tag refers to the receipt time. Thus, Eq. (2) is rewritten

\[ \phi_j^i(t_k) = \phi_j^i(t_k) - f/c \rho_j^i - \phi_j^i(t_k) + N_j^i \]  

(3)

where \( k \) refers to the \( k \)-th sample time.
1.1 Differenced data types

Geodesists are mainly interested in using these phase observables for base line vector recovery. This information is extracted only from the term \(-f/c \rho_2^9\) in Eq. (3) which is a function of the satellite and receiver locations. The other terms in Eq. (3) are the clock difference term, \(\Phi^9(t_k) - \Phi^9_2(t_k)\), and the initial integer ambiguity, \(N_{1,2}^9\). Selectively differencing these phase observables is one way to eliminate these non-geodetic terms and thus simplify the reduction algorithms. Goad and Remondi\(^1\) first showed results from differencing phase observables over a common satellite to eliminate the satellite phase (or clock) term. For example, two stations, 1 and 2, at the (almost) instant, \(t_k\), measure the phase difference between satellite 9 and the local oscilators. Using Eq. (3), this is expressed as

\[
\Phi^9_{2,1}(t_k) = \Phi^9_1(t_k) - \Phi^9_2(t_k) + \frac{f}{c} (\rho_1^9 - \rho_2^9) + N_{1}^9 - N_{2}^9.
\]  

(4)

Equation (4) is now referred to as the single difference observable between stations 1 and 2 at the k-th epoch to satellite 9. As anticipated, this observable does not contain the satellite phase value \(\Phi^9(t_k)\) but now contains differences in phase (or clock) values between ground receivers 1 and 2, initial ambiguity differences, and distance differences. The clock differences do not necessarily have a known relationship from one epoch to the next and thus must be estimated on an epoch-by-epoch basis. Furthermore, the receiver clock difference \(\Phi^9_1(t_k) - \Phi^9_2(t_k)\) will be the same should another satellite participate in the difference processing.

The term involving the \(\rho_1^9 - \rho_2^9\) can be calculated based on a knowledge of the station and satellite positions. \(N_{1}^9 - N_{2}^9\) is the same for all epochs. Thus these two terms are "common" contributors and need not be considered on an epoch-by-epoch basis. The least-squares normal matrix possesses a structure which allows one to solve for the epoch receiver clock difference values rather easily\(^2\). For a given base line, no phase observables appear more than once, thus the measurements are uncorrelated and no special treatment are required. But when more than one base line are considered, the phase observables will appear more than once, and thus these measurements will, in general, be correlated. Some investigators have given results using double differences (for example, Goad and Remondi\(^1\), Bock et al.\(^3\)). These are obtained by taking differences of two single difference observables at a common epoch from two receivers to two different satellites. For an example, receivers 1 and 2 and satellites 4 and 9 at epoch \(k\) are chosen. Then

\[
\Phi^{4,9}_{2,1}(t_k) = \Phi^{4}_{2,1}(t_k) - \Phi^{9}_{2,1}(t_k)
\]
\[ \frac{t}{c} \left( \rho_1^4 - \rho_2^4 - \rho_1^2 + \rho_2^2 \right) + N_1^0 - N_2^0 - N_2^0 + N_1^4. \] (5)

Now all satellite and receiver clock differences are removed. The right side contains four ranges and a linear combination of four ambiguities. Unlike the single-difference observables, the double differences are less complicated to handle (no epoch-to-epoch clock differences to model or estimate), but the data sets are correlated. To optimally process this information one must decide whether to go with a more complex model as given by Eq. (4) for the single differences, or to choose the simpler double difference model given by Eq. (5) and then be forced to model the correlated nature of the double difference observables. Triple differences are generated by differencing two successive double differences in time. This has the effect of removing the bias and increasing the number of ranges to compute in the model to eight. Losses of lock will cause a spike in the triple-difference data rather than a step as in double-difference processing. These spikes (or outliers) are easily removed by an automatic editor or outlier detection scheme. This is an extremely desirable characteristic, especially when looking at the data initially or when the most precise results are not required. This was discussed by Goad and Remondi\(^1\)) where actual comparisons between single-, double-, and triple-differencing results were given.

In Eq. (5) the double difference is seen to possess a very desirable quality—the integer nature of the bias. Since all clock terms are removed in the differencing process, only base line and integer bias effects remain. The results of processing data from many base lines whose lengths are less than, say, 30 km are that the estimates of the biases are indeed very close to integers. This integer nature will deteriorate as the base line lengths are increased and non-cancelling effects such as ionospheric and tropospheric refraction and satellite orbit errors contaminate the solutions.

If the integers can be determined, then these biases should be forced to keep their integer values and be excluded from the set of solution parameters. This greatly strengthens the recovery of base line components. As carrier phase tracking receivers become more available, the opportunity to simultaneously collect data from several locations will become more routine. Not only is this desirable from an economical point of view, but with 3 or more receivers collecting data at the same time, vector closure is an important characteristic of the data reduction scheme when all data are processed together.

One data scheme whose least-squares weight matrix will remain uncorrelated (diagonal) no matter how many receivers are collecting data simultaneously is one which analyzes the undifferenced phases. It is quite obvious that a phase-only algorithm will consider uncorrelated data, but how does one exploit the integer nature of the biases? This question will now be answered.
1.2 The base station - base satellite concept

The base station - base satellite concept can best be presented by an example. Suppose that at the k-th epoch a sufficient number of phase measurements are collected to constitute a double difference. Again let us assume there are two stations, 1 and 2, and two satellites, 4 and 9. Then from Eq. (3), the measurements are given as follows:

\[ \phi_1^4(t_k) = \phi^4(t_k) - \frac{f}{c} \rho_1^4 - \phi_1(t_k) + N_1^4 \]
\[ \phi_2^4(t_k) = \phi^4(t_k) - \frac{f}{c} \rho_2^4 - \phi_2(t_k) + N_2^4 \]
\[ \phi_1^9(t_k) = \phi^9(t_k) - \frac{f}{c} \rho_1^9 - \phi_1(t_k) + N_1^9 \]
\[ \phi_2^9(t_k) = \phi^9(t_k) - \frac{f}{c} \rho_2^9 - \phi_2(t_k) + N_2^9. \]

Again it should be emphasized that these four phase measurements could be used to generate one double-difference observable. Now the above relations are rewritten substituting \( \tilde{N}^i_j = N^i_j + \phi^i - \phi_j \) in each equation. One gets

\[ \phi_1^4(t_k) = - \frac{f}{c} \rho_1^4 - \tilde{N}_1^4(t_k) \]
\[ \phi_2^4(t_k) = - \frac{f}{c} \rho_2^4 + \tilde{N}_2^4(t_k) \]
\[ \phi_1^9(t_k) = - \frac{f}{c} \rho_1^9 + \tilde{N}_1^9(t_k) \]
\[ \phi_2^9(t_k) = - \frac{f}{c} \rho_2^9 + \tilde{N}_2^9(t_k). \]  

(6)

 Arbitrarily, station 1 and satellite 4 will be chosen as the base station and base satellite, respectively. Then let us define

\[ k_2^0 = \left[ \tilde{N}_2^9(t_k) - \tilde{N}_2^4(t_k) \right] - \left[ \tilde{N}_1^9(t_k) - \tilde{N}_1^4(t_k) \right]. \]

Notice that no time argument is included in the K variable. This is intentional. After substituting for the N's, one gets

\[ k_2^0 = N_2^9 - N_2^4 - N_1^9 + N_1^4. \]  

(7)

That is, the K variable consists of the linear combination of the initial integer ambiguities. It remains the same throughout an observation session if lock is maintained. This is the essence of the algorithm. Now the four measurements are rewritten with only the fourth measurement actually changed.
\[ \phi_1^4(t_k) = -\frac{f}{c} \rho_1^4 + \overline{N}_1^4(t_k) \]
\[ \phi_2^4(t_k) = -\frac{f}{c} \rho_2^4 - \overline{N}_2^4(t_k) \]
\[ \phi_1^9(t_k) = -\frac{f}{c} \rho_1^9 + \overline{N}_1^9(t_k) \]
\[ \phi_2^9(t_k) = -\frac{f}{c} \rho_2^9 + K_2^9 - \overline{N}_2^9(t_k) + \overline{N}_1^9(t_k) - \overline{N}_1^4(t_k). \] (8)

The key idea is that \( \phi_2^9(t_k) \) is the only phase observable which has neither a base station subscript nor a base satellite superscript. When this occurs, the \( \overline{N} \) is replaced with the \( K \) formulation. The \( K \) values are indeed the same integer biases one would obtain if double differences were generated. But this formulation allows the data to remain uncorrelated and thus simplify the overall process, no matter how many stations collect data simultaneously. Some base station and base satellite must be chosen. This selection is not at all crucial to the data reduction.

Now the technique can be stated formally. (1) Choose the base station and the base satellite. (2) When a phase measurement is encountered, if it involves either the base satellite or the base station, then the mathematical formulation free of the \( K \)'s is used. Otherwise the \( K \) formulation given in the last equation of (8) is used. Epoch-to-epoch values of \( N \) must be estimated. The station positions and \( K \) variables are common contributors to all epochs. This formulation also allows for a very particular structure of the set of least-squares normal equations 2).

1.3 Ionospheric refraction

At radio frequencies the effect of the ionosphere on group velocity is to retard it inversely proportionally to the square of the frequency. A corresponding advance is created in the phase. Thus Eq. (3) must be augmented to account for this effect. The GPS satellites transmit at two frequencies, \( L_1 \) and \( L_2 \). The \( L_1 \) frequency is 1575.42 MHz which is exactly 154 times the fundamental P-code chipping rate of 10.23 MHz. The \( L_2 \) frequency is 1227.6 MHz which is exactly 120 times the P-code rate. Equation (3) is now augmented to account for the ionospheric effect,

\[ \phi_2^i(t_k) = \phi_1^i(t_k) - \frac{f}{c} \rho_2^i - \phi_2^i(t_k) + N_2^i + A/f. \] (9)

Equation (9) must be written twice, once each for the \( L_1 \) and \( L_2 \) frequencies. At this time we recognize that both the station and satellite \( L_2 \) phase values should be equal to the corresponding \( L_1 \) phase values scaled by the frequency ratio \( f_2/f_1 \) since they are both based on the same fundamental rate of 10.23 MHz.
\[ \phi_j^1(t_k)_{L_1} = \phi_j^1(t_k) - \phi_j(t_k) - f_1/c \cdot \rho_j^1 + N_j^1(L_1) + A/f_1 \]
\[ \phi_j^1(t_k)_{L_2} = f_2/f_1 \left[ \phi_j^1(t_k) - \phi_j(t_k) - f_1/c \cdot \rho_j^1 \right] + N_j^1(L_2) + A/f_2 \]  

(10)

We desire to eliminate the ionospheric effect by combining the two observations above in a linear combination as follows:
\[ \phi_j^1(t_k) = \alpha_1 \phi_j^1(t_k)_{L_1} + \alpha_2 \phi_j^1(t_k)_{L_2} \]

Choosing
\[ \alpha_1 = \frac{f_2^2}{f_1^2 - f_2^2} \]
\[ \alpha_2 = -\frac{f_1 f_2}{f_1^2 - f_2^2} \]  

(11)

yields the following "corrected" observation:
\[ \phi_j^1(t_k) = \phi_j^1(t_k) - f_1/c \cdot \rho_j^1 - \phi_j(t_k) + \alpha_1 N_j^1(L_1) + \alpha_2 N_j^1(L_2). \]  

(12)

Equation (12) now looks exactly as the original phase equation given in (3) except the integer bias term is replaced with the linear combination of the \( L_1 \) and \( L_2 \) integer biases. Thus when ionospherically corrected data are processed, no programming changes except to linearly combine the dual frequency data into one measurement are required. Once combined, the data reduction continues without change.

2. DATA PROBLEMS

The most prevalent problem associated with GPS carrier phase observation is the occurrence of cycle slips. This occurs when the satellite signal is lost and reacquisition is required. Upon reacquiring the carrier, the integer ambiguity almost always takes on a different value which appears in the data as an abrupt jump in the carrier phase history. This jump must be accommodated in the analysis. It can be manually fixed. Additional differences can isolate the jump as is done in triple differencing, or, additional unknowns can be introduced whenever a jump occurs. As is usually the case, the technique which yields the most precise baseline determination requires the most labor, and the most automatic procedure does not make use of the integer nature of the ambiguity. One possible exception will be the analysis of data over short baseline lines where the integer values of the biases are easily obtained. For these, cycle slips can be accommodated automatically by simply ignoring the integer cycle count. By using only the fractional part of the measurement residuals (rounded to zero), phase measurement processing with biases set to integers should proceed easily after a good estimate has been made available from some other technique such as triple differencing.
The cycle slips can occur on either the \( L_1 \) channel or the \( L_2 \) channel. They can occur at the same times or different times. The sources can be in the hardware or due to signal blockages such as trees, buildings, mountains, etc.

Large cycle slips can be identified when looking at single station data, but to date these jumps usually cannot be repaired to better than a cycle or two except when the data are combined interferometrically. The optimum solution is to fix the cycle slips on a station-by-station basis (unless the slip is due to an abnormally large gap in the data history), and a possible solution to this problem will be discussed next.

3. SINGLE SITE CYCLE FIXING

A possible technique for repairing cycle slips at a single station is based on the availability of dual frequency data. We start with Eq. (10) and note that the clock-based terms (receiver phase, satellite phase, and range term) are the same for \( L_1 \) and \( L_2 \) except for the factor \( f_2/f_1 \). The \( L_1 \) and \( L_2 \) integer ambiguities are not related, and the ionospheric terms are related by \( f_1/f_2 \). Thus when the linear combination

\[
\delta_k = \phi_k^1(t_k) L_1 - f_1/f_2 \phi_k^2(t_k) L_2
\]

is formed, all the clock terms drop out and then

\[
\delta_k = N_j^1(L_1) - \frac{f_1}{f_2} N_j^2(L_2) + (1-\frac{f_1^2}{f_2^2})
\]

Since the integers \( N_j \) should remain the same, epoch-to-epoch changes in this quantity are only due to the ionosphere. The term involving \( A \) represents 65% of the contribution to the \( L_1 \) phase measurement (with opposite sign). Figure 1 was generated from a dual frequency phase history while collecting data in Alaska in 1984. The epochs were separated by 120 seconds, and the data were collected at approximately 9:00 a.m. local time. This data set was chosen for its obvious cycle slips at epochs 6, 7, and 8. A listing of the raw phase data is given in Table 1 along with the differenced data. The Doppler contribution has been removed based on the pseudo-range determination of absolute position. What is noticed immediately is the difference in dynamic range between the Doppler corrected \( L_1 \) and \( L_2 \) phase values and differenced data. The average change from epoch to epoch of the differenced data is 0.1 cycles while it is 17 cycles for \( L_1 \) and 13 cycles for \( L_2 \) (disregarding the cycle slips). Also remember that these epoch-to-epoch differences are over a time interval of 120 s. What is proposed here is to track the difference signal and use it to detect and fix short term cycle slips. The \( L_1 \) coefficient in Eq. (13) is 1.0 and the \( L_2 \) coefficient is -1.28. Thus an isolated cycle slip would cause the \( \delta_k \) to change by either 1.0 or -1.28. If the same slip occurred on both channels, the \( \delta_k \) would change by 0.28, much greater than the uncertainty that one could predict. Also, if data closer in time were collected, the
epoch-to-epoch changes would be reduced proportionally to the decreased interval, but the cycle slip changes would remain at the same large values. Another advantage to tracking the difference quantity is being able to follow the contribution of the ionosphere (see Bender and Larden\textsuperscript{4}).

![Graph](image)

**Fig. 1** Raw Phase Difference \([L_1-\text{FL}/\text{F2})L_2\]

4. **FILTERING TECHNIQUES**

A Kalman filter is easily implemented to follow the signals and the difference quantity. I have chosen to represent the state transitions of the signals as a third-order process (quadratic polynomial). This model is linear in the polynomial coefficients with state transitions computed as follows:

\[
\begin{align*}
\begin{bmatrix}
X(t_{n+1}) \\
\dot{X}(t_{n+1}) \\
\ddot{X}(t_{n+1})
\end{bmatrix} &=
\begin{bmatrix}
1 & \Delta t & 1/2 \Delta t^2 \\
0 & 1 & \Delta t \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
X(t_n) \\
\dot{X}(t_n) \\
\ddot{X}(t_n)
\end{bmatrix}
\end{align*}
\]

or

\[
X(t_{n+1}) = \Phi(\Delta t) X(t_n)
\]

where the prime denotes previously updated values. The covariance of the prediction is
Table 1

$L_1/L_2$ phase differences and Doppler-corrected phase values.

<table>
<thead>
<tr>
<th>Epoch</th>
<th>Difference (cycle)</th>
<th>$L_1$ Phase (cycle)</th>
<th>$L_2$ Phase (cycle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.14</td>
<td>.99</td>
<td>.66</td>
</tr>
<tr>
<td>2</td>
<td>.11</td>
<td>-17.00</td>
<td>-13.33</td>
</tr>
<tr>
<td>3</td>
<td>.11</td>
<td>-34.49</td>
<td>-26.96</td>
</tr>
<tr>
<td>4</td>
<td>.10</td>
<td>-51.47</td>
<td>-40.19</td>
</tr>
<tr>
<td>5</td>
<td>.10</td>
<td>-67.58</td>
<td>-52.74</td>
</tr>
<tr>
<td>6</td>
<td>39129.58</td>
<td>39045.94</td>
<td>65.17</td>
</tr>
<tr>
<td>7</td>
<td>261022.78</td>
<td>197508.41</td>
<td>-4941.71</td>
</tr>
<tr>
<td>8</td>
<td>794221.62</td>
<td>79303.44</td>
<td>91.77</td>
</tr>
<tr>
<td>9</td>
<td>-.51</td>
<td>-134.33</td>
<td>-104.27</td>
</tr>
<tr>
<td>10</td>
<td>-.51</td>
<td>-149.09</td>
<td>-115.77</td>
</tr>
<tr>
<td>11</td>
<td>-.52</td>
<td>-163.76</td>
<td>-127.19</td>
</tr>
<tr>
<td>12</td>
<td>-.57</td>
<td>-178.04</td>
<td>-138.28</td>
</tr>
<tr>
<td>13</td>
<td>-.56</td>
<td>-192.07</td>
<td>-149.22</td>
</tr>
<tr>
<td>14</td>
<td>-.56</td>
<td>-206.28</td>
<td>-160.30</td>
</tr>
<tr>
<td>15</td>
<td>-.59</td>
<td>-220.72</td>
<td>-171.53</td>
</tr>
<tr>
<td>16</td>
<td>-.58</td>
<td>-235.21</td>
<td>-182.82</td>
</tr>
<tr>
<td>17</td>
<td>-.60</td>
<td>-249.48</td>
<td>-193.93</td>
</tr>
<tr>
<td>18</td>
<td>-.63</td>
<td>-264.19</td>
<td>-205.36</td>
</tr>
<tr>
<td>19</td>
<td>-.65</td>
<td>-279.17</td>
<td>-217.02</td>
</tr>
<tr>
<td>20</td>
<td>-.67</td>
<td>-294.03</td>
<td>-228.58</td>
</tr>
<tr>
<td>21</td>
<td>-.68</td>
<td>-309.23</td>
<td>-240.42</td>
</tr>
<tr>
<td>22</td>
<td>-.74</td>
<td>-324.26</td>
<td>-252.09</td>
</tr>
<tr>
<td>23</td>
<td>-.75</td>
<td>-339.50</td>
<td>-363.96</td>
</tr>
<tr>
<td>24</td>
<td>-.83</td>
<td>-355.06</td>
<td>-276.02</td>
</tr>
<tr>
<td>25</td>
<td>-.85</td>
<td>-371.08</td>
<td>-288.49</td>
</tr>
<tr>
<td>26</td>
<td>-.86</td>
<td>-387.03</td>
<td>-300.91</td>
</tr>
<tr>
<td>27</td>
<td>-.92</td>
<td>-403.58</td>
<td>-313.76</td>
</tr>
<tr>
<td>28</td>
<td>-.96</td>
<td>-420.36</td>
<td>-326.80</td>
</tr>
<tr>
<td>29</td>
<td>1.03</td>
<td>-437.15</td>
<td>-339.83</td>
</tr>
<tr>
<td>30</td>
<td>1.04</td>
<td>-454.38</td>
<td>-353.25</td>
</tr>
<tr>
<td>31</td>
<td>1.17</td>
<td>-472.02</td>
<td>-366.93</td>
</tr>
<tr>
<td>32</td>
<td>1.15</td>
<td>-489.78</td>
<td>-380.74</td>
</tr>
<tr>
<td>33</td>
<td>1.18</td>
<td>-507.66</td>
<td>-394.65</td>
</tr>
<tr>
<td>34</td>
<td>1.24</td>
<td>-526.18</td>
<td>-409.04</td>
</tr>
<tr>
<td>35</td>
<td>1.25</td>
<td>-543.86</td>
<td>-422.81</td>
</tr>
<tr>
<td>36</td>
<td>1.29</td>
<td>-561.57</td>
<td>-436.58</td>
</tr>
<tr>
<td>37</td>
<td>1.38</td>
<td>-579.04</td>
<td>-450.12</td>
</tr>
<tr>
<td>38</td>
<td>1.45</td>
<td>-596.62</td>
<td>-463.76</td>
</tr>
<tr>
<td>39</td>
<td>1.54</td>
<td>-614.30</td>
<td>-477.47</td>
</tr>
<tr>
<td>40</td>
<td>1.62</td>
<td>-631.96</td>
<td>-491.17</td>
</tr>
<tr>
<td>41</td>
<td>1.72</td>
<td>-649.58</td>
<td>-504.82</td>
</tr>
<tr>
<td>42</td>
<td>1.79</td>
<td>-667.15</td>
<td>-518.48</td>
</tr>
<tr>
<td>43</td>
<td>1.90</td>
<td>-684.36</td>
<td>-531.78</td>
</tr>
<tr>
<td>44</td>
<td>1.95</td>
<td>-701.65</td>
<td>-545.21</td>
</tr>
<tr>
<td>45</td>
<td>2.01</td>
<td>-718.30</td>
<td>-558.15</td>
</tr>
<tr>
<td>46</td>
<td>2.16</td>
<td>-735.47</td>
<td>-571.40</td>
</tr>
<tr>
<td>47</td>
<td>2.27</td>
<td>-752.41</td>
<td>-584.52</td>
</tr>
<tr>
<td>48</td>
<td>2.34</td>
<td>-769.38</td>
<td>-597.68</td>
</tr>
<tr>
<td>49</td>
<td>2.42</td>
<td>-786.50</td>
<td>-610.97</td>
</tr>
<tr>
<td>50</td>
<td>2.55</td>
<td>-803.56</td>
<td>-624.16</td>
</tr>
</tbody>
</table>
calculated as

\[
\begin{bmatrix}
P_{n+1}
\end{bmatrix} = \phi(\Delta t) \begin{bmatrix}
P_n
\end{bmatrix} \phi^T(\Delta t) + [Q]
\] (17)

where the matrix [Q] represents the uncertainty in the predictions due to errors or omissions in the transition matrix. The [Q] matrix can take on a rather dynamic behavior whereby the magnitude of the elements can be chosen to be small when smooth behavior is encountered, and large when the quadratic model cannot adequately model the state transitions. The measurements (L1, L2 phase; phase differences) are then used to update the predicted quantities by the standard Kalman formulae. Finally the covariance matrices are updated.

Again, the vectors X represent the values of measured carrier phase difference between the satellite transmitted phase fronts and the internal receiver oscillator values or the differenced quantities \( \phi(L_1) - (f_1/f_2) \phi(L_2) \). The latter values are the least changing quantities (see Figs. 2, 3, and 4) and are used for deciding when lock was not maintained. If a loss-of-lock occurs, the differenced quantity and the \( L_2 \) prediction are used to obtain integer corrections to the \( L_1 \) and \( L_2 \) phase measurements. And the process continues.

Fig. 2 Filtered Phase Difference \( [L_2-(f_1/f_2)L_2] \)
Fig. 3 Filtered Phase Data ($L_1$)

Fig. 4 Filtered Phase Data ($L_2$)

* Doppler contribution has been removed.
Experience with several data sets has revealed that the corrections are good to 1 cycle or better. This alone justified preprocessing the data prior to interferometric analysis. However, should a data sampling interval much smaller than 120 seconds be chosen, then the predictor/filter performance would improve. This does not imply that the increased sampling must be used in the interferometric processing step. The single site, preprocessed data could be sampled to reduce the computational workload in subsequent data reductions. A typical sampling rate might be from 3 to 30 seconds.

5. APPLICATION TO DYNAMIC PROBLEMS

Returning to Eq. (14), it is again emphasized that the term containing the receiver to satellite range value has been removed in the differencing process. Obviously, if the receiver was in motion (say in an airplane), the term would not appear and thus the same gradually changing signal as given in Fig. 2 would be expected. However, the ability to predict the \( L_2 \) phase values will be degraded by the arbitrary motion of the platform. But still, this is one very important quantity that can be used to measure whether or not lock has been maintained.

6. UNMODELED ERRORS

No matter which technique one chooses, it is clear that unmodeled errors are always present and will contaminate the data reductions to some (hopefully small) level. When combining data from more than two stations, it is also clear that cancellation of these unmodeled errors, such as environmental effects, will be more complete for the closer stations. Expressed mathematically we say that the measurement errors are correlated based on station separation. Given some gain matrix, \( K \), the linear estimate can be written

\[
\hat{x} = K y
\]

where \( x \) is the vector of estimated parameter values, \( y \) is the vector of phase measurements. The key idea here is to realize that the vector \( y \) is composed of more than just parameters \( x \) and noise \( e \). Let us express \( y \) as

\[
y = A x + B z + e
\]

where the design matrices \( A \) and \( B \) map the effects of the parameters to be estimated, \( x \), and the unadjusted parameters (but which have error), \( z \) respectively. Usually \( K = (A^T W A)^{-1} A^T W \) where the \( W \) matrix is the inverse of the "noise" covariance matrix. But in this case, the properly modeled covariance matrix consists of both noise and contribution of the unadjusted errors. The proper weight matrix should be modeled as
\[ W^{-1} = B \text{cov}(z) B^T + \text{cov}(e) \]

where \( \text{Cov}(e) = E(ee^T) \),
\( \text{Cov}(z) = E(zz^T) \), and the mean values of \( z \) and \( e \) are assumed to be zero.

This is almost identical to the approach used in collocation\(^5\). By properly modeling the correlation between phase measurements, the estimation of base line components and the integer ambiguities could accommodate systematic error contributions. Put a different way, modeling correlations based on station separation may keep errors present on long base lines from contaminating integer bias recovery on short base lines when simultaneous solutions with many stations' data are attempted. More rigorous error statistics should also be available from the data reduction.

Most often residual errors due to environmental sources such as the troposphere or ionosphere are difficult, if not impossible, to model in the matrix \( B \). Thus, some rule is usually used and is expressed mathematically as a function of distance between stations. In this regard, any rule that is used should not destroy the positive-definite structure of the weight matrix, \( W^0 \). This is identical to the standard collocation implementation.

REFERENCES


GPS ORBIT DETERMINATION USING THE DOUBLE DIFFERENCE PHASE OBSERVABLE

G. Beutler
Astronomical Institute, University of Berne, Switzerland

ABSTRACT

After discussing orbital precision requirements for GPS space vehicles a review of simple but correct orbit determination (and orbital biases elimination-) procedures is given. The 1984 Alaska GPS experiment and the 1985 High Precision Baseline Test (HPBL-Test) are presented as typical examples where orbit determination techniques are mandatory. The results indicate that we may expect accuracies of 1 part in $10^7$ in the near future for GPS-derived regional and continental networks -- provided orbit determination techniques are combined with the network processing mode.

1. INTRODUCTION

In this article we summarize theory and applications of GPS orbit determination facilities developed at the Astronomical Institute of the University of Berne, Switzerland. Most of the material presented here may be found in Refs. 1-3.

Although only partially deployed, the Global Positioning System (GPS) is already being exploited for high precision geodetic surveys. The Macrometer® V-1000, a first generation interferometric receiver using only the L1 carrier phase of the GPS signal, has clearly demonstrated its capability in differential positioning. In 1984 the Alaska GPS Experiment Ref. 2, in 1985 the March HPBL-test (High Precision Baseline-Test. see e.g. Ref. 3, clearly demonstrated the potential of measuring networks with GPS on a regional or even continental scale with an accuracy of a few centimeters - provided the orbits are known or are being estimated with an accuracy of 1-2 meters.

These orbital precision requirements suggest that an extended worldwide tracking network equipped with high-precision receivers will be required. Such a network is not now available to civilian users nor will one be available in the immediate future. Plans for several regional civilian tracking networks are presently being developed and/or imple-
mented, and these may provide acceptable orbits over certain regions. However, the geodesist using GPS for large-scale, high-precision surveys may not be able to assume that the GPS orbits are known to sufficiently high accuracy. He may have to estimate so-called orbital biases along with the parameters in which he is actually interested, i.e., the relative coordinates of the receivers.

Methods to estimate such orbital biases have been developed and widely used in the processing of Transit Doppler data. With the exception of some of the so-called short arc procedures, these algorithms do not describe the orbits by physical parameters. For example, one technique is to parallely shift and rotate the orbit. What usually results is non-physical in the sense that the resulting orbit is not a particular solution of the equations of motion of the satellite.

Here we attempt to demonstrate that there is no need for such non-physical methods in the determination of GPS satellite orbits. As a matter of fact, it is quite simple to model the orbital biases for these satellites to any precision required in a purely physical way. Because of the orbital characteristics of GPS satellites (almost circular orbits with semi-major axes of about 26,500 km), we may assume the earth’s gravity field to be known. As a rigorous modelling of the gravitational attraction due to the sun and the moon is not a serious problem, the only significant external force on the satellites which is not adequately known a priori is radiation pressure.

The subsequent discussion is divided into three sections. Orbital precision requirements for geodetic applications are discussed in Section 2. Then the necessary theory is developed in Section 3. In Section 4 we present applications.

2. PRECISION REQUIREMENTS

The accuracy of orbits needed to obtain baseline estimates of a certain accuracy depends mainly on the length of the baseline (see Ref. 4, Eq. (84)):

\[
\frac{db}{b} = \frac{dr}{\rho}
\]  

(1)

where \( b \) is the length of the baseline
\( \rho \) is the range (receiver to satellite)
\( dr \) is the orbit error
\( db \) is the induced baseline error.

For GPS satellites we have approximately

\[
\rho = 25000 \text{ km}
\]  

(2)
Accepting, for example,

\[ db = 1 \text{ cm} \]  \hspace{1cm} (3)

as a maximum for the baseline error introduced by the orbit, we obtain the values in Table 1 for the maximum orbit error, \( dr \), allowed.

**Table 1**

Maximum permissible orbit error, \( dr \), of 1 cm for an accuracy in a baseline of length \( b \).

<table>
<thead>
<tr>
<th>( b ) (km)</th>
<th>( dr ) (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2500.0</td>
</tr>
<tr>
<td>1.0</td>
<td>250.0</td>
</tr>
<tr>
<td>10.0</td>
<td>25.0</td>
</tr>
<tr>
<td>100.0</td>
<td>2.5</td>
</tr>
<tr>
<td>1000.0</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Rather than assuming a specific value for a baseline error independent of baseline length, we may wish to talk about a relative error expressed as parts per million (ppm) of the baseline length. Table 2 gives the values of \( dr \) for a number of relative baseline errors.

**Table 2**

Maximum permissible orbit error, \( dr \), for a certain relative accuracy in the baseline

<table>
<thead>
<tr>
<th>( db/b ) (ppm)</th>
<th>( dr ) (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>125.0</td>
</tr>
<tr>
<td>1</td>
<td>25.0</td>
</tr>
<tr>
<td>0.5</td>
<td>12.5</td>
</tr>
<tr>
<td>0.1</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Summarizing Tables 1 and 2:

a) The orbital accuracy required for surveying depends highly on the length of the base lines to be measured.

(b) for "local surveys" (diameter of surveyed region smaller than 100 km), an orbital
accuracy of 5 m to 10 m will be sufficient in most cases.

(c) for regional or even continental surveys, orbital accuracies of 1 m to 4 m will be required.

3. PRINCIPLES OF ORBIT DETERMINATION

The orbit of every satellite is a particular solution of a system of second-order differential equations:

$$\ddot{\mathbf{r}}(t) = \mathbf{r}(t; \dot{\mathbf{r}}(t), \mathbf{r}(t), \mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n)$$

(4)

where $\mathbf{r} = \mathbf{r}(t)$ is the position of the satellite in a nonrotating (with respect to inertial space), geocentric coordinate system.

$\ddot{\mathbf{r}}(i), i=1,2,$ are the first and second time derivatives of $\mathbf{r}(t)$

$\mathbf{p}_i, i=1,2,\ldots,n$ are parameters defining the forces acting on the satellite ("dynamical" parameters, e.g. parameters describing radiation pressure).

To define an orbit uniquely (one particular solution of Eq. (4)), additional information has to be supplied. Normally the problem is formulated as an initial value problem:

$$\mathbf{r}(t_0) = \mathbf{r}_0(k_1, k_2, \ldots, k_6)$$

$$\dot{\mathbf{r}}(t_0) = \mathbf{r}_0(1)(k_1, k_2, \ldots, k_6)$$

(5)

where $k_i, i=1,2,\ldots,6$ are six parameters uniquely specifying the vectors on the right-hand sides of Eq. (5). Possible choices for these parameters are:

- Components of vectors $\mathbf{r}_0, \mathbf{r}_0(1)$
- Osculating orbital elements at time $t_0$.

If we know the right-hand sides of Eq. (5) (parameters $k_i, i=1,2,\ldots,6$) and the dynamical parameters $\mathbf{p}_i, i=1,2,\ldots,n$, the orbit of a satellite is uniquely defined. We therefore may state:

(a) Orbit determination in its usual, more restricted, sense is defined as the problem of determining the six parameters $k_i, i=1,2,\ldots,6$ defining the initial values on the right-hand sides of Eqs. (5).

(b) Orbit determination in its most general sense is the problem of determining the six parameters $k_i, i=1,2,\ldots,6$ defining the initial values or boundary values and
the dynamical parameters \( p_i \), \( i=1,2,\ldots,n \).

For the application we have in mind here (modelling the orbits of GPS satellites) most of the dynamical parameters in Eq. (4) may be assumed to be known (coefficients of the earth’s gravity field, gravitational force of the sun and the moon). For utmost accuracy, however, we will have to estimate some of those parameters defining radiation pressure.

To solve an orbit determination problem, at some point one needs observations. An observation may be defined as a value of a function of satellite positions, ground positions, and nuisance parameters, like clock offsets or clock drifts of satellite and receiver clocks. More specifically, the following GPS observables have been instrumented in presently available receivers:

- pseudoranges using the C/A-code
- pseudoranges using the P-code
- Doppler measurements
- phase measurements of carriers.

The carrier phase measurements are potentially the most powerful measurements as they can be made with the greatest precision. We therefore assume that we will be dealing with measurements of this kind for orbit determination problems.

Every orbit determination is actually an orbit improvement process using observations such as those given above. These observations are nonlinear functions of the satellite position \( \mathbf{r}(t_i) \) (and possibly of the velocity \( \mathbf{r}^{(1)}(t_i) \)) at observation time \( t_i \). \( \mathbf{r}(t_i) \) in turn is a nonlinear function of the parameters of the orbit. The linearization of the orbit determination problem is therefore done in two steps: (a) the observation has to be approximated as a linear function of \( \mathbf{r}(t_i) \), which is straightforward; and (b) \( \mathbf{r}(t_i) \) has to be represented by a linear function of the unknown parameters \( k \) (and possibly some of the \( p \)). This is done in the following way. Since Eq. (4) is nonlinear, we must linearize it and determine our orbit iteratively, where in each iteration step we assume we have a known approximate orbit \( \mathbf{r}_a(t) \) at our disposal.

\[
\begin{align*}
\mathbf{r}_a(2) &= \mathbf{r}(t_i; \mathbf{r}_a, \mathbf{r}_a^{(1)}, p_1, p_2, \ldots, p_n) \\
\mathbf{r}_a(t_0) &= \mathbf{r}_a(0) = \mathbf{r}_a^{(1)}(0) \approx \mathbf{r}_a^{(1)}(t_0) + \mathbf{r}_a^{(2)}(t_0) = \mathbf{r}_a^{(1)}(k_{a1}, k_{a2}, \ldots, k_{a6}) = \mathbf{r}_a^{(1)}(t_0) + \mathbf{r}_a^{(2)}(t_0) \\
\mathbf{r}_a^{(1)}(t_0) &= \mathbf{r}_a^{(1)}(k_{a1}, k_{a2}, \ldots, k_{a6})
\end{align*}
\]

where the \( k_{a1} \), \( i=1,2,\ldots,6 \) are approximate values of the unknown parameters.

The true, initially unknown, orbit is now assumed to be a linear(ized) function of the parameters \( k_{a} \), \( i=1,2,\ldots,6 \):
\[ \vec{r}(t) = \vec{r}_a(t) + \sum_{i=1}^{6} \vec{z}_1(t)(k_i - k_a) \]  \hspace{1cm} (7)

where

\[ \vec{z}_1(t) = \left. \frac{\partial \vec{r}_a}{\partial k_i} \right|_{k=k_a} \quad i=1,2,\ldots,6 \]

If we deal with the more general problem, additional terms involving the dynamical parameters appear on the right-hand side of Eq. (7):

\[ \vec{r}(t) = \vec{r}_a(t) + \sum_{i=1}^{6} \vec{z}_1(t)(k_i - k_a) + \sum_{i=1}^{n} \vec{z}_1^*(t)(p_i - p_{a_i}) \]  \hspace{1cm} (8)

where

\[ \vec{z}_1^*(t) = \left. \frac{\partial \vec{r}_a}{\partial p_i} \right|_{p=p_{a_i}} \quad i=1,2,\ldots,n \]  \hspace{1cm} (9)

The functions \( \vec{z}_1(t) \), \( \vec{z}_1^*(t) \) are solutions of an initial (or boundary) value problem, which follow from the primary problem (6) by taking the (total) derivatives of all the equations in (6) with respect to \( k_i \), \( i=2,3,\ldots,6 \) and \( p_i \), \( i=1,2,\ldots,n \). The resulting set of differential equations are usually called the systems of variational equations. It is elementary to show (see Ref. 1) that each of the \( \vec{z}_1 \) or \( \vec{z}_1^* \) is a particular solution of a linear second order differential equation system.

Let us summarize: In every iteration step of the orbit improvement process, we have to solve one system of nonlinear differential equations of type (6), and, for each orbit parameter we have to estimate (for each of the \( \vec{z}_1 \), \( \vec{z}_1^* \)) one linear differential equation system.

In our GPS software system we solve the nonlinear systems (Eqs. (4), (5) with the technique of numerical integration, the partial derivatives \( \vec{z}_1 \), \( i=1,2,\ldots,6 \) are computed approximately: (a) the \( k_i \), \( i=1,2,\ldots,6 \) are chosen to be the osculating elements pertaining to the initial epoch \( t_0 \); (b) in formula (9) the approximate orbit \( \vec{r}_a \) is replaced by the Kepler orbit defined by these osculating elements. It is then possible to compute the partials analytically using the well known formulae of the two-body problem.

Let us conclude this section by briefly reviewing our modelling of the force field acting on GPS-satellites. (In Ref. 2 the influences of the constituents of the force field on the GPS orbits are discussed in some detail):

- The earth's gravity field is expressed by spherical harmonics, where we may select the coefficients of either the GEM-10 or the GRIM-3L1 model. (Usually we cut off the development after degree and order 8).
- Point mass attractions from sun and moon are taken into account.

- The radiation pressure model we use at present is very simple from the point of view of application: Only two parameters define the model, only two parameters may be estimated from observation data. In Fig. 1 we define three unit vectors: \( \hat{e}_1 \) is the antenna axis, \( \hat{e}_2 \) is the solar panels axis and \( \hat{e}_o \) is the unit vector pointing to the sun.

Fig. 1 Model of GPS Space Vehicle

Legend

\( \hat{e}_1 \) : Antenna axis
\( \hat{e}_2 \) : Axis of solar panels
\( \hat{e}_o \) : Unit vector to sun

We model the acceleration \( \ddot{a} \) exerted by radiation pressure as

\[
\ddot{a} := -p_o \cdot \hat{e}_o + p_a \cdot \hat{e}_2
\]  

(10)

The first term is called direct solar radiation pressure (it would be the only one if the space vehicles were spherically symmetric or if the surfaces were totally absorbing the radiation), the second is usually referred to as "y-bias". Several effects may contribute to this latter term. Misalignment of the \( \hat{e}_2 \)-axis (\( \hat{e}_a \) not perpendicular to \( \hat{e}_o \)) certainly contributes to this term. Fliegel et al. (see Ref. 5) discuss thermal radiation of the space vehicle as another candidate.
Parameters $p_a$ and $p_b$ may be estimated from observation data. This means that the partial derivatives of the orbit with respect to these parameters have to be computed too (see Eqs. (9)). At present we do that by numerical integration.

Very often it is useful to introduce a priori knowledge for some parameters, into the adjustment. If we have e.g. precise ephemerides at our disposal as a priori orbit source, it would be a pity not to use the fact, that these orbits are good to - let us say - 10 meters along track and 2 meters in the other directions. In our parameter estimation program we take a priori information for parameters into account by adding an artificial observation equation for the (improvement of) parameter under consideration. To this observation equation a weight proportional to the inverse of the (estimated) a priori variance of the parameter is assigned.

4. APPLICATIONS

The results for two campaigns are summarized here:

(a) 1984 Alaska GPS Campaign (see Ref. 6).
(b) March 1985 High Precision Baseline Test (HPBL-Test) (Ref. 3).

In the first case we are looking at a typical regional network (~ 2500 km x 1000 km) where orbit improvement had to be done as a byproduct of the analysis. the main interest lying on the estimation of site coordinate. It became clear that network-processing mode and the estimation of "orbital biases" were the two keys to the success of this particular analysis. The situation was quite different for the second analysis. Here orbit estimation was the main topic, the quality of receiver coordinates "only" served to prove the quality of the orbits estimated from GPS observations from four fiducial sites whose coordinates were known from VLBI observations.

4.1 The 1984 Alaska GPS Campaign

During the summer of 1984, the U.S. National Geodetic Survey (NGS) and the Jet Propulsion Laboratory operated Mobile VLBI systems in Alaska and Canada as part of a series of measurements by the National Aeronautics and Space Administration's Crustal Dynamics Project. Given the great expense and complex logistics of moving mobile VLBI systems to and around such a distant location, NGS decided to measure the same baselines by GPS in the hopes that as GPS technology and techniques mature and a history of GPS/VLBI comparisons emerge, these measurements may be performed solely by the less expensive GPS systems. The sites occupied by both the mobile VLBI and GPS systems are shown in Fig. 2, the approximate distance between the sites may be found in Table 3.

During this GPS/VLBI intercomparison, five TI-4100 GPS receivers were available. The original GPS observing plan was to occupy Nome, Fairbanks, Sourdough, Yakataga, and
Whitehorse for 2 consecutive days. The second scheduled session kept the same receivers at Nome and Fairbanks and moved the other receivers to Sandpoint, Kodiak and Yellowknife for 2 additional observing days. The principal observing session on each day consisted of two satellite scenarios. The first began in early afternoon local time, included SV’s 6, 8, 9 and 11 and lasted almost 3 hours. This was immediately followed by the second scenario which included SV’s 4, 9, 11, and 13 and lasted about 1 hour. A second observing session was scheduled for each day approximately 12 hours after the primary one, but it turned out that these sessions did not produce useful results.

The first observing occupation scheme went as planned and produced data that contained only a few cycle slips. The second occupation scheme had to be severely modified as it lasted for the next 2 weeks due to receiver failures in Alaska. Coincidentally these data are heavily burdened with cycle slips.

Fig. 2 The 1984-Alaska-GPS-Campaign
Table 3

Approximate distances between stations in kilometers

<table>
<thead>
<tr>
<th>Station</th>
<th>Kodiak</th>
<th>Nome</th>
<th>Sand</th>
<th>Sour</th>
<th>White</th>
<th>Yaka</th>
<th>Yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair</td>
<td>849</td>
<td>1025</td>
<td>1006</td>
<td>1179</td>
<td>591</td>
<td>414</td>
<td>1510</td>
</tr>
<tr>
<td>Kodiak</td>
<td>848</td>
<td>557</td>
<td>1004</td>
<td>1598</td>
<td>329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nome</td>
<td>1285</td>
<td>671</td>
<td>1591</td>
<td>1598</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand</td>
<td>276</td>
<td>1044</td>
<td>1276</td>
<td>1598</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sour</td>
<td>789</td>
<td>632</td>
<td>2135</td>
<td>2460</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>603</td>
<td>1726</td>
<td>2681</td>
<td>1575</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yaka</td>
<td>1631</td>
<td>2135</td>
<td>2460</td>
<td>1575</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fair : Fairbanks  White : Whitehorse
Sand : Sandpoint   Yaka : Yakataga
Sour : Sourdough   Yellow : Yellowknife

After the pre-processing step (see Ref. 6) roughly 20'000 double difference observations (8000 from scheme 1, 12000 from scheme 2) could be processed by the parameter estimation program. In the program run combining the two observation schemes, thus giving the entire Alaska network with eight sites as main result, a total of 330 parameters had to be estimated of which 180 were orbit parameters. Since we had precise ephemerides at our disposal, very small variances could be associated with the artificial observations constraining the orbit parameters (lo errors of 15 m along track and 3 m in the other directions were used, see Ref. 6 for more information).

Consistency and integrity of both, the GPS and VLBI solutions could be demonstrated by a 7-parameter Helmert transformation between the two final networks. Table 4 gives residuals and transformation parameters of this transformation.

4.2 The March 1985 High-Precision Baseline Test (HPBL-Test)

This campaign took place from March 29 to April 5 in the United States of America.

In March/April 1985 space vehicles 4, 6, 8, 9, 11, 12 and 13 could be observed. The daily observation period started approximately at 3 a.m. and lasted till 11 a.m. Nine TI-4100 receivers, three AFGL dual frequency instruments and two SERIES-X receivers were distributed over nine sites (see Fig. 3 and Table 5), among them Haystack, Fort Davis, Richmond and Mojave which we used as fiducial points (coordinates known from VLBI-surveys). (For Hat Creek and Owens Valley VLBI results are available too.)
Table 4

Residuals of a Helmert Transformation between the GPS Solution for the 1984 Alaska GPS Campaign and the VLBI-Solution transformed into WGS-72 Observations Schemes 1 and 2

<table>
<thead>
<tr>
<th>Station</th>
<th>VLBI - GPS (m)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td>YAKATAGA</td>
<td>0.146</td>
<td>-0.027</td>
<td>-0.147</td>
</tr>
<tr>
<td>FAIRBANKS</td>
<td>0.023</td>
<td>0.098</td>
<td>-0.106</td>
</tr>
<tr>
<td>KODIAK</td>
<td>-0.301</td>
<td>0.039</td>
<td>-0.038</td>
</tr>
<tr>
<td>Nome</td>
<td>-0.085</td>
<td>0.083</td>
<td>0.115</td>
</tr>
<tr>
<td>Sandpoint</td>
<td>0.132</td>
<td>-0.153</td>
<td>0.253</td>
</tr>
<tr>
<td>Sourdough</td>
<td>0.157</td>
<td>0.030</td>
<td>-0.173</td>
</tr>
<tr>
<td>Yellowknife</td>
<td>-0.112</td>
<td>-0.106</td>
<td>0.237</td>
</tr>
<tr>
<td>Whitehorse</td>
<td>0.041</td>
<td>0.036</td>
<td>-0.140</td>
</tr>
</tbody>
</table>

RMS of Transformation = 0.16 m
Rotation around x-axis = 0.18 ± 0.03 arc sec
Rotation around y-axis = -0.01 ± 0.03 arc sec
Rotation around z-axis = 0.04 ± 0.01 arc sec
Scale factor = -0.36 ± 0.07 mm/km

Fig. 3 Receivers and sites in the 1985 HBPL Test

- Hat Creek
- Mammoth Lake
- Owens Valley
- Mojave
- Fort Davis
- Austin
- Richmond
- TI-4100
- Series-X
- Macrometer II
Table 5

Distances in km between sites of the March 1985 HPIL-Test

<table>
<thead>
<tr>
<th></th>
<th>BICP</th>
<th>FTDA</th>
<th>HAYS</th>
<th>HATC</th>
<th>MAMM</th>
<th>MOJA</th>
<th>NSWC</th>
<th>RICH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUST</td>
<td>2035</td>
<td>598</td>
<td>2678</td>
<td>2417</td>
<td>2100</td>
<td>1866</td>
<td>2081</td>
<td>1778</td>
</tr>
<tr>
<td>BICP</td>
<td>1508</td>
<td>3930</td>
<td>484</td>
<td>71</td>
<td>245</td>
<td>3557</td>
<td>3742</td>
<td></td>
</tr>
<tr>
<td>FTDA</td>
<td>3136</td>
<td>1933</td>
<td>1577</td>
<td>1314</td>
<td>2586</td>
<td>2363</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAYS</td>
<td>4033</td>
<td>3959</td>
<td>3904</td>
<td>669</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HATC</td>
<td>416</td>
<td>729</td>
<td>3731</td>
<td>4066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAMM</td>
<td></td>
<td>313</td>
<td>3597</td>
<td>3800</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOJA</td>
<td></td>
<td></td>
<td>3498</td>
<td>2046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSWC</td>
<td></td>
<td></td>
<td></td>
<td>1442</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Station names:

AUST Austin ARL
BICP Owens Valley Big Pine
FTDA Fort Davis
HAYS Haystack Obs., Westford
HATC Hat Creek, CA
MAMM Mammoth Lake, Casa 1956, Ca
MOJA Mojave, Barstow, Ca
NSWC Dahlgren, Va
RICH Richmond, Perrine F

Table 6

Observation schedule for TI-4100 (TI) and AFCL (MM) receivers

<table>
<thead>
<tr>
<th>Day</th>
<th>AUST</th>
<th>BICP</th>
<th>FTDA</th>
<th>HAYS</th>
<th>HATC</th>
<th>MAMM</th>
<th>MOJA</th>
<th>NSWC</th>
<th>RICH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 29</td>
<td>TI</td>
<td>TI</td>
<td>MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
</tr>
<tr>
<td>Mar 30</td>
<td>TI</td>
<td>TI</td>
<td>TI.MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
</tr>
<tr>
<td>Mar 31</td>
<td>TI</td>
<td>TI</td>
<td>TI.MM</td>
<td>TI.MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
</tr>
<tr>
<td>Apr 01</td>
<td>TI</td>
<td>TI</td>
<td>TI.MM</td>
<td>MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>MM</td>
</tr>
<tr>
<td>Apr 02</td>
<td>TI</td>
<td>TI</td>
<td>TI.MM</td>
<td>MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
</tr>
<tr>
<td>Apr 03</td>
<td>TI</td>
<td>TI</td>
<td>MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
</tr>
<tr>
<td>Apr 04</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI.MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
</tr>
<tr>
<td>Apr 05</td>
<td>TI</td>
<td>TI</td>
<td>TI.MM</td>
<td>TI.MM</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
<td>TI</td>
</tr>
</tbody>
</table>

Station names: see Table 5

Because we did not have a complete set of SERIES-X data at our disposal, only Macro-meter and TI-observations were used for the investigation. The observation schedule for the
two instrument types are given in Table 6. The TI-receivers recorded data each 30 seconds. (which leads to impressive numbers of observations (5000 to 12000 double differences could be used per day), the Macrometer recorded data at a much lower rate (one epoch per 6 minutes).

The present analysis was severely handicapped by a number of circumstances, the most important being that

(a) No surface weather data had been available to us when we processed the campaign.

(b) The GESAR software implemented in the TI-4100 receivers was not working properly. (The consequence was that one clock offset had to be estimated for each TI-receiver for each session.)

In order to study repeatability of the estimated parameters the campaign was artificially divided into two parts:

Part A: March 29 - April 1
Part B: April 2 - April 5

Either all observations of part A or all observations of part B were processed. Tests were made with 1 day arcs, 2 day arcs, 4 day arcs. "Only" broadcast ephemerides were available as a priori information. Each arc was characterized by six osculating elements and by two radiation pressure parameters (see Eq. (10)). For the latter parameters the following a priori values were used for all satellites

\[ p_o = 0.94 \times 10^{-7} \text{ m/s}^2 \]
\[ p_s = 0 \text{ m/s}^3. \]  \hspace{1cm} (11)

The perigee was constrained by

\[ \omega = 0.1 \text{.} \]  \hspace{1cm} (12)

No constraints have been imposed on the other elements.

Results:

(a) Radiation Pressure Results

It soon became evident that for short arcs (1 day) it is not necessary (and does not make sense) to improve the a priori values for \( p_o \), \( p_s \) given in (11). For 2-day arcs only \( p_o \) could be estimated with a satisfactory reliability. Only for 4-day arcs the \( y \)-bias (\( p_s \)) gave a significant contribution.
The estimates for $p_e$ were quite consistent: The results obtained in parts A and B of the campaign differ from each other by only $1 \times 2 \times$. A similar statement holds for the different arc-lengths used. The situation is somewhat less satisfactory for $p$, the $y$-bias parameter (see Ref. 3).

(b) Osculating Elements

Two facts are worth being mentioned:

1. During the campaign, the broadcast ephemerides were of a surprisingly high quality: the positions derived from our a posteriori orbit estimates differed only for $10 - 25$ m from the broadcast predictions. (The exception was satellite 4 where differences of up to $100$ m occurred.)

2. Fig. 4 illustrates the "state of the art" of our orbit modelling capabilities: There a two-day arc (April 2, 3, solution B21) is compared with a four day arc (April 2-5), solution B42. In solution B21 for each arc seven parameters were estimated (six osculating elements and $p_e$), in solution B42 8 parameters (six elements and $p_e$, $p$). As one may see the differences between the two solutions are of the order of a few meters, which indicates that the orbits should be good to a few parts in $10^{-7}$ (see Section 2).

Fig. 4

SOLUTION B21 - SOLUTION B42

SUM* 8  DATE(START)= 1986.4 2.3.8
DIFF: RADIAL(+), ALONG TRACK(X), OUT OF PLANE(X)
Table 7

Differences in the estimated station coordinates with respect to solution B21 (only stations other than fiducial stations)

<table>
<thead>
<tr>
<th>Station</th>
<th>Difference</th>
<th>d(latitude)</th>
<th>d(longitude)</th>
<th>d(height)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cm</td>
<td>cm</td>
<td>cm</td>
</tr>
<tr>
<td>Hat Creek</td>
<td>A10 - B21</td>
<td>-2.4</td>
<td>0.3</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>A21 - B21</td>
<td>-4.9</td>
<td>-8.0</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>A42 - B21</td>
<td>-1.6</td>
<td>3.9</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>B10 - B21</td>
<td>-2.4</td>
<td>0.3</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>B42 - B21</td>
<td>0.1</td>
<td>2.3</td>
<td>-0.6</td>
</tr>
<tr>
<td>Owens Valley</td>
<td>A10 - B21</td>
<td>1.2</td>
<td>-1.7</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>A21 - B21</td>
<td>1.2</td>
<td>-5.1</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>A42 - B21</td>
<td>0.8</td>
<td>0.6</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>B10 - B21</td>
<td>1.1</td>
<td>-1.4</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>B42 - B21</td>
<td>0.0</td>
<td>1.3</td>
<td>-0.4</td>
</tr>
<tr>
<td>Mammoth Lake</td>
<td>A10 - B21</td>
<td>-1.3</td>
<td>-0.4</td>
<td>-2.3</td>
</tr>
<tr>
<td></td>
<td>A21 - B21</td>
<td>-1.7</td>
<td>-4.8</td>
<td>-0.8</td>
</tr>
<tr>
<td></td>
<td>A42 - B21</td>
<td>-1.2</td>
<td>2.2</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>B10 - B21</td>
<td>-1.3</td>
<td>-0.4</td>
<td>-2.3</td>
</tr>
<tr>
<td></td>
<td>B42 - B21</td>
<td>0.0</td>
<td>1.6</td>
<td>-0.5</td>
</tr>
<tr>
<td>Austin</td>
<td>A10 - B21</td>
<td>2.1</td>
<td>5.6</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>A21 - B21</td>
<td>2.3</td>
<td>5.3</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>A42 - B21</td>
<td>4.6</td>
<td>-0.6</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>B10 - B21</td>
<td>-2.1</td>
<td>5.6</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>B42 - B21</td>
<td>0.2</td>
<td>-2.0</td>
<td>2.5</td>
</tr>
<tr>
<td>Dahlgren (NSWC)</td>
<td>A10 - B21</td>
<td>4.0</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>A21 - B21</td>
<td>2.4</td>
<td>0.5</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>A42 - B21</td>
<td>4.9</td>
<td>-1.5</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td>B10 - B21</td>
<td>4.0</td>
<td>-0.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>B42 - B21</td>
<td>-3.1</td>
<td>7.0</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Where: Aik:  
- Solution using all observations from March 29 to April 1.
- \( i \) = arc length in days.
- \( k \) = number of radiation pressure parameters estimated.

Bik:  
- Solution using all observations from April 2 to April 5.
- \( i \) = arc length in days.
- \( k \) = number of radiation pressure parameters estimated.
(c) Coordinates

An independent check of the quality of the estimated orbits is possible by comparing the coordinates (as obtained in the various solutions) for the stations Hat Creek, Owens Valley, Mammoth Lake, Austin and Dahlgren, because these receiver locations were estimated without constraints in all program runs.

If our orbits would be good to – let us say – 1 ppm, we would expect differences of up to 4 meters in the different solutions, since we have baselines of up to 4000 km. The results in Table 7 show differences that are almost two orders of magnitude smaller!

Table 7 also shows, that the differences between the three different orbit types (1-, 2-, 4-day arcs, different radiation pressure models) cause only coordinate changes on the sub-decimeter level, which indicates adequate modelling.

We also see that the repeatability “A-solutions minus B-solutions” is of the same order of magnitude. (Remember that solutions A and B are independent in the sense that they have no observations in common and that the orbits are completely independent from each other.)

It is interesting (but not surprising), that the largest differences occur in height – hopefully surface weather data and water vapour radiometer data will improve the situation somewhat.

* * *

REFERENCES


ACCURACY PROBLEMS WHEN COMBINING TERRESTRIAL AND SATELLITE OBSERVATIONS

W.M. Welsch
Institut für Geodäsie, Universität der Bundeswehr München, D-8014 Neubiberg

ABSTRACT
The combination of terrestrial and satellite aided network observations requires hybrid models of two kinds: a hybrid functional model and a stochastic model being able to process the heterogeneous observations appropriately. An appropriate combination of stochastic information means to adjoin proper weights to the various types of observations on one hand and on the other hand to separate the relevant internal accuracy from that accuracy which is affected by a number of network external influences. The first task can be overcome by the technique of estimating variance components of the individual observation groups, the second one can be handled by applying S-transformations to covariance matrices. The paper deals with problems of the hybrid stochastic model. Examples are given.

1. INTRODUCTION

The combination of classically performed terrestrial and satellite aided geodetic network observations rank among the many tasks being attended with the introduction of the new observation techniques. Doing so, several aspects are important. One of them is the functional connection of the heterogeneous observations. This is the problem of knotting together different reference systems whose interrelations are not clearly defined. As a rule the task is mastered by similarity transformations. It has often been treated, as an extensive bibliography proves.

The stochastic hybrid model, however, has been hardly ever dealt with. Therefore, the following representation intends to make a contribution, first by analyzing the accuracies of individually adjusted terrestrial and satellite networks. These accuracies are of two kinds: accuracies which are influenced by numerous network external effects, and internal accuracies which are relevant for relative accuracy statements within the network. The combination of those internal and external accuracies in a hybrid stochastic model has to be treated with care. This is one problem. The other problem is harmonizing the stochastic model by adjoining proper weights to the individual and heterogeneous groups of terrestrial and satellite aided observations.

Therefore, when comparing or combining terrestrial and satellite aided network observations the first task is to get a deeper insight into the accuracies of the individual and heterogeneous observations.
2. **THE ACCURACY OF SATELLITE AIDED NETWORK OBSERVATIONS**

First information about the accuracies of satellite aided network observations can be taken from the various data processing software packages. The covariance matrices provided are, however, affected with all kinds of network external effects such as instrumental and propagation medium influences, ephemeris errors etc. In many cases, especially in cases of small networks and short observation periods with multi-station instrumentations, those effects exercise a systematic influence by shifting, rotating and scaling the network - both the coordinates and the elements of the covariance matrix $C_x$. In case only the mutual or relative positions of the network points are of interest, this "external" covariance matrix $C_x$ does not provide an appropriate measure of the relative or internal accuracies of the point positions. To achieve such a measure the external effects have to be split off. According to the "Internal Error Theory of Networks" (Ref. 1; Appendix A) this can be done by translations and rotations and a change of scale common to all network points. That means the covariance matrix $C_x$ of the network points provided by those data processing software packages and representing externally affected accuracies has to undergo a similarity transformation (S-transformation) in order to reveal the "real" internal accuracy $C_i$ of the network point determination.

The diagonal block structure of the covariance matrix of the GPS measurements of the INN VALLEY Network (Appendix C) shows the following characteristics:

$$
S_{x_i}^C : \ s_x = \pm 1.5 \text{ cm}, \ s_y = \pm 1.3 \text{ cm}, \ s_h = \pm 2.0 \text{ cm}
$$

(subscript $s_*$ for satellite). These values contain external effects. After their removal by S-transformation the internal accuracies are

$$
S_{x_i}^i : \ s_x = \pm 1.3 \text{ cm}, \ s_y = \pm 1.2 \text{ cm}, \ s_h = \pm 1.6 \text{ cm}
$$

3. **THE ACCURACY OF TERRESTRIAL NETWORK OBSERVATION RESULTS**

If, for instance, horizontal directions, spatial distances and zenith angles are introduced into a free three-dimensional network adjustment, freedom of the geodetic datum exists only with respect to 3 translations and the rotation around the $h$-axis. The remaining 3 datum dispositions are established by the scale of distances and the orientation of the zenith angles (plumbline or normal of the ellipsoid). Due to these dispositions of 3 datum constraints which are network external effects the network is somehow restrained from being absolutely free. The restraints are reflected in the covariance matrix of the adjusted coordinates, too. Thus the covariance matrix provided by a "free" network adjustment of this kind provides external accuracies which are, in the case of the terrestrial INN VALLEY Network,
\( \tau^{C}_{x} : s_{x} = \pm 0.4 \text{ cm}, \quad s_{y} = \pm 0.4 \text{ cm}, \quad s_{h} = \pm 3.6 \text{ cm} \) \hspace{1cm} (3)

(subscript \( \tau^{*} \) for terrestrial). After S-transformation the corresponding results of the internal accuracies are

\( \tau^{C i}_{x} : s_{x} = \pm 0.4 \text{ cm}, \quad s_{y} = \pm 0.4 \text{ cm}, \quad s_{h} = \pm 2.1 \text{ cm} \). \hspace{1cm} (4)

As expected, the S-transformation influences especially the accuracy in heights by rotations around the x- and y-axes since strong datum effects are filtered off.

4. COMPARISON OF SATELLITE AND TERRESTRIAL OBSERVATION RESULTS

The first possibility to achieve assertions on accuracies of satellite aided network observations in comparison to terrestrial network observations is the simple comparison of specific results obtained from the individual adjustments of the terrestrial and satellite observations, respectively.

The coordinates of those adjustments can be compared with each other after a 7-parameter similarity transformation which splits off network external effects caused for instance by a different geodetic datum.

\[
\begin{align*}
\tau^{R}_{S} dx &= \tau^{X} - m R^{X} s_{S} \\
\tau^{R}_{S} dy &= dy \\
\tau^{R}_{S} dh &= dh \\
\end{align*}
\hspace{1cm} (5)

coordinate differences to be investigated after transformation

\( \tau^{X} \) centre of gravity related terrestrial coordinates

\( s^{X} \) centre of gravity related GPS coordinates

\( m \) scale factor

\( R \) rotation matrix.

Applied to the example of the INN VALLEY Network the following results have been obtained (Ref. 2):

The coordinate differences in \( dx \) range from \(-2.4 \text{ cm} \) to \(+3.0 \text{ cm} \), in \( dy \) from \(-3.2 \text{ cm} \) to \(+2.8 \text{ cm} \), in height \( dh \) from \(-12.8 \text{ cm} \) to \(+12.6 \text{ cm} \). The r.m.s. error in position is \( s_{p} = \pm 2.6 \text{ cm} \), while the r.m.s. error in height is \( s_{h} = \pm 8.0 \text{ cm} \). \( s_{p} \) is an excellent result, it corresponds quite nicely to what can be derived from Eq. 1. The uncertainty in height \( s_{h} \) can be questioned, it is much higher than the r.m.s. value \( s_{h} = \pm 2.0 \text{ cm} \) of the GPS derived heights and also higher than \( s_{h} = \pm 3.6 \text{ cm} \) of the terrestrial heights. The conclusion may be drawn that the GPS heights cannot be burdened with the discrepancy. It is most likely that the refraction influenced terrestrial heights are much weaker and not so consistent; therefore, they cannot compete with the satellite aided height observations.
Most suitable for comparison purposes are also those elements which are by themselves independent of network external influences and geodetic datum invariant, i.e. spatial distances.

The differences

$$\delta d = \delta T - \delta S$$

(6)

of the terrestrial and satellite network distances are, with only one exception, positive, ranging from +0.2 cm to +9.8 cm; related to the length of the distances a scale factor of $m_T = +2.47$ ppm can be derived. The differences of the scaled distances range from -2.6 cm to +6.6 cm having a relative accuracy of ±2.0 ppm.

The above-mentioned statements are the usual ones when terrestrial and satellite observations are compared. However, the rigorous combination of both the observation types yields accuracy statements which go much further.

The first step is some considerations on the functional and stochastic hybrid network models.

5. THE COMBINATION OF TERRESTRIAL AND SATELLITE AIDED NETWORK OBSERVATIONS

- THE FUNCTIONAL MODEL

The observation equations of a combined or hybrid adjustment model have to be formulated with respect to a common reference system. For the following models the "satellite" coordinates, i.e. the network point coordinates determined by satellite aided observations, and their covariance matrix were transformed onto the terrestrial coordinate system referring to the national geodetic datum.

5.1 Gauss-Markov model

A relatively simple model regards the satellite aided observations as direct coordinate observations which are added to the observation equations of the terrestrial measurements. Therefore, the combined system of observation equations is

$$\begin{align*}
T^T + T^V &= Ax, \\
T^P_1 &= T^{C_1} \\
S^T + S^V &= Ix, \\
S^P_x &= S^{C_x}
\end{align*}$$

( $T^T$ terrestrial observations, $T^V$ satellite coordinates, $T^{P_1}$, $T^{C_1}$ and $S^P_x$, $S^{C_x}$ weight and covariance matrices respectively of the observations).
The vector $x$ of unknowns contains, among others, the point coordinates $x, y, h$ common to both types of networks.

In case the satellite coordinates are used as approximate coordinates, $S\ell = 0$. Then the normal equations are

$$(A^T T_1 P_A + S^T P_x)x = A^T T_1 P_{\ell} \ell, \tag{8}$$

where the satellite coordinates are only represented by their weight matrix $S^T P_x$. The cofactor matrix of the adjusted coordinates

$$Q_x = (A^T T_1 P_A + S^T P_x)^+ = (T^T P_x + S^T P_x)^+ = T^+ S^+ Q_x \tag{9}$$

demonstrates that the weights of the terrestrial and satellite coordinates are directly added. The mutual interaction will depend on the size and on the "nature" of the two partners. With point coordinates observed by satellite aided techniques, $Q_x$ is regular in general.

5.2 Gauss-Helmert model

Both the sets of terrestrial and satellite coordinates are considered observations within a system of condition equations with unknowns. The unknowns are the transformation parameters $p$ (Eq. A.1). The system of equations is

$$T^x + T^u - (S^x + S^u) + G^p = 0 \tag{10a}$$

or

$$S^u + G^p + \omega = 0 \tag{10b}$$

with

$$B = \begin{bmatrix}
1 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & -1 & 0 \\
& & & & & & \\
& & & & & & \\
\end{bmatrix},$$

$$v^T = \begin{bmatrix}
T^x & T^y & T^h & S^x & S^y & S^h \\
\end{bmatrix},$$

$$w^T = \begin{bmatrix}
- S^x & - T^y & - S^y & T^h & - S^h \\
\end{bmatrix}.$$
\[ P_x = \begin{bmatrix} A^T & P_1 & 0 \\ 0 & 0 & S_x \end{bmatrix} = \begin{bmatrix} T_x & 0 \\ 0 & S_x \end{bmatrix}. \] (11)

The solution is according to the well-known algorithms. Attention has to be paid to the inversion of \( P_x \) which is possibly positive semi-definit.

If one transforms model Eq. 10 owing to

\[ Bv = \tilde{v} \] (12)

into the Gauss-Markov model

\[ w + \tilde{v} + Gp = 0, \] (13)

the weight matrix \( P_x \) results from

\[ \tilde{P}_x = T_x + S_x, \] (14)

if the \( Q \)-matrices are the inverses of the corresponding \( P \)-matrices.

6. THE COMBINATION OF TERRESTRIAL AND SATELLITE AIDED NETWORK OBSERVATIONS

- THE STOCHASTIC MODEL

There are two major problems involved in the proper combination of the stochastic information of terrestrial and satellite aided network observation.

6.1 The combination of heterogeneous observations

The stochastic information contained in geodetic observations can often not be utilized in its entirety in "usual" adjustments especially in cases of heterogeneous observation material since with the unit variance \( s_0^2 \) only one single quantity of the stochastic model is estimated on a global scale. If, however, some knowledge about the stochastic structure of the observations is pre-given, the estimation of the variance components \( s_i^2 \) of individual groups of observations is made possible. This a posteriori estimation technique (Appendix B) is suitable to adjoin proper weights to the various observation groups.

Due to the linear structure of the stochastic model in Eq. 7

\[ C_1 = \sigma_0^2 Q_1 = \sigma_{TT}^2 T_1 + \sigma_{SS}^2 Q_1 \] (15)

\[ \sigma_0^2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \sigma_T^2 \begin{bmatrix} T_1 & 0 \\ 0 & 0 \end{bmatrix} + \sigma_S^2 \begin{bmatrix} 0 & 0 \\ 0 & S_x \end{bmatrix}. \]
the unit variance $\sigma_0^2$ as well as the variance components $\sigma_T^2$ and $\sigma_S^2$ can be estimated (Eq. B.16). Even a refinement is possible if one splits $\sigma_0^2$ into its components

$$\sigma_T^2 \sigma_1^2 = \sigma_{\text{dir}}^2 \sigma_{\text{dir}}^2 + \sigma_{\text{dis}}^2 \sigma_{\text{dis}}^2 + \sigma_{\text{zen}}^2 \sigma_{\text{zen}}^2$$

(dir directions, dis distances, zen zenith angles)
and estimates the respective variance components.

For the combination of the terrestrial observations with the internally highly accurate GPS measurements of the INN VALLEY Network there were 6 separate groups of observations established for the VCE: horizontal directions, zenith angles, 3 groups of spatial distances and the GPS observations. By this means the interaction of the satellite measurements and the individual groups of terrestrial observations can be studied thoroughly (Ref. 3).

Applying VCE to the Gauss-Markov model (Eq. 13) the linear structure (Eq. 14) to be considered is

$$\begin{align*}
\tilde{\sigma}_1^2 &= \sigma_0^2 \sigma_1^2 = \sigma_T^2 \sigma_x^2 + \sigma_S^2 \sigma_x^2 \\
\sigma_0^2 \sigma_x^2 &= \sigma_T^2 \sigma_x^2 + \sigma_S^2 \sigma_x^2
\end{align*}$$

(17)

The estimation procedure is according to Eq. B.13.

6.2 Combination of internal and external stochastic information

There are several possibilities of combining the stochastic information in terms of internal and external accuracies.

<table>
<thead>
<tr>
<th>terrestrial accuracies</th>
<th>internal</th>
<th>external</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>internal</th>
<th>external</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>satellite accuracies</th>
</tr>
</thead>
</table>

Combination of stochastic information
The combination of only internal accuracies provides, after proper weighting by VCE, internal accuracies within the combined network. An analogous conclusion can be drawn if exclusively external accuracies are combined. It can be questioned, of course, whether such a combination is reasonable in this case — are apples and oranges assembled? The same question arises if internal accuracies are combined with external ones. However, after S-transformation in all cases the same internal accuracy of the combined network can be revealed. This is illustrated in the following (see Appendix A):

\[
\begin{align*}
Q_{T,x}^I &= H_{T,x}O_{T,x}H, & Q_{S,x}^I &= H_{S,x}O_{S,x}H \\
Q_{T+S,x}^I &= H_{T,x}O_{T,x}H + H_{S,x}O_{S,x}H = H_{T,x}O_{T,x}H + H_{S,x}O_{S,x}H \\
&= H_{T,x}O_{T,x}H + O_{S,x}H.
\end{align*}
\]  

Equation 19 shows that the internal covariance matrix of the combined network can be achieved either by the sum of the internal covariance matrices of the terrestrial and satellite networks or by the S-transformed sum of the external covariance matrices. Due to the idempotence of \( H \) the same is valid if an external covariance matrix is combined with an internal one.

Applied to Eq. 9

\[
\begin{align*}
Q_{T+S,x}^I &= H_{T+S,x}O_{T+S,x}H = (H_{T,x}O_{T,x}H + H_{S,x}O_{S,x}H)^+ = (H_{T,x}O_{T,x}H + H_{S,x}O_{S,x}H)^+ ,
\end{align*}
\]

and applied to Eq. 14

\[
\begin{align*}
Q_{x}^I &= H_{T,x}O_{T,x}H = H_{T,x}O_{T,x}H + H_{S,x}O_{S,x}H = H_{T,x}O_{T,x}H + H_{S,x}O_{S,x}H
\end{align*}
\]

the same statement holds (Eq. A.11).

7. THE COMBINATION OF TERRESTRIAL AND SATELLITE AIDED NETWORK OBSERVATIONS

RESULTS

The INN VALLEY Network serves again as an example for numerical experiments (Appendix C).

7.1 The accuracy of coordinate observations

As the first measure the accuracy of the satellite aided coordinate observations as obtained from a hybrid network adjustment is investigated.

Introducing the external GPS covariance matrix \( S_{x}^* \) (Tab. 1, row 2) leads to coordinate accuracies of the combined network which cover the terrestrial accuracies (Tab. 1,
row 1) totally. After applying VCE the a priori weights of the GPS coordinate observations turn out to be by a factor 4.5 less accurate than before (Tab. 1, row 3 vs. 2). The accuracy situation of adjusted coordinates of the combined network (Tab. 1, row 4) which represents at the same time the accuracy of the adjusted coordinate observations is governed by the external accuracy of the GPS coordinates (Eq.1). For further processing, see next paragraph.

Table 1
Accuracies of coordinate observations

<table>
<thead>
<tr>
<th>ROW</th>
<th>status</th>
<th>accuracies of coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$S_x$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[cm]</td>
</tr>
<tr>
<td>terrestrial network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>combined terrestrial and GPS ($S_{C_x}$) network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6.7</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>combined terrestrial and GPS ($S_{C_x}$) network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.3</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>5.2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Status: 1 = a priori input standard deviations of coordinate observations;
2 = a priori output standard deviations after VCE;
3 = standard deviations of the adjusted coordinate observations after VCE; they represent also the standard deviations of the adjusted netpoint coordinates.
All standard deviations are r.m.s. values.

Using the internal GPS covariance matrix (Eq. 2; Tab. 1, row 5) one obtains after VCE the internal observational accuracies of the GPS measurements (Tab. 1, row 6) which lead to standard deviations of adjusted coordinate observations and coordinates (Tab. 1, row 7) being the same as the terrestrial ones. An exception is the standard deviation $S_h$ demonstrating the good influence of GPS measurements to an increased (external) accuracy of heights in this combined network.
7.2 Accuracy of adjusted coordinates

The results of the combined terrestrial and GPS-network are:

\[ s_{C_x}^T \text{ (Eq. 1) and } t_{C_x}^T \text{ (Eq. 3) lead to } s_{T_x}^T \text{ after VCE (Tab. 1, row 4):} \]

\[ s_{T_x}^T : x = \pm 1.7 \text{ cm, } y = \pm 1.5 \text{ cm, } h = \pm 3.4 \text{ cm} \]  \hspace{1cm} (22)

\[ s_{C_i}^T \text{ (Eq. 2) and } t_{C_i}^T \text{ (Eq. 4) lead to } s_{T_x}^C \text{ after VCE:} \]

\[ s_{T_x}^C : x = \pm 0.4 \text{ cm, } y = \pm 0.4 \text{ cm, } h = \pm 1.6 \text{ cm} \]  \hspace{1cm} (23)

S-transformation of Eq. 22 leads to Eq. 23 as does S-transformation of any other combination of external and internal accuracies, for instance of the results as given in Tab. 1, row 7. Note that the internal accuracy of heights has been improved by the introduction of GPS-measurements.

A more detailed investigation of the behaviour of the accuracies of the original and adjusted terrestrial observations and the coordinate observations is given by Ref. 3.

8. Changes in positions and heights due to the network combination

The coordinates based on satellite aided observations were transformed onto the adjusted coordinates of the terrestrial network prior to the network combination. The terrestrial network coordinates were serving as approximate coordinates in all cases of hybrid adjustments. However, in contrast to the terrestrial adjustment with a datum definition based on the national geodetic datum, the satellite coordinates and their covariance matrices determine the geodetic reference system within the combined adjustment models.

If one considers the mean values of coordinate unknowns resulting from the various combinations of stochastic information, one can realize that systematic changes in positions and heights result from the application of the external covariance matrices. This effect is due to the external influences which govern the external covariance matrices; in contrast, the use of the internal covariance matrices leaves on an average the hybrid network coordinates in the position of the terrestrial network.

This effect is the clearest with the height coordinates. Therefore, the behaviour of the changes in heights will be discussed in the following.

The observed GPS heights differ from the terrestrial heights between -12.9 cm ... +12.3 cm (mean ± 0 cm) before the combination. After combination with \( s_{C_x}^T \text{ (Eq. 1) and VCE, height changes occur from } -7.4 \text{ cm ... +3.6 cm (mean } -2.5 \text{ cm). Particularly interesting are the results of the combination with the internal GPS covariance matrix } s_{C_x}^C \text{ (Eq. 2): they reach from } -11.0 \text{ cm ... +6.5 cm (mean } ± 0 \text{ cm).} \)
Considering the essential results of the combination of terrestrial with GPS heights, the distinct influence of the GPS heights to the height determination of the network points becomes clear. Especially the extreme values -11.0 cm and +5.5 cm are highly significant improvements of two, terrestrially only weakly determined heights.

9. CONCLUSIONS

An appropriate combination of the stochastic information obtained from terrestrial and satellite aided network observation requires

- a hybrid functional and stochastic model of adjustment
- variance component estimation techniques
- S-transformation of covariance matrices.

Some results are

- various combinations of external and internal covariance matrices can be applied;
- after variance component estimation an identical internal covariance matrix of the adjusted coordinates of the hybrid network can be obtained from different combinations by S-transformations;
- the covariance matrices of satellite aided coordinate observations as obtained from various standard software packages are strongly affected by external influences;
- for comparison of accuracies only internal covariance matrices (or invariant functions) are suitable;
- GPS observations can contribute to an improvement of the internal accuracy of even very precise local terrestrial networks.
Appendix A

S-TRANSFORMATION

The S-transformation parameters \( p \) are 3 translations, 3 rotations and 1 scale factor for a three-dimensional network

\[
p^T = \begin{vmatrix} t_x & t_y & t_z & e_x & e_y & e_z \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \end{vmatrix},
\]

(A.1)

The coefficient matrix \( G \) is for \( n \) points

\[
G^T = \begin{vmatrix} 1 & 0 & 0 & 1 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & \cdots & 0 & 0 & 1 \\ 0 & -z_1 & y_1 & 0 & -z_2 & y_2 & \cdots & 0 & -z_n & y_n \\ z_1 & 0 & -x_1 & z_2 & 0 & -x_2 & \cdots & z_n & 0 & -x_n \\ -y_1 & x_1 & 0 & -y_2 & x_2 & 0 & \cdots & -y_n & x_n & 0 \\ x_1 & y_1 & z_1 & x_2 & y_2 & z_2 & \cdots & x_n & y_n & z_n \\ \end{vmatrix}.
\]

(A.2)

The transformation of a \( C_x \) covariance matrix is carried out with

\[
H = (I - G(G^T G)^{-1} G^T)
\]

(A.3)

according to Ref. 1 to

\[
C_x^{-1} = HC_x H^T.
\]

(A.4)

The matrices \( G \) and \( H \) have the following important properties:

\[
r(G) = d
\]

(A.5)

\((r(*)\) rank), if the number of transformation parameters applied is \( d \).

\[
r(G^T G) = d,
\]

(A.6)

so that \((G^T G)\) is regular and \((G^T G)^{-1}\) exists.

\[
\bar{G} = G(G^T G)^{-1} G^T = \bar{G}^T
\]

(A.7)

is symmetric and idempotent.
\[ \tilde{G} \tilde{G} = \tilde{G} , \quad r(\tilde{G}) = \text{tr}(\tilde{G}) = d \quad (A.8) \]

(C russe trace).

The transformation matrix \( H \) is symmetric and idempotent

\[ H = H^T , \quad HH = H , \quad r(H) = \text{tr}(H) = \text{tr}(I) - \text{tr}(\tilde{G}) = 3n-d , \quad (A.9) \]

and the Moore-Penrose generalized inverse \( H^+ \) is the matrix itself

\[ H^+ = H . \quad (A.10) \]

From the idempotence (Eq. A.9) it follows that a S-transformation can only once be applied to a covariance matrix \( C_x \); a second S-transformation of the same matrix would not change it.

The rank defect of the resulting matrix \( C_x^1 \) is \( d \), if \( r(H) = 3n-d \leq r(C_x) \); \( d = 7 \) at the most in case of a three-dimensional network. On this condition, with \( P_x \) as the inverse of \( C_x \) and with Eq. A.10 also

\[ C_x^1 = H C_x H = (HP_x H)^+ \quad (A.11) \]

holds (Ref. 4).
Appendix B

VARIANCE COMPONENT ESTIMATION

In the following only a short review is given. For detailed studies see for instance Ref. 5 - 7.

The stochastic part of the Gauss-Markov model

\[ \ell + v = Ax, \quad C_1 \]  \quad (B.1)

\( \ell \) vector of observations \((n\times 1)\)
\( v \) vector of residuals \((n\times 1)\)
\( A \) design matrix \((n\times u)\)
\( x \) vector of unknowns \((u\times 1)\)

Consists of the covariance matrix \( C_1 \) of the observations which can be divided in two

\[ C_1 = \sigma_0^2 Q_1 \]  \quad (B.2)

\( \sigma_0^2 \) theoretical unit variance
\( Q_1 = P_1^{-1} \) cofactor matrix of the observations

\( Q_1 \) is considered pre-given, \( \sigma_0^2 \) is after the adjustment \( (a \ posteriori) \) unbiasedly estimated by the empirical unit variance \( S_0^2 \), i.e.

\[ S_0^2 = \frac{v^T P_1 v}{r} \]  \quad (B.3)

\( r \) redundancy of the observations

\[ r = \text{tr}(P_1 Q_1^{vv}) = n - u \]  \quad (B.4)

(\text{tr}(\cdot) \text{ trace operator}),

\[ Q_1^{vv} = Q_1 - A(A^T P_1 A)^+ A^T = Q_1 - A Q_1 A^T \]  \quad (B.5)

\( Q_1^{vv} \) cofactor matrix of the residuals \((n\times n)\),

\[ v = (AQ_1 A^T - I) \ell = - Q_1^{vv} P_1 \ell \]  \quad (B.6)

and
\[ v^T P_1 v = \ell^T P_1 Q^{vv} P_1 \ell = s^2_0 \text{tr}(P_1 Q^{vv} P_1 Q^{vv}) \]  \hspace{1cm} (B.7)

Frequently observations of different types have to be adjusted in one hybrid model, i.e. horizontal directions, spatial distances, zenith angles drop in the adjustment model of a terrestrial network, or even terrestrial and satellite aided network observations have to be combined. In such a case the original Gauss-Markov model is split into several parts corresponding to the different groups of observations

\[
\begin{bmatrix}
\ell_1 \\
\vdots \\
\ell_c
\end{bmatrix} +
\begin{bmatrix}
v_1 \\
\vdots \\
v_c
\end{bmatrix} =
\begin{bmatrix}
A_1 \\
\vdots \\
A_c
\end{bmatrix} \cdot x .
\]  \hspace{1cm} (B.8)

Also the stochastic model falls into groups

\[
C_1 = \sigma_0^2 Q_1 = \sigma_1^2 Q_1 + \ldots + \sigma_c^2 Q_c = \sigma_1^{2p-1} + \ldots + \sigma_c^{2p-1} .
\]  \hspace{1cm} (B.9)

While the model was homogeneous, only the estimation of the unit variance \( \sigma_0^2 \) was of interest. In the inhomogeneous case, however, the wish may exist to estimate the variance components \( \sigma_1^2 \ldots \sigma_c^2 \) as well.

This can be achieved by dissecting Eqs. B.5 and B.7 according to

\[
Q^{vv} =
\begin{bmatrix}
Q_{11}^{vv} & \ldots & Q_{1c}^{vv} \\
\vdots & \ddots & \vdots \\
Q_{c1}^{vv} & \ldots & Q_{cc}^{vv}
\end{bmatrix}
\]  \hspace{1cm} (B.10)

with

\[
Q_{ii}^{vv} = Q_i - A_i Q_{x_i}^T A_i^T \quad i = 1 \ldots c
\]

\[
Q_{ij}^{vv} = - A_i Q_{x_j}^T \quad i,j = 1 \ldots c , \quad i \neq j
\]

and

\[
v^T P_1 v = \sum_{i=1}^{c} v_i^T P_1 v_i
\]  \hspace{1cm} (B.11)

with

\[
v^T P_1 v = \sum_{j=1}^{c} \text{tr}(P_j Q^{vv} P_j Q^{vv}) s_j^2
\]

\[
= \sum_{j=1}^{c} t_{ij} s_j^2 .
\]  \hspace{1cm} (B.12)
After composing to one system of equations

\[
\begin{bmatrix}
    u_i^T P \, u_i \\
    \vdots \\
    u_c^T P \, u_c
\end{bmatrix}
\begin{bmatrix}
    t_{11} & \cdots & t_{1c} \\
    \vdots & \ddots & \vdots \\
    t_{c1} & \cdots & t_{cc}
\end{bmatrix}
\begin{bmatrix}
    s_i^2 \\
    \vdots \\
    s_c^2
\end{bmatrix}
\]

(B.13)

the variance components \( s_i^2 \) can be solved for by inversion. The solution has to be found iteratively until all components converge \( s_i^2 \rightarrow s_j^2 \). The final estimation result is unbiased.

Eq. B.13 is the general formula which has to be applied if the linear structure of the \( Q_1 \) matrix is "overlapping", i.e.

\[
\sigma_0^2 \begin{bmatrix}
    0_1 \\
    \vdots \\
    0_2
\end{bmatrix} = \sigma_1^2 \begin{bmatrix}
    0_1 \\
    \vdots \\
    0_2
\end{bmatrix} + \sigma_2^2 \begin{bmatrix}
    0_1 \\
    \vdots \\
    0_2
\end{bmatrix}
\]

(B.14)

or "sequential"

\[
\sigma_0^2 \begin{bmatrix}
    0_1 \\
    \vdots \\
    0_2
\end{bmatrix} = \sigma_1^2 \begin{bmatrix}
    0_1 \\
    \vdots \\
    0_2
\end{bmatrix} + \sigma_2^2 \begin{bmatrix}
    0_1 \\
    \vdots \\
    0_2
\end{bmatrix}
\]

(B.15)

In this case the equation system B.13 can be simplified. The estimation formula is

\[
u_i^T P \, u_i = s_i^2 \text{tr}(P \, Q_{11})
\]

(B.16)

for each variance component \( s_i^2 \) independent of the others. The estimation process is iterative, too.

One very important fact of the estimation of variance components is its independence of the geodetic datum or other network external effects.
THE INN VALLEY NETWORK

Terrestrial observations

The INN VALLEY Network of the Institute of Geodesy of the Bundeswehr University Munich covers a space of 25 km × 15 km × 1.3 km. All points but one are monumented by concrete pillars. The terrestrial coordinates were determined by 48 spatial distances (three classes of accuracies), 44 horizontal directions, 44 zenith distances and 8×2 components of the deflections of the vertical.

![Diagram of the INN VALLEY Network](image)

\[ \text{Error ellipse} \quad 1 \text{cm} \]

\[ m_H = 36 \text{mm} \quad m_p = 5.9 \text{mm} \]

Fig. 1 The INN VALLEY Network: terrestrial observations
GPS observations

In November 1984 coordinate differences of the INN VALLEY Network points were determined by Macrometer V 1000 measurements (Ref. 2).

Fig. 2 The INN VALLEY Network: Macrometer measurements
( 1 simultaneous observations)

* * *
REFERENCES


3) W. Welsch, W. Oswald, The hybrid adjustment of terrestrial and satellite aided network observations - investigation of accuracies, XVIII. International FIG Congress, invited paper 503.1, Toronto 1986


THE LEP TRILATERTATION NETWORK

J. Gervaise and J. Olsfors
CERN, Geneva, Switzerland

ABSTRACT
The installation and the alignment of a particle accelerator require a very high accuracy. For the new CERN project, the Large Electron Positron ring (LEP) of 27 km circumference, a reliable control network has been created. Measurements have been made with the two-colour Terrameter, giving an accuracy of $10^{-7}$, and precision levelling. Computations took into account a very fine assessment of the geoid, through collocation on a large mass model and astro-geodetic measurements with a zenithal camera. Resulting local coordinates were good to $\sigma \leq 1.5$ mm for planimetry and $\sigma \leq 5$ mm for altimetry above a local ellipsoid. In December 1984, seven points of this network were observed using GPS (Global Positioning System) satellite receivers. The measurements were carried out by GEO-HYDRO Inc. with three Macrometer V-1000 Surveyor instruments. Six independent baselines of this GPS network were measured twice during three consecutive days. Although the observation constellation was rather poor, a common evaluation of all observations (with the Bernese GPS Software Package) gave rms-errors of roughly 1 - 2 mm per coordinate, showing an excellent internal consistency of the GPS solution. The comparison with the terrestrial solution was established through a Helmert transformation. The overall rms error of the transformation is 4 mm.

1. INTRODUCTION

LEP (Large Electron-Positron Collider) is intended to develop more exactly the extensive new synthesis of electro-weak interaction born out of the discoveries of the $W$ and $Z_0$ bosons. In such a machine, electrons and their antiparticles, positrons, are accelerated simultaneously. Whilst turning in opposite directions, they collide head-on and are annihilated and transformed into energy which, in turn, results in new particles and antiparticles. LEP is a 27 km-long underground ring located in the Pays de Gex, France, and partially in Switzerland, with eight experimental areas. The existing CERN machines will act as injectors to LEP, which represents an important economy. This implies that the LEP geodetic network should be adapted conveniently to the previous networks (SPS).
Although geodesy, topography and cartography may be but a small part of the preliminary studies for a large civil engineering project, they are an indispensable part, and the work of the geodesist must precede that of the other engineers. This is all the more true for the construction of a 27 km particle collider, a project requiring the highest precision. In search for greater accuracy, the geodesist must look beyond the causes and effects that are already appreciated to other influences, albeit small ones, at the very limits of measurement.

The complexity of the geodesy has grown with each new accelerator, notably due to certain influences which become more significant with the increasing size of accelerators. The influence of the earth's curvature is a first example of this ever-increasing complexity. For LEP there is another influence to consider, that of the deviation of the vertical, which produces effects that were not foreseen at the start of the project.

Due to the curvature of the earth, the normal height of the centre \( P_0 \) of the Proton Synchrotron (PS) is 0.8 mm above the horizontal plane of the synchrotron. Because the Linac is aligned in the horizontal plane of the injection point, the origin of the levelling had to be 0.3 mm above the height of this injection point. For the Intersecting Storage Rings (ISR) transfer tunnels the corrections reached 9.5 mm for TT1 and 15.8 mm for TT2. The ISR being virtually circular, as is also the PS, there were no height corrections around the ring itself. The corrections for the Super Proton Synchrotron (SPS) transfer tunnels were made in like manner. However, due to the existence of straight sections of 256 m, the machine can no longer be considered circular. To keep the accelerator in the plane, the height corrections in these straight sections reached 3.5 mm. The geometry of LEP has been calculated in the CERN tridimensional system \((x, y, z)\) but, as \(z\) is not measurable, all \( z\)-coordinates have been converted to normal heights \( h\) above the reference ellipsoid.

The surrounding topography gives rise to distortions of the equipotential surfaces. The extent of these distortions has been calculated by theory and confirmed by measurements of the deviation of the vertical. This has allowed the definition of the theoretical altitudes \( h\) of all the points of LEP. The deviation of the vertical, caused by the non-homogeneity of the surrounding topography, necessitates a correction for the determination of points at the bottom of the pits in relation to the geodetic points on the surface. For a 140 m pit at the foot of the Jura, this correction is 6 mm.

The deviation of the vertical also affects the gyroscope measurements which provide the orientation for the excavation of the LEP tunnel.

The precision required during the various steps of the construction of LEP necessitates, as was the case for the SPS, a surface geodetic network of very high standard. This is why, as soon as the project was approved, the decision was taken to equip the Applied Geodesy Group of CERN with an LDM 2 Terrameter. This two-wavelength electromagnetic distance measurement device arrived at CERN in Autumn 1982.
2. **PRINCIPLE OF MEASUREMENT WITH INSTRUMENTS USING TWO DIFFERENT WAVELENGTHS IN THE VISIBLE SPECTRUM**

The use of two-colour electromagnetic distance measuring instruments will modify considerably the approach to the problems of very high accuracy geodesy.

Rather than consider the time dt taken by the light to cover a length element dl, we use the proportional quantity c dt, called the optical path, the length covered in the vacuum during a time dt. The ratio of the speed of an electromagnetic wave in a vacuum to that in matter is known as the absolute index of refraction and given by \( n = c/v \). This index is actually frequency-dependent. The dependence of \( n \) on the wavelength of light is a well-known effect called dispersion. The optical path becomes \( n \cdot v \cdot dt = n \cdot dl \). Thus, the optical path \( L \), going from A to B, is defined by:

\[
L_{(AB)} = \int_A^B n \, dl
\]

(1)

The principle of two-colour electromagnetic distance measuring instruments was proposed by Prilepin in 1957 and Bender & Owens in 1965, in order to make the distance measurements less dependent on the refractive index of air. This index is essentially variable for long paths through the atmosphere and impossible to measure accurately with one-frequency EDM instruments. Along the same path, simultaneous measurements of the optical path with different frequencies allow the desired correction for the refractive index to be largely determined.

![Fig. 1 Definition of group velocity u](image-url)
The two-colour instruments measure two different optical paths due to the dispersion of the refractive index in relation to the wavelength. We assume a geometrical path length \( L \) of 15 km between light source and reflector, i.e. 30 km outgoing and incoming. The one-way optical length will be \( L + S \), where \( S \) is the additional contribution due to the atmosphere. Light being modulated, \( L + S \) is inversely proportional to the group velocity \( u \). To study the addition of waves of slightly different frequency, it is sufficient to consider two waves \( \lambda_1 \) and \( \lambda_2 \) travelling at velocities \( v_1 \) and \( v_2 \). The total disturbance arising from the combination of these waves having equal amplitude and zero epoch angles may be regarded as a travelling wave of frequency \( f_m \) with modulated amplitude \( E_0^2 (x,t) \) (Fig. 1).

The irradiance is proportionnal to \( E_0^2 \) and oscillates about a value of \( 2 E_0^2 \). The disturbance \( E \) consists of a high frequency \( (\omega) \) carrier wave, modulated by a cosine function. The rate at which the modulation envelope advances is known as the group velocity \( u \). In a non-dispersive media, in which the phase velocity \( v \) is independent of wavelength, \( u = v \). Specifically, in a vacuum, \( v = u = c \).

For optical media in regions of normal dispersion, the refractive index increases with frequency inasmuch as \( u < v \) \((dv/d\lambda)\); \( u \) is smaller than \( v \). Clearly, one should define also a group index of refraction \( n_g = c/u \) which must be carefully distinguished from \( n \). So, \( S \) can be defined by:

\[
S = \int_0^L (n_g - 1) \, dl. \tag{2}
\]

For red light alone, \( S \) is about 400 cm. If the red line of a helium-neon laser (632.8 nm) and the blue line of a helium-cadmium laser (441.6 nm) are chosen, the extra optical paths \( S_B \) and \( S_R \) for the blue and red respectively will differ by about 10%, giving a difference \( \Delta S = S_B - S_R \) of about 40 cm. By convention, the dispersive power of the atmosphere is a number \( A \) without dimension:

\[
A = (n_g^B - n_g^R) / (n_g^R - 1), \tag{3}
\]

its inverse \( 1/A \) is called the constriengy of the atmosphere.

One can thus write the difference of the optical path:

\[
\Delta S = \int_0^L A (n_g^2 - 1) \, dl. \tag{4}
\]

\( A \) being independent of the atmospheric density along the path and only weakly dependent on the atmospheric composition, we can replace it with a good approximation by its average value \( \langle A \rangle \) and take it outside the integral sign:

\[
\Delta S = \langle A \rangle \int_0^L (n_g^2 - 1) \, dl = \langle A \rangle S_R, \tag{5}
\]
then

\[ S_R = S / (A) \] (6)

and

\[ L = L_R - S_R = L_R - (L_B - L_R) / (A), \] (7)

which shows that the true distance is obtained by measuring only the optical lengths \( L_B \) and \( L_R \) over the same path thus removing the effect of the atmosphere from the optical measurement.

We call refractivity the value \( N = (n - 1) \times 10^6 \) and \( N_{go} \) the group refractivity in standard atmospheric conditions, temperature being 15°C and atmospheric pressure being 760 mm of mercury.

We can calculate \( (A) \) for the wavelengths used in the Terrameter - 632.8 nm in the red, 441.6 nm in the blue. The Barrell & Sears formula is particularly convenient for visible wavelengths:

\[ N_{go} \times 10^6 = 287.604 + 3 (1.6288/\lambda^2) + 5 (0.0136/\lambda^4), \] (8)

where \( \lambda \) is the wavelength of the radiation in nm. This allows the constringency to be calculated: \( 1/A = 21.12 \).

If we use Edlen's equation with the Cauchy form for visible and near infra-red light,

\[ N_{go} \times 10^6 = 272.600 + 4.608/\lambda^2 + 0.066/\lambda^4. \] (9)

In this case, \( 1/A = 21.16 \).

In the Terrameter, the following simplified formula assumes a 0.03% CO₂ atmosphere, with \( P_w \) the partial pressure of water vapour, and \( P_s \) the partial pressure of dry air. \( P \) is atmospheric pressure in millibars; \( T \) is temperature in degrees Kelvin.

\[ N_{go} = (n-1) \times 10^8 = (C_1 D_s + C_2 D_w) \times 10^8 \] (10)

\[ D_s = P_s/T \left[ 1 + P_s (57.9 \times 10^{-8} - 9.325 \times 10^{-4}/T + 0.25844)/T^2 \right] \]

\[ D_w = P_w/T \left[ 1 + P_w (1 + 3.7 \times 10^{-4} \cdot P_w) (-2.37321 \times 10^{-3} + 2.23366/T - 710.792/T^2 + 7.75141 \cdot 10^4/T^3) \right] \]

\[ C_1 = (\text{He-Cd}) \ 80.874255 \times 10^{-6} \] (He-Ne) \ 84.735399 \times 10^{-6}

\[ C_2 = (\text{He-Cd}) \ 69.094806 \times 10^{-6} \] (He-Ne) \ 73.700935 \times 10^{-6}
The value of $P_w$ can be determined from the relative humidity RH and the saturated water vapour pressure; i.e.

$$P_w = RH \times P \text{ saturated.} \quad (11)$$

In this case, $1/A = 20.86$.

The measurement of $\Delta S$ gives the atmospheric correction to the optical path length. To measure $\Delta S$, it is necessary to modulate the polarisation of the red and blue light simultaneously at frequencies of 3 GHz. The signal-to-noise ratio allows a great accuracy to be obtained and also the reduction of systematic errors. $\Delta S$ is determined with a precision equal to a small fraction of the modulation wavelength under good atmospheric conditions.

3. LDM 2 TERRAMETER (Terratechnology Corporation)

The first prototype, presented in June 1969 at Boulder, Colorado, at the "International Symposium on Electromagnetic Distance Measurements", is now produced in a limited series. Unlike conventional EDM instruments which approximate the refractive index of air along the measurement path by sample measurements of temperature, air pressure and water vapour, the Terrameter makes a direct and precise measurement of the refractive index by using simultaneous distance measurements at two optical wavelengths, one in the red part of the spectrum and the other in the blue. The instrument calculates correction terms from the optical path length difference between the two wavelengths and computes the corrected base line distance. This automatically eliminates the first-order effects of temperature, air pressure and water vapour.

The basic principle of operation is the same as that of the Fizeau velocity of light experiment. In that experiment, light was returned from a distant retro-reflector to the photodetector only if there existed a proper temporal and spatial relationship between the outgoing and incoming light. Light returned to the detector only if the transit time of the light was exactly equal to the integral number of modulation periods. The principle of the Fizeau cog-wheel was adapted first by Michelson and then by Bergstrand in 1930. A KD*P (potassium dideuterium phosphate) crystal cell is used in the Terrameter to modulate the polarization of light by a linear electro-optical effect, known as Pockel's effect, inasmuch as the induced birefringence is proportional to the applied electrical field and, therefore, the applied voltage. Pockel's effect exists only in certain crystals which lack a centre of symmetry. The operating principle for such a device is that the birefringence is varied electrically by means of a controlled applied electric field. The retardance can be altered as desired, thereby changing the state of polarization of the incident linear wave. In this way, the system functions as a polarization modulator. A Pockel cell is simply an appropriate non-symmetric, oriented, single crystal immersed in a controllable electric field.
Fig. 2 Scheme of the LDM2 Terrameter

In the Terrameter (Fig. 2), red and blue light from the He-Ne and He-Cd lasers enters a Rochon prism at the proper angle and polarization to make the outgoing beams co-linear. The light passes through a microwave modulator that modulates the ellipticity of the polarized light at 3 GHz. The light, transmitted by a Cassegrainian telescope, traverses the path being measured and is returned by the retro-reflector. The beam is received by the same optics used for transmission, and passes through the modulator a second time, where the ellipticity of the polarization is increased or decreased depending upon the phase of the modulator excitation. The beams emerging from the prism are separated by colour and directed to the photodetectors.

The analogue outputs of the photodetectors are processed and used to control one VCO (voltage controlled oscillator) for each colour. The VCO output can be varied between 4 and 5 MHz, depending on input frequency giving a 2995.5 MHz signal which after amplification to 4 Watts drives the optical modulator.

Because there is only one modulator for the two colours, it must be switched alternatively between the red and blue frequencies.

In order to adjust the modulating frequency so that a minimum of light is received at the photodetectors, each VCO is switched between two frequencies, one below and the other above the centre frequency (Fig. 3).

As can be seen on the figure, the minima are much better defined than the maxima. The frequency difference between the minima is:

$$ f_c - f_{c'} = \frac{\nu}{2D}, \quad (12) $$
where \( v \) is the speed of light in the air, and \( D \) is the one-way distance. The amount of VCO frequency shift \( \Delta f \) (determined by the "Approximate distance" set by the operator on the control panel) is about one quarter of \( f_c - f_c' \).

Fig. 3  Received light intensity vs. modulator frequency

The internal frequencies are measured and these values used in the microcomputer to calculate the true distance. Since \( A \) is a weak function of the meteorological parameters, rough end-point values of temperature, pressure and relative humidity are entered using thumbwheels on the instrument panel (Fig. 4).

Fig. 4  Back view of the Terrameter
With these data the microcomputer determines the true distance, which is digitally displayed on a liquid crystal readout on the back panel of the instrument. A hard copy of all input data, internal frequencies and true distance may be obtained every ten seconds by connecting a terminal or a printer to the Terrameter's computer interface. A rubidium frequency standard is used as the reference for counters that measure the internal frequencies that are used for the distance calculations. This standard is several orders of magnitude more accurate than ever required and no subsequent recalibration is necessary. The repeatability is $2 \times 10^{-10}$ after 10 minutes of warm-up, with a long term stability of $4 \times 10^{-11}$.

The laser beams from the instrument are directed to a distant retro-reflector (Fig. 5). An aligning telescope is used to acquire the reflected beam, and the operator can then adjust the fine pointing of the instrument by viewing the frosted glass aperture mounted on the rear.

![Fig. 5 Retroreflector](image)

The Terrameter's range is from 350 m to 15 km. The outgoing power of the He-Ne laser is 5 mW and that of the He-Cd laser is 10 mW. The polarisation is modulated at 3 GHz, i.e. a modulation wavelength of 10 cm. The beam dispersion in the atmosphere is limited to $2 \times 10^{-4}$ radians.
The frequencies are servo-controlled in order that the return signal arrives in opposite phase to the outgoing signal. Thus the number \( N \) of wavelengths on the outward and return path is an integer plus a half. The red and blue frequencies are adjusted such that \( N \) is equal for the two colours. The ratio of the two indices is equal to the ratio of the two frequencies. Due to the dispersion, it is possible to calculate \( n_B \) and \( n_R \) indices, all the other parameters being known except \( N \). It is thus necessary to determine an approximate value of the distance to better than half a wavelength (0.05 m). The instrument is automatic, apart from the introduction of the approximate distance and meteorological values. However, due to the elaborate optical and electronic systems, it needs maintenance and internal adjustments which require the presence of an electronics engineer throughout the observation period. Also the bulkiness and weight inhibit easy use in all field geodetic measurements. It has been used principally to measure the deformations of the earth's crust, the stability of large construction sites and landslips. It can, with difficulty, be adapted to the measurement of particular geodetic networks such as those of CERN.

4. Measurement of the LEP Geodetic Network with the Terrameter

The first critical operation was to determine the instrumental constant of the device. The Applied Geodesy Group has at its disposal a 500-m outdoors baseline which has a pillar with forced centring at every 50 m. With the new model of the Distinvar, the absolute accuracy over a distance of 50 m is 0.03 mm, which means, on the 500 m of the base, a precision of 0.1 mm (relative precision of \( 2 \times 10^{-7} \)). It has been necessary to increase the length of this base by building two more pillars, one located 478 m to the north and the second 549 m to the south. Figure 6 shows all the measurements carried out with the Terrameter. Values measured with the Distinvar between pillars 2 and 12 have been included in the adjustment, the results of which have given a constant for the instrument-reflector set of 0.0275 m.

![Diagram](attachment:image.png)

Fig. 6 Determination of the constant of the Terrameter
The constant of the instrument-reflector set having been determined with the maximum accuracy, it was possible to measure a geodetic figure, the nucleus of the LEP network, in order to carry out the acceptance tests of the Terrameter.

Figure 7 shows the location of the geodetic points which surround the LEP ring. Measurements were made from all points except the pillar of point 232 located on the water tower of Collex-Bossy. The wooded, hilly and rather dense urbanization of the area meant that not all pillars could be positioned at ground level.

At each point the pillar is equipped with a forced centring reference socket. The Terrameter has a locating cylinder with an expanding locking system. Each distance was measured 100 times. One measurement is displayed every ten seconds, which means a little more than a quarter of an hour per distance. Measurements were taken at different times of the day and over several days.
Table 1

Comparison of reciprocal observations with the Terrameter
(without addition of the instrument constant)

<table>
<thead>
<tr>
<th>Pillars</th>
<th>Observed distances (m)</th>
<th>Distance difference (mm)</th>
<th>Precision x 10^-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>226 - 230</td>
<td>7482.33805, 7482.33803</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>226 - 231</td>
<td>7314.84645, 7314.84691</td>
<td>0.46</td>
<td>0.6</td>
</tr>
<tr>
<td>226 - 232</td>
<td>8808.13369</td>
<td></td>
<td></td>
</tr>
<tr>
<td>226 - 233</td>
<td>13663.40556, 13663.40524</td>
<td>0.32</td>
<td>0.2</td>
</tr>
<tr>
<td>226 - 234</td>
<td>7755.90733, 7755.90860</td>
<td>1.27</td>
<td>1.6</td>
</tr>
<tr>
<td>230 - 231</td>
<td>3543.57321, 3543.57230</td>
<td>0.91</td>
<td>2.6</td>
</tr>
<tr>
<td>230 - 232</td>
<td>8110.89990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>230 - 233</td>
<td>8869.18119, 8869.18148</td>
<td>0.29</td>
<td>0.3</td>
</tr>
<tr>
<td>230 - 234</td>
<td>11001.47017, 11001.46911</td>
<td>1.06</td>
<td>1.0</td>
</tr>
<tr>
<td>231 - 232</td>
<td>4593.22481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>231 - 233</td>
<td>6541.98778, 6541.98669</td>
<td>1.09</td>
<td>1.7</td>
</tr>
<tr>
<td>231 - 234</td>
<td>8080.72270, 8080.72221</td>
<td>0.49</td>
<td>0.6</td>
</tr>
<tr>
<td>233 - 232</td>
<td>6458.90541</td>
<td></td>
<td></td>
</tr>
<tr>
<td>233 - 234</td>
<td>11294.64935, 11294.64892</td>
<td>0.43</td>
<td>0.4</td>
</tr>
<tr>
<td>234 - 232</td>
<td>4861.91680</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Pillar 232, water tower of Collex-Bossy, non-stationned

Table 1 shows the comparison of all reciprocal distance measurements. The r.m.s. value of these differences, expressed as a relative value is $0.9 \cdot 10^{-7}$. Therefore, the relative precision of independent measurements carried out under different conditions is as stated by the manufacturer.

The network has been computed by a tridimensional least squares adjustment programme constraining at each point the measured altitude above the XY plane of the CERN cartesian system. This transformation makes use of a local reference ellipsoid with the parameters recommended by IUOG (Canberra 1979) and includes the computation of geoidal heights.
The residuals between measured and adjusted distances are less than 2.5 mm and the r.m.s. values of these residuals is 1.3 mm. The axes of the confidence ellipses obtained from the covariance matrix are less than 1.5 mm (Table 2).

Table 2
Results of the LEP geodetic measurements

Fixed point and orientation point

<table>
<thead>
<tr>
<th>Pillar</th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>DX (mm)</th>
<th>DY (mm)</th>
<th>RDX (mm)</th>
<th>RDY (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>226</td>
<td>1288.2656</td>
<td>723.7339</td>
<td>2500.8843</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>233</td>
<td>-1150.7310</td>
<td>14165.7641</td>
<td>2729.9655</td>
<td>-13.0</td>
<td>71.5</td>
<td></td>
<td>1.5</td>
</tr>
</tbody>
</table>

Fixed points in Z

<table>
<thead>
<tr>
<th>Pillar</th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>DX (mm)</th>
<th>DY (mm)</th>
<th>RDX (mm)</th>
<th>RDY (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>230</td>
<td>-4176.6449</td>
<td>5828.7078</td>
<td>2745.4066</td>
<td>-29.0</td>
<td>13.3</td>
<td>1.7</td>
<td>1.8</td>
</tr>
<tr>
<td>231</td>
<td>-1130.9828</td>
<td>7626.9147</td>
<td>2527.4834</td>
<td>-15.4</td>
<td>48.7</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td>232</td>
<td>3132.0103</td>
<td>9336.7623</td>
<td>2491.8598</td>
<td>32.0</td>
<td>40.3</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>234</td>
<td>6815.9375</td>
<td>6164.0805</td>
<td>2456.1840</td>
<td>37.5</td>
<td>0.5</td>
<td>1.6</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Distances

<table>
<thead>
<tr>
<th>Pillar</th>
<th>Observed distances</th>
<th>Residues (mm)</th>
<th>Compensated distances</th>
</tr>
</thead>
<tbody>
<tr>
<td>231 - 226</td>
<td>7314.87420</td>
<td>-0.60</td>
<td>7314.87360</td>
</tr>
<tr>
<td>231 - 230</td>
<td>3543.60020</td>
<td>-0.82</td>
<td>3543.59938</td>
</tr>
<tr>
<td>231 - 232</td>
<td>4593.25230</td>
<td>-1.01</td>
<td>4593.25129</td>
</tr>
<tr>
<td>231 - 233</td>
<td>6542.01470</td>
<td>-1.22</td>
<td>6542.01348</td>
</tr>
<tr>
<td>231 - 234</td>
<td>8080.75000</td>
<td>-0.67</td>
<td>8080.74933</td>
</tr>
<tr>
<td>230 - 232</td>
<td>8110.92740</td>
<td>1.83</td>
<td>8110.92923</td>
</tr>
<tr>
<td>226 - 230</td>
<td>7482.36550</td>
<td>0.24</td>
<td>7482.36574</td>
</tr>
<tr>
<td>226 - 232</td>
<td>8808.16120</td>
<td>1.72</td>
<td>8808.16292</td>
</tr>
<tr>
<td>226 - 234</td>
<td>7755.93550</td>
<td>-0.57</td>
<td>7755.93493</td>
</tr>
<tr>
<td>234 - 230</td>
<td>11001.49710</td>
<td>2.51</td>
<td>11001.49961</td>
</tr>
<tr>
<td>234 - 232</td>
<td>4861.94430</td>
<td>-0.01</td>
<td>4861.94429</td>
</tr>
<tr>
<td>233 - 226</td>
<td>13663.43290</td>
<td>-1.44</td>
<td>13663.43146</td>
</tr>
<tr>
<td>233 - 230</td>
<td>8869.20880</td>
<td>2.05</td>
<td>8869.21085</td>
</tr>
<tr>
<td>233 - 232</td>
<td>6458.93290</td>
<td>0.90</td>
<td>6458.93380</td>
</tr>
<tr>
<td>233 - 234</td>
<td>11294.67660</td>
<td>-0.69</td>
<td>11294.67591</td>
</tr>
</tbody>
</table>

The actual surface geodetic network for LEP (Fig. 8) was measured in October 1983. It is not possible to publish here the comparison of reciprocal distance measurements between all the pillars of the network, but the r.m.s. difference is 1.3 mm and the axes of the confidence ellipses are less than 1.5 mm. These results are the same as those obtained in the measurement of the test network.
This network was adapted to that of the SPS, which shares five geodetic points including 226. The displacement vectors range from 0.25 mm to 6.10 mm (8 m pillar), showing the quality of the two networks and the stability of the points, as the SPS surface network has not been measured since 1976.

In summer 1985, a new measurement of the LEP geodetic network (Fig. 9) has been made, in order to check the stability of points and to determine new pillars at the vicinity of access pits to the tunnel.
Fig. 9  LEP geodetic network (1985)

The main results of this trilateration campaign are shown in Tables 3 and 4. The comparison with 1983 network is given in Table 5.
Table 3
LEP test geodetic network, 1985
Free figure adjustment

Fixed point: 226
Orientation point: 233
Number of observations: 31
Number of unknowns: 18
All points constrained at their measured altitude H

<table>
<thead>
<tr>
<th>Point</th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>SX (mm)</th>
<th>SY (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>215</td>
<td>1200.67301</td>
<td>2077.57816</td>
<td>2484.5225</td>
<td>1.23</td>
<td>.32</td>
</tr>
<tr>
<td>223</td>
<td>1821.90857</td>
<td>4555.77730</td>
<td>2476.2915</td>
<td>1.10</td>
<td>1.53</td>
</tr>
<tr>
<td>226</td>
<td>1288.26220</td>
<td>723.73090</td>
<td>2500.8724</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>230</td>
<td>-4176.63933</td>
<td>5828.70653</td>
<td>2745.4962</td>
<td>.88</td>
<td>.92</td>
</tr>
<tr>
<td>231</td>
<td>-1130.98254</td>
<td>7626.91154</td>
<td>2527.4999</td>
<td>1.01</td>
<td>.73</td>
</tr>
<tr>
<td>232</td>
<td>3132.00429</td>
<td>9336.76067</td>
<td>2491.8226</td>
<td>.80</td>
<td>1.19</td>
</tr>
<tr>
<td>233</td>
<td>-1150.72829</td>
<td>14165.76389</td>
<td>2729.9498</td>
<td>.15</td>
<td>.84</td>
</tr>
<tr>
<td>234</td>
<td>6815.93367</td>
<td>6164.07295</td>
<td>2456.1707</td>
<td>1.02</td>
<td>1.50</td>
</tr>
<tr>
<td>236</td>
<td>-3886.99035</td>
<td>1148.46141</td>
<td>2602.9765</td>
<td>.98</td>
<td>1.14</td>
</tr>
<tr>
<td>353</td>
<td>-2796.83180</td>
<td>4648.59131</td>
<td>2527.0570</td>
<td>1.17</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Columns [2] and [3]: adjusted coordinates
Columns [5] and [6]: standard-deviation (1σ) on adjusted coordinates
Table 4

LEP test geodetic network, 1985
Adjusted Distances of LEP network

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Point 1</td>
<td>Point 2</td>
<td>Observed (m)</td>
<td>Sigma (mm)</td>
<td>Adjusted (m)</td>
</tr>
<tr>
<td>223</td>
<td>230</td>
<td>6138.0287</td>
<td>1.43</td>
<td>6138.0287</td>
<td>- .01</td>
</tr>
<tr>
<td>223</td>
<td>231</td>
<td>4260.7574</td>
<td>.84</td>
<td>4260.7574</td>
<td>.02</td>
</tr>
<tr>
<td>223</td>
<td>353</td>
<td>4619.9517</td>
<td>.93</td>
<td>4619.9517</td>
<td>- .05</td>
</tr>
<tr>
<td>226</td>
<td>215</td>
<td>1356.8156</td>
<td>.38</td>
<td>1356.8156</td>
<td>- .02</td>
</tr>
<tr>
<td>226</td>
<td>230</td>
<td>7482.3613</td>
<td>1.43</td>
<td>7482.3613</td>
<td>-2.34</td>
</tr>
<tr>
<td>226</td>
<td>231</td>
<td>7314.8718</td>
<td>1.25</td>
<td>7314.8723</td>
<td>- .51</td>
</tr>
<tr>
<td>226</td>
<td>234</td>
<td>7755.9323</td>
<td>2.94</td>
<td>7755.9314</td>
<td>.86</td>
</tr>
<tr>
<td>226</td>
<td>236</td>
<td>5193.6560</td>
<td>1.37</td>
<td>5193.6558</td>
<td>.25</td>
</tr>
<tr>
<td>226</td>
<td>353</td>
<td>5665.0880</td>
<td>1.05</td>
<td>5665.0867</td>
<td>1.20</td>
</tr>
<tr>
<td>230</td>
<td>215</td>
<td>6561.7170</td>
<td>1.69</td>
<td>6561.7173</td>
<td>- .28</td>
</tr>
<tr>
<td>230</td>
<td>232</td>
<td>8110.9225</td>
<td>.58</td>
<td>8110.9226</td>
<td>- .13</td>
</tr>
<tr>
<td>230</td>
<td>233</td>
<td>8869.2101</td>
<td>1.46</td>
<td>8869.2111</td>
<td>- .93</td>
</tr>
<tr>
<td>230</td>
<td>353</td>
<td>1828.7323</td>
<td>1.22</td>
<td>1828.7308</td>
<td>1.50</td>
</tr>
<tr>
<td>231</td>
<td>230</td>
<td>3543.5988</td>
<td>.79</td>
<td>3543.5983</td>
<td>.48</td>
</tr>
<tr>
<td>231</td>
<td>215</td>
<td>6019.4543</td>
<td>1.79</td>
<td>6019.4545</td>
<td>- .18</td>
</tr>
<tr>
<td>231</td>
<td>226</td>
<td>7314.8740</td>
<td>2.73</td>
<td>7314.8723</td>
<td>1.68</td>
</tr>
<tr>
<td>231</td>
<td>232</td>
<td>4593.2468</td>
<td>.65</td>
<td>4593.2465</td>
<td>.32</td>
</tr>
<tr>
<td>231</td>
<td>233</td>
<td>6542.0165</td>
<td>1.75</td>
<td>6542.0154</td>
<td>1.10</td>
</tr>
<tr>
<td>231</td>
<td>234</td>
<td>8080.7475</td>
<td>2.62</td>
<td>8080.7464</td>
<td>1.11</td>
</tr>
<tr>
<td>231</td>
<td>236</td>
<td>7040.7109</td>
<td>1.85</td>
<td>7040.7096</td>
<td>1.30</td>
</tr>
<tr>
<td>233</td>
<td>226</td>
<td>13663.4334</td>
<td>1.43</td>
<td>13663.4330</td>
<td>.36</td>
</tr>
<tr>
<td>233</td>
<td>232</td>
<td>6458.9298</td>
<td>.98</td>
<td>6458.9299</td>
<td>- .07</td>
</tr>
<tr>
<td>233</td>
<td>234</td>
<td>11294.6758</td>
<td>1.96</td>
<td>11294.6764</td>
<td>- .65</td>
</tr>
<tr>
<td>233</td>
<td>236</td>
<td>13302.3839</td>
<td>1.11</td>
<td>13302.3839</td>
<td>.07</td>
</tr>
<tr>
<td>234</td>
<td>215</td>
<td>6944.8709</td>
<td>4.13</td>
<td>6944.8711</td>
<td>- .25</td>
</tr>
<tr>
<td>234</td>
<td>223</td>
<td>5246.6471</td>
<td>1.30</td>
<td>5246.6472</td>
<td>- .08</td>
</tr>
<tr>
<td>234</td>
<td>232</td>
<td>4861.9497</td>
<td>.43</td>
<td>4861.9496</td>
<td>.02</td>
</tr>
<tr>
<td>236</td>
<td>215</td>
<td>5173.2320</td>
<td>1.40</td>
<td>5173.2318</td>
<td>.16</td>
</tr>
<tr>
<td>236</td>
<td>353</td>
<td>3666.7583</td>
<td>1.61</td>
<td>3666.7586</td>
<td>- .28</td>
</tr>
<tr>
<td>236</td>
<td>234</td>
<td>11820.7631</td>
<td>3.84</td>
<td>11820.7654</td>
<td>-2.25</td>
</tr>
<tr>
<td>236</td>
<td>232</td>
<td>10785.4941</td>
<td>.72</td>
<td>10785.4942</td>
<td>- .16</td>
</tr>
</tbody>
</table>

Mean value of residuals : 0.07 mm
Standard-deviation of residuals : 0.9 mm
Column [4] : standard-deviation (1 o) of measurements
Column [6] : difference between observed and adjusted values
Table 5
Comparison between 1985 and 1983 networks (Helmert Transform)

<table>
<thead>
<tr>
<th>Scale Factor</th>
<th>Rotation</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ech = 1.000000</td>
<td>( \text{Alpha} = -0.00001 ) (gon)</td>
<td>( \text{TX(m)} = -0.000 ) ( \text{TY(m)} = 0.001 )</td>
</tr>
</tbody>
</table>

Transformed coordinates of active (stable) points

<table>
<thead>
<tr>
<th>Point</th>
<th>( X ) (m)</th>
<th>( Y ) (m)</th>
<th>( dX ) (mm)</th>
<th>( dY ) (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>215</td>
<td>1200.6729</td>
<td>2077.5789</td>
<td>-1.05</td>
<td>0.35</td>
</tr>
<tr>
<td>223</td>
<td>1821.9083</td>
<td>4555.7780</td>
<td>0.91</td>
<td>0.74</td>
</tr>
<tr>
<td>226</td>
<td>1288.2622</td>
<td>723.7316</td>
<td>0.25</td>
<td>-1.80</td>
</tr>
<tr>
<td>231</td>
<td>-1130.9831</td>
<td>7626.9120</td>
<td>0.45</td>
<td>-0.56</td>
</tr>
<tr>
<td>233</td>
<td>-1150.7293</td>
<td>14165.7644</td>
<td>0.73</td>
<td>0.99</td>
</tr>
<tr>
<td>234</td>
<td>6815.9333</td>
<td>6164.0741</td>
<td>-1.42</td>
<td>-0.70</td>
</tr>
<tr>
<td>236</td>
<td>-3886.9903</td>
<td>1148.4617</td>
<td>1.06</td>
<td>1.19</td>
</tr>
<tr>
<td>353</td>
<td>-2796.8321</td>
<td>4648.5917</td>
<td>0.53</td>
<td>-0.22</td>
</tr>
</tbody>
</table>

Transformed coordinates of passive (unstable) points*

<table>
<thead>
<tr>
<th>Point</th>
<th>( X ) (m)</th>
<th>( Y ) (m)</th>
<th>( dX ) (mm)</th>
<th>( dY ) (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>230</td>
<td>-4176.6397</td>
<td>5828.7068</td>
<td>2.40</td>
<td>-0.30</td>
</tr>
<tr>
<td>232</td>
<td>3132.0036</td>
<td>9336.7615</td>
<td>-6.00</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Columns [4] and [5] : \( dX \) and \( dY \) with respect to 1983 coordinates

* The unstable character of some points results from both external considerations and iterative runs of the Helmert transform program. The point 232 is at the top of a 36 m high water-tower. The point 230 is located on a steep side of the Jura mountain and its displacement vector, oriented along the line of greatest slope may correspond to a slippage of the ground.
5. **CONCLUSION**

The results shown in Table 5 are self-explanatory and show the quality of the LEF trilateration network. These results will be compared with those obtained from the satellites of the Global Positioning System but, to be able to make this comparison it is necessary to take into account the Newtonian attraction of the nearby and distant mountains around the site so allowing the calculation of the true heights above the reference ellipsoid.

* * *

**BIBLIOGRAPHY**


DEVIATION OF THE VERTICAL

Werner Gurtner
Astronomical Institute, University of Berne, Switzerland

Beat Burk
Institute for Geodesy and Photogrammetry, ETH Zurich, Switzerland

ABSTRACT
The presentation gives a short overview of the most important formulae, definitions used in the computation of vertical deflections and geoidal undulations, a summary of the known computation methods. As an example the results of the respective computations for the CERN LEP control network are presented.

1. INTRODUCTION

The ultimate goal of the control survey for a particle accelerator is to ensure that the particle beam will follow a path as close to the theoretical figure as possible.

One critical point is the realisation of a true geometrical plane of the size necessary for the Large Electron Positron Collider (LEP, diameter ca. 8.5 km) using classical geodetic methods as precise levelling: Since levelling is related to the equipotential surfaces of the earth's gravity field, any undulations and divergences of these surfaces will show up in a distortion of the plane deduced from the levelled heights, if they are not known or not taken into account.

In the presented paper we first summarize the most important definitions and formulae of the physical geodesy used or referred to later on. Then we will give an overview of the methods of observation and computation of the vertical deflections and the geoidal heights. As an example we will show the main results of the respective computations performed for the control network of the LEP.

2. THE GRAVITY FIELD OF THE EARTH

The notations used here are mostly taken from Ref. 1.

The potential $W$ of the gravity is the sum of the potential $V$ of the gravitational force and the potential $\xi$ of the centrifugal force:

Potential of the gravitational force:
\[ V = k \cdot \iiint \frac{\rho}{r} \cdot dv. \] (1)

Potential of the centrifugal force:
\[ \ddot{z} = \frac{1}{2} \omega^2 \left( x^2 + y^2 \right). \] (2)

Potential of the gravity:
\[ W = W(x,y,z) = V + \ddot{z} \] (3)

with: \( k \): gravitational constant
\( \rho \): density of the volume element dv
\( l \): distance between dv and the point \( P(x,y,z) \)
\( \omega \): angular velocity of the earth
\( \sqrt{x^2+y^2} \): distance from the axis of rotation.

The gradient \( \ddot{g} \) of the gravity field is called gravity, its direction the direction of the plumb line ("Vertical").
\[ dW = \text{grad} \ W \cdot dx = \ddot{g} \cdot dx \quad \text{with:} \quad dx = (dx,dy,dz). \] (4)

The equipotential surfaces \( W(x,y,z) = \text{const} \) are called level surfaces, so:
\[ dW = \ddot{g} \cdot dx = 0 \quad (dx \in \text{level surface}) \]
i.e. \( \ddot{g} \) is perpendicular to the level surface.

The geoid is a unique level surface - coinciding with the surface of the oceans.

The direction of the plumb line in a Point \( P \) is usually expressed by two angles, the latitude \( \phi \) and the longitude \( \lambda \). (Since the plumb line is slightly curved, \( \phi \) and \( \lambda \) change with the height of \( P \).)

The height of \( P \) above the geoid is called orthometric height \( h \).

The difference between the values of the gravity potential at \( P \) and at its footprint in the geoid is called geopotential number. It can be computed as:
\[ C = W_0 - W = \frac{H}{g} \cdot \mathrm{d}h \quad (W_0 \text{ at the geoid}) \quad (5) \]

or

\[ h = - \int_{W_0}^{W} \frac{\mathrm{d}W}{g} = \int_{0}^{C} \frac{\mathrm{d}C}{g} \quad (6) \]

\( \phi, \lambda, h \) or \( \phi, \lambda, C \) are called natural coordinates.

A first approximation of the earth by a mathematical surface is a sphere. A much better approximation is an ellipsoid of revolution. We can look for a "best" fitting ellipsoid either locally (e.g. "Bessel-Ellipsoid" in Switzerland) or globally (e.g. NWL8-E, WGS-72).

![Fig.1 Approximation of the Geoid](image)

If the ellipsoid is an equipotential surface of a normal gravity field it is called level ellipsoid:

\[ U = U(x, y, z) : \text{Potential of normal gravity} \]
\[ \gamma : \text{Normal gravity}. \]

\( U \) is completely determined by:

- the ellipsoidal dimensions \( a, b \)
- the total mass \( M \)
- the angular velocity \( \omega \)

Ellipsoidal Coordinates:

The reference ellipsoid be given by its semi-major axis \( a \) and the excentricity \( e \), with:

\[ e^2 = (a^2 - b^2)/a^2 \]
Fig. 2 Ellipsoidal coordinates

\[ P(x, y, z) = P(B, L, H) \]

with:
- \((B, L)\): Direction of the normal \(n\)
- \(H\): Ellipsoidal height

\[
\begin{align*}
x &= (N + H) \cdot \cos B \cdot \cos L \\
y &= (N + H) \cdot \cos B \cdot \sin L \\
z &= (N \cdot (1-e^2) + H) \cdot \sin B
\end{align*}
\]

\[ N = a \cdot \left(1 - e^2 \cdot \sin^2 B\right)^{\frac{1}{2}} \]

Undulations, deflections, anomalies

\[ \omega = \omega_p \]

- \(h\): Orthometric height

Geoid: \( \omega = \omega_0 \)

Ref. Ellipsoid: \( U = \omega_0 \)

Fig. 3
Some definitions:

\[ N \quad : \quad \text{geoidal height (undulation)} \]

\[ \Delta g = \vec{g}_{P_0} - \vec{g}_{Q_0} \quad : \quad \text{gravity anomaly vector} \]

\[ \Delta g = g_{P_0} - g_{Q_0} \quad : \quad \text{gravity anomaly} \]

\[ \theta_0 , \theta \quad : \quad \text{deflection of the vertical} \]

\[ \theta - \theta_0 \quad : \quad \text{curvature of the plumb line} \]

\[ T = W - U \quad : \quad \text{disturbing potential} \]

\[ \xi = \phi - B \quad \{ \text{components of the deflections} \} \]

\[ \eta = (\lambda - L) \cdot \cos B \quad \{ \text{in NS-, EW-direction} \} \]

An important relation between the disturbing potential \( T \) and the geoidal height is the so called Bruns Theorem:

\[ N = \frac{T}{r} \quad (8) \]

**Divergence of the Equipotential Surfaces**

![Figure 4](image)

\[ W = W_A = W_B \]

\[ \Rightarrow \Delta h = h_B - h_A \neq 0 \]

\[ W = W_0 \quad \text{(Geoid)} \]

\[ \Delta W = W_B - W_A = 0 \]

Fig. 4

Important consequences of the non-parallelity of the equipotential surfaces are the dependency of levelled height differences on the chosen path of the levelling: the mis-closure of a levelled loop:

\[ \int dh \neq 0 \]

and the curvature of the plumb lines:

\[ (\phi_B, \lambda_B) \neq (\phi_B, \lambda_B) \]

\[ (\xi_B, \eta_B) \neq (\xi_B, \eta_B) \]
3. SOME EXAMPLES

3.1 Detailed structure of the geoid

The first example shows a small part of the geoid in the Bernese Oberland. Undulations on the decimeter level over a few kilometers can be found.

(Equidistance: 5 cm, altitudes in m above sea level)

![Geoidal undulations](image)

Fig. 5 Geoidal undulations

3.2 Deflections in a profile

The figure shows a north-south profile of the deflections ($\xi$-component) at the earth's surface and in the geoid. Variations of 10 seconds in the surface values can be found within a few kilometers only.

![North-South profile "Gotthard" (Deflections $\xi$)](image)

Fig. 6: North-South profile "Gotthard" (Deflections $\xi$)
3.3 Gravity anomalies

The largest gravity anomalies in Switzerland are caused by the Ivrea mass anomaly.

**Fig. 7: Gravity anomaly caused by the Ivrea Body**

Why do we need to know the geoid and the deviations of the vertical? Most of all because of the difference between the "physical" and the "geometrical world". A few examples:

- Levelling is related to equipotential surfaces
- Zenith and horizontal angles measured with a theodolite are related to the local directions of the plumb line
- Astroseodetic point positioning gives the direction of the physical plumb line
- Network adjustments are done in the geometrical space: The natural coordinate systems have to be bound together!
- Only distances are not influenced by the disturbing potential
- Every site has its own local (natural) coordinate system
- GPS-results are given in a pure three dimensional coordinate system. GPS "sees neither geoidal undulations nor deviations of the vertical".
- Particles in an accelerator will not follow the equipotential surface but they have to orbit in a true plane.
4. **DETERMINATION OF \((\phi, \eta), A_g, N\)**

4.1 **By Observation**

4.1.1 **Overview**

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Instrument</th>
<th>Reference</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\phi, \lambda)</td>
<td>Theodolite Zenith camera</td>
<td>B, L</td>
<td>(\xi_A, \eta_A)</td>
</tr>
<tr>
<td>(g)</td>
<td>Gravimeter</td>
<td>(\gamma)</td>
<td>(A_g)</td>
</tr>
<tr>
<td>(\Delta H)</td>
<td>Doppler, GPS Levels</td>
<td>(R^\prime, R^R)</td>
<td>(\Delta N)</td>
</tr>
<tr>
<td>(\Delta h)</td>
<td>Levels</td>
<td>(M, M^R)</td>
<td></td>
</tr>
</tbody>
</table>

4.1.2 **Example**

As an example we give a short description of the transportable zenith camera used by the Institute for Geodesy and Photogrammetry of the ETH Zurich.

Principles of operation: By taking a photograph of the fixstars by means of a special camera in a vertical setup the astronomical parameters \(\phi\) and \(\lambda\) of the observation point can be determined. Two essential quantities must be known to fulfill the requirements of the method:

- The epoch (local sidereal time) corresponding to the exposures (start and end of each single exposure as fed back from the shutter)
- The readings of the electronic levels mounted in perpendicular position on the camera tube to determine the spacial orientation of the camera and its optical axis with respect to the local plumb line.

The camera as it can be seen in Fig. 8 and 9 is equipped with a set of electronic levels, a high precision internal clock, an electronic shutter and a dedicated microprocessor system for the monitoring of the camera and the data access and storage.

After development, the films (size 6.5 x 9 cm) are measured on a high precision comparator to obtain the plate coordinates of the star images. Applying an appropriate transformation model between the theoretical and the measured plate coordinates of the known stars one can determine a set of astronomical parameters \(\phi\) and \(\lambda\) referring to the camera position. The components of the vertical deflection are easily obtained by the equations

\[ x = \phi - B, \]
\[ \eta = (\lambda - L) \cos B. \]
Fig. 8

Transportable zenith camera of the Institute of Geodesy and Photogrammetry, ETH Zürich. Beside the optical tube a pair of electronic levels of 'Talyvel 2' type is mounted in perpendicular orientation.

Fig. 9

Complete camera system in "ready-to-use" position. The electronics on the right hand side contains the power supply, the clock with time-signal receiver, the level electronics, the hardware to monitor the shutter and a microprocessor system for the data handling and storage.
4.2 By computation

4.2.1 Overview

<table>
<thead>
<tr>
<th>Input</th>
<th>Method</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta g$</td>
<td>$\left{ \frac{1}{4\pi G} \int_0^\infty \frac{\Delta g \cdot ds}{d\varphi} \cdot \left{ \cos \varphi \right} \cdot \sin \varphi \right} \cdot \Delta I$</td>
<td>$\xi_I, \eta_I$</td>
</tr>
<tr>
<td>$\Delta g$</td>
<td>Stokes</td>
<td>$N_I$</td>
</tr>
<tr>
<td>$\xi_A, \eta_A$</td>
<td>Astrogeodetic Levelling $\Delta N = - \int_0^\infty \theta \cdot ds$</td>
<td>$\Delta N_I$</td>
</tr>
<tr>
<td>$z$</td>
<td>Reciprocal Zenith Distances</td>
<td>0</td>
</tr>
<tr>
<td>$\xi_A, \eta_A, \Delta g$</td>
<td>Astrogravimetric Levelling: Combination of Vening-Meinesz and Astrogeod. Levelling</td>
<td>$\xi_I, \eta_I, \Delta N_I$</td>
</tr>
<tr>
<td>$\xi_A, \eta_A, (\Delta g, N)$</td>
<td>Mass Model Indirect Methods using Collocation</td>
<td>$\xi_I, \eta_I, \Delta N_I$</td>
</tr>
</tbody>
</table>

4.2.2 Example: Indirect Computation Using Collocation

This method has been described in more detail e.g. in Ref. 2 or Ref. 3: Provided a mass model of (a part of) the earth's crust be given, it is possible to compute the differences between observed derivatives of the disturbing potential (deviations of the vertical, gravity anomalies etc.) and the corresponding values defined by the mass model. These differences show a much smoother behaviour than the original observations. Therefore they are more suited for interpolation. Using collocation (appropriate covariance functions between all the quantities are necessary) we can not only interpolate between the given types of observation, but compute e.g. equipotential surfaces $N^I$ of the reduced disturbing potential. Adding the geoidal heights given by the mass model ($N^M_I$) leads us to the final geoidal heights $N^I$. 
The astroseodetic geoid in Switzerland has been computed using (Ref. 2):

- a set of more than 200 components of observed deviations of the vertical
- a mass model of the earth's crust in and around Switzerland, containing information about the topography, the depth of the Mohorovicic discontinuity, and the most prominent mass disturbance, the "Ivrea Body".

5. APPLICATIONS ON THE CERN-LEP CONTROL NETWORK

In order to increase the relative accuracy of the interpolated deviations and geoidal heights, the Geodetic Institute of the ETH Zürich observed additional deviations of the vertical in the LEP area using the above described transportable zenith camera. The field measurements within the LEP-area took place in August 1983. All observations were carried out in only two nights. This particular work within the LEP project was the first field experiment with the camera, which was built at the Institut für Erdmessung, Technical University Hannover. Meanwhile the system has been used in the southern part of Switzerland and Northern Italy where, so far, a set of about 100 vertical deflection points have been measured (c.f. Ref. 6).
These additional vertical deflections in the LEP area have been introduced as additional reference values into the computation program. The results of the computations can be found in detail in Ref. 4 and 5. They can be summarized as follows:

- The vertical deflections range from 0 to 15 arc seconds at surface level and from 0 to 9 arc seconds at zero level.
- The resulting separation between the local reference ellipsoid and the geoid reaches 200 mm at 10.5 km from the origin.
- The altimetric corrections along the LEP machine - to obtain a true plane in space - vary from -40 mm to +100 mm.

Figures 10, 11, and 12 show the locations of the astrogeodetic stations, the total deflection intensities, and the equipotential heights in the LEP area. (The local reference ellipsoid coincides with the geoid at $\rho_0$.)

![Diagram of astrogeodetic stations]

Fig. 10: Astrogeodetic stations
Fig. 11: Total deflection intensity (arc seconds)

Fig. 12: Equipotential heights (mm)
REFERENCES


4) B.A. Bell, A Simulation of the Gravity Field around LEP, Private communication, 1985.


COMPARISON BETWEEN TERRAMETER AND GPS RESULTS -- AND HOW TO GET THERE

G. Beutler
Astronomical Institute, University of Berne, Switzerland.

ABSTRACT
After a short historical introduction accuracy limiting influences on GPS-derived engineering type networks are discussed in section 1, sections 2 and 3 are devoted to the 1984 Macrometer- and the 1985 Sercel- GPS Campaigns conducted on the CERN - LEP control network.

1. MEASURING ENGINEERING TYPE NETWORKS

In the early days of GPS (1982-1984) measurement campaigns usually were organized with two receivers operating simultaneously. The processing software was baseline- and session-oriented, which means that one three-dimensional vector (from the first to the second receiver) was estimated for each session. The scientific objectives may be summarized as follows:

- Repeatability demonstrations of baseline measurements
- Study of figure misclosures
- Accuracy demonstrations over very short baselines (invar-wire measurements)

Accuracy demonstrations of the order of 3 to 5 ppm were considered a success. The Ottawa-Macrometer campaign\(^1\) may serve as a typical example of a "pioneer-age" campaign.

If a GPS campaign is organized today, it is almost certain that more than two receivers are operated simultaneously. Also, the processing software went through a remarkable development. The First International Symposium on Precise Positioning using GPS in Rockville 1985 and the Fourth International Geodetic Symposium on Satellite Positioning in Austin 1986 clearly showed the trend from baseline- and session- to network- and campaign- oriented software systems\(^2\-\text{a})\).

The scientific objectives may be summarized as follows:
- Repeatability demonstrations on entire networks (no figure misclosures)
- Subdivision of a large network: If a network has to be observed in several sessions, how many points of the network should be in common from session to session?
- Accuracy demonstrations on terrestrial high precision networks.
- Studies of accuracy-limiting effects such as troposphere, ionosphere, orbits and receiver clocks.

The ultimate accuracy-goal for GPS-established engineering type networks certainly is 0.1 ppm -- but it is doubtful whether this will be achievable at all. What certainly is possible on a routine basis with good processing software are accuracies of the order of 1 ppm in networks of the order of 10 km x 10 km.
Let us now have a look at the accuracy-limiting influences on GPS derived small scale networks.

1.1 Orbits

The influence of orbit errors on baseline and network estimates is one of the topics discussed in the contribution "GPS Orbit Determination" also included in these proceedings. It is therefore sufficient to summarize here our experiences with different orbit and processing types in small scale applications.

- Two orbit types are -- more or less -- readily available: Broadcast ephemerides transmitted by the satellites and precise ephemerides produced by the U.S. Department of Defence in an a posteriori analysis. It has been our experience during the past few years that broadcast ephemerides usually are of a surprisingly good quality. Generally one may expect orbit accuracies of the order of 20 meters or better, the exceptions being satellites with oscillator problems (such as space vehicle 4) where errors of the order of 100 meters may be present in broadcast ephemerides. The few specimens of precise ephemerides we had at our disposal do not allow a general judgement of their orbit quality, but there are indications that the accuracies are of the order of 10 meters for all satellites.

- Our parameter program has the capability of estimating orbits\(^1\)). According to our experience one may not expect significantly better network coordinates if this program option is used if, at least, orbits of the broadcast quality are available. It must be pointed out that this statement does not hold for networks larger than ~ 10km x 10km.

- For the 1984 CERN Macrometer campaign we had both orbit types at our disposal. No significant difference in the quality of the GPS solution could be detected when replacing one orbit type by the other. This again demonstrates the high standard of the broadcast orbits.

- This comfortable situation could change considerably if, in future, broadcast ephemerides were artificially deteriorated. Then it would be mandatory to estimate at least some of the orbital elements in order to produce high quality results.

1.2 Ionosphere

As probably first pointed out by Campbell et al\(^1\)) the main systematic effect of neglecting a "well behaving" ionosphere (no dramatic changes in space and time) when processing GPS carrier phase data (single frequency) is a scale-factor in the GPS-derived baselines or networks. Baselines measured with single frequency GPS receivers tend to be short by a few tenths to -- may be -- three ppm depending on the electron content in the ionosphere. It should be mentioned that an agitated ionosphere (e.g. travelling ionospheric disturbances) may have an unpredictable influence on the coordinates derived from single frequency GPS observations. We had a closer look at this scale-phenomenon when we processed the 1984 CERN Macrometer campaign. This part of the 1984 CERN GPS analysis may be summarized as follows:
- The study was motivated by an unexplained scale factor of 1.8 ppm between the GPS and the terrestrial solutions when the ionosphere was completely neglected.
- It seemed appropriate (night time observations in moderate latitude) to model the ionosphere for the Cern campaign with the single layer model, assuming that all electrons are concentrated in a thin layer at a height $H$ above ground. Moreover it was assumed that the electron density $E$ was a constant (in time and within the layer).
- Four program runs were made using electron densities of $E = 0 \times 10^{16}$, $10 \times 10^{16}$, $20 \times 10^{16}$, $30 \times 10^{16}$ electrons per m$^2$.
  and it could be demonstrated that each change of $\Delta E = 10 \times 10^{16}$ introduced a change of 0.5 ppm in the scale of the GPS solution.
- It should be stated clearly that we had no hope of estimating the electron content from the Maccrometer observations (only marginal differences in the estimated rms errors of double differences for different electron contents occurred). It seemed most unlikely however, that an electron content of more than $E = 10 \times 10^{16}$ should have been present.

More conclusive statements concerning the ionosphere could be made when analyzing the material of the 1984 Alaska GPS Campaign$^*$. The situation was quite different in this case:
  (a) the size of the net was $\sim 1000\text{km} \times 2000\text{km}$.
  (b) the observations were made during the afternoon.
  (c) both carrier frequencies L1 and L2 were measured by the TI-4100 receivers.

It turned out that the L2 derived solution had a scale of 1.9 ppm and the L1 1.4 ppm with respect to the solution using the so called ionosphere-free linear combination of L1 and L2. (For the sake of completeness I should mention that the scale of this combined solution was identical with the scale of VLB1). From our point of view these results show that an ionosphere induced scale factor has to be expected if baselines or networks are estimated from single frequency phase data and if the ionosphere is neglected.

### 1.3 Troposphere

When processing GPS observations, usually the models developed for the TRANET - Doppler system are used; the best known ones may be the Hopfield- and the Saastamoinen models. In order to give an impression of the order of magnitude of the tropospheric refraction correction we reproduce here the Saastamoinen formula$^3$:

$$\Delta r = 2.28 \times (P + (1255/(273+T) + 0.05) \times e)$$

where $\Delta r$ is the tropospheric refraction correction in zenith direction in mm and P,T,e are atmospheric pressure (mbar), temperature (°C) and water vapour pressure (mbar).

The water vapour pressure in turn may be expressed as a function of relative humidity $\%$ (expressed in $\%$) and temperature:
\[ e = \frac{H}{100} \times \exp(-37.2465 + 0.21366 \times (T+273) - 0.000256988 \times (T+273)^2). \]

Using the two formulae it is quite easy to compute the bias \(\Delta \delta r\) introduced into the GPS observable (in the zenith direction) by an error in temperature of 1°C, in pressure of 1 mbar and in humidity of 1%, at 0°C, 1000 mbar and 50% humidity:

\[ \Delta \delta r = 3 \text{ mm}, 2 \text{ mm}, 0.6 \text{ mm} \text{ for errors of 1°C, 1 mbar, 1}. \]

These biases are even more pronounced if the surface temperature is rising. If we have for example T = 30°C, an error of 1°C would introduce a bias of 14 mm while an error of 1% in humidity would introduce 14 mm into the observable in the zenith direction. It may be shown that these meteorological biases in the observable induce biases in the estimated height coordinates, where the above figures will be amplified by a factor of about 3. (For a more detailed treatment see Ref. 9).

Effects of this order of magnitude are of no importance in the "TRANET-Doppler world", but they are of vital importance in GPS where we are hunting the millimeter! Keeping in mind the accuracy limits of standard meteor equipment and the severe problems involved with this kind of measurement (e.g. surface effects), it is not really surprising that in GPS processing of small scale networks it is usually preferable not to use the weather data recorded at the receiver sites directly, but to use them to compute some kind of local "standard atmosphere". This has been our experience and that of others too\(^9\). During the last few months we have tested atmosphere models of the kind used for classical EDM - measurements (e.g. model by Essen and Froome). Although first applications are looking promising\(^1\), we are not yet in a position to draw any conclusions.

For the two CERN campaigns to be discussed in the subsequent pages we used the surface weather data to generate a local atmosphere, where temperature, pressure and humidity gradients were assumed to be known as a function of the heights of the sites. Consequently the meteorological measurements uniquely served to define temperature, pressure and humidity at a reference height \(H_0\).

2. THE 1984 CERN MACROMETER CAMPAIGN

A detailed report concerning this campaign was presented at the 1985 FIG meeting in Munich\(^9\). Therefore I only summarize the main characteristics and results.

Points P226, P230, P231, P232, P233, P234 and P236 of the CERN-LEP control network (see Fig. 1) were observed with three Macrometer V-1000 Surveyor instruments from 11 December to 13 December 1984. The daily observation periods started at ~ 0h and ended at ~ 3h30m (UT).
Each observation period was divided into two observation sessions: The first lasted for two hours, the second for 80 minutes. Therefore two independent baselines could be measured per session, four per day and twelve during the entire campaign. Roger Jones from GEOHYDRO Inc. organized the campaign in order to measure twice those six baselines starting from the central point P231. (This implies that P231 was permanently occupied by one of the receivers).

Since Navstar 4 was moved by the U.S. DoD during the week preceding the campaign, the Macrometer instruments could not find this satellite. The consequence was that usually only three satellites could be observed simultaneously, a fourth satellite being visible only during 10-15 minutes at the end of the first and during the first 10-15 minutes of the second session. Considering five to six simultaneously observed satellites to be an optimum, it is clear that the Macrometer Campaign was severely handicapped by the "absence" of Navstar 4.
The observations were processed with the Bernese GPS Software Package\(^2\). The main characteristics of the processing step may be summarized as follows:

- Double differences were used as basic observables.
- Precise ephemerides could be used to produce the final solution. No attempt was made to improve these orbits.
- The central point P231 was held fixed. Its WGS-72 coordinates were transformed from the Swiss Datum "CH - 1903" using TRANSIT-Doppler derived parameters.
- Initially 54 parameters (6*3 coordinates, 36 ambiguities) had to be estimated.
- Of these 36 ambiguities 33 could be resolved (related to integer numbers) by our automatic ambiguity resolution algorithm.
- After ambiguity resolution the adjustment was repeated with only 21 unknowns (6*3 coordinates, 3 ambiguities) to give the final solution.
- Four different program runs (with different electron contents) gave four different solutions. In this report we only present the solution for \(E = 10 \times 10^{16}\) electrons per m\(^2\); which seems to be a hypothesis that makes sense.

In Table 1 we reproduce what we believe is "the best" GPS solution. The 1 o errors for the coordinates (actually the coordinate differences with respect to point P231) are not given in this Table. They are typically 1 mm for latitude, 2 mm for longitude and 3 to 4 mm for height.

### Table 1
Results of 1984 CERN Macrometer Campaign
Datum: CH 1903

<table>
<thead>
<tr>
<th>Station</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>P226</td>
<td>46 13 36.24140</td>
<td>6 01 54.44865</td>
<td>498.821</td>
</tr>
<tr>
<td>P230</td>
<td>46 17 32.26122</td>
<td>6 00 36.07847</td>
<td>747.743</td>
</tr>
<tr>
<td>P231 *)</td>
<td>46 17 25.41472</td>
<td>6 03 21.02502</td>
<td>428.774</td>
</tr>
<tr>
<td>P232</td>
<td>46 16 54.05284</td>
<td>6 06 50.77772</td>
<td>494.033</td>
</tr>
<tr>
<td>P233</td>
<td>46 20 21.28416</td>
<td>6 06 11.26072</td>
<td>740.512</td>
</tr>
<tr>
<td>P234</td>
<td>46 14 22.10559</td>
<td>6 07 50.34666</td>
<td>457.019</td>
</tr>
<tr>
<td>P236</td>
<td>46 15 21.30264</td>
<td>5 58 45.18231</td>
<td>603.758</td>
</tr>
</tbody>
</table>

*) Fixed point

This brings us to the last and most interesting part of the analysis, the comparison between the GPS and the Terrameter results. In order to make a meaningful comparison, the terrestrial coordinates -- given in a local Cartesian system -- had to be transformed into the WGS - 72 coordinate system. The parameters of this transformation are:

(a) astronomical latitude and longitude of the z-axis,
(b) astronomical azimuth of the x-axis and
(c) WGS - coordinates (Doppler derived) of the origin of the local system.

For a more explicit definition of these parameters see Ref. 12.

The transformation between the GPS- and the Terrameter-derived WGS - 72 coordinates was then performed through a seven-parameter Helmert transformation (three translation parameters, three rotation parameters, one scale factor). The latter four parameters (transformation GPS --> Terrameter) were estimated as:

\[ \begin{align*}
\omega_{ns} &= 0.25'' \pm 0.08'' \text{ (rotation about ns-axis)} \\
\omega_{ew} &= 0.52'' \pm 0.08'' \text{ (rotation about ew-axis)} \\
\omega_z &= 0.72'' \pm 0.05'' \text{ (rotation about z-axis)} \\
s &= 1.3 \text{ mm/km} \pm 0.3 \text{ mm/km} \text{ (scale factor).}
\end{align*} \]

The three rotation angles are all smaller than 1'', which means that the GPS based orientation and the orientation deduced from the terrestrial observations match quite nicely. The scale factor of 1.3 ppm still lacks an explanation.

Table 2 gives the residuals in the coordinates that remain after the Helmert transformation between the terrestrial and the GPS - solutions. The figures in Table 2 demonstrate the high consistency and integrity of both the Terrameter and the GPS solution.

**Table 2**

<table>
<thead>
<tr>
<th>Station</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>P226</td>
<td>3</td>
<td>-3</td>
<td>6</td>
</tr>
<tr>
<td>P230</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>P231</td>
<td>3</td>
<td>3</td>
<td>-6</td>
</tr>
<tr>
<td>P232</td>
<td>-3</td>
<td>-8</td>
<td>-1</td>
</tr>
<tr>
<td>P233</td>
<td>-3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P234</td>
<td>1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>P236</td>
<td>-2</td>
<td>2</td>
<td>-4</td>
</tr>
</tbody>
</table>

3. THE 1985 CERN SERCEL CAMPAIGN

Eight points of the CERN - LEP control network -- actually the seven points already surveyed by Macrometer in 1984 plus point "PM 15" (see Fig. 1) -- were equipped with three Serce TR55 receivers \(^{13}\) from 14-18 October, 1985. The campaign was organized in a way very similar to that of the CERN Macrometer: The daily observation period (from 3h30m to 7h UT) was divided into two sessions, each lasting for approximately 90 minutes. P231, the central
point, was always occupied and always by the same receiver unit. As opposed to the 1984 campaign, the Sercel measurements lasted for five nights. Since the receivers were not moved between the two daily sessions, three points could be surveyed per night (namely P231, P230, P236 on the 14th, P231, P226, P234 on the 15th, P231, P232, P233 on the 16th, P231, P236, PM 15 on the 17th and finally P231, P226, PM 15 on the 18th of October).

Although the three Sercel receivers had no problems tracking all GPS satellites available at the time of the campaign, the configuration was not much better than during the 1984 campaign, due to low maximum elevations of two of the satellites.

Moreover it became obvious — unfortunately only during the processing phase four months after the campaign — that one of the receivers had had serious oscillator problems. As a matter of fact these problems turned out to be so serious that its phase observations could not be used to produce high quality coordinates. Being familiar with E. Murphy’s law I was not surprised to see that the faulty receiver was the one always used at the central point P231! Therefore it is clear that the Sercel campaign could not produce a network comparable to the Macrometer net.

What remained is summarized in the following Table, where we only compare baseline lengths (established with Terrameter, Macrometer and Sercel) from that part of the net which could be covered by the two properly working Sercel receiver units.

The conclusions are quickly drawn:

(a) If the Sercel receivers are working properly they promise to produce high quality results (the formal errors being of the same order of magnitude as those for Macrometer).

(b) In order to demonstrate the Sercel receivers’ capability to produce GPS derived networks of a consistency and integrity comparable to the Macrometer V-1000 receivers it is certainly necessary for Sercel to repeat the CERN campaign (or a campaign in a net with a comparable ground truth).

Table 3

Comparison of Baseline lengths (meters)
(T = Terrameter, M = Macrometer, S = Sercel)

<table>
<thead>
<tr>
<th>Baseline</th>
<th>S</th>
<th>S - T</th>
<th>S - M</th>
<th>M - T</th>
</tr>
</thead>
<tbody>
<tr>
<td>P226-P234</td>
<td>7755.932</td>
<td>-0.02</td>
<td>0.017</td>
<td>-0.019</td>
</tr>
<tr>
<td>P226-P236</td>
<td>5193.633</td>
<td>-0.02</td>
<td>-0.012</td>
<td>-0.010</td>
</tr>
<tr>
<td>P232-P234</td>
<td>6458.931</td>
<td>-0.04</td>
<td>0.003</td>
<td>-0.007</td>
</tr>
<tr>
<td>P234-P236</td>
<td>11820.738</td>
<td>-0.32</td>
<td>0.000</td>
<td>-0.032</td>
</tr>
<tr>
<td>P226-PM15</td>
<td>2204.744</td>
<td>-0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P234-PM15</td>
<td>5757.046</td>
<td>-0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P236-PM15</td>
<td>6248.884</td>
<td>-0.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


3) C.C. Goad. Ibid., pp. 347-356.


7) W. Gurtner. GPS papers presented by the Astronomical Institute of the University of Berne in the Year 1985. Mitteilungen der Satellitenbeobachtungsstation Zimmerwald No. 18. 53-70, University of Berne (1985).


10) Y. Bock. private communication.


HIGH PRECISION GEODESY APPLIED TO CERN ACCELERATORS

J. Gervaise and E.J.N. Wilson
CERN, Geneva, Switzerland

ABSTRACT
After examining the whys and wherefores of applied geodesy in high energy particle accelerator construction, the authors show how the transverse beam size and hence the dimension of the magnets are related to positional tolerances and review the methods and instruments developed for this purpose. Three accelerators have successively been built at CERN, the 28 GeV Proton Synchrotron (PS), The intersecting Storage Rings (ISR), the Super Proton Synchrotron (SPS), and a fourth one is under construction, the Large Electron Positron Collider (LEP). Since 1954, date of the PS construction, many difficulties have been encountered and the authors show how applied geodesy has been able to overcome them. From the traditional geodetic methods available in the 1950's, used to build a 200-m diameter accelerator, up to the 1980's, many methods and instruments ranging from invar wire to Terrameter, have been developed. These have improved the reliability, accuracy and speed of the geodetic measurements and have given a philosophy and a strategy for the Applied Geodesy of a 27-km circumference collider such as LEP.

1. INTRODUCTION

Over the last three decades, particle accelerator energies have grown from several tens of GeV to several hundred GeV, and unavoidably the diameters of circular accelerators have similarly increased from two hundred metres to ten kilometres. Despite this escalation, it has proved possible to achieve tolerances on magnet positioning which are not only as tight as in the past, but even more precise due to the survey experience already gained.

The sequence of large projects undertaken by CERN, the 28 GeV Proton Synchrotron (PS, 1954-1959), the Intersecting Storage Rings (ISR, 1966-1971), the 400 GeV Super Proton Synchrotron (SPS, 1971-1976) and the Large Electron Positron Collider (LEP, 1981-), has ensured continuity in the work of the CERN Applied Geodesy Group, and each project has been in itself a major challenge. The Group has thus been constantly searching to improve
reliability and speed of measurement without sacrificing in any way the accuracy essential to proper accelerator operation.

A circular accelerator consists basically of magnets arranged in series according to a certain configuration. The magnets are separated by straight sections where there is no magnetic field, and form a lattice having a certain periodicity \( P \). To an observer standing outside, the magnetic structure resembles an underground train in a circular tunnel. However, for the accelerated proton, which sees things from the inside, the picture is quite different. It has the sensation that it is running round a magnetic channel, and this channel must be as smooth as possible to ensure that no oscillations are induced which would broaden the particle beam to which it belongs. The protons are sensitive to any bumps occurring in the channel; these bumps may be due to three causes, each independent of the other, but which have the same effect on the magnetic field that guides the beam around the machine. They are:

- slight movements of the foundation rock or of the concrete structures supporting the magnets,
- random errors in magnet alignment,
- errors in the beam guidance field itself.

These errors perturb the ideal closed orbit which passes down the axis of the magnets. The geodesist measures angles, distances, offsets and tilts. After least squares adjustment, he provides the beam orbit specialists with the following values:

- \( \Delta r \) : the radial deviation of the synchrotron proton orbit,
- \( \Delta z \) : the vertical deviation from the mean plane selected for the closed orbit along which the accelerated protons circulate,
- \( \Delta t \) : the deviation in the transverse tilt of the magnets, which introduces coupling between the horizontal and vertical planes and makes orbit corrections difficult.

As an illustration of the effect of these errors and their correction, Fig. 1 shows the strong correlation between the survey results and the local orbit corrections in the SPS ring. Also shown is the effect of a correction procedure by a deliberate movement of a few quadrupoles by a few tenths of millimetres, calculated from orbit measurements made with beam position sensors. The amplitude of the closed orbit can thus be reduced by a factor 2 or 3.

It is a feature of machines such as the SPS that the relative errors of quadrupole focusing magnets are the most critical: the radial and vertical variations from one quadrupole to the two adjacent quadrupoles must remain as close as possible to the ideal value given by the theoretical geometry. Next, the curve formed by the successive positions of the quadrupoles has to be made as smooth as possible along the entire perimeter of the machine. Thus, the two basic concerns in accelerator metrology are, first to carry out an overall survey that is close to the theoretical geometry and then to smooth the curve in order to minimize the errors locally. We shall discuss this aspect in later sections.
Fig. 1 Beam position measurements before and after correction

A large fraction of the considerable investment in capital and technological ingenuity necessary to build a modern synchrotron finds its way into the magnetic fields which guide and focus the beam in its circular path. Their alignment is crucial to the performance of the machine and stretches survey techniques to the limit. In the following sections, we first show how the transverse beam size and hence the dimensions of the magnets is related to positioned tolerances and go on to show how the tolerances may be realised.

2. TRANSVERSE COORDINATES

The bending fields of a synchrotron are usually vertical directed, causing the particle to follow a curved path in the horizontal plane (Fig. 2). The force acting on
the particle is horizontal and is given by:

\[ F = e \mathbf{v} \times \mathbf{B} \]  

(1)

where: \( \mathbf{v} \) is the velocity of the charged particle in the direction tangential to its path, 
\( \mathbf{B} \) is the magnetic guide field.

\[ y \] is used as general transverse coordinate for both \( x \) and \( z \)

Local centre of gyration

Central orbit

\( s \) (Tangential to beam direction)

(Fig. 2 Charged particle orbit in a magnetic field)

If the guide field is uniform, the ideal motion of the particle is simply a circle of radius of curvature \( \rho \). We can also define a local radius of curvature \( \rho(s) \), to describe motion in a non-uniform field. We shall suppose that it is possible to find an orbit or curved path for the particle which closes on itself. We define this as the equilibrium orbit or closed orbit. The machine is usually designed with this orbit at the centre of its vacuum chamber.

Of course, a beam of particles enters the machine as a bundle of trajectories which are centred on the ideal orbit. At any instant a particle may be displaced horizontally or vertically from the ideal position. The transverse displacements are \( x \) and \( z \), respectively. The particles may also have a divergence angle:

\[ x' = dx/ds \quad \text{or} \quad z' = dz/ds \]  

(2)

This would cause them to leave the vacuum pipe if it were not for the carefully shaped field which turns them back towards the beam centre so that they oscillate about the ideal orbit. The design of the restoring fields determines the transverse excursions of the beam and the size of the cross section of the bending magnets. The restoring or focusing magnets have four poles and are usually embedded in the lattice of bending magnets in an alternating pattern. Half focus the beam while, the other half defocus. In Fig. 3 we see an example of such a magnet pattern, which is one cell, or about 1% of the circumference of the 400 GeV SPS at CERN. This focusing structure is called FODO. For the purpose of this paper, it is sufficient to grasp the idea of a beam of particles oscillating from side to side of the centre of the machine in the focusing channel of quadrupoles. The envelope of these oscillations follows the function \( \beta(s) \) which has waists near each defocusing magnet and a maximum at the centres of F quadrupoles. Since F quadrupoles in the horizontal plane are D quadrupoles vertically, and vice versa, the two functions \( \beta_H(s) \) and \( \beta_V(s) \) are one half-cell out of register in the two transverse planes.
For formal reasons $\beta$ has the dimensions of length but the units bear no relation at this stage to physical beam size.

![Diagram of Bending Magnets and Focusing Magnets](image)

Fig. 3 One cell of the CERN 400 GeV Super Proton Synchrotron representing 1/108 of the circumference. The pattern of dipole and quadrupole magnets (F and D) is shown. Beam particles make betatron oscillations within the shaded envelopes.

Particles do not follow the $\beta(s)$ curves but oscillate within them in a form of modified sinusoidal motion whose phase advance is $\varphi$. The phase change per cell in the example shown is close to $\pi/2$ but the phase advance is modulated throughout the cell.

Later, we shall discuss how the functions $\beta$ and $\varphi$ are related to the focusing pattern and to beam sizes.

3. **BENDING MAGNETS AND MAGNETIC RIGIDITY**

Most of the ring of a synchrotron consists of a guide field of bending magnets.

Suppose the particle has a relativistic momentum vector $p$ and travels perpendicular to $B$ into the plane of the diagram (Fig. 4). After time $dt$ it has followed a curved path of radius $\rho$ whose length is $ds$ and its new momentum is $p + dp$. Since we may equate the force and rate of change of momentum:

$$ e \bar{v} \times \bar{B} = \frac{dp}{dt} $$

and we see from resolution of momenta that:

$$ \frac{dp}{dt} = \frac{|p| \, ds}{\rho \, dt} = \frac{|p| \, ds}{\rho \, dt} \cdot $$

(3)
On the other hand, if the field and plane of motion are normal, the magnitude of the force may be written:

\[ e \left( \vec{v} \times \vec{B} \right) = e |B| \frac{ds}{dt} \quad (5) \]

Equating (4) and (5) we find we can define a quantity known as magnetic rigidity:

\[ (B\rho) = \frac{P}{e} \quad (6) \]

Strictly we should use the units Newton.second for \( p \) and express \( e \) in coulombs to give \((B\rho)\) in tesla.metres. But in charged particle physics a more common convention is to quote \( p/c \) in units of electronvolts. Whereupon:

\[ (B\rho)[\text{T.m}] = \frac{pc[\text{eV}]}{C[\text{m.s}^{-1}]} = 3.3356(pc)[\text{GeV}] \quad (7) \]

![Diagram](image)

Fig. 4 Vector diagram showing differential changes in momentum for a particle trajectory.

Figure 5 shows the trajectory of a particle in a bending magnet or dipole of length \( \rho \). Usually the magnet is placed symmetrically about the arc which is the particle's path. One may see immediately that:

\[ \sin \left( \frac{\theta}{2} \right) = \frac{\rho}{2\rho} = \frac{kB}{2(B\rho)} \quad (8) \]

and if \( \theta \ll \pi/2 \),

\[ \theta \approx \frac{kB}{(B\rho)} \quad (9) \]

The bending magnet aperture must be wide enough to contain the sagitta of the beam which is:

\[ \pm \frac{\rho}{2} \left( 1 - \cos(\theta/2) \right) = \pm \frac{\rho\theta^2}{16} = \frac{\rho\theta^2}{16} \quad (10) \]

The ends of bending magnets are often parallel but in some machines are designed to be normal to the beam. There is a focusing effect at the end which depends on the angle of these faces.
4. QUADRUPOLE MAGNETS AND FOCUSING

The principal focusing element in a modern synchrotron is a quadrupole magnet. The poles of a quadrupole alternate and the field shape (Fig. 6) is such that it is zero on the axis of the device but its strength rises linearly with distance from the axis. This can be seen from a superficial examination of Fig. 6, remembering the product of field and length of a field line joining the poles is a constant. Symmetry tells us that the field is vertical in the median plane (and purely horizontal in the vertical plane of asymmetry). The field must be downwards on the left of the axis if it is upwards on the right.

This last observation ensures that the horizontal focusing force, evB_z', has an inward direction on both sides and, like the restoring force of a spring, rises linearly with displacement, x. The strength of the quadrupole is characterized by its gradient dB_z/dx, normalized with respect to magnetic rigidity:

$$ k = \frac{1}{(B_0)} \frac{d B_z}{d x} \quad \ldots \quad (11) $$

The angular deflection given to a particle passing through a short quadrupole of length L and strength k at a displacement x is therefore:

$$ \Delta x' = k k \cdot x \quad \ldots \quad (12) $$

If we compare this with a converging lens,

$$ \Delta x' = x/f \quad \ldots \quad (13) $$
we see that the focal length of such a quadrupole is:

\[ f = \frac{1}{k^2} \]  

(14)

The particular polarity of quadrupole shown in Fig. 6 would focus positive particles coming out of the paper, or negative particles going into the paper in the horizontal plane. A closer examination reveals that such a quadrupole is defocusing in the vertical plane and deflects particles with a vertical displacement away from the axis. We shall see later that in spite of this seemingly damming characteristic a FODO pattern of alternating gradient has a net focusing effect in both planes.

Fig. 6 Components of field and force in a magnetic quadrupole. Positive ions approach the reader on paths parallel to the y axis (from Cyclic Particle Accelerators by J.J. Livingood).

5. THE GUTTER ANALOGY OF FOCUSING

It is important to start with an almost tangible concept of focusing. If we are prepared to ignore vertical defocusing for a moment and consider an infinitely long quadrupole we find that a particle oscillates within it exactly like a small sphere rolling down a slightly inclined gutter with constant speed. Figure 7 shows two views of this motion and from the right hand view we recognize the motion as a sine wave: note too that the sphere makes four complete oscillations along the gutter. In accelerator terms its motion has a wave number \( Q = 4 \).
Figure 7  Two views of a sphere rolling down a gutter as it is focused by the walls.

Now extend this analogy by bending the gutter into a circle rather like the brim of a hat. Suppose we provide the necessary instrumentation to measure the displacement of the sphere from the centre of the gutter each time it passes a given mark on the brim. We also have a means to measure its transverse velocity which, with the aid of a microcomputer, we might convert into its divergence angle:

\[
\frac{dx}{ds} = \frac{v_1}{v_\parallel}
\]

Suppose we make the hat out of a slightly different length of gutter than shown so that Q is not an integer. If we plot a diagram of \(x'\) against \(x\) we can plot a point for each arrival of the sphere. This is called a phase space diagram. The sphere may have a large transverse velocity as it crosses the axis of the gutter or it might have almost zero transverse velocity as it reaches its maximum displacement.

The locus will be an ellipse (Fig. 8) and the betatron phase will advance by \(Q\) revolutions each time the particle returns. Of course, only the fractional part of \(Q\) may be deduced.

The time has come to define some of the beam dynamical quantities more rigorously. The area of the ellipse is a measure of how far away from an ideal trajectory the particle is:

\[
\text{Area} = \pi c \ [\text{mm.mrad}]. \quad (15)
\]
The maximum amplitude, or major axis, of the ellipse is defined:

\[ x = \sqrt{e\beta}, \]  

so that to satisfy equation 15:

\[ x' = \epsilon/\beta. \]  

The quantity \(\beta\) is a property of the focusing system, not the beam, and is the function plotted in Fig. 2. In an alternating gradient focusing system such as Fig. 1 the brim of the hat will vary its width and curvature around the crown and \(\beta\) will follow this variation in some way. Note that the aspect ratio of the ellipse is just \(\beta\).

6. **ALTERNATING GRADIENT FOCUSING**

Fig. 9 Ray diagrams to show the existence of a contained trajectory in an alternating gradient optical system. The upper diagram shows the horizontal motion and lower shows the vertical.
The principle is best illustrated in Fig. 9 which shows an optical system in which each lens is concave in one plane while convex in the other. It is possible, even with lenses of equal strength to find a ray which is always on axis at the D lenses in the horizontal plane and therefore only sees the F lenses. The spacing of the lenses would have to be 2f. If the ray is also central in the lenses which are vertically defocusing the same condition will apply simultaneously in the vertical plane. At least for one particular particle or ray the trajectory will be contained.

Now let us return to the FODO cell structure. We shall see that by suitable choice of lens strength, the function \( \beta(s) \), which follows the peak excursion of a beam, can be made to be periodic in such a way that it is large at all F quadrupoles and small at all D quadrupoles. Symmetry will ensure this is true also in the vertical plane. Particles oscillating within this envelope will always tend to be further off axis in F quadrupoles than in D quadrupoles and there will therefore be a net focusing action. Earlier we said that \( \beta \) is the aspect ratio of the phase space ellipse. At F quadrupoles the ellipse will be squat and at D quadrupoles it will be tall.

![Diagram](image)

**Fig. 10** The paths of particles within a FODO lattice are within the envelope of betatron motion and, like the rays of Fig. 8, are always closer in the D quadrupoles so they receive a net focusing effect. The phase space ellipse is tall and narrow at the D lens where the beam has a large divergence spread.

7. **The Equation of Motion**

The motion of the particle in the transverse direction is described by Hill's Equation. In the vertical plane, this is written:

\[
2'' + k(s)z = 0, \\
k(s) = \frac{1}{b_p} \frac{dB(s)}{dx}. 
\]

This is the famous Hill's Equation, a linear equation with periodic coefficient, \( k(s) \), which is the distribution of focusing strength around the ring.
In the horizontal plane, we must include a term for the curvature ρ:

\[ x'' + \left( \frac{1}{\rho(s)^2} - k(s) \right) x = 0. \]  \hspace{1cm} (19)

The equation is reminiscent of simple harmonic motion but with a restoring constant k(s) which varies around the accelerator. We assume that k(s) is periodic on the scale of one turn of the ring or some smaller unit, a cell or period, from which the ring is built.

The solution of Hill's Equation is not unlike simple harmonic motion:

\[ y = \sqrt{c} \cdot \sqrt{\beta(s)} \cos [\varphi(s) + \lambda], \]  \hspace{1cm} (20)

where \( y \) is a general transverse coordinate (either x or z),
\( \sqrt{c} \) is the beam's contribution to the amplitude (\( c = \text{emittance} \)),
\( \sqrt{\beta(s)} \) is a periodic function which describes the modulation of the amplitude due to the structure of the ring.

Note that the envelope of the motion of a beam of particles with random starting phase, \( \lambda \), will be the function:

\[ \sqrt{c\beta(s)}. \]

The term \( \cos[\varphi(s) + \lambda] \) describes the oscillation of an individual particle within this envelope. The phase must satisfy the condition:

\[ \varphi(s) = \int \frac{ds}{\beta(s)}, \]  \hspace{1cm} (21)

or Eq. (17) will not satisfy the equation of motion, Eq. (19).

8. **Floquet Coordinates**

In this section we shall consider imperfections and to do this we introduce a transformation into Floquet coordinates. Earlier we derived Hill's Equation:

\[ \frac{d^2y}{ds^2} + k(s)y = \frac{\Delta B(s)}{(B_0)} = F(s) \]  \hspace{1cm} (22)

Here we have included a driving term on the right hand side which represents a perturbation due to a field error.

The solutions of Hill's Equation are:
\[
y = \sqrt{\epsilon} \beta(s) \sin(\varphi(s) + \lambda),
\]
\[
y' = \sqrt{\epsilon} \beta(s) \left[ \cos(\varphi(s) + \lambda) - \alpha \sin(\varphi(s) + \lambda) \right],
\]
\[
\varphi(s) = \int ds / \beta;
\]
which form an invariant:
\[
Y y^2 + 2 \alpha yy' + \beta y'^2 = \epsilon.
\]  \(24\)

The equation and its solutions are so reminiscent of a forced harmonic oscillator that it is tempting to seek a transformation which removes the \(s\) dependence of \(k(s), \beta(s)\) and \(\varphi(s)\). Such a transformation, named after Floquet, involves a change to new variables:
\[
\eta = \frac{x}{\sqrt[3]{\beta(s)}}, \quad \psi = \int \frac{ds}{Q \beta(s)}.
\]  \(25\)

Substitution in Eq. (22) yields:
\[
\frac{d^2 \eta}{d \psi^2} + Q^2 \eta = g(\psi) = Q^2 \beta^{3/2}(s) f(s) .
\]  \(26\)

We have reduced the problem to that of a forced harmonic oscillator at the expense of a little physical reality. Transverse displacements are scaled by the local beta function and \(\psi\) does not correspond to azimuth. However, \(\psi\), like azimuth, \(\theta\), advances \(2\pi\) per turn and is equal to azimuth at the quadrupoles of a FODO lattice. We shall make use of these new Floquet coordinates in the next section where the effect of errors is often treated as a forcing term. Note that errors are weighted with \(\beta(s)^{3/2}\) on the right hand side.

9. ORBIT DISTORTION FROM ONE KICK

The size of the vacuum chamber and magnet aperture can become uneconomically large if field imperfections which can come from misalignments are allowed to distort the perfection of the central synchronous orbit which does not hit the walls. In extreme cases there may be no single closed orbit. To be able to calculate and correct this effect of orbit distortion we must first understand the effect of adding a short dipole of integrated field \(\delta (B\delta)\) somewhere in the machine. Note that in the jargon of the accelerator designer the effect of such a dipole is a "kick" though the dipole deflects by the same angle on each turn. A "bump" might be a better word but this is reserved for another use.
The effect of the kick is best visualized in the Floquet coordinates of a machine unwound as in Fig. 11. We know the orbit must be a solution of:

$$\frac{d^2 \eta}{d\psi^2} + Q^2 \eta = 0$$ (27)

and if we place the kick at $\psi = \pi$, symmetry tells us the solution must start as:

$$\eta = \eta_o \cos Q\psi$$ (28)

with a maximum diametrically opposite the kick.

Fig. 11 A closed orbit for one turn of a synchrotron plotted in Floquet coordinates. The deformation is due to a single dipole at the centre of the diagram.

At the location of the kick the slope of the orbit:

$$- \eta_o Q \sin Q$$ (29)

changes sign and this change must therefore equal half the kick angle:

$$\delta(Bk)/\beta_p$$ (30)

transformed into $d\eta/d\psi$ coordinates.

To perform this transformation, we remember:

$$\frac{d\psi}{ds} = \frac{1}{QB(s)}, \quad \frac{dx}{ds} = \beta(s), \quad \frac{d\eta}{ds} = \beta(s)$$ (31)

and can deduce that at the kick:

$$x' = \frac{dx}{ds} = \beta_k \frac{d\psi}{ds} = - \frac{\eta_o}{\sqrt{B_k}} \sin Q = \frac{\delta(Bk)}{2\beta_p}$$ (32)
Thus, the amplitude of the distortion is:

\[
\eta_o = \frac{v\beta_k}{2|\sin \psi|} \frac{\Delta(B\|$ 

and elsewhere the physical distortion scales with \(\beta(s)\) modulated as \(\cos \psi(s)\):

\[
x = \left[ \frac{v\beta(s)\beta_k}{2|\sin \psi|} \cdot \frac{\Delta(B\|$ 

10. SOURCES OF CLOSED ORBIT DISTORTION

Of course, no one in their right mind would place such a dipole magnet in an otherwise perfect machine. In practice there are a large number of uncorrelated kicks due to misalignments and errors in the magnets of the ring.

<table>
<thead>
<tr>
<th>Type of element</th>
<th>Source of kick</th>
<th>rms value</th>
<th>(\frac{\Delta B/B_p}{B_p})</th>
<th>Plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient magnet</td>
<td>Displacement</td>
<td>(\Delta y)</td>
<td>(\Delta y)</td>
<td>x,z</td>
</tr>
<tr>
<td>Bending magnet</td>
<td>Tilt</td>
<td>(\Delta \theta_p)</td>
<td>(\theta \Delta \theta_p)</td>
<td>z</td>
</tr>
<tr>
<td>(Bending angle = 0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bending magnet</td>
<td>Field error</td>
<td>(\Delta B/B)</td>
<td>(\theta \Delta B/B)</td>
<td>a</td>
</tr>
<tr>
<td>Straight sections</td>
<td>Stray field</td>
<td>(\Delta B_p)</td>
<td>(\theta [\Delta B_p/(\rho B)])</td>
<td>x,z</td>
</tr>
</tbody>
</table>

Table 1 shows some of the common sources of such kicks. Apart from obvious field errors from geometry and remanent field variations among the magnets, there are two sources of distortion related to alignment; the displacement of a quadrupole produces a dipole kick for a particle on the central orbit, and a tilted bending magnet will deflect vertically. The fourth column of Table 1 gives the angular deflection from which Eq. (35) gives:

\[
y_1 = \frac{v\beta(s)\beta}{2|\sin \psi|} \cdot \Delta y_1 \cdot \cos \psi(s)+\lambda_1 \]

where \(\lambda\) is a random phase.

Taking the rms of all these uncorrelated distortions (including the phase \(\lambda\)) gives:

\[
y_{rms} = \frac{\beta(s)}{2v \sqrt{2|\sin \psi|}} \left[ \sum \beta_1 (\Delta y_1)^2 \right]^{1/2}.
\]
In making allowances for the fact that chance may not be on the side of the accelerator designer, it is traditional to quote twice this value as an allowance which ensures that in more than 90% cases a large aperture allowance is often considered unnecessary luxury. Proven techniques exist to apply correction to obtain the first turn and to the circulating beam orbit.

11. FOURIER ANALYSIS OF ORBIT DISTORTION

Certain frequencies in the azimuthal pattern of errors contribute most to the distortion. This response of the accelerator to these "resonant" frequencies is an important concept which reappears in the theory of gradient errors and later in treating nonlinear effects. Again, we use Floquet’s transformed betatron motion to see the driving effect of the error pattern, $F(s)$. If:

$$F(s) = \frac{\Delta B(s)}{B_0}$$

then the motion is described by the differential equation:

$$\frac{d^2 \eta}{d\psi^2} + Q^2 \eta = g(\psi) = Q^2 \beta^{1/2} F(s) .$$

(38)

We can think of $g(\psi)$ as the new driving term and Fourier analyse it:

$$g(\psi) = Q^2 \sum_{k} f_k e^{ik\psi}$$

(39)

where

$$f_k = \frac{1}{2\pi} \int_{0}^{2\pi} B^{1/2}(\psi) F(\psi) e^{ik\psi} d\psi .$$

(40)

since

$$d\psi = \frac{ds}{QB}$$

(41)

$$f_k = \frac{1}{2\pi Q} \int_{0}^{2\pi R} \beta^{1/2}(s) F(s) e^{ik\psi(s)} ds .$$

(42)

The differential equation now becomes:

$$\frac{d^2 \eta}{d\psi^2} + Q^2 \eta = \sum_{k=1}^{\infty} Q^2 f_k e^{ik\psi}$$

(43)
and has the solution:

$$\eta = \frac{v^2}{k=1} \frac{f_k e^{ik\psi}}{Q^2 - k^2}$$  \hspace{1cm} (44)

We see that the machine will be particularly sensitive to Fourier components where $k$ is close to the $Q$ value of the ring. An example of this is to be seen in Fig. 12 which shows an uncorrected closed orbit in the FNAL main ring. Each bar is the position with respect to the centre line of one of the 100 or so position pick-ups around the ring. The ring has been uncoiled into a straight line. By counting the peaks one can see that the response is rich in frequencies 19 and 20. This is not surprising since $Q = 19.25$. Even if the error spectrum has the characteristics of white noise (all $f_k$'s equal) the denominator of Eq. (44) will be much smaller for $k = 19$ or 20.

![Image of beam position monitors at each quadrupole of the FNAL Main Ring](image)

**Fig. 12** Beam position monitors at each quadrupole of the FNAL Main Ring display the shape of the orbit before correction (200 units is the width of the vacuum chamber)

We could turn this to our advantage. We could correct such an orbit error by installing 38 small dipoles of alternating sign around the ring to cancel the 19th harmonic. A more economic way would be to simply install one of each sign at opposite ends of a diameter. These dipoles, effective and function perturbations, generate all odd harmonics in the driving function, with equal strength but only the 19th will have a major effect on the orbit distortion and can be used to cancel the 19th harmonic pattern in the distortion (Fig. 13).

![Diagram showing the sign of correction dipoles necessary to excite or compensate even or odd Fourier components of distortion around the ring of a synchrotron](image)

**Fig. 13** Diagram showing the sign of correction dipoles necessary to excite or compensate even or odd Fourier components of distortion around the ring of a synchrotron
There are few machines where one would be satisfied just to correct these leading harmonics since the sum of the residue is still about 30% of the distortion before correction. Nevertheless, this principle of harmonic correction is sometimes used to correct a machine without realigning all the magnets. Displacing a quadrupole produces a dipole field on the central orbit. Displacing four dipoles at quadrants will produce the desired pattern.

12. LOCAL ORBIT CORRECTION

It is often necessary to steer the beam from side to side or up and down in the vacuum chamber, either to centre it or, to move it close to some injection or extraction device. Clearly one wants to localize this distortion as much as possible.

The most convenient way of making such a bump, the half-wave bump, is shown in real space in Fig. 14 and in phase space in Fig. 15 where $\beta_0$ is the $\beta$ at the peak and $\beta_k$ is the $\beta$ at the kick.

![Fig. 14 A triad bump of dipoles producing a local orbit distortion.]

![Fig. 15 Triad bump seen in phase space.](image)

The change in divergence at $C$ is produced by small DC bending magnets which bend through an angle $\delta_1$, $\delta_2$. If one is fortunate enough to find two places separated by a betatron phase advance of $\pi$ the bump will be a simple sine wave of amplitude:

$$\Delta y = \beta_0 \beta_k \delta_1.$$  \hspace{1cm} (45)

If both dipoles are at the same beta value their strength can be equal.

This bump has the characteristic that it leaves the beam centred elsewhere in the machine. However, if we are forced to some phase difference other than $\pi$ the bump must contain a third element to make a cusp and satisfy the boundary conditions $(x, x') = 0$ at entrance and exit. Such a bump is called a triad.
Triads may be superimposed, each calculated to restore the local measurement of distortion (the bar in Fig. 11) to zero. They will overlap, but when added algebraically and applied to a set of dipoles as numerous as the measurements in Fig. 11, they can correct distortion to a small percentage of its initial value. Such a scheme is frequently used at injection in a synchrotron where the strength of dipoles needed is rather small and where the beam is large. In some machines triad bumps are made by displacing quadrupoles from their ideal position.

13. CERN PROTON SYNCHROTRON (CPS) [1954-1959]

Fig. 16: Basic survey configuration for the 28 GeV Proton Synchrotron
The problem of aligning the magnets of a circular particle accelerator was first encountered in 1954, following the decision to build a 28 GeV proton synchrotron in Switzerland, at Meyrin. The selection of survey instruments, and therefore the techniques available for meeting the tolerances required were fairly limited. At that time, the logical consequence was to use traditional geodetic methods in order to work out a possible approach to this problem. The estimated values of the achievable tolerances were introduced in the computations for the closed orbit, in order to define the size of the magnet aperture. This aperture determined the dimensions of the magnets and therefore the dimensions of the tunnel and, in part, the overall cost of the accelerator.

What tools did the surveyor have at his disposal in 1954? Computers had only just made an appearance and least squares adjustment had still to be made on calculating machines. The first direct consequence was that the geodetic reference figure had to be as simple as possible to avoid major adjustments which were always time-consuming with calculating machines. Figure 16 illustrates the octagon \((P_1P_2\ldots)\) with a central point \((P_0)\), which was the basic network of the synchrotron's geodetic system.

It also shows the elements \((r, \omega, \alpha)\) had to be measured so that the magnets were properly aligned on the calculated orbit. The second consequence was that, as calculations had to remain simple, maximum accuracy had to be achieved in the measurements.

At that time, geodesy relied essentially on angle measurement using a theodolite (Fig. 17). A better accuracy could be obtained by using forced centring when locating the various instruments, and by building the sighting targets with the same precision as the reference holes of the forced centring system.

Distance measurement was the second method available. In 1954, length measurement by electromagnetic methods was just making its appearance, with instruments such as the geodimeter and tellurometer. The original model of the geodimeter was tested in 1947 with light modulated at 10 MHz. The standard deviation of this instrument \((5 \text{ mm} + 10^{-6} \text{ of the distance})\) makes it unsuitable for very short distances. The same also applies to the tellurometer, which is based on the use of transmitters and receivers operating on very short wavelengths (3000 MHz and above). The only available tool was, therefore, the invariant wire (or tape) used for measuring the bases at the end of primary triangulations (Fig. 18).

The third of the three tools was the Wild N3 bubble level used with Taylor spheres placed on the bevel edges of the sockets. This method offers a standard deviation of 0.3 mm/km.

To correct the position of the magnets according to the results of measurement, they were installed on three jacks. This enabled movement in three directions (translation) and movements through three angles (rotation), corresponding to the six degrees of freedom of all non-deformable bodies.
Fig. 17 Invar wire reading system
(microscope, theodolite &
target on the same axis)

Fig. 18 Knife-edge pulley

Over a ten-year period, repeated measurements of the monument positions in the PS reference figure showed the long-term consistency taking into account movement of the molasse, movement of the monuments, and random errors in measurements to be 0.1 mm per 100 m per year. A measurement carried out independently of any triangulation confirmed this figure. From 24 August 1965 to 13 February 1968 a pair of Narussi horizontal pendulums were mounted on the central pillar $P_0$ anchored in the molasse 10 m below ground level. These instruments measure the variations of their support in relation to the direction of the vertical, and therefore of the movement of the vertical axis of the 10 m pillar. Fig. 19 shows that the over-all movement of the molasse, and of the monument itself, scarcely exceeded 0.15 mm in a North/South direction over a period of two and a half years.
The requirements of the CERN Survey Group have been such that more than 10 km of invar wire have been used to date. When the PS was being built, wires of 81 and 105 m were used. It was necessary to use frictionless pulleys, because an excess tension of 1 g on wires of such length resulted in variations δl of the order of 10 μm. The development of pulleys in which the ball-bearings were replaced by balanced knife-edges enabled the excess tension to be reduced to 0.002 g. During microscopic measurements of the scales fixed at each end of the wires, it was noted that they were constantly becoming elongated under a tension of 196.18 N (Newton). As we were the first to use wires longer than 100 m with pulleys that were virtually frictionless, we were naturally in a position to detect non-elastic elongations. The Bureau International des Poids et Mesures in Sèvres (Paris) was immediately informed, and as a result carried out a series of tests on a 24 m invar wire which subsequently confirmed our results.

Figure 20 shows the elongation of invar wire and tape when subjected to tension for a prolonged period (Proceedings of the 1960, 1962, 1963 Meetings of the Bureau International des Poids et Mesures, Paris). Figure 21 shows the elongation of the invar wires as a function of time, Fig. 22 as a function of length.
Fig. 20 Elongation of invar wire and tape as a function of the number of years of use

Fig. 21 Elongation of invar wire as a function of time
Fig. 22  Elongation of invar wire as a function of its length

These elongations were detected because CERN possesses, for standardization of its wires, a 64 m bench constructed by the Société Genevoise d'Instruments de Physique and located, between 1959 and 1969, in one of the radial tunnels of the CPS. After 1969 (Fig. 23) the bench was installed in a specially equipped tunnel near the ISR. Only the 4 m rule and the microscopes were retained. The rest of the equipment was modernized. Until the advent of laser interferometers (Fig. 24 & 25), this standardization bench was the only one in the world capable of calibrating invar wires of any length between 0.40 and 64 m.

Fig. 23  Calibration bench
Fig. 24  Interferometer and display

Fig. 25  Corner cube reflector mounted at the end of a 4 m rule, with microscope to read the rule
14. INTERSECTING STORAGE RING (ISR) [1966-1971]

The use of computers for geodetic calculations brought about a fundamental change in the philosophy of measurement and, to a certain extent, in accelerator design. The PS and the machines whose design was based on it were the last to have radial tunnels. The studies carried out at that time for the ISR and the SPS already illustrated a basic survey system based on a chain of braced quadrilaterals. In the case of the PS, the centre is real and physically exists, whereas in the case of the ISR and SPS the centre is virtual. It was therefore necessary to find a way of superimposing two sets of successive measurements. Furthermore, the elimination of all angle measurement and the sole use of distance measurement modified the observation equations in the computations, and obliged us to reconsider the entire problem of trilateration adjustments. The studies were carried out simultaneously for the ISR and the SPS.

The ISR lattice consists of 48 magnet periods divided into four super-periods, whereas the complete geodetic figure comprises 32 quadrilaterals. Figure 26 shows one of the octants of the overall framework.

![Fig. 26 Braced quadrilateral survey system of one octant of the ISR](image)

Starting with one monument as reference, the error ellipses were calculated for each monument up to those diametrically opposite (Fig. 27). In this way, it is possible to judge the distortion of the reference figure. The value of the semi-major axis of the largest error ellipses is 0.2 mm for a probability of 0.40, and 0.6 mm for a probability of 0.99.

This means that from one set of alignments to the next, the variation δr in the radius of the orbit is of the order of 0.08 mm in relation to the theoretical orbit, but its position in space has a 99% chance of being inside a circle of errors of 1.2 mm in diameter. The precision in the relative position of magnets in relation to another one, inside two adjacent quadrilaterals, is better than 0.1 mm.

To achieve this result, it was decided to eliminate systematically all angle measurement. The distances were measured with invar wire, using the "distinvar" for the reference geodetic figure and for positioning the magnets. Only two days were needed to
measure the 32 quadrilaterals of the reference figure and to calibrate the invar wires before and after measurement in the ring.

![Diagram of error ellipses](image)

**Scale of error ellipses**

**Fig. 27** Ellipses of errors computed when taking monument 11 as the starting point and direction 11-51 as orientation vector. They show the way in which errors accumulate around the Storage Rings' circumference.


On 27 January 1971, beam collisions were produced for the first time in the ISR. On 19 February the construction of the SPS was approved by the majority of Member States and immediately there was a new challenge for the CERN Applied Geodesy Group. The PS and the ISR were surface structures built by the "cut and fill" method with prefabricated concrete tunnels. Unlike the NAL site (Batavia, USA), which is horizontal, flat and free from vegetation, woods or forest, the SPS tunnel could only be located underground in the Chattian molasse plateau of the Geneva Basin and be bored with a tunnelling machine.
Three major problems were successively encountered during construction of the SPS machine:

- the gyroscopic traverse for guiding the boring machine,
- the installation of magnet supports before the final completion of the tunnel, i.e. without a complete underground geodetic figure,
- the final installation of quadrupole and dipole magnets and auxiliary components within the required tolerances.

Each of these three problems was a challenge to the geodesists, and could be successfully solved only by making a series of extremely precise measurements, based on a highly accurate geodetic network on the surface.

The special feature of the SPS tunnel was that the work started by the digging of six shafts located at regular intervals around the 6,911 m circumference. The main tunnel was thus divided into sextants representing the six super-periods of the accelerator. When making the underground traverse to guide the boring machine, it was not possible to measure a geodetic bearing at each end of one sextant of the accelerator. The surface coordinates could be transferred down each shaft to the tunnel, but not accurate bearings.

![Diagram showing distance and alignment measurements](image)

---

*Fig. 28 Geodetic reference figure of the SPS*

The geodetic reference figure of the SPS (Fig. 28) is a modified version of that of the ISR. In the latter, because of the relatively favourable ratio between the width and length of a braced quadrilateral (1 to 2.5) distance measurements alone were used for the ISR framework consisting of 32 braced quadrilaterals. In the SPS, this ratio is approximately 1 to 10. To keep the deviation of the radius within the tolerances, the reference figure had to be stiffened by measuring the offsets of the monuments of three consecutive quadrilaterals with nylon wire equipment. As the SPS is a separated-function machine, the components which must be accurately positioned are the quadrupoles situated at one end of each semi-period of the accelerator's lattice, every 32 m. In this case, therefore, it was necessary to adapt the metrology system to the periodicity of the quadrupoles. Opposite each quadrupole, two brackets were fixed, one on the outer wall, the other on the inner wall. In addition to the distance measurement of each braced quadrilateral, nylon wire alignment inside three consecutive quadrilaterals ensured the precise shape of the geodetic framework in each sextant.
To make an a posteriori precision estimate it is worthwhile to recall the notion of the redundancy coefficient:

$$\tau = \frac{N_{\text{observation}}}{n_{\text{unknowns}}}$$

The value of $\tau$ for the ISR is 1.28, and 1.93 for the SPS.

In the ISR, the statistical study of the residuals after adjustment gives a standard deviation $\sigma = 18 \, \mu m$, which with $\tau = 1.28$ corresponds to an estimated accuracy of 34 $\mu m$.

In the SPS, the same study of length measurements gave a standard deviation $\sigma = 33 \, \mu m$, which, with $\tau = 1.93$, provides also an estimated accuracy of 34 $\mu m$.

These results, obtained from tens of thousands of measurements made with the distinvar over 15 years, show the very great homogeneity and reliability of this process in CERN's specific operating conditions. The above results obtained with this type of metrology are leading to a limit of precision of between 30 and 40 $\mu m$ for distances ranging up to 50 m.

On the 2 January 1976, the coordinates of the brackets used for the underground survey chain were calculated. Without having to move any support already fixed on the foundation raft, it was possible to perform the fine adjustment of all the magnets and beam instrumentation, and on the 3 May 1976, protons were circulating in the SPS for the first time.

The fact that the protons were able to circulate round the machine without any correction to the orbit showed that the degree of accuracy required from the geodesists at the design stage had been kept. In fact, we already knew this from a detailed analysis of the survey measurements and radial smoothing results. This analysis was based on the rms sum of all the random deviations on the radial position of the quadrupoles.

For the six points obtained from the surface network, the radial standard deviation was 1.3 $\mu m$. With regard to the geodetic network of the brackets, which used these points as reference, an examination of the ellipses shows that, after overall adjustment of the circumference, the radial standard deviations are basically similar for each sextant. They steadily increase from the reference pillar, reach a maximum at the middle of the sextant, after which they decrease and finally disappear at the next reference pillar; in each sextant, the pattern of the error ellipses is that of a very long cigar. For the ultimate measurement made, this maximum represents a standard deviation of 1.2 $\mu m$.

The complete set of radial uncertainties in the surface network and underground framework provides, for each sextant, a cigar-shaped envelope. This envelope provides, among other things, a picture of the possible variations in the accelerator radius, which has an uncertainty varying from 1.3 $\mu m$ at the reference points to 2.5 $\mu m$ in the middle of
each sextant. However such large mid-sextant values have no direct influence on the correct operation of the accelerator provided that the relative radial error $dr$ remains within the narrow limit imposed by the machine's parameters ($0.15$ mm), thus compelling the curve of the successive quadrupole positions to remain continuous, with discrete increments.

Fig. 29 Combination of random errors and curve obtained after quadrupole smoothing

Its general appearance will be that of a curve of the second or third degree. This curve is contained in a random manner inside the envelope shown in Fig. 29 and constitutes an approximation of the theoretical line among an infinite number of possible locations within this envelope.

Smoothing was done by measuring, with a nylon wire device, the local curvature of successive sets of three adjacent quadrupoles. (Fig. 30). A comparison of this curve with the theoretical curve gave local information on the positions of the quadrupoles in relation to a mean curve, for which the absolute position did not need to be known. This was done, in the least squares adjustment, by solving the observation equations with the double condition $v^2$ and $dr^2$ minimum, $v$ being the residuals of the measurements and $dr$ the deviation in relation to the curve.

This measurement was made for checking purposes only, but it provided a direct estimate of the relative residual errors. On the basis of the smoothing results, the calculated value of the parameter $dr$ (the radial deviation of a quadrupole in relation to the two adjacent quadrupoles) was $0.08$ mm for the overall circumference of the synchrotron.
Levelling gave rise to the usual problems. The overall accuracy obtained is 0.5 mm/km, and a relative accuracy of 0.10 mm was attained for one quadrupole in relation to the two adjacent quadrupoles.

Fig. 30 Smoothing principle

16. INSTRUMENTATION

To obtain such an accuracy it is obvious that the entire geodetic system uses precision-type reference sockets for centring the measuring instruments (Fig. 31). It is possible to correct the verticality of the sockets at any time.

Fig. 31: Precision-type reference socket

To facilitate and speed up the precise positioning of the quadrupoles, special jacks with polyurethane pads were developed by the CERN Survey Group (Fig. 32).

Among the new devices developed and used at CERN for applied geodesy, it seems worthwhile to describe four of the most typical belonging to the special CERN Survey instrumentation.
17. LENGTH MEASUREMENT DEVICES

The self-aligning reflector makes it possible to carry out distance measurements with a laser interferometer without using mechanical guidance. In good conditions, measurements can be made up to 60 m with an error less than 0.01 mm (Fig. 33).

![Fig. 33 Laser interferometer](image)

The beam from the laser is split in two by the interferometer prism, one part is reflected back to the prism by the fixed reflector and the other by the mobile reflector. The two beams are recombined and reach the receiver, where the phase difference between them is detected. This difference is proportional to the distance over which the mobile reflector has been moved. The receiver signal is fed to the display unit, which shows the measured distance in mm, 1/4 wavelengths or inches.
The reflector is mounted on a carriage, which can be moved in the horizontal and vertical directions by two servo-motors. The servo-motors are controlled by a servo-amplifier, which receives the signal from the four-quadrant photocell. This forms a servo-loop, which keeps the returned laser beam centred on the photocell by moving the reflector. For the vertical adjustment of the XY-support at the reference sockets, there is a bubble level and two adjustment screws.

The Distinvar (Fig. 34), developed by the CERN Applied Geodesy Group in 1962, has been used for thousands of measurements.

![Distinvar Diagram](image)

**Fig. 34** Distinvar

The Distinvar consists of three parts: the 1.65 mm geodetic invar wire with its plug-in fittings, the fixed reference point and the measuring head. The latter two components have a male cylinder for precise centring inside the 30 mm diameter sockets. These sockets must be adjusted for verticality and serve as the origins for the length measurements.

The equipment is a precision balance fitted on a carriage. The balance serves to apply a constant horizontal tension on the invar wire via a system of levers attached to a fixed weight (1.5 kg). The carriage seeks the centre of equilibrium of the instrument. This centre of equilibrium is determined by a system for detecting the position of the balance-beam. According to the information provided by the detection system, the carriage is automatically moved by a reversible motor via a micrometer screw until it reaches the centre of equilibrium. A mechanical counter integral with the screw displays the longitudinal position of the carriage. The length measurement is obtained by subtracting the dial reading from that obtained during the calibration of the wire.
The total travel of the carriage is 50 mm. Depending on the invar wires used, the Distinvar can measure lengths between 0.40 and 50 m. repeatability of the reading is 0.01 mm. Measurement accuracy is not a function of the length of the wire but of the sensitivity of the balance. Thousands of measurements made with the Distinvar over many years show the standard deviation to be 0.03 mm.

18. OFFSET MEASUREMENT DEVICES

The laser alignment system with tracking receiver is an active target system using a laser reference beam. It is used at CERN for offset measurements which arise when establishing the geodetic framework of accelerators. This system is particularly useful in turbulent air conditions where the stretched wire method described later is unsuitable. It can measure up to 600 mm over a distance of 100 m with a resolution of 0.01 mm.

The apparatus consists of a laser source, with its special optical system, and a differential photo-cell receiver.

![Diagram of laser alignment system with tracking receiver](image)

**Fig. 35** Laser alignment system with tracking receiver

The laser is a commercial He-Ne unit of 1 mW beam power, followed by an optical system which reduces the divergence by a factor of ten and increases the beam diameter by the same factor. The laser beam increases from 11 mm at the output of the optical system to
only 14 mm at 100 m. A 25 mm diameter diaphragm is located at the focus of the system. The diaphragm serves as a spatial filter and cleans up the beam; it eliminates non-parallel rays and assures a Gaussian distribution of light intensity across the beam diameter. The transmitter is mounted on a adjustable plate which allows the beam to be steered onto the receiver. Once adjusted, the plate can be locked (Fig. 35).

The carriage of the photo-electric receiver is driven by a precision screw perpendicular to the laser beam, and is servo-controlled to remain centred on the beam. The position of the carriage along the screw is measured by an encoder which feeds 100 pulses per millimetre to an up-down counter. This counter stores the instantaneous position of the carriage relative to a “zero”, established by means of a reference cursor attached to the carriage and passing between a fixed LED and a photo-cell pair, resetting the counter to zero at the “zero” position. Instantaneous positions are read 10 times per second via a sampling gate and added together in a summing counter. A manual selector allows the choice of 10, 100 or 1000 samples.

The nylon wire alignment system is designed to replace angular observations by precise measurement of the shortest distance (offset) from one point to a straight line used as a reference. This line consists of a taut nylon wire. The vectors thus obtained are introduced into the planimetric computations with other measurements.

Once the wire has been stretched between the extreme points, the measurement system is installed on the intermediate point to be determined. The offset vector to the nylon wire is then measured. The mobile detector passes a reference point, thus resetting the distance counter, and then continues its travel until it locates the wire. The detector then centres under the wire and the value of the measured offset is displayed (Fig. 36).

The method of operation consists of measuring the various offsets at least twice with different origins, in order to obtain redundant and independent observations. The calibration of the instrument is carried out by turning the detector system through
180°. The maximum measurable offset is 520 mm, the accuracy of the detector position is 0.01 mm. The automatic system has now been used intensively with a standard deviation of 0.05 mm over 100 m long chords. The observation conditions must be stabilized to ensure that even very slight draughts do not introduce systematic errors into the measurements. In such conditions, better results are obtained by laser alignment.

19. CONCLUSION

By way of conclusion, it is useful to summarize the evolution of the Applied Geodesy Group from 1954 to 1982 and to emphasize the mile stones which have been passed.

For the PS, in 1954, the improvement of the first order geodetic instruments and the use of primitive computing machines then available inevitably had a strong influence on the methods used for the positioning of the magnets. The development of these devices, the understanding between geodesists and mechanics and the good integration of the Applied Geodesy Group in the CERN structures made it possible to carry out an absolute metrology.

In 1961, the quite fortuitous invention of the Distinvar completely transformed the measurement methods. This permitted the elimination of the construction of radial tunnels in accelerator buildings and the angle measurements in geodesy. The appearance of more sophisticated computers opportunely added to these developments and led to the success of the ISR metrology. Collaboration between geodesists and specialists in mechanics and electronics became more and more essential.

Since 1972, the surface geodesy required by the SPS dimensions has called for the use of E.D.M. instruments the accuracy of which was about $10^{-6}$ of the measured distance. The appearance of the laser interferometers made the calibrations easier. For the underground metrology, the development at CERN of the laser and nylon wire alignment measuring devices allowed the measurement of small angles to be replaced by offset measurements. Here, too, the success of work was shared equally between geodesists and mechanical and electronic engineers.

For LEP, the automation of the measuring devices will again modify the practices of the technicians. The positioning of the magnetic components will result more from the smoothing of the quadrupole curve than from the computation of their coordinates. The use of the interferometer with a self-aligning reflector will avoid repeated visits to the calibration base during the positioning operations.

In spite of all these successive developments, including the automation of the instruments, which increases the speed and reliability of the measurement, the role of the individual people remains preponderant. The geodesist must still remain very close to the reality of the quantity being measured. Whether automatic or not, the most important and time consuming factor in the use of the distinvar will always be the correct rolling and
unrolling of the wire. To achieve a precision of $10^{-7}$ with the Terrameter, i.e. 0.1 mm/km, will never be an easy task. The Terrameter is a Formula 1 car with its team of mechanical and electronic technicians. It is clear that the technology and equipment developed to meet the unusual CERN requirements could have a fruitful spin-off in all the fields where accuracy and speed of measurement are essential to the success of the work being carried out. At the human level, one of these spin-offs has been that many young surveyors and engineers have spent up to two years at CERN gaining a knowledge of, and experience in, this new technology with which they can profitably return to their home countries.

* * *

BIBLIOGRAPHY

PARAGRAPHS 2 TO 13


PARAGRAPHS 1 & 14 to 19


GYROSCOPE TECHNOLOGY, STATUS AND TRENDS

Wilhelm F. Caspary
Institut für Geodäsie, Universität der Bundeswehr München

ABSTRACT
After a sketch of history of gyroscope technology, the dynamic law of a gyroscope and its specialization to the suspended gyrocompass are outlined. The equations of motion of a north-seeking gyro are reviewed forming the base for observation techniques. Following a state-of-the-art report, the relations between the observables and the final azimuth are fully discussed and actions and precautions required for precise results are proposed. With modern instruments and observation methods it is presently possible to determine an azimuth with a standard deviation of 3" to 6" within less than one hour. The principles of non-rotating optical gyroscopes are presented and their potential as alternatives to conventional instruments is discussed.

1. INTRODUCTION

Gyroscopes as instruments for measuring azimuths have always had their place on the fringe of geodesy. In the early days of this technology it really was a venture to carry out observations, since the equipment was voluminous and heavy, the procedure time consuming and the accuracy of the results unsatisfactory. In addition to that, the average surveyor did not understand the way the gyroscope works. Not that the theory is difficult, but it is beyond every day experience. By reason of this situation gyro technology acquired the reputation of being an exotic subject hardly of use for surveying applications. This still exists although the scene has changed completely.

Modern mechanical gyro's are handy and provide the azimuth with an accuracy of a few seconds of arc in less than one hour. The dynamics of gyroscopes has been treated at length in numerous papers, admittedly sometimes more confusing than enlightening, but the interested surveyor can easily find suitable presentations of the theory. In spite of these significant advances the acceptance has remained on a low level.

Presently optical gyroscopes have attracted widespread attention as an alternative to mechanical gyro's in navigation platforms. Their performance has already reached a level which makes them rate as potential candidates for a completely new generation of north-seeking instruments.
2. HISTORICAL BACKGROUND

The dynamic behaviour of a spinning wheel had been known for a long time, when the French physicist L. FOUCAULT made his famous experiment in 1852. He suspended a gimballed rotor from an earth-fixed support using a thread, such that the axis of the rotor was forced to remain in a horizontal plane. The objective was to prove that the earth rotates and that therefore, caused by the torque due to the precession in the earth's gravitational field, the spin axis would turn towards true north. Unfortunately the experiment failed because it was impossible at that time to maintain the required angular velocity (LAUF 1963)). Nevertheless, this was the birth of a north-seeking instrument for which FOUCAULT coined the name gyroscope and which has become increasingly important for navigation and geodesy.

FOUCAULT's experiment turned the attention of many famous scientists of the last century to the theory of the gyroscope. Authors like CAYLEY, EULER, KELVIN, KLEIN, SOMMERFELD and GRAY made valuable contributions to the dynamics of this instrument. But it took until 1908 before working gyrocompasses were constructed by ANSCHÜTZ-KÄMPFE in Germany and SPERRY in the USA. These single-gyro compasses were designed for the navigation of vessels, but it turned out that they were too sensitive to the ship's rolling. Further research and experiments led to two- and three-gyro compasses showing much better performance. The final breakthrough was achieved by ANSCHÖTZ in 1922 with a floated spherical gyroscope meeting all needs for use at sea.

Paralleling this development first studies and experiments were conducted aiming at the use of gyroscopes for bore-hole and tunnel measurements. These efforts were supported and advocated by the mining surveying authorities in Germany. The first experimental instrument was constructed by SCHULER in 1921 and tested in coal mines. Already in 1926 an improved gyroscope was built conjointly by ANSCHÜTZ and BREITHAUPT and in 1936/37 ANSCHÜTZ presented the third generation. All these instruments remained on the prototype level and were not adopted by the surveying community, because the accuracy was unsatisfactory, the weight was too heavy and the observational procedures were too complicated.

In 1948 a new start was made at the Mining University Clausthal by RELLENSMANN. The following period is noted by a steady progress and by noteworthy contributions of different companies, having led to the current arsenal of gyroscopic instruments for azimuth determination. There are basically two types of instruments: the gyroscope attachments and the gyrotheodolites. In Western Europe mainly three companies offer gyrocompasses for precision observations on the civil market which are suitable for geodetic applications:

Institute of Mining Surveying (WBK), Bochum, FRG
Hungarian Optical Company (MOM), Budapest
Wild, Heerbrugg, Switzerland

The instruments which are best documented and probably of most widespread usage are the attachment GAK 1 of Wild, the gyrotheodolites G1-B1 and G1-B11 of MOM and the MW 50 (MW 50 a) and Gyromat of WBK.
3. THE FREE GYROSCOPE

The free gyroscope consists of a fast spinning rotor which is mounted in a set of gimbals such that it is decoupled from all rotations of the support. Under the assumption that the gimbal bearings are free from friction and that the instrument is balanced and strictly symmetrical so that the three axes coincide with the principle axes of inertia and intersect in the centrepoint of the rotor, the spin axis will indefinitely maintain its orientation in inertial space. To an observer on the earth the spin axis apparently moves on a cone with an axis parallel to the rotation axis of the earth. Figure 1 shows a gimbal mounted elementary (free) gyro.

![Diagram of gimbal mounted free gyroscope](image)

Figure 1  Schematic of a gimbal mounted free gyroscope (FABECK 1980)

If a torque is applied in the direction of the 2-axis then the angular momentum of the rotor changes its direction and the 3-axis begins a motion about the 1-axis which is known as precession of the gyro. The free gyro tries to align the spin axis with the direction of the applied torque.

In the English literature this type of gyroscope is usually denoted as a two degrees of freedom (TDF) gyro. It represents the basic concept of the gyros used to control the attitude of inertial platforms.

The suppression of one degree of freedom of the rotor relative to the frame yields the single degree of freedom (SDF) gyro. A special version of this type is the north-seeking gyroscope or gyrocompass. The gyroscope of Figure 1 could be converted into a
gyrocompass if the 1-axis is kept in the direction of the local vertical and the 3-axis in the horizontal plane, i.e. the rotational degree of freedom about the 2-axis is suppressed.

4. **THE SUSPENDED GYROCOMPASS**

If a gyroscope is suspended from an earth fixed support such that the spin axis is constrained to remain in a horizontal plane (Figure 2), then a torque caused by the force of gravity acts upon the instrument. The force changes its direction relative to an inertial coordinate system with the angular velocity of the rotating earth. The rotor reacts with an angular motion (precession) towards the direction of the earth rate vector. The inertia of the system will prevent the spin axis from stopping when it reaches the meridian. Instead it will move further until the directional force exceeds the inertia and causes the system to swing back. Thus an oscillation

![Figure 2](image1)

**Figure 2** Suspended meridian-seeking gyroscope

![Figure 3](image2)

**Figure 3** Decomposition of the earth rate vector in a local vertical and horizontal component as functions of the latitude

![Figure 4](image3)

**Figure 4** The horizontal component of Figure 3 resolved in directions of the 3- and 2-axis of the gyro
is built up, which would last indefinitely if the system was free of friction. Since the motion is constrained to the horizontal plane, only the component $\omega \cos \varphi$ of Figure 3 is effective, where $\omega$ is the earth rate and $\varphi$ the latitude. Figure 4 shows the view from space at the horizontal plane. The horizontal component of the earth rate is decomposed in two components, the one pointing in the direction of the 3-axis (spin axis) is $\omega \cos \varphi \cos \alpha$ and the other one pointing opposite to the direction of the 2-axis is $\omega \cos \varphi \sin \alpha$ with $\alpha$ being the momentary angle between spin axis and local meridian. The exact equation of the precession of the gyroscope is rather involved. It expresses the resultant effect of the inertia of the rotor, the rotating gravity vector and the torsion of the suspension band. Since it is well documented in several text books e.g. DEIMEL (1950)²

1 Suspension Tape (band)
2 Mast
3 Gyro Motor (rotor)
4 Illuminated Moving Mark
5 Collimator
6 Reading Scale
7 Forced Centering to the Theodolite
8 Clamping Ring
9 Upper Tape Clamp

Figure 5 Cross-section of a gyro attachment (Wild GAK 1)
The moving mark (4) oscillates with the rotor (3) about the vertical axis. The centre line of the oscillation is the direction of the meridian except for some corrections explained later. The observation aims at the determination of this centre line, which is achieved by taking time and/or position information of the moving mark with respect to reading scale.
and FABECK (1980) in this paper no derivation is given. The discussion of the equation in the next section follows basically the approach selected by VANICK (1972).

For a better understanding at first, an actual gyroscope attachment is presented in Figure 5, showing some constructive details and introducing the relevant terms.

5. EQUATIONS OF MOTION

The motion of a spinning suspended gyroscope can be described by a pair of coupled differential equations. The 3-axis (spin axis) follows the curve of an elliptic spiral of the kind depicted in Figure 6. In reality the proportion $\beta : \alpha$ is much less and the damping much smaller, so that the position of equilibrium is practically never attained. If the $\beta$-component of the motion and other small effects are neglected and if the $\alpha$-component is restricted to small angles by a good pre-orientation then the equation of motion has the simplified form

$$(J + \frac{N^2}{G}) \ddot{\alpha} + k \dot{\alpha} + (F + B) \alpha - B \alpha_t = 0$$

with

- $J$ (g cm$^2$) moment of inertia toward 1-axis
- $N$ (g cm$^2$s$^{-1}$) gyrosopic moment torque of the rotor
- $G$ (g cm$^2$s$^{-2}$) moment of gravity
- $k$ (g cm$^2$s$^{-1}$) damping constant
- $F$ (g cm$^2$s$^{-2}$) directive moment torque of the earth's rotation
- $B$ (g cm$^2$s$^{-2}$) directive moment torque of the suspension
- $\alpha_t$ zero-torque direction of the suspension tape

Equation 1 can be decomposed yielding

$$J \ddot{\alpha} + k \dot{\alpha} + B (\alpha - \alpha_t) = 0$$

$$(\frac{N^2}{G}) \ddot{\alpha} + F \alpha = 0$$

where Eq. 1a describes the torsional pendular motion of the non-spinning gyro about the 1-axis and Eq. 1b expresses the motion of the torsion-free suspended spinning gyroscope.
The solution of Eq. 1 results in the expression for the horizontal pendular motion of the real spinning gyroscope

$$\alpha = \alpha_0 + C e^{-\lambda t} \cos(\omega_p t + c)$$  \hspace{1cm} (2)

where $C$ and $c$ are integration constants depending on the initial conditions. The equilibrium position deviates from true north due to the twist of the suspension tape by

$$\alpha_0 = \frac{B}{B+F} \alpha_t.$$  \hspace{1cm} (3)

The damping factor $\lambda$ is given by

$$\lambda = \frac{k}{2(J + \frac{N^2}{G})}$$  \hspace{1cm} (4)

and the frequency $\omega$ of the swing is

$$\omega_p^2 = \frac{B+F}{J + \frac{N^2}{G}} - \lambda^2.$$  \hspace{1cm} (5)

These equations describe the horizontal component of the oscillation of the suspended gyroscope with an accuracy better than required for the determination of azimuths, but they have to be modified for practical use, since $\alpha$, $\alpha_0$, and $\alpha_t$ refer to true north, while the observations are carried out with respect to zero of the reading scale of the gyro or of the theodolite.

6. OBSERVATION METHODS

There are six methods of finding north using a suspended gyroscope (THOMAS 1982). The oscillation curve of the moving mark as observed through the collimator is depicted in Figure 7 as a function of time.

The equation of this damped simple harmonic motion is given by

$$\alpha = \Delta N' + A e^{-\lambda(t-t_0)} \sin((t-t_0)\frac{2\pi}{T})$$  \hspace{1cm} (6)

where $\Delta N'$ is the required angle between the zero line $N'$ of the observation scale and the centre line $M$ of the oscillation. $A$ is the amplitude and $T$ the period of the motion. Since the damping factor $\lambda$ is usually negligibly small, three independent observations suffice to uniquely determine $\Delta N'$, $A$, and $T$. The classical method is the observation of three consecutive reversal points, $a_1, a_2$, and $a_3$ for example. The well-known SCHULER Mean is then employed to calculate the required scale value of the centre line of oscillation.
\[ \Delta N' = \frac{a_1 + 2a_2 + a_3}{4} \quad \text{(7)} \]

The observation of transit times was first proposed by Schwendener (1964)\textsuperscript{6).} If three consecutive transits through a selected graduation line \( \bar{a} \) are timed using a stop watch with lap timing facilities, then the difference \( \Delta \bar{a} \) between this line and the centre line of oscillation is computed from

\[ t_{12} = t_2 - t_1 \quad , \quad t_{23} + t_{12} = T \]
\[ t_{23} = t_3 - t_2 \quad , \quad t_{23} - t_{12} = \Delta t \]

where again linear damping is assumed. The amplitude \( A \) is read at the turning point.

A combination of two transit times and one reversal point reading can be used to determine \( \Delta N' \) as has been shown by Thomas (1982)\textsuperscript{5).} Let \( t_1 \) be the first transit time, next the turning point \( a_1 \) has to be read and finally the second transit time \( t_2 \) at the same selected graduation line \( \bar{a} \) is taken. If the period \( T \) is known then

\[ t_{12} = t_2 - t_1 \quad , \quad p = 2 \sin^2 \frac{\pi}{2T} t_{12} \]

\[ \Delta \bar{a} = (\bar{a} + a_1[p-1]) / p \quad \text{(9)} \]

can be used to localize the centre line of oscillation on the reading scale, where linear damping has been assumed.
All three basic methods can be used in the clamped and in the tracking mode.

The tracking mode requires a special theodolite with a double tangent screw, so that the observer can rotate the theodolite continuously such that the gyro mark remains in coincidence with the zero line of the reading scale. The reversing point readings and/or the transit times are taken at the horizontal circle of the theodolite. This observation method has the advantage that no torsion is applied to the suspension tape, so that the motion is governed by Eq. 1b. On the other hand the tracking requires much skill and patience and leads to systematic errors if not performed perfectly. The firm MOM has developed an automatic tracking system which is incorporated in the gyroscope series Gi-B2. In spite of the reportedly good results the manufacturing of these instruments has been given up.

It seems that generally the observations in the clamped mode are favoured. The disadvantage of having to consider the torsional oscillation of the system is outweighed by the absence of any interference of the observer with the oscillating system. To improve the accuracy of the results the basic methods as outlined above have been further developed. The reversal point method is usually employed taking more than the minimum of three turning point readings. The accuracy of the readings can be upgraded by using special optical micrometers and the redundant observations are evaluated in a strict least squares adjustment.

The same trend can be observed for the transit method. Precise recording watches are used which make it possible to take the times of transit at all graduation lines, so that as many as 60 observations become available for each swing. The times are taken to one hundredth of a second. A rigorous adjustment on the basis of Eq. 6 yields the required results (CASPARY/SCHWINTZER 1981)²).

7. AUTOMATED OBSERVATION

Since the observation of transit times and turning points is wearisome, time-consuming and requiring experienced observers, efforts were soon made to automate the measuring procedure. First experiments were made in Canada with a MOM Gi-B1 and a WILD GAK 1 and similarly at CERN with the WILD attachment. The results were very encouraging as a considerable gain in accuracy and a reduction of observation time could be achieved. Today two automated gyroscopes of high precision are on the market.

MOM offers the Gi-B1 gyrotheodolite, which is equipped with two photodiodes. The transit of the moving mark through these diodes is automatically timed by a quartz-stabilized impulse counter. The results appear on a display and can be recorded or directly transferred to the computer (HP 41 C) if connected. The position of the centre line of the oscillation is computed from Eq. 10 requiring the observation of one complete period.
The observation has to be repeated at least once. Figure 8 explains the terms of Eq. 10.

\[ \Delta N' = a_0 \frac{\cos \tau_1 - \cos \tau_3}{\cos \tau_1 + \cos \tau_3}, \quad \tau_i = \frac{t_i - t}{T}, \quad i = 1, 3 \]  

(10)

Figure 8  Automated time reading, MOM Gi-B11

WBK has automated the gyrotheodolite MW 77 and sells it under the name Gyromat. For the determination of the centre line of oscillation a novel integration method is used. The swing is continuously sensed by a photo-diode equipped with an opto-electronic pick-off. The signal of one period only is used. It is filtered and integrated, the result is a measure for the required angle \( \Delta N' \) as can be seen from Figure 9. The observation process is automatically temperature corrected and controlled by certain functions.

On a lower level of accuracy, furnishing military requirements mainly, BODENSEEWerk GERAETETECHNIK (BGT) has developed an automatic gyro attachment. The directive moment of the spinning gyro is sensed and converted into a proportional current which generates a counter moment suppressing the precession. The same current drives a torquer which directs a reference mark towards true north to be picked off by the theodolite.

Figure 9  Integration method of signal sensing in the Gyromat (EICHROLZ 1978)
8. **CALCULATION OF THE AZIMUTH**

The astronomic azimuth $A'$ relative to the true meridian is computed from Eq. 11. Figure 10 shows the meaning of the terms which are discussed in due detail later.

$$A' = a_0 + \Delta N' + E + \beta_p + \beta_f$$  \hspace{1cm} (11)

Figure 10  Graph of the quantities required to compute the azimuth

The zero-torque direction of the tape according to Eq. 1 is given by

$$a_t = a_0 + \Delta N' - \varepsilon$$  \hspace{1cm} (12)

Substituting $a_0$ of Eq. 3 yields

$$a_t - \frac{B}{B+F} a_t = \Delta N' - \varepsilon$$

$$a_t = \frac{B+F}{F} (\Delta N' - \varepsilon)$$

Which finally results in

$$a_0 = \frac{B}{F} (\Delta N' - \varepsilon)$$  \hspace{1cm} (13)

Hence the computation requires the knowledge of $\varepsilon$ and of the ratio $c = B/F$. 

- $M$  Centre Line of Oscillation
- TZP  Tape Zero Position
- $N'$  Zero of the Reading Scale
- PO   Pre-orientation Reading
- ZHC  Zero of the Horizontal Circle
- —  Angle on the Horizontal Circle
- —  Angle on the Reading Scale
- —  Computed Angle
\( c \) is the zero-torque direction of the tape on the reading scale. It can be determined by any of the methods outlined in sections 5 and 6 observing the oscillation of the non-spinning gyro. Since the tape zero position is not very stable it is usually measured before and after the observation of the spinning gyro. If the difference is small, then the mean is used for the correction of Eq. 13 otherwise the last value should be preferred. The ratio \( c \) is a constant which varies with latitude. The manufacturers provide a table of \( c \) as the result of a calibration for each individual instrument. As SCHWENDENER (1964)\(^6\) has shown, the factor \( c \) is a direct function of the periods of the oscillations of the gyro in the tracking mode \( T_t \) and in the clamped mode \( T_c \).

\[
\frac{B}{F} = c = \frac{T_t^2 - T_c^2}{T_c^2}
\]  \hspace{1cm} \text{(14)}

Other methods of determination are discussed in VANCEK (1972)\(^4\).

The angle \( E \) in Eq. 11 is an instrument constant which relates the zero line of the reading scale to the horizontal circle of the theodolite. This constant is usually determined on an astronomical reference line. The determination has to be repeated at regular intervals if accurate results are required. Modern instruments e.g. the MOM Gi-Bll and the Gyromat, possess an instrument fixed reference mirror enabling the measurement of changes of \( E \) together with each azimuth observation.

The last two terms of Eq. 11 are theodolites readings, \( \beta_r \) after sighting the reference object, and \( \beta_p \) after the pre-orientation of the gyroscope.

Since \( A' \) refers to the momentary direction of the earth's rotation axis, all standard reductions known from astronomic azimuth determination have to be applied to convert \( A' \) into a bearing defined in a selected coordinate system.

9. **ACCURACY CONSIDERATIONS**

On assessing the performance of gyroscopes it is particularly important to distinguish precision and accuracy and to consider the environmental circumstances at the observation site. Measurements in a laboratory or a tunnel yield more consistent results than those in the field. Most of the standard deviations being reported on in the literature are computed from repeat measurement and hence measures of precision. Figure 10 shows the number of elements, all affected by random and systematic errors, which are put together to make up the azimuth. It is obvious that the result cannot have an accuracy of one second of arc or better. But it is possible to create favourable conditions, to apply well designed procedures and to exercise care in order to get optimal results.

Two theodolite readings \( (\beta_p, \beta_r) \) belong to each azimuth, they should be taken twice at the beginning and twice at the end of each azimuth determination. Differences indicate
instabilities of the set-up. Optimal results require observation pillars, tripods enable only second best results.

No azimuth can be more accurate than the instrument constant $E$. The methods of determination and of checking $E$ have been outlined already in section 8. Experiments have shown that $E$ changes with temperature, therefore a calibration in a temperature controlled laboratory should be carried out if field measurements are considered.

The determination of the centre line in the non-spin mode ($c$) should be carried out before and after the observation of the equilibrium position of the spinning gyro ($\Delta N'$). If a transit method is employed, then three full swings for $c$ and two for $\Delta N'$ are adequate.

The constant $c = B/F$ of Eq. 13 should be determined for each observation site. If a chronometric method is applied, then the period $T_C$ of Eq. 14 is available without extra measurements, only $T_t$ has to be measured, which can be done by timing four to five reversals points.

Some further rules to be exercised in order to get good results are:

Swinging of the gyro without spinning for half an hour eliminates torsions and deformations of the tape being imposed by clamping for transport.

After spin-up of the gyro a certain time is needed for the instrument to attain temperature equilibrium. Dependent on the air temperature this can require up to one hour for MOM and WILD instruments. The Gyromat needs less time since it only slightly warms up due to the low spinning rate.

The tape zero position ($c$) should be regulated near to zero to minimize systematic effects on the oscillation. For the same reason the pre-orientation should be performed as well as possible to get a small value of $\Delta N'$.

Some authors recommend measuring with two different pre-orientations symmetrical to the zero line.

After these actions the actual observations can start. Regular checks and corrections of the vertical axis of the instrument contribute to the quality of the results.

A protection from certain environmental factors like magnetism, movement of air by ventilation or wind, and vibrations caused by heavy machinery or traffic is desirable. This can be achieved by carefully selecting the site and the time of measurements and by shielding the instrument.

If these rules are obeyed and the best observation procedures are employed then the observed azimuths have typically the following standard deviations. The WILD GAK I mounted on a single second theodolite using extended chronometric methods or special devices for reading the scale provides the azimuth with 5" to 6" standard deviation. The automated gyrotheodolites Gyromat and Gi-Bill perform slightly better, namely with an accuracy of 3" to 4".
The precision of results for measurements in a lab are usually better by a factor of two.

10. NON-CONVENTIONAL GYROSCOPES

Since the early 1960s optical gyroscopes have been under development. These instruments do not have moving parts, are of apparently simple design and seem to have the potential of better performance in respect to resolution and dynamic range, as compared with the conventional mechanical gyroscopes. But the first prognoses that these new instruments would soon completely replace the rotating machines had to be revised, since numerous problems were encountered in the development of prototypes. Today the laser gyro is on an operational level. It is used in strap-down platforms for aircraft navigation (Boeing 757, 767 and Airbus 310), where a wide measuring range is more important than a high precision.

The basic idea of optical gyroscopes is due to the French physicist SAGNAC, who built in 1913 the first ring interferometer to study the nature of light. If a ring interferometer is mounted on a turntable and rotated then a light beam travelling with the rotation will need more time to arrive back to the source than a beam travelling against the rotation. This is the so-called SAGNAC effect, which can easily be computed if minor corrections due to the general relativity theory are neglected. The difference $\Delta L$ of the optical path length of the two beams is

$$\Delta L = \frac{4F}{c \omega}$$

(15)

where $F$ is the area enclosed by the path, $c$ the velocity of light and $\omega$ the angular velocity of the ring with respect to inertial space. The path difference can be measured indirectly by observing the fringe pattern produced by the interfering beams, see Figure 11.

![Diagram of a ring interferometer](image)

Figure 11  Schematic of a ring interferometer according to SAGNAC (RODLLOFF 1982).
Since $c$ is constant, the path difference $\Delta L$ is directly proportional to the area $F$ and the angular velocity $\omega$ of the set-up. The technical problem is to generate a measurable effect, which is achieved by two different approaches.

The fiber-optic gyroscope attains the required sensitivity by use of a large area $F$. Optical fiber with a length up to 1000 m is wrapped into a coil of very low volume. The first rotation rate sensors of this type are now being readied for application. Figure 12 shows the concept of a fiber-optic gyro as developed by SEL. It has a dynamic range of $400^\circ/s$, a repeatability of 50 ppm and a drift of less than $3^\circ/h$.

The laser gyro consists of a cavity as depicted in Figure 13. It converts the path length difference in a frequency signal, which can be observed more easily. The mirrors create a closed path calibrated to be an integral multiple of the wave length $\lambda$ thus forming an optical resonator. Rotation of the ring laser induces a change of the eigenfrequencies of the two beams travelling in opposite directions. Their beat frequency is proportional to the rate of the resonator. It is measured by observing the fringe pattern produced by superimposing the beams. The earth's rate of $15^\circ/h$ generates a beat frequency of 1 Hz in a laser gyro of $F = 100 \text{ cm}^2$ and $\lambda = 0.633 \mu\text{m}$, being easily measurable.

1. Mirror
2. Dielectric Mirror
3. Anode
4. Cathode
5. Length Control
6. Cervit Block
7. Corner Prism
8. Readout Detector

Figure 13 Schematic of a gyro block assembly (RODLOFF 1982)
The main problem with the laser gyro is the so-called lock-in effect, which prevents the measurement of small angular velocities. The reason for this effect is the tendency of coupled oscillators to swing on a common frequency if the eigenfrequencies differ only a little. Thus input rates below a certain threshold produce a zero output. A remedy is to mechanically dither the laser block with a frequency above the lock-in rate.

As already mentioned, laser gyros are operational as attitude sensors for inertial platforms. They have an effective range of $10^{-2}$ $^0$/h to 1000$^0$/s. Prognoses of today are that the sensitivity to smaller rates will improve and will become comparable to the best mechanical gyroscopes. Since their production is less expensive, they may possibly one day compete with conventional suspended gyroscopes. Suitably mounted, so that rotations about a horizontal and a vertical axis are possible, optical gyro are capable of sensing the true north direction and the geographic latitude.

11. CONCLUSION

Over the last two decades significant advances have occurred in the technology of conventional suspended gyroscopes. The accuracy of determining north has improved considerably due to the advent of new instruments, the introduction of more effective chronometric observation methods and to automating the observation of the centre line of oscillation. The same actions have simplified the measurements and reduced the time required, so that it has become possible to measure an azimuth within 15 to 60 minutes with a standard deviation of 3" to 6".

In future, optical gyroscopes may replace the conventional instruments. Within 25 years laser gyros have reached a level of performance rendering them suitable for application in strap-down platforms. Fiber-optic gyros have been under development for 10 years, seemingly meeting now the performance requirements of inertial navigation. Both realizations of the Sagnac effect hold the promise of giving higher performance at lower cost.
REFERENCES


BIBLIOGRAPHY


UNDERGROUND GEODESY

J.-C. Fischer, M. Hayotte, M. Mayoud, G. Trouche
CERN, Geneva, Switzerland

ABSTRACT

The tunnelling work, for a particle accelerator, must be carried out within very strict and tight tolerances. The CERN Applied Geodesy Group has the task to regularly control the geometry of the galleries and caverns constructed for the LEP project. For this purpose, different means and methods have been developed and implemented, related to the link between the surface and underground geodetic networks, the guiding control with accurate gyro traverses and the check of cross sections with photo-profiling or profile scanners. Instruments and methods are described by the authors and the results presented.

1. INTRODUCTION

The LEP tunnel constitutes an exceptional underground construction, due to its size (27 km circumference) as well as its position in a tilted plane. Its role is to accommodate a particle accelerator and this function involves some unusual geometric constraints.

The high precision required for the alignment of accelerator components 1, 2), compared with the common accuracies of civil engineering, is such that the geometry of the accelerator must be considered as absolute, i.e. rigidly fixed by particle-beam optics. An accelerator tunnel being rather strictly dimensioned for economical reasons, its axis must follow the theoretical axis of the machine within tight tolerances.

These tolerances (Fig. 1) are derived from CERN experience in tunnelling. They fix the allowable limits of the errors which can affect the position and shape of the galleries 3):

- errors of the surface geodetic network,
- errors in the determination of control points at the tunnel level,
- errors of the gyro-controlled traverse,
- errors in transfers from pillars to brackets for guidance by laser,
- deviations of the tunnel boring machines.
First stage

Molasse

\[ R_1 = 2.00 \quad R_2 = 2.16 \]

\[ +8\text{cm} \]

MAIN TUNNEL

Limestone

\[ R_1 = 2.2 \ (2.42) \]
\[ R_2 = 2.36 \ (2.58) \]

Second stage (concreting)

\[ R = 1.88 \pm 1\text{cm} \]

Caverns

\[ R \text{ mini} = R + e \]
\[ e = \text{minimal thickness of concrete} \]

\[ R_1 \leq R \leq R_2 \]

Fig. 1 Tolerances of the tunnels and caverns
The tolerances also ensure a minimum thickness of concrete everywhere along the
tunnels and caverns.

To avoid costly mistakes, conflicts or disputes, it is necessary to make controls at
different stages of the construction, and to finally check the completed lining of the
galleries and caverns. For such a long and difficult task, it was explicitly specified
that the construction firms must have a competent team of surveyors. Respective roles of
the firms and of the CERN Applied Geodesy Group have been defined and a close
collaboration has been established. Thus, the main targets of underground geodesy are:

- to provide the basic geodetic framework defined in the specifications,
- to preserve the interests of CERN in the geometrical conformity of the
  construction,
- to limit the risks of errors by preventive operations and to facilitate the
  task of firms by giving convenient data.

2. LINK BETWEEN SURFACE AND UNDERGROUND GEODESY

2.1 Basic framework

For both the civil engineering works and the installation of the machine, CERN has
implemented a surface geodetic network which covers the LEP site and provides facilities
for the determination of control points near the eight pit areas. This double affectation
of the network and the fact that it was measured with the Terraneter 4, 5) ensure that
pillar locations are known to within a millimetric accuracy.

From that network, three top points are determined above each pit. They are used for
a vertical transfer, down to tunnel level, in order to link the first reference pillars of
the main gallery which constitute the basic underground framework that CERN was
contractually engaged to provide.

2.2 Scheme of operations

The reference points above the pits are in the form of a conical hole in an aluminium
plate, three of these plates being screwed to the curb-stone of each pit (Fig. 2). The
conical hole has a centring function for either a standard CERN reference target or for
half-spherical holders used for zenithal sights.

Generally, the nearest geodetic pillar of the surface network is always located
between 100 and 500 m away from a pit. Linking measurements are made with a well
calibrated AGA 140 total station and a WILD Na2 level. The three top points remain tied
by accurate measurements in order to keep a good local homogeneity. The accuracy of this
link is estimated to be $\sigma_{xy} \leq 2$ mm and $\sigma_H < 1$ mm.
Fig. 2  Reference points above pits

Fig. 3  Linking measurements at tunnel level
The second operation is the vertical transfer to tunnel level. It consists of the simultaneous determination of three points on tripods with respect to top points, completed with linking measurements to the first tunnel pillars (Fig. 3).

This operation involves setting up theodolites and their (calibrated) EDM, one MOM or WILD gyro, one WILD Na2 level and the special target or reflector holders. All horizontal and vertical angles are measured, as well as distances, orientations and height differences. Observations are processed as a spatial block by means of the CERN 3-D adjustment program 6). Deflections of the vertical are taken into account to correct gyro measurements and to express the correcting vector due to the difference between the physical plumbline and the normal to the reference ellipsoid. For a pit as deep as 140 m, this correction can reach 6 mm. The accuracy of this second operation is estimated to be \( \sigma_y \leq 1 \text{ mm} \) and \( \sigma_H \leq 2 \text{ mm} \).

3. GUIDING CONTROL WITH GYRO TRAVERSE

3.1 General principles

Having guided the main and transfer tunnels of the SPS (more than 10 km of total length), for which the first semi-automatic gyroscopes were developed and used, the CERN Applied Geodesy Group has gained good experience in tunnelling control. Even with free traverses 3.3 km long for the LEP, instead of 1.1 km for the SPS, it was evident that the use of gyroscopes was the only reliable method to ensure the required guiding accuracy.

Control pillars are regularly spaced along the LEP tunnel every 39.50 m corresponding to the lattice of the machine (Fig. 4). Over the 3.3 km of an octant - from one geodetic point to the next - and with an orientation error \( \sigma_a = 40 \text{ cc} \) (0.0040 gon) on each gyro measurement on each pillar, the expected radial misclosure will not exceed 23 mm r.m.s. if distances between the pillars are carefully measured with invar wires. When adding the transfer and acquisition errors for laser piloting and the normal oscillations of the tunnel boring machines, it can be estimated that the resulting radial errors will be, at most, in the vicinity of 8 cm r.m.s. for the excavation stage. These figures are derived from CERN experience in using WILD GAK 1 gyro (manual or semi-automatic) and they have been used to define the tolerances in the specifications issued to contractors.

However, CERN development of fully automatic computerised gyro as well as new commercially available instruments promised better accuracies. This fact was not a sufficient argument to change the already tight tolerances, but it had a good practical consequence. The radial errors of a gyro traverse follow a quadratic law with respect to the number of stations, hence a better accuracy has the advantage of allowing an increase in the distance between measuring points (Fig. 5), thus giving a noticeable gain in the efficiency of such long and delicate survey operations.
Fig. 4 Tunnel cross section and control pillar

POLYGONALE GYRO (Fereture 22.8 mm)

Fig. 5 Optimisation graph
In practice, considering the severity of the tolerances, the special experience of CERN in the domain and the interest of the contractors, a compromise has been reached between the two parties. The CERN Applied Geodesy Group carries out and computes the gyro traverses in a free and open collaboration with the firms, under the latter's responsibility. While CERN is engaged to closely follow the progress of the tunnel boring machines, the transfer of the geometry from control pillars to wall brackets, for laser guiding, remains the task of the firms' surveyors.

3.2 Automation of a WILD GAK 1 gyro

The automation of the WILD GAK 1 gyro revolves around four adaptations (Fig. 6, 7):

- the replacement of the graduated scale by a CCD line of 1024 photodiodes, spaced at 13 µm, with some subsequent modifications of the optical system,
- the replacement of the manual adjustment knob by a motor-encoder set in order to achieve computer-controlled release of the spinning motor,
- the development and integration of a miniature microcomputer - also used in the new Distinvar 7) - based on a 6502 processor and constructed on a 177 mm x 55 mm multi-layer printed circuit board. This computer can act on all electro-mechanical functions of the instrument and leaves an 8 kb space for the BASIC programs which control the measuring procedures and process the data,
- the replacement of the GKK3 converter by a small, flat 400 Hz power supply housed inside the cover of the instrument.

Fig. 6 Schema of the modified GAK 1 gyro
Fig. 7 The modified GAK 1 gyro

The aim of these modifications was to produce a compact and fully automatic gyro. A single 4-pin jack links the instrument to any terminal display or portable computer: two pins for the power supply and two for RS232 communication. Technological details will be given in a future report on this instrument 9).

The computation of North with the in-board computer is based on the least-squares fitting of a damped sinusoid onto a sample data set of the motion, provided by the CCD line. This slow but efficient method has been finally retained after unsuccessful attempts at numerical integration.

The integration of all CCD data 9) would certainly have been the best algorithm; fast and exhaustive. Nevertheless, when looking for North with the motor on, mechanical
micro-oscillations tend to generate a lot of parasitic data. Buffering and filtering problems related to CPU and RAM limits made real-time processing impossible.

The calibration of the instrument must be carefully and rigourously carried out. For the highest accuracy, this operation is not so simple because the torque of the suspension tape cannot be neglected \footnote{10}. Starting from the basic equation of gyros, the ratio between the torque \( M_r \) of the tape and that of the spinning motor can be expressed as:

\[
k = \frac{M_r}{B \theta \Omega \cos \varphi} \tag{1}
\]

where \( B \) is the inertial momentum of the spinning motor with respect to its axis, \( \theta \) the rotation speed of the motor and \( \Omega \) that of the earth, and \( \varphi \) the latitude of the observation station.

![Diagram](image)

**Fig. 8 Parameters of a gyro-theodolite measurement**

The above figure shows the angles and directions which need to be considered for the expression of the azimuth \( A \) of a target point:

\[
A = l - (l_0 + E + \alpha). \tag{2}
\]

Starting from Eq. 1, the value of \( \alpha \) can be expressed in linear readings of the CCD line:

\[
\alpha = (1 + k)Q - k.ZP. \tag{3}
\]
Introducing a conversion factor $\mu$, this value is turned into angular units:

$$\alpha = \mu \left(1 + k\right)Q - \mu kZP.$$  \hspace{1cm} (4)

Its solution requires three steps:

1) successive measurements around North give a collection of linear values $Q = f(\theta)$ with slope $-1/\mu \left(1 + k\right)$;

2) the $k$ ratio must be determined by measuring the periods $T$ and $T_0$ of the motion, respectively with and without torque in the suspension tape. The last operation is made by following-up the reference index with the slow motion screw of the theodolite. It can be shown that $k = \left(T_0^2/T^2\right) - 1$;

3) the angle $\varepsilon$, external constant of the gyro, is obtained by outdoor calibrations with respect to the local bearings in the CERN system. A special calibration hut has been made over a pillar of the geodetic network for this purpose.

A strong dependency of the GAK 1 performance on the internal temperature of the instrument has been observed, and this phenomena has been modelled (Fig. 9). The improvement on the results is shown in Figs 10 and 11. The standard deviation of an identical set of calibration measurements is reduced from 55 cc to 11 cc.

![Graph](image.png)

Fig. 9  Variations of angle $\varepsilon$ with respect to internal temperature of the gyro
The in situ performance of the modified GAK 1 varies from 11 to 18 cc (1 o). After a warming-up period, a ZP - MOTOR - ZP - MOTOR - ZP sequence is observed at each station, for less than two hours. No significant improvement is gained when repeating this sequence. Each observation in the tunnel is always framed by two outdoor calibrations, one before and one after. The optimal working length has been fixed at 79 m, i.e. one pillar in two.

Fig. 10 Calibration of a GAKI gyro without temperature correction

Fig. 11 The same data set as Fig. 10 but after correction
3.3 The MOM Gi-B11 gyro

As the development and testing of the modified WILD gyro was not completed in time, there was no choice but to start with a commercially available one. Only two accurate and automatic instruments were on the market: the Gyromat, from the Mining Institute of Bochum, and the Gi-B11, from MOM. The general performance of these two gyros is rather similar and the MOM was chosen because of shorter delivery times.

The MOM Gi-B11 gyro-theodolite is a semi-automatic instrument, in which the precession of the spinning motor is observed by two photocells, spaced with a well calibrated interval, according to the transit method (Fig. 12). Data acquisition and processing is achieved by means of an HP 41-CV pocket calculator through an HP-IL interface. An RS 232 option is also available and allows control by any portable computer.

The instrument has two external units, for power supply and control. It is designed to be set-up on a tripod and a special centring device has been introduced for CERN pillars. The reliability of the Gi-B11 has lead to the choice of the "accuracy option" in a long sequence of measurements, with sights of 158 m, i.e. one pillar in four. To eliminate the eccentricities of the theodolite and of the centring device, observations are made in four positions and two complete cycles (Motor + ZP) are measured at each position. These eight independent determinations of North take 3½ h. They are also framed by two outdoor calibrations.

The MOM gyro is bigger than the WILD, hence a more favourable k ratio allows a simple and direct calibration procedure. If no temperature effects have been noted, a regular and unexplained linear drift is nevertheless observed and corrected (Figs. 13 and 14).

The final accuracy of this instrument when used according to the above procedures is \( \sigma = 7 \) cc.

3.4 Processing and results

Before computation, gyro measurements are corrected for convergence of meridians and the effects of vertical deflections, using the parameters of the local geoid previously determined and modelled \(^{11, 12}\).

All measurements of the control traverses are processed in the CERN 2-D or 3-D adjustment programs \(^{5}\):

- corrected gyro observations,
- short- and long-side angular measurements,
- long-range EDM measurements,
- invar distances between pillars,
- direct and indirect levelling data.
Fig. 12  Layout of the MOM G1-B11 gyro
Derivation = 41 cc / 22 measures

Derivation = 38 cc / 34 measures

Fig. 13 Calibration for piloting pits P1 - P2

Fig. 14 MOM 2 calibration pits P8 - P7
When making successive least-squares adjustments as the boring advances, the progressive addition of overlapping data causes some slight changes in the computed coordinates. However, these millimetric variations have no practical consequences for the guidance of the boring machines.

Until now, three octants (3330 m) of the LEP tunnel have been completed and the radial misclosures of the control traverses have been 4 mm at P2 (Fig. 15), 22 mm at P7 (Fig. 16) and recently 14 mm at P5.

Fig. 15 Misclosure at control point P2

Fig. 16 Misclosure at control point P7
4. TRANSVERSE GEOMETRICAL CONTROLS OF UNDERGROUND WORKS

4.1 General remarks

An essential point in the General Specifications issued to contractors is that the thickness of concrete in vaults, tunnels or caverns, must not be less than the specified value. In order to make sure that this minimum thickness is respected, tight controls have been undertaken at the semi-finished stage of the construction.

These controls had to take account of several criteria and constraints:

- in an underground environment, measurements must be fast and reliable,
- the instrumentation must be simple, robust and easy to operate,
- remembering the 27 km length of tunnelling, the collecting and processing of the data must be very efficient,
- the final accuracy of the controls must be close to 1 cm,
- the instruments and methods used must be applicable to small or large galleries (from 4.4 m diameter to 25 m high) and be adapted to the state of surfaces which may be rough when made with such techniques as explosives, punchers, or road-headers, or smooth when bored with a full-face machine.

Two methods have been retained: the photo-profiling method which permits very fast data acquisition, and the profile scanner (EDM measurements without reflector) for the control of large caverns.

4.2 Photo-profiling method

4.2.1 Qualities of mono-photogrammetry

The main advantages of photo-profiles are that operations in the field are fast and that the original data (the photograph) gives a view of the entire section without having to make, in situ, the difficult choice of characteristic points. The accuracy depends on the performance of the instruments, the size of the reference frame with respect to that of the measured section, and the observation and computation methods used to process the data. A compromise can always be found among these parameters to produce the required accuracy.

The only inconveniences are the intermediate steps required between acquisition and observation of the data such as processing of the films, enlargement and printing. Nevertheless, a photograph constitutes a permanent and incontestable proof of the data, which can be useful in case of dispute.
4.2.2 Instrumentation

As the method was mainly designed for the profiling of small galleries (from $\varnothing$ 4.40 to $\varnothing$ 6.00 m), it was decided beforehand to operate with non-metric $6 \times 6$ cameras. These are cheap and light, their distortion is well within the accuracy required in civil engineering and they can be equipped with a large changing-box in order to gain time. We have selected the Hasselblad SWC/K camera as presenting the best optical and mechanical qualities for our application since it has a wide-angle with low distortion, fixed-lenses (more robust), an external view-finder and a spherical bubble.

Profiles are illuminated with the "Photosect 40" flash-device from Rockset AB - Sweden (Fig. 17). The original equipment has been slightly modified in order to provide a reference frame of five known points instead of two. For this purpose, four rods with reflecting tips are mounted around the central plate of the Photosect and a fifth target is attached to the centre of the plate, together with a flat retro-reflector for the direct measurement of distances.

Fig. 17 A photo-profile in limestone section of the tunnel
The camera and the Photoset are each set up on special wheeled tripods which can be easily drawn along the tunnel rails. These two carriages are linked with a rope to ensure the correct framing distance. The positions of the profiles to be measured are first marked on the tunnel walls and the reference plane of the photoset is visually orientated using these marks. The analytical method used for computation of the profiles does not require rigorous perpendicularity of the optical axis of the camera with respect to the transverse section. This simplifies the operating procedure since the camera need only be roughly pointed towards the central target, using the external view-finder.

Absolute position in the tunnel is determined by means of a tacheometric station AGA 140, set up on a known reference pillar of the gyroscopic traverse used for guiding the boring machine. The angles and distance to each station of the photoset are recorded in the "Geodat" solid-state memory.

To simplify measurement of the photographs, films are enlarged by a factor three; positive prints being made on special paper with a plastic core for stability and rigidity, its flatness being ensured by suction plate. The orthoscopy of the enlarger has been verified with a test-grid and proved to be good, strict parallelism between the film-holder and the printing table not being critical since pictures are analytically rectified. Unfortunately, it was not possible to find a small electrostatic digitising table with a real accuracy of $10 = 0.1\text{mm}$. Instead, the Digicon B polar digitizer CORADI was adopted. With polar and angular resolutions of 0.01 mm and 0.0040 gon respectively, the maximum r.m.s. error on coordinates is 0.03 mm. Although more accurate than initially needed, this device is well adapted to the observation of photographs and gives an additional safety margin. The measured data is collected on an EPSON HX 20 micro-computer.

4.2.3. Data processing

For efficiency as well as for accuracy, it was decided to process the data in an analytical way. The analogical methods are in fact too heavy and too slow; they require the use of telescopes to ensure rigorous orthogonality of the optical axis with respect to the object plane, and the scale of the enlargement must be visually controlled on the printing table. Two methods of analytical processing have been considered. One is by space resection which computes the position, angular parameters and calibration of the camera, the other being by "rectification" of the photograph, which expresses the projective transform between the object plane and the image plane.

Due to the fact that the object is a plane in space and the camera is non-metric, it was not advantageous to choose the space resection method. In the relation between image and object coordinates:

$$\begin{pmatrix} x - x_s \\ y - y_s \\ c \end{pmatrix} = k \cdot R \cdot \begin{pmatrix} x - x_s \\ Z - Z_s \\ D \end{pmatrix}$$  \hspace{2cm} (5)
where \( k \) is a scale factor, \( R \) an orthogonal \( 3 \times 3 \) rotation matrix and \( s \) the perspective centre, the unknown focal length \( C \) cannot be determined if the distance \( D \) between the camera and one point of the reference frame is not measured.

It was more convenient — again for efficiency — to formulate the problem as a projective transform:

\[
x_1 = \frac{A x_1 + B z_1 + C}{G x_1 + H z_1 + 1} \quad (6)
\]

\[
y_1 = \frac{D x_1 + E z_1 + F}{G x_1 + H z_1 + 1} \quad (7)
\]

In this way, with one more free parameter in the equations, no special care other than framing is necessary when setting up the camera. Any position of the image plane is related to the object. With eight unknowns, four known points in the object reference frame would be sufficient to obtain a solution. In order to improve the solution and to check the computation, the fifth control point (central plate of the Photosect) is used and also gives redundancy. The projective equations are linearised and solved by the least-squares method.

These computations are made with a HP 9845 desk computer. Coordinates of the gyroscopic traverse and tacheometric data are taken into account for the absolute positionning of the measured profiles. A comparison with theoretical profiles is then made and the final results are drawn on a BENSON plotter (Fig. 18).

4.2.4 Performance

All error sources of the photo-profile method have been investigated including mechanical errors of the rods making up the reference frame, misalignment of the Photosect device due to the inaccuracy of the bubble, conicity of the light-flux and random observation errors. All these factors give a total predicted error on coordinates of \( \sigma \approx 1 \) cm for \( \theta = 4.40 \) m profiles, including linkage to the gyroscopic traverse.

An experimental check has been made on the 4 m x 5 m frame of a hall door. Eight targets were placed on the frame and then measured with an accurate micro-triangulation. Eleven photographs were taken in the same manner as those in the tunnels. A statistical study of the residuals left with respect to the theoretical test network gave \( \sigma x = 6.7 \) mm and \( \sigma z = 6.5 \) mm, in good agreement with the previous predicted errors.

To give an idea of the efficiency of the method, with which more than a thousand measured photo-profiles have been obtained, the operation times — excluding travel to the work area — are the following:

- 2 minutes for the full data acquisition of one Photosect station,
- 18 minutes per photo-profile for observation, computation and plotting.
Initially developed to check the limestone section of the tunnel excavated with explosives, the method proved to be so fast, easy and accurate that it has been extended to the whole tunnel. In the molasse, excavated with full-face boring machines, the tunnel is protected by a concrete segmental lining. The transverse controls here are also made with photo-profiles every 9.80 m (coffering length) to facilitate the positioning of the concrete segments (Fig. 19). This was the simplest and quickest way to make a preliminary check, to provide useful data to the civil-engineering technicians and to guarantee that the tunnel will have the right dimensions and correct location.
4.3 Electro-optical profiles with a profile scanner

4.3.1 Necessity of another method

As already stated, the accuracy of photo-profiling is dependent on the ratio between the reference frame and tunnel dimensions. One other limitation, mainly with flash systems, comes from the difficulty to illuminate the vault. For the large caverns which will house the huge experimental equipment of LEP, it was necessary to find a profiling method able to measure up to 25 m high with again an accuracy of about 1 cm. The equipment of the physics experiments fits into these caverns so tightly that construction tolerances have to be very strict and photo-profiling is no longer suitable. Triangulation would be a long and costly way to make the measurements, with many targeting problems. The solution is to use an electro-optical distancemeter which can work without a retro-reflector. Such an instrument exists and is based on the measurement of the transit time of a pulsed laser light, directly reflected by the object being measured.
4.3.2 The profile scanner Geo-Fennel PFT 2

This instrument is a "total station", i.e. an electronic theodolite, in which the
distancemeter is different from the normal devices used in geodesy and surveying. Instead
of measuring the phase difference between a reference beam and the return beam of
modulated light travelling forwards and backwards over the distance to be determined, the
instrument measures the transit time of the pulses emitted by an infra-red (λ = 905 nm)
laser diode at a rate of $10^7$ Hz:

$$D = \frac{1}{2} \frac{v}{2} t.$$  \hspace{1cm} (8)

The accuracy of such a measurement depends on the precision of the oscillator and on
the degree to which the speed of light $v$ can be assessed. The manufacturer does not
give details about the technology and the components of this instrument, but claims an
accuracy of $\sigma = \pm 1$ cm for distances up to 2500 m.

The main advantage of this difference principle is that the monochromaticity and the
coherence of the laser light gives a sufficient identification of the return signal peaks
even in the case of imperfect reflections of the light beam. This property allows
measurements to be made without retro-reflectors prisms, directly onto rock or concrete
walls. An additional advantage is that cyclic error linked to phase measurements are
suppressed.

The other functions of the instrument are that of an electronic theodolite: measurement of
azimuthal and zenithal angles (with automatic collimation) using incremental encoders. To facilitate profile scanning, the vertical motion is fitted with
a rest disk allowing for stepwise measurements every 5 gon. The instrument is also
equipped with a built-in data memory with a storage capacity of 1000 full measurements.

4.3.3 Calibration of the distancemeter

Several tests have been made at the CERN calibration base in order to verify the
performance of the instrument 13). The first calibration consisted of 20 measurements
every meter, over distances ranging from 1 to 30 m, on an aluminium plate used as a
reflector. In the resulting calibration curve (Fig. 20), a singular shift can be seen at
about 7 m, but despite this fact, the discrepancies remain within less than $\pm 1$ cm at
the two mean levels of the curve. After having chosen another "zero" constant arising
from the low level of the previous curve, a second set of calibration measurements has
produced the curve of Fig. 21. Here measurements over 7 m are all contained in an error
band of $\pm 5$ mm. Errors on short distances seem to have a linear drift from $+30$ mm to
zero.

Other experiments have been carried out to study the influence of the reflectance of
the materials. Using white or dark paper as a reflector, approximately the same shape of
calibration curves have again been found but with shifts between them. Using a plate of dark rough concrete a similar curve has also been produced between 4 and 18 m. Here the discrepancies are more spread out, showing the influence of the surface irregularities.

Fig. 20 Calibration curve of the FENNEL distancemeter for short distances

Fig. 21 Calibration curve of the FENNEL distancemeter for long distances
The conclusion of these experiments is that the measurements are strongly influenced by the nature of the materials. The best way to cancel these singular shifts is to make in situ calibrations of the distancemeter. Several distances must be measured against a graduated ribbon between chosen points and the instrument, in order to fix an appropriate constant before making further measurements. After this preliminary operation, it can be expected that the accuracy will remain at about ± 1 cm r.m.s.

4.3.4 Profile measurements in tunnels

Fig. 22 FET 2 profile plot
The standard handle of the PET 2 instrument has been replaced by a special bridge which permits the force-centring of CERN device combining a light target and a retro-reflector prism. In this way, the profile station can be accurately linked to a reference pillar of the gyro traverse, using the AGA 140 tacheometric station. The local orientation of the instrument can be estimated or determined by angular measurements referring to the bearing determined with the AGA station.

After a control of the calibration, the direct survey of the profile is processed stepwise by means of a ratchet mechanism on the vertical motion. However, special attention must be paid to characteristic points, in order not to miss them.

A comparison on six profiles with the photo-profiling method has given satisfactory results. Figure 22 shows the same profile as that processed by photoprofile and shown in Fig. 18 but this time using the PET2; the two results being almost identical.

5. CONCLUSION

The geometric control of underground construction is a difficult and critical task due to specific technical problems, environmental constraints and the heavy responsibility. To avoid complaints and conflicts, the credibility of these controls must be impeccable. Consequently, a high level of accuracy and reliability have been sought.

A major consideration was to ensure that the methods adopted were highly efficient, i.e. that measurements could be made rapidly and in a limited number of operations. This is extremely important since not only is access to the boring area difficult and extremely confined, but also the survey measurements cannot be allowed to unduly perturb the excavation work.

It is hoped that in the near future the technology of optical gyroscopes will provide a faster, simpler, but nevertheless accurate way to oriente tunnels.

* * *

REFERENCES


2) M. Mayoud, Applied geodesy for CERN accelerators, Part II, Ibid.

3) M. Hayotte, Tolérances, implantation et contrôle de la géométrie, Annexe 9 de l'Appel d'Offres CERN I-1061/SB-LEP.

5) J. Gervaise, J. Olsfors, The LEP trilateration network, these proceedings.

6) J. Illife, Three-dimensional adjustment in a local reference system, Ibid.


8) G. Bain et al., Le Gyroscope asservi – automation par micro-ordinateur, to be published.

9) J. Illife, M. Mayoud, Computation of the free oscillation of the GAK 1 CERN gyro by means of numerical integration, CERN, LEP-SU/int. 84-01, (1984)


12) S. Burki, W. Gurtner, Deviation of the vertical, these proceedings.

ABSTRACT

The aim of this report is to describe the means by which the metrology work load needed to place the LEP accelerator in position is accommodated.

1. INTRODUCTION

This report follows the complete process of data communication between the measuring instruments and the central computer system, as illustrated in Fig. 1.

The measuring instruments, excluding one, which are used in the LEP surveys are built using the same principles as those developed and implemented over several years by the Applied Geodesy Group.

However, to accommodate the LEP work load and to reduce the measurement time, whilst taking special care in data acquirement and treatment, these instruments have been automated and computerized here at CERN.

Three different types of data acquisition devices are now used:

- the first simply displays the measurements,
- the second is a portable microcomputer,
- the third is a portable computer conceived and constructed at CERN, specially for the survey work in LEP.

For the office computer all the materials come from a well known firm while data treatment is performed on the CERN central computer system.
2. INSTRUMENTS

2.1 Description

2.1.1 Distinvar (Fig. 2)

The distinvar is designed to make accurate differential measurements (± 0.015 mm) of short distances (from 0.40 to 50 m). It consists of three parts: the invar geodetic wire with its clip-in fittings, the remote attachment head and the measuring instrument. The latter is a precision balance mounted on a movable carriage. The principle of the measurements is to find the position of the movable carriage when the balance is in equilibrium. Absolute lengths can be obtained, provided that the complete system (remote attachment head - invar geodetic wire - measuring instrument) is calibrated on a primary base line.

2.1.2 Laser interferometer with self-aligning reflector (Figs. 3 & 4)

At CERN, the primary base line for distance measurements is determined with laser interferometry. Therefore, we make use of a calibration room where its possible to make absolute measurements of distances from 0.40 to 64 m. Obviously, to gain accuracy and time, the primary base line should be used in the field as often as possible.

Since the interferometer only measures change in phase, any interruption in the reception of the laser beam results in a break in the counting system and, hence, a complete loss of the information.

The main difficulty in the use of the movable reflector is that it must be guided within half the diameter of the laser beam. In the calibration room, the movable beam is mounted on a carriage which moves along a rail. Clearly, this cannot be used in the field. Hence, a portable set has been developed, containing the reflector prism which aligns itself on the laser beam. The reflector is mounted on a carriage, which can be moved in the horizontal and vertical directions by two servo-motors controlled by a servo-amplifier receiving signals from a four-quadrant photo-cell. This forms a servo-loop which, by moving the reflector, keeps the reflected laser beam centred on the photocell. In good conditions, measurements can be made up to 50 m with a standard error of less than 0.01 mm.

2.1.3 Horizontal offset measuring device (Fig. 5 & 6)

This device is designed to measure the shortest offset distance from one point to a vertical plane which contains a straight line used as a reference.

The distance range is 0.60 m. and the accuracy is ± 0.03 mm. According to the way the reference line is physically produced, two different types of devices are used.
Nylon wire reference line. - In addition to the 0.25 mm diameter nylon wire, the nylon set consists of three parts each fitted with a locating cylinder to ensure centring in the reference sockets. These parts are: the offset measurement device and two remote devices to attach the wire of which one is equipped with an automatic winder which acts as a wire reserve and applies a constant tension on the wire. The offset measurement device contains a movable carriage driven by a micrometric screw which is driven by a motor and coupled to a displacement counter. An electro-optical system detects the wire and controls the driving motor of the screw in order to centre the movable carriage under it. The distance measured is the displacement of the carriage from a reference point where the shifting counter is reset to the centring position under the wire. The electro-optical system can, to a certain extent, measure the angle between the rule and the reference plane and automatically calculate the reduced distance.

Laser beam reference line. - The laser set consists of two parts: first, the transmitter with its laser optics and power supply, mounted on a locating cylinder to ensure centring in the sockets and to orientate the beam, second, an offset measurement set which has the same function as that of the nylon set except that the wire detecting system is replaced by a photo-cell set for the detection of the laser beam.

2.1.4 Vertical offset measuring device (Figs. 7 & 8)

This device is new in the range of metrology instruments at CERN. It detects on a vertical active rule a horizontally adjusted laser beam. Apart from the position of the laser beam detector, the rule is a copy of the horizontal offset laser measuring device, but in a vertical position. The active rule gives very good results, but the horizontal laser emitter is unsatisfactory and two devices are under development.

In the first, the laser beam is emitted by a GLO laser from WILD and optical fiber directs it to the eyepiece of a WILD ZL zenithal telescope. An automatic device makes the optical axis of the telescope vertical with a precision of one in two hundred thousand. A pentaprism made at CERN horizontally adjusts the laser beam and can rotate, thus making the beam sweep out a horizontal plane (Fig. 9). The use of the zenithal telescope makes two-face observations possible by rotating the telescope through 200 grades without moving the upper prism. This system is easy to handle, but the quality of the beam is inadequate with respect to the required accuracy. The laser spot is not homogenous and we do not obtain a parallel beam with a diameter consistent with the detection cells of the active rule. We are therefore considering the possibility to modify the optics of the ZL telescope.

In the second, the emitter is a classical laser tube centred on a cylindric mechanism. An optical system reduces the divergence by increasing the diameter of the beam. The laser and optics set is mounted on an adjustable support, levelled by an electronic clinometer mounted on two dogs gripping the cylinder of the laser. The average of the measurements obtained by two-face observations of the laser and clinometer eliminates instrumental errors. The quality of the beam is excellent, but on the other
hand the handling is arduous and brings back memories of the "NIVEAU LEMOIR". However it is presently used for testing the active rule (Fig. 10).

2.1.5 Clinometer (Figs. 11 & 12)

This instrument is designed to accurately measure the tilt in one direction only and it can be used either for absolute or differential measurements. The instrument detects the transverse position of a pendulum suspended by means of five copper wires from a rigid support. The position is read by an electro-optical system using differential photodiodes and amplifiers.

2.2 Automatization and computerization (Fig. 13)

All the instruments described above are specially designed to be flexible, adaptable to the needs of the user, and easily interfaced. They communicate information via a unique link containing the necessary power cables, as well as permitting standard serial communication. In each instrument a microprocessor circuit board, through adapted incorporated interfaces, controls the electro-optical and electro-mechanical elements and the user dialogue.

Two microprocessor circuits boards have been developed by the Applied Geodesy Group. The first one is truly a microcomputer, programmable in machine code and in basic language, which can be programmed in the field by the user and is specially adapted to the more complicated instrument calculations. The second circuit board is simpler but well adapted to control measuring systems. It is programmable in machine code and the programs cannot be modified by the user in the field. The control programs are interactive and only use the three ASCII characters [CR], [LF], [ESC]. These instruments are not equipped with control keyboard and display, but are operated with the aid of a control box, a terminal or a computer. All the necessary electronics are contained within the instrument thus allowing the user the freedom to choose the terminal or the measurement acquisition system.

3. TERMINALS AND FIELD COMPUTERS (Fig. 14)

In the laboratory, for test measurements, the instrument can be connected to a standard character terminal and a 12 V power supply.

For the unrecorded field measurements, we have developed a control and power box which relays information to the instrument via a unique cable containing power and communication links. A 12 V battery is incorporated in the box, also equipped with a display panel and four control buttons; one being the on/off switch, the other three representing the ASCII characters recognized by the interactive programs incorporated in the instruments.
With a single instrument the collection of measurements is made by connecting a power supply and a portable microcomputer. This type of machine is now available commercially and we have chosen the EPSON PX4 which can also be used as a single terminal.

For the power supply operation control and simultaneous data collection from several instruments, we have developed and built a field computer (Fig. 15), whose main characteristics are the following:

- easily transportable
- can be powered by internal or external battery or main power supply
- possesses sufficient memory allocation to store the data and measurement acquisition programs, the metrological operation control files and the data storage files
- includes an additional stable support memory which can be disconnected from the main system for transportation of data (RAM cartridge of 64 Kbytes)
- offers the option to connect several measurement instruments and a terminal at the same time (five RS232 lines)
- possesses a printer with 40 characters per line
- is compatible with the CERN network of computers, permitting the exchange of data contained in the field computer memory and the central system
- is user programmable with PILS an interacting language developed at CERN
- incorporates transducers for the measurement of temperature, humidity, pressure and the voltage of the battery. This information is accessible from the user program
- the terminal used in the system is the EPSON PX4.

4. THE OFFICE COMPUTER (Fig. 16)

The office machine must complement the data acquisition of the field computer to provide a complete and efficient chain of measurement and calculation. It acts as an intermediary between the field machine and central computer data bases and must allow:

- temporary and local storage of data allowing the user direct and instantaneous access
- local treatment in order to quickly prepare the installation files and to control and validate the measurements made in the field
- easy recuperation of the necessary elements from the central calculation data bases to prepare the installation files
- assistance in the development and maintenance of the field computer operating system
- formulation and transfer of field data to the main computer system for the calculation of the geodetic and metrological networks, or for adding to the data bases.
The system chosen to ensure the above properties is an HP 9540 accommodating several users and tasks. It functions under the UNIX operating system and can be programmed in FORTRAN, PASCAL and C language. It works in conjunction with a hard disc, a rapid input/output tape drive for back-ups, a printer, a double flexible disc drive and six HP 150 microcomputers used as work stations. The computer is linked to the CERN main computer system by four serial lines and an ETHERNET connection. The work stations are equipped with a printer, two floppy disc drives and two RS232 lines for communications with the HP 9540, the field computers and the CERN network.

5. CERN COMPUTER CENTRE

The main computer centre, one of the biggest scientific calculation centre in Europe, houses three large systems: the first using large CDC machines, the second based on large scale IBM and SIEMENS equipment, the third using several VAX machines.

There are many other computers, essentially 32-bit machines shared throughout the CERN site for allocated local tasks. All are connected via CERNET-ETHERNET networks to allow very high speed data transfer between them. For the users, the computers are accessible by a remote batch service named INDEX, which works at 4800 bauds, manages 1300 terminal and also allows connection with the European and American scientific nets, EARN, and BITNET.

We use the facilities of the computer centre for all large calculations generated by the problems of geodesy, metrology and topography, and also to update and consult the data bases.
Fig. 1 Process of data communication

Fig. 2 Distinvar
Fig. 3 Scheme of the laser interferometer

Fig. 4 Self-aligning reflector
Fig. 5  Profile section of the horizontal offset measuring device

Fig. 6  Plan of the horizontal offset measuring device
Fig. 7  Front view of the vertical offset measuring device

Fig. 8  Profile section of the vertical offset measuring device
Fig. 9 Laser beam - automatic level

Fig. 10 Laser beam - manual level
Fig. 11 Cross-section of the clinometer

Fig. 12 Diagram of tilts

Fig. 13 Scheme of an instrument
Fig. 14  Terminals, control base and field computers
**Fig. 15** Scheme of the field computer

**Fig. 16** Scheme of the office computer
APPLIED METROLOGY FOR LEP

Part I : COMPUTING AND ANALYSIS METHODS

M. Mayoud
CERN, Geneva, Switzerland

ABSTRACT
The search for accuracy in geodetic metrology demands that the processing of the data always remains rigorous and takes into account any factor which can affect the precision of the measurements or that of the results. This constant care leads to very precise considerations in defining the absolute positioning of a new accelerator within an area submitted to strong deflections of the verticals, in processing the data of long and flexible linear networks, or in estimating the real accuracy of the alignment of machine elements. The basic principle of the methods developed at CERN for the geodetic positioning of LEP, the processing and smoothing of the alignment data and the stochastic analysis of networks are described in this third part. An estimate of the predicted errors of LEP metrology is given as a conclusion.

1. INTRODUCTION

For each new project, the size of the accelerator to be built has led to reconsideration of several aspects of the methodology. The major changes in concept have been introduced in designing the SPS (2.2 km diameter) and LEP (8.6 km diameter) control networks.

First, in both cases, the computation of the theoretical XYZ coordinates of the machine has involved more and more detailed considerations of the geometry of the earth. For the SPS, a spherical approximation was sufficient to express the effects of the earth's curvature in computing the Z ordinates, correcting the vertical "descent" of geodetic points along the shafts or properly tilting the magnets, in order to obtain a real plane in space. With the LEP project, which partly lies under the Jura mountains, a further step has been to determine the vertical deflections generated by gravity disturbances, and then to express the separation between a reference equipotential surface and a reference (local) ellipsoid. This knowledge provides the necessary corrective factors to convert measured altitude into ellipsoidal heights in 3-D computations, to correct the coordinates of bottom points from the effects of vertical deflections or to reduce the gyro measurements.
One other change in the methodology is that repetitive measurements of the SPS or LEP control networks could no longer be thought of and managed as "absolute" surveys. For such long and flexible ring-shaped figures, the variations of the coordinates issuing from different sets of comparable measurements have no physical meaning for the particles. The trajectory of a beam within an accelerator is mainly sensitive to short-range errors. Survey or alignment are thus "seen" as local imperfections of the guidance magnetic field while long range errors have less effect but are not negligible. Then, the major requirement for the geometry of an accelerator is that relative errors must be as small as possible. In other terms, the figure must be smooth. This smoothing concept is fundamentally involved in a particular refinement process used for the first installation of a large machine and for any new partial or global survey when a re-alignment of components is to be done.

It must also be remembered that the certitude in any accuracy problem cannot be acquired without a thorough knowledge of the stochastic behaviour of the measured networks. Although this statement sounds self-evident, its reality depends on the effective means for estimating the actual errors and deformations that a given network may undergo, due to the effects of both random and systematic errors in the measurements. For that purpose, a simulation method has been developed at CERN on the basis of a statistical analysis with controlled perturbations.

2. THEORETICAL POSITIONING OF LEP

2.1 Transformation of machine coordinates

The theoretical coordinates of the components of the machine are first produced by the MAD program (Methodical Accelerator Design) in an arbitrary plane system. The positioning of this theoretical machine is then made according to local constraints (Fig. 1 & 2) including:

- feasibility of a future link between the SPS and LEP for e-p physics, this fixing the first interaction point
- joining the existing reconnaissance gallery
- confinement within the molasse bank
- optimisation of the depth of pits.

Successive unitary transforms are expressed in relation to these constraints and their final product gives the theoretical geodetic coordinates of the machine (Ref. 1).

2.2 The CERN coordinate system

As the particles practically ignore gravity, the CERN coordinate system is a cartesian system. For historical and logical reasons, its vertex is the centre of the first proton synchrotron built at CERN, the point $P_0$ of the PS machine.
Fig. 2 Geological cross-section of the LEP site
The geographical coordinates of this origin are:

\[
\begin{align*}
\phi_P &= 51.36920 \text{ gon} = 46^\circ 13' 56'' .2 \text{ N} \\
\lambda_P &= 6.72124 \text{ gon} = 6^\circ 02' 56''.8 \text{ E}
\end{align*}
\]

While the azimuth of the Y CERN axis with respect to geographical North is:

\[
\gamma_P = 37.77864 \text{ gon} = 34^\circ 02' 02''.8
\]

The Z CERN axis is collinear with the vertical at \( P_0 \). To avoid any confusion between the \( Z_c \) ordinate and the altitude \( H \), a constant of 2000 m was added to the \( H \) values of the PS machine. The local coordinates of the vertex \( P_0 \) are:

\[
\begin{align*}
X_P &= 2000.00000 \text{ m} \\
Y_P &= 2097.79265 \text{ m} \\
Z_P &= 2433.66000 \text{ m} \\
H_P &= 433.65921 \text{ m}
\end{align*}
\]

If, in the past, a spherical model of the earth was sufficient to express the geodetic data of accelerators (Ref. 2), this was no longer the case with respect to the size of LEP (Fig. 3).

The reference system has therefore been re-defined to accommodate a projection onto an ellipsoid, the radius of curvature of which is dependent upon the azimuth. The figure used at CERN is the IUGG 1980 ellipsoid with the following parameters:

- semi-major axis \( a = 6378.137 \text{ m} \)
- semi-minor axis \( b = 6356.752 \text{ m} \)
- polar flattening \( f = 1/298.26 \) where \( f = (a-b)/a \)
- first eccentricity \( e^2 = 0.006 694 380.022 \) where \( e^2 = (a^2-b^2)/a^2 \)
- second eccentricity \( e'^2 = 0.006 739 496.775.5 \) where \( e'^2 = (a^2-b^2)/b^2 = e^2/(1-e^2) \)

This ellipsoid is considered tangential to the geoid at \( P_0 \), and hence the normal to the ellipsoid is collinear with the vertical at \( P_0 \). The radii of curvature of the ellipsoid in the meridian \( \rho \) and in the prime vertical \( \nu \) at latitude \( \phi \) are:

\[
\begin{align*}
\rho &= a(1-e^2)/(1-e^2 \sin^2 \phi)^{3/2} \\
\nu &= a/(1-e^2 \sin^2 \phi)^{1/2}
\end{align*}
\]

At \( P_0 \) the two principle radii are:

\[
\begin{align*}
\rho_{P0} &= 6368.761.40 \text{ m} \\
\nu_{P0} &= 6389.299.67 \text{ m}
\end{align*}
\]
The consideration of the osculating circle in $P_0$ (vertex at zero level) is sufficient, within the CERN area, to express the geometrical relations allowing the conversion between $Z$ ordinates and ellipsoidal heights $HE$. The radius of this circle, for an azimuth $\alpha$, is:

$$\rho_\alpha = \frac{v}{1+e^2 \cos^2 \phi \cos^2 \alpha}$$  \hspace{1cm} (3)

Knowledge of the separation $S_o$ between the ellipsoid and the geoid is derived from physical considerations: computation of equipotential surfaces in a mass model and astrogeodetic measurements using a zenith camera (Refs 1, 3, 4). This quantity is the necessary link between the observable heights $HG$ above the geoid and the ellipsoidal heights, which are purely geometric (Fig. 3).

**Fig. 3 Ellipsoidal and geoidal heights**

### 2.3 Theoretical altitudes at LEP level

Derivation of ellipsoidal heights from measured altitudes needs two additive terms. The first gives the position of the "zero" equipotential surface - the geoid - with respect to the reference volume $V$ (Fig. 4). If $T$ is the perturbing potential generated by the topographic masses and $\gamma$ the normal gravity of the reference volume, the separation $S_o$ at point $M$ is expressed by the Bruns' formula:

$$S_o = \frac{T_M}{\gamma_M}$$  \hspace{1cm} (4)

The second term is the orthometric correction of the measurements, giving the "normal" altitudes in the Molodensky free air model. For instance, from $L$ to $M$, this correction would be:
\[ C_n = \frac{B}{L} g \cdot dh / \gamma \]  

where \( g \) is the actual measured gravity along the traverse.

Fig. 4  Equipotential surfaces and vertical deflections

This last term corrects, in fact, the inner divergence of equipotential surfaces which reflects the curvature of the plumblines within the irregularities of the gravity field inside the crustal part of the earth. The values of these curvatures are also derived from the mass model. When no gravity measurements can be made, this being the case before tunnelling of the LEP project started, these values give the possibility to anticipate the evaluation of the normal correction and therefore to compute the predicted altitudes of the project in order to get a true plane in space.

For that purpose, the deflections \( \theta \) of the verticals have been expressed along the machine, onto its tilted plane, using the computed gradients of curvature. A similar reduction has been applied to astro-geodetic measurements. Integrating these values with a least-squares fitting, the corrective paraboloid so obtained expresses the sum of the two
additive terms \( S_{o+n} \) and allows the computation of the theoretical altitudes of the project elements.

3. **ADJUSTMENT OF OBSERVATIONS**

3.1 **Adjustment programs**

Many adjustment programs have been successively written to satisfy the geodetic needs for CERN accelerators: planimetric (XY) or altimetric (H) programs, tridimensional adjustment strictly limited to micro-geodesy, processing of large matrices (Ref. 5, 6), Helmert transforms, etc.

The size of the LEP project called for a new and rigorous computational tool, fully adapted to the processing of all kinds of geodetic data in a local system and whose main features are the following:

- local 3-D adjustment on the ellipsoid GRS 80. Altitudes are referred to the local geoid, which has been investigated and parametrized
- generalised least-squares processing of all types of available data, some being very peculiar to the metrology of large objects
- statistical and variance analysis of the results - generation of random and/or systematic perturbations for simulations
- flexibility in data handling: free formatting and intensive use of keywords.

A detailed description of the program can be found in the contribution from J.C. Illiffe (Ref. 7). The principles of the CERN simulation method will be explained in Section 4.

3.2 **Radial and vertical smoothing**

When installing the machine components, the first determination of the control network gives the displacement vectors between their actual "rough" position and their theoretical one. In fact, magnets are positioned around an unknown mean trend curve (one among an infinity) contained within the envelope of maximum errors. The polynomial degree of the curve depends on redundancy, overlap of measurements, and the bridge distance between control (fixed) points. The final relative errors are a quadratic combination of those of the network itself and those of the positioning, i.e. installation errors. Their statistical nature is essentially gaussian: the aligned elements are randomly and normally distributed around this mean trend curve (Fig. 5).

As the major requirement for the geometry of an accelerator is that the relative errors must be very small (\( \sigma \leq 0.1 \text{ mm} \)), an obligatory step for surveyors is to check the installation by measuring and - if needed - improving the smoothness of the machine alignment.
Fig. 5 Position of magnets with respect to theoretical orbit

Another important consideration is that, when making successive surveys of long and flexible figures, absolute comparisons would be a nonsense. The difference between trend curves, corresponding to each survey, must be analytically eliminated without inducing systematic or harmonic errors, which are critical for an accelerator. Therefore, a special smoothing algorithm has been developed in order to process the local data in a purely relative way, i.e. without any absolute involvement in coordinates (Refs. 8, 9).

The smoothing process consists of a set of radial - or vertical - measurements (the last measured values after installation) which are treated in the following way:

Considering for instance the radial observation scheme (Fig. 6), the relationship between three points H, I and J with an off-set measurement on I is:

\[
\frac{1}{1+k} \cdot dR_h - dR_i + \frac{k}{1+k} \cdot dR_j + St - Sm = v_i
\]

where:

\[k = \frac{proj(HI)}{proj(HJ)}\]

\[dR = \text{unknown radial discrepancy with respect to mean curve}\]

\[St = \text{theoretical sagitta of these three points}\]

\[Sm = \text{measured sagitta}\]

\[vi = \text{residual of the measurement}\]

Each point receives three overlapping measurements resulting in a good redundancy. Nevertheless, these measurements only cover six points (i.e. three quadrupoles) and the global system would be rather poorly conditioned if a determination of coordinates was required. But as the purpose is to get local and purely relative information, this fact is not critical if the normal system is solved under the double condition \[||dR||\] and \[||v||\] minimum.

The condition \[||dR||\] minimum constrains the reference line of the ordinates \(dR\) to be the mean curve. This condition cannot be directly expressed in the adjustment. Neither
can it be linearized, since the differential increments, the unknowns \( dR \), and the residual \( v \) are, nearly, of the same order of magnitude. The convergence would consequently be slow and the results uncertain.

![Diagram of smoothing measurements](image)

**Fig. 6** Scheme of smoothing measurements

The only way which has been found for solving such a system is to move \( dR_i \) onto the right-hand side of the equation, introducing for each central measurement on \( I \) a new observation equation:

\[
\frac{1}{\sqrt{3}} \left( \frac{1}{1+k} dR_i + \frac{k}{1+k} dR_j + St - Sm \right) = dR_i \quad (\text{v})
\]

(6)

The weight \( \frac{1}{\sqrt{3}} \) comes from the fact that each point receives three measurements.

The normalisation of these additional equations is equivalent to \( \text{I}(dR + v)^2 \) minimum, i.e.:

\[
(\text{I}dR_i^2 + \text{I}v_i^2 + 2\text{I}dR_i v_i) \text{ minimum}
\]

(7)

\( Iv_i^2 \) is set to its minimum by the normalisation of the first equation, for each measurement. These residuals \( v_i \) constitute a Gaussian distribution \( [0, \sigma_i^2] \) and the radial ordinates \( dR \) are also expected to be a set of Gaussian variates with, of course, a zero mean-value. \( dR_i^2 \) will consequently be a minimum if the quantity \( (\text{I}dR_i v_i) \) tends to zero.

For two Gaussian samples \( U(0, \sigma_1) \) and \( V(0, \sigma_2) \) of the same dimension \( N \) in their initial (random) order, it can be easily proved that if, in general, \( X \) and \( Y \) are two random variables with density functions \( a(X) \) and \( b(Y) \), the expectation of the product is:

\[
E(XY) = \sum_{i,j} p_{ij} X_i Y_j = \sum_{i,j} p_{ij} X_i Y_j = \sum_{i,j} p_{ij} X_i Y_j
\]

\[
E(XY) = \int f(X,Y).X . Y dX dY = \int a(X) . b(Y) . X . Y dX dY = \int a(X) . X dX . b(Y) . dY
\]

whence \( E(XY) = E(X) . E(Y) = X . Y \)

for \( X = Y = 0 \),
then \( E(XY) = 0 \).
If the sample \( V \) has dimension \( n = N/3 \), this is more difficult to prove. But it can be checked easily with a program using the computer's Gaussian generator.

This global procedure is a kind of compromise, which is not quite rigorous but nevertheless satisfactory. Simulations and real computations have given acceptable results both for checking the first installation and for successive and comparative surveys of the machines.

For vertical positions, smoothing has been first made "visually", by drawing a mean curve on a plot of differences \( H \) measured - \( H \) theoretical. A similar procedure has been tried by generating pseudo-observations of vertical off-set values, derived from adjusted altitudes, with the same overlap. Results are, here also, satisfactory:

- unfavourable sign sequences are located,
- outstanding \( dR \) values can be pointed out.

4. **EXPECTED RESULTS FOR LEP METROLOGY**

4.1 **The CERN simulation method**

Starting from the a priori standard deviations on measurements, derived from experience, the well-known tools of stochastic analysis are:

- unit weight variance: \( \sigma^2 = V'PV / n_0 - n_u \),
- covariance matrix: \( C_x = \sigma^2 N^{-1} \), which gives expression of absolute and relative error ellipses,
- histogram of residuals, estimated accuracy of (groups of) observations, elimination or refinement methods,
- confidence intervals on estimates.

Nevertheless, this classical way of analysis may leave some interpretation problems on mixed networks. Significant distortions can occur on a posteriori estimates and it is difficult to appreciate the relative "strength" of each group of observations. Furthermore, no signal is given to detect the systematic errors, which are not modelled in this process. This kind of simulation does not give a clear view of the behaviour of a perturbed network and one must remain rather wary when interpreting the resulting statistical figures.

A pragmatic way to obtain a picture of the true situation is to simulate on a computer all the perturbations which can affect the geometry of a figure: Gaussian errors in measurement, artificial generation and addition of systematic errors, controlled constraints, etc. (Ref. 9).
For this purpose, the provisional coordinates (or theoretical ones for accelerators) are taken as ideal. Measurements are supplied in the input file as for a normal computation of the network. The program computes the ideal measurements and a gaussian generator adds random errors scaled on a-priori variances. The data is then processed as usual. Repeating these operations gives a "Monte-Carlo" generation of 'n' sets of hypothetical measurements of the same network.

Empirical statistics carried out on the results give very interesting estimates to be compared with the known (and controlled) a priori values of the variance of each group of observation or, even, to allow a direct analysis of the effect of errors on the coordinates. Tests can also be made to appreciate the agreement of actual results with predicted values. Such a method gives a clear idea of the "response" of a complex network to random errors and provides some corrective factors to apply to the various estimates in order to make a correct scaling of the predicted errors.

When adding systematic errors and/or controlled constrains, the simulations also give a true image of the distortions suffered by the network. The effect of each constraint can then be evaluated. The resulting shifts of the mean values of the residuals, with respect to any selected systematic error in each group of measurements, gives an idea of the "warning lights" to watch for when actual measurements are to be processed.

This pragmatic method has been used for years at CERN and it is a very helpful tool for the engineer who needs a real knowledge of the network he has to design and optimise, and then observe and compute.

4.2 Preliminary study of the LEP network

The LEP underground network is a 27 km long chain (Fig. 7) determined with invar lengths and off-set measurements, which is constrained by eight control points coming from the surface geodetic network. Each octant, between two control points, is therefore 3.3 km long.

![Fig. 7 Scheme of measurement](image-url)
The preliminary study of the error envelope of this network has been made with the following assumptions and means:

a) Radial and longitudinal relative errors at control points are taken as $\sigma_r = \sigma_\ell = 2$ mm. These figures come from the simulations of the surface network and experience in "descending" points.

b) Along the figure, the radial errors $c_n$ and $c_{n+1}$ of the ends generate radial discrepancies which can be estimated at positions 1/4, 1/2 and 3/4 of the octant (Fig. 8).

$$c_n = (1/2) (c_n + c_{n+1}) \Rightarrow \sigma_n = \sqrt{(1/4) (2^2 + 2^2)} = 1.4 \text{ mm}$$

$$c_{n+1} = c_n + (1/4) (c_{n+1} - c) \Rightarrow \sigma_{n+1} = \sqrt{(9/4) x 2^2 + (1/16) x 2^2} = 1.6 \text{ mm}$$

Fig. 8 Effects of radial errors

c) A longitudinal error $dL$ in control points has a distorting effect resulting in radial errors $dR$, because of the flexibility of the figure. For LRP networks, this effect has been simulated and, at the middle of one octant, it produces an error $dR = 1.8 \text{ dL}$ (Fig. 9a).

Fig. 9a Longitudinal flexibility

The estimate of $dL$ comes from longitudinal errors in ends $n$ and $n+1$:

$$dL = c_n + c_{n+1} \Rightarrow \sigma_{dL} = \sqrt{2^2 + 2^2} = 2.8 \text{ mm}$$

The resulting $dR$ along the figure are expressed in Fig. 9b.

Fig. 9b Effects of longitudinal errors

d) The radial errors of the metrological chain itself are evaluated through repeated simulations of the underground network. Initial assumptions are $\sigma = 0.05$ mm for invar measurements and $\sigma = 0.08$ mm (short) or $\sigma = 0.15$ mm
(long) for off-set measurements. The values issued from confidence 
ellipses and confirmed by statistics on coordinates are summarized in 
Fig. 10.

\[ \begin{array}{c|c|c|c}
(N) & 1/4 & 1/2 & 3/4 & (N+1) \\
3.3 & \ & \ & \ & 3.3 \\
\end{array} \]

Fig. 10

It should be noted that the bumps are caused by a higher degree of 
flexibility near the ends, even with the overlapping measurements 
which provide the continuity and homogeneity from one octant to 
the next. The final r.m.s. envelope of the radial errors is the 
quadratic composition of these independent errors (Fig. 11).

\[ \begin{array}{c}
(13) \quad (14) \quad (13) \\
S_1 \quad S_3 \quad S_1 \\
2 \quad Pn \quad (13') \quad (14') \quad Pn+1 \\
\end{array} \]

Fig. 11 Final errors before and after link

4.3 Full simulation of ten hypothetical surveys of LEP

In order to have a general check on these estimates, ten simulations 
of the whole machine network have been carried out with all random 
perturbations (errors in control points and in measurements). The 
values found at positions 0, 1/4, 1/2, 3/4 and 1/1 were 
respectively 2.1, 3.9, 5.2, 4.7 and 2.1 mm. A Fischer test of the 
comparison between these empirical estimates and the predicted 
one gives a full agreement at 95% confidence level 
(Fig. 12).

Further studies have been made on the polynomial degree and the 
harmonic Fourier content of these ten simulated surveys of the 
machine. The aim was to extract the trend of the mean curve of 
the positions in order to discern the contributions from long-range
(absolute) and short-range (relative) errors. The resulting data has been introduced into the program simulating the orbits of the particles. Analysis of these simulated orbits has shown that the perturbations generated by the alignment errors can be damped by the planned correcting magnets. It has also established the fact that absolute long-range errors have more effect than previously thought in accelerator theory.

![Graph](image)

**Fig. 12** Empirical estimates derived from simulations

* * *

REFERENCES


2) J. Gervaise, M. Mayoud, E. Menant, Système tridimensionnel de Coordonnées utilisé au CERN, CERN 76-03, (1976).


7) J. C. IIiffe, Three-dimensional Adjustments in a Local Reference System, these proceedings.


APPLIED METROLOGY FOR LEF

Part II: DATA LOGGING AND MANAGEMENT OF GEODETIC MEASUREMENTS WITH A DATABASE

J.-P. Quesnel
CERN, Geneva, Switzerland

ABSTRACT
This paper describes the database which has been created, with the ORACLE system, to manage the data from the positioning and surveying of the CERN particle accelerators in order to check the true shapes of the machines and the alignment of their components. A comparison between data handling with the database and with a classical file architecture is given.

1. INTRODUCTION

Today, some 8000 elements, laid end to end, make up CERN's particle accelerators. This represents a total beam line length of more than 20 km. These quantities will be doubled when the LEP machine starts. To manage all the data connected with the geometry of the accelerators, the Applied Geodesy Group has now established a relational database. The data architecture will be first described and then its efficient use examined.

2. DATA ARCHITECTURE

In the following sub-paragraphs, we shall describe the data which is stored and then how it is structured.

2.1 Stored data

All the data related to the geometry of the particle accelerators and the survey measurements is stored in one of the following categories:

Geometric definition of the accelerators. Different algorithms allow the accelerators to be defined, in accordance with the required characteristics and the chosen strategies. Programs describe the beam lines as a succession of elements laid end to end along the reference orbit. Each point of entrance and exit of the beam in or out of the
element is defined by coordinates, and the three angular parameters of the orbit at these points are given in the reference coordinate system used at CERN.

**Geometric definition of the elements.**— Each element to be aligned is equipped with two sockets which guarantee the centring of all the metrological instruments and whose positions are known exactly in relation to the theoretical beam path in the element. Thus, with a three-dimensional rotational matrix, we can compute the coordinates of these targets in the CERN system.

**Deliberate misalignments.**— The accelerators are designed for specific characteristics of the particle beams. If these characteristics are modified, or if the mode of use is changed (for instance, accelerator mode changed to collider mode for the SPS), it is sometimes necessary to modify the alignment of some quadrupoles, in order to adjust the orbit. These deliberate misalignments are stored in the database.

**Geodetic network coordinates.**— To ensure the correct absolute and relative positioning of several accelerators and each of their beam lines, CERN is covered by a geodetic network which extends on the surface and also underground. All the points are known either in ellipsoidal coordinates or in the CERN coordinate system. The altitudes are defined with reference to the ellipsoid or the geoid.

**Measurements.**— It is important to distinguish between two sorts of measurements. On one side, there are those usually made for the determination of a geodetic network, while on the other, are the measurements taken during the component's alignment. The former have to be used together, the latter are independent and connected only to the element itself.

The measurements of a geodetic network or a machine are of several types:

- distance, made with an electro-optical distancemeter, a distinvar or a calibrated tape
- vertical or horizontal angles
- offset
- levelling.

Each type of measurement has a code and is stored with the date, the number of the instrument and the place where it has been made.

When an element has just been aligned, we immediately make a complete survey of its position in relation to the reference points used. Each set of measurements is dated, stored and archived.

**True position of the elements.**— All the measurements described above can be computed to give the true position of each component. In fact, we store the comparison between the true position and the theoretical one, along and radially to the beam.
2.2 Structure of the Tables

Numerology.— A strict and precise numerology had to be defined for all the CERN machines. All the machines or transfer lines, all the components to be aligned, and all the targets have names which satisfy the rules adopted.

Tables.— All the information contained in the database is stored in the form of tables, each made up of a row for each entry and as many columns as required. The rows are not ordered in any way and when deciding the structure of the tables, some columns can be chosen as indices in order to help select the data. A unique index is also needed to be sure that a row is not recorded twice in the table.

In addition, the tables are independent in the database, but their architecture has to satisfy the relational language which allows access to several tables at the same time, and also the unique storage of the data. All around this main architecture are other tables which contain data devised to aid the connections between tables and the selection of data. Thus, we have tables which contain the name of the machines, the list of different sorts of targets, the calibration values of the invar wire, etc. (Fig. 1).

---

**Fig. 1 Organigram**
3. USING THE DATABASE

A database is useful only if it is active. It must allow anyone access at any time. Further, it must not be necessary to copy the information, in order to use it, and updating must be safe and easy. The database of the Applied Geodesy Group has been established with the ORACLE system \(^1\). It is closely connected with a specific software tool kit developed according to the needs of the Group.

3.1 Updating

To ensure that the procedure of updating remains simple, only the "primary" data are stored. These are theoretical data on accelerators and non-derived measurements. All the data resulting from a computation of data stored in the database is obtained by activating the corresponding programs. For theoretical data, the tables are updated by completely clearing the rows related to the modified machine, and recording the new values; even if only one row is to be changed. This procedure is the safest and the fastest for this purpose. It is driven by programs which can handle errors and do not allow an update when somebody else is using the table.

The measurements are added directly to the tables, and it is possible to check if the row is not already there. They are transferred to the database either directly from the field microcomputer or by hand from the keyboard of the terminal. A "menu" enables access to several updating programs.

3.2 Data Access

ORACLE has three tool kits with which forms can be defined on the terminal screen. These forms are used to present the information usually needed for consultation.

Moreover, the user has a direct access to the system and, at any time, can give the commands he wants in SQL language.

3.3 The Transition from database to data processing

However complete a database may be, it is of little benefit if consulted only as a dictionary. It is very important not only to get the data on the screen, but also to be able to use it directly in existing FORTRAN or PASCAL programs by writing selecting commands in the programs. This possibility exists with ORACLE system (Fig. 2) which also permits the storage format of the measurements of a geodetic network to be changed before computation.

Figures 3 and 4 give an example of the storage of the measurements.
1. SELECT * FROM CALIBRATION
   2. WHERE NO_FIL='EPAS'

<table>
<thead>
<tr>
<th>NO_FIL</th>
<th>REPORT</th>
<th>APPROX JOUR</th>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPAS</td>
<td>16757</td>
<td>2108 07-JAN-86</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIEU</th>
<th>JOUR</th>
<th>NO_APPAREIL</th>
<th>CODE POINT1</th>
<th>POINT2</th>
<th>VALEUR</th>
<th>Ecart_type</th>
<th>NO_FIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPA</td>
<td>04-FEB-86</td>
<td>INVO2</td>
<td>5 EPA</td>
<td>0000006 EPA00000001</td>
<td>4531</td>
<td>.00005</td>
<td>EPAS</td>
</tr>
</tbody>
</table>

IF1) SELECT POINT1,POINT2,ROUND(((REPORT/1000)+(VALEUR/1000000)-(APPROX/1000000)),5)
2. FROM MEASURES M_CALIBRATION C
3. WHERE C.NO_FIL='EPAS'
4. AND M.NO_FIL='EPAS'
5. /

Fig. 2 Use Example

Fig. 3 Classic handling
When a database is not used, a multitude of files is created each day, containing the data for each sort of measurement. All these files have to be handled and linked together and the space needed is usually double that necessary because of the copies. Furthermore, after checking the measurements - which usually follows the transfer of the data from the field microcomputer into the main computer - it is often necessary to repeat some measurements, for instance if they are wrong or not precise enough. These new data are not usually presented in the same form as the others since the procedure of the second measurement is not so systematic. Thus it is difficult to manage with an automatic procedure. When a database is used, this management is much easier. All the measures are stored as they arrive in only one table. An interactive program of selection gives access to all the data at the time of the computation.

The selection program can be included in the computing program. Thus, the data are used directly; it is not necessary to copy them. For the compensation programs, which are important, it is preferable that this search of data takes place before the computing, and separately. The data is then grouped into files to be used normally by the computing.
program, which saves a lot of time for the machine. In fact, it is quite expensive to run large programs interactively.

4. CONCLUSION

Today's instruments, with the ease of data transmission and automatic collection are greatly increasing the quantity of data. However, a relational database simplifies enormously all the problems linked to the management of the ever more numerous measurements. In the case of the information relating to the geometry of the CERN accelerators, it is the guarantee of safe storage and an easy and rational manipulation of the information.

A database must be evolutive and able to adapt itself to new methods of measurement or computing, which can be introduced at any time. It should no longer be regarded as a background technical issue, but should be approached as a pragmatic strategy for surveying and implementing integrated data-oriented systems.

The survey database fulfils exactly these requirements.

*   *   *

REFERENCE

METROLOGY FOR EXPERIMENTS

C. Lasseur
CERN, Geneva, Switzerland

ABSTRACT
This paper outlines the approach to metrology of the Experiments Section of the Applied Geodesy Group at CERN. This procedure has been developed to deal with the continuous evolution in size and complexity of modern particle physics experiments. The precision in construction and the resolution of these experiments necessitates the adoption of geodetic methods to achieve the required sub-millimetric positioning accuracies. The philosophy behind the adopted approach and the methodologies applied are discussed with respect to collider physics in general and the future LEP experiments in particular. The instruments employed are briefly outlined stressing the usage of a complete and versatile chain of data capture and calculation.

1. INTRODUCTION

In October 1984, the Nobel Prize for physics went to Carlo Rubbia and Simon van der Meer for the discovery of the W and Z bosons, which produced experimental proof of the electroweak theory. These discoveries rank among the greatest achievements in the history of science and were the climax of technological excellence and teamwork on a scale never before seen in the field of pure science. Among the many elements required to facilitate this discovery was the design and construction of a new machine, the Antiproton Accumulator, the adaptation of existing machines, namely the SPS and PS, to fulfill roles they were not designed for, and also the construction of general purpose detectors of hitherto unseen size and complexity.

The Applied Geodesy Group was fully involved at all stages of this challenging and exacting project; in the extension of the geodetic network required by the new works, in the civil engineering work both above and below ground, in the metrology of the accelerators new and old, and finally in the construction of the two new experiments UA1 and UA2 in underground caverns on the main SPS tunnel. These experiments had to be assembled with great accuracy (often sub-millimetric) in a confined space and this challenging precision was demanded for experiments of an unparalleled size and complexity; in the case of UA1, a large box 10 x 5 x 7 m which weighs over 2600 tons (Fig. 1), and for UA2 a mere 600 tons but more complex and innovatory in structure.
The purpose of CERN is to provide opportunities for research in particle physics. Accelerators give the particles the required kinetic energy (hence the name high energy physics) before being brought into interaction with other particles. The properties of particles are deduced by looking at the decays of collision by means of equipment called detectors. The experiments can be divided into two groups, following the way the accelerator is used (Fig. 2):

- fixed target experiments: beams are ejected tangentially from the accelerator and strike a target. The particles so produced continue (for the most part) in the same direction as the beam, hence the detectors are grouped downstream of the target;

- collider experiments: in this case two beams travelling in opposite directions are brought into collision. By this means, the highest collision energy ever achieved in controlled conditions is realised. The secondary particles produced in such a collision fly out in all directions, hence the detectors surround the interaction point.
The LEP project will operate in collider mode. The experiments under construction will surpass those of the SPS in size, accuracy and complexity. For example L3, a five storey structure weighting over 12000 tons and permanently placed on the LEP ring is predicted to produce a resolution of several hundreds of microns. The three other experiments (DELPHI, ALEPH and OPAL) will be mobile, less massive (around 10000 tons) but produce a similar resolution.
2. HOW EXPERIMENTS ARE CARRIED OUT AT CERN

To analyse what happens, the experimental physicist must be able to measure, with great precision, the properties of the particles emerging from the collisions, such as their direction, time of passage (speed), energy, electric charge and mass. A variety of techniques can be applied.

A selection of these is summarized in the following table:

<table>
<thead>
<tr>
<th>PURPOSE</th>
<th>TYPE OF DETECTOR</th>
<th>PROCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of passage measured to the</td>
<td>Scintillator</td>
<td>Charged particles produce tiny flashes of light; when they cross blocks of plastic and this is detected by photomultipliers.</td>
</tr>
<tr>
<td>accuracy of the best electronics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mm survey precision)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direction/location accuracies of</td>
<td>Multiwire proportional</td>
<td>Ionisation caused by charged particles passing through a gas is picked up by an electric field between closely spaced planes of wires.</td>
</tr>
<tr>
<td>a few hundreds of microns</td>
<td>chamber</td>
<td></td>
</tr>
<tr>
<td>(sub-mm survey precision)</td>
<td>Drift chambers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time projection chamber</td>
<td></td>
</tr>
<tr>
<td>Mass/momentum sign of charge</td>
<td>Magnetic field (cryogenics)</td>
<td>Magnetic fields curve the path of charged particles in opposite directions depending upon the sign of their electric charge. The amount of bending measures the particle's momentum.</td>
</tr>
<tr>
<td>(mm survey precision)</td>
<td>Magnet (a few tesla)</td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>Calorimeters</td>
<td>The energy deposited in a gas by a charged particle can be deduced from the amount of ionisation it produces. Iron plates can be interspaced with apparatus for measuring energy deposited by absorbed particles.</td>
</tr>
<tr>
<td>(sub-mm survey precision)</td>
<td>(hadron and electro-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>magnetic)</td>
<td></td>
</tr>
</tbody>
</table>

Every one of the CERN experiments involves complex detection systems of the types mentioned above; they surround the region where the collisions or the hits occur and record properties of the emerging decay particles. One can describe proton-antiproton detectors (UA1 and UA2) and the future LEP experiments as sets of Russian dolls, each being designed to detect a certain phenomena such as those described in the above table. Experiments of this type are designed to cope with large numbers of particles, collecting unbiased information from collision products collected over a solid angle as large as possible. UA1 contains 6000 sense wires, 30 km of extruded aluminium required by special positioning detectors; UA2 includes 240 cells pointing towards the centre of the interaction region and 5000 sense wires. To sift and analyse the massive amounts of data produced requires the speed and power of digital computers. The wealth of information recorded makes for prodigious data handling problems bearing in mind that each collision produces enough information to fill a telephone directory.
3. THE SURVEYOR'S ROLE

At the outset, the support given to the experimental teams in setting up physics experiments was purely empirical.

In order to obtain good results in the analysis of events (reconstruction of particle paths), a precise knowledge of detector positions, both with respect to the beam and respect to each other, is mandatory. Experience has shown that doing this using off-line alignment programs based on cosmic events or track fitting is tedious.

The large and complex structure of a modern physics experiment, made up of many individual modules, requires the use of a careful and systematic surveying and alignment procedure (often sub-millimetric) to derive the full benefit from the quality of the detector components and off-line software. The use of coordinates provided by the Experiment Section of the Applied Geodesy Group as approximate coordinates (raw data) in the off-line alignment programs allows a considerable saving in computer time.

The surveyor has two main responsibilities:

- towards the physics collaboration to ensure that the dimensional parameters of the experiment are fulfilled during fabrication and assembly and to give the final position once the whole detector is set in the particle beam (data-taking position).

- towards the machine group to ensure that common equipment (accelerator / experiment) such as experimental targets, magnets and virtual interaction points are positioned inside the required tolerances with respect to the theoretical beam position.

3.1 Geometrical approach of building a physics experiment

The use of increasingly complex detectors precludes a "Do it yourself" approach. For the physicist, the definition of the accuracy of the results is of primary interest. The global precision for an experiment is a combination of, firstly, the internal precision of each of the detectors and, secondly, the relative positioning accuracy. The surveyor's service is of importance in the second category.

The individual components of an experiment at CERN are rarely manufactured in the same place and are first fitted together on site during the final assembly of the experiment. Thus good accuracy of fit is required to avoid last minute machining and modification which is costly both in time and money.

There are a large number of measuring devices available for small items. When large objects have to be measured the erection or creation of large (often specialised) measuring devices has to be envisaged. These devices often require a considerable associated infrastructure, can be the cause of undesirable side effects (interference, work interruption, deformation) and do not give a global view of the object in its correct position.
The 3-D metrology method gives a picture of an object in a certain position at a given instant. This picture is defined in a referential frame whose main directions are known either in a general reference system or one related to the object itself and the coordinates so produced can be compared to the theoretical values, so ensuring that tolerances are respected. The method is both direct and of universal application; spatial coordinates of specified points on an object providing a suitable basis for the computation of spatial distances, body shapes and dimensions.

The methods and principles that will be explained here are, for the most part, based on our experience with UA1 and UA2 and our preparation for the LEP collider experiments. If we continue with the analogy of detectors as Russian dolls, the scheme of survey operations is that as each is inserted, one inside the other, their relationship is established geometrically. Once assembled, only the largest one is visible; the position of the smaller ones is reconstructed, assuming the geometrical relationship established during assembly is maintained.

The definitive precision of the experiment must include the various precisions in the geometrical relationship whose inaccuracies must be masked by the "noise" in the definitive precision in the working position. Reproducable accuracy (when remeasuring) must be equal to stability (of the object and the reference frame) accuracy, both of which have to be less than or equal to the relative precision between the components. This requires:

- the use of high precision instruments and methods at all stages (manufacturing, assembling, definitive survey).
- comprehensive surveying computation methods to determine an object by spatial coordinates.

We will particularly focus on these features of our work.

3.2 Necessity of good understanding

An agreement between the physicist and surveyor has to be built up from an early stage so that they understand the service we offer and, of course, the limits of that service. Basic working premises have to be clearly expressed, such as the global positioning precision required by the experiment, the precision and position of certain critical elements and, of course, working conditions. Every step should be defined in advance, in writing, and be included in the planning, from the initial to final stages. A comprehensive agreement has to be established covering all steps of the assembly and subsequent measurements. Any weakness in the overall project coordination can create problems, such as invisible references, means of adjustment unattainable, lack of preparatory measurements; the result being detectors either unmeasurable or only with reduced precision.

Prior to the installation of any element, be it a detector module or even a major support structure, it is mandatory that a computer file is prepared on the internal survey
data. This gives a complete description of the unit (position, shape, dimensions) with respect to a set of fiducial marks. The latter are physical and exterior points attached mechanically to the object itself, whose position (rectangular or polar coordinates) are known with respect to a reference of use to the physicist (for a wire chamber this could be the principle sense wires). There must be enough points to fully describe the detector.

The external survey data file is a set of geometrical parameters describing the spatial position of a detector with respect to the theoretical particle beam position when in the data-taking position. It is an adaptation of the internal survey data file following measurements when the experiment is on-line. The definitive document is called the external survey data-base and it is the surveyor's role to determine i.e. to provide the spatial coordinates of the whole set of fiducial marks. Therefore, for the surveyor, an experiment is a list of XYZ coordinates. He has to ensure the real significance of this data, keep them up to date and be able to define the quality of the results.

4. SURVEYING PROCEDURE

The most frequently used procedure in surveying for the determination of spatial distances, body shapes and dimensions is the so-called micro triangulation-trilateration method. Theodolites or length measuring devices are set up at the end points of a baseline whose length and height difference are known to the required degree of accuracy. Precise measurements of the horizontal and vertical angles and/or distances permit determination of the X, Y and Z coordinates of the reference points on the object. If the measurements have to be repeated or if several sides of the object have to be surveyed (i.e. additional bases are necessary), it is worthwhile establishing a network of fixed points whose coordinates are known from a preliminary survey. Fixed points can be installed either as instrument stations, pillars, brackets on walls, or as sighting marks. The establishment of a network has the advantage that the location of the instrument can be freely selected, depending on the shape, size and position of the object to be measured.

To achieve sub-millimetric precision, the Survey Group has made certain choices with regard to instrumentation and methods. These include the intensive uses of distance measurements, calibrated instruments, force centring, well structured reference networks, redundancy, rigorous computer programs and a geodetic approach to the problem.

We will describe certain instruments and methods used by the CERN Survey Group: not all of them are universally applicable, indeed often they were specially designed for high-precision measurements in the experiments.

4.1 Precision requirements

The accuracy of the position of an intersected point depends on the accuracy of the calibration scale, from which the baseline data are derived, and from the observation
distances (when the coordinate computation is based on the measurement of angles). Further accuracy determining factors include the magnitude of the intersection angle, the quality of the object point marking, the stability of the instruments, illumination conditions and of course the shape of the object itself.

4.1.1 Difficulties in measuring angles

The prerequisite for measuring angles is the absence of temperature gradients along the optical path; curvature of a light ray varies with the angle between its trajectory and the normal to the surfaces of equal temperatures: over a distance of 100 m and for a gradient of 1°C, it is deviated by 1 mm. Gradient of temperature, grazing sights and unequal lengths of sight all make optical measurements unreliable.

An ordinary theodolite telescope is unsuitable because of its unstable collimation which can change by as much as 30 seconds of arc throughout its range of focus, usually 1.8 m to infinity. When one seeks very high accuracies (0.1 mm or better), the position of the line of sight must be known very accurately with respect to some physical point. Moreover, the three axes, nominally orthogonal, do not, except by chance, pass through a common point. The optical tooling telescope seems to be essential for all serious accurate work (Taylor-Hobson: line of sight central to the cylindrical tube to an accuracy of 0.005 mm and collimation maintained within 2 arc seconds). However this equipment needs special features (top plates, trivet stands) which are not flexible enough for our purposes.

For all these reasons precise surveying networks have to be formed as much as possible of linear measurements.

Angle measurements are still used for the "details" in experiment metrology because the sighting distances are very short – a few metres and most of the halls at CERN have a temperature regulation system. Also in some cases, pure triangulation is the only suitable method because the object to be measured is not accessible directly or cannot be subjected to undesirable loads (deformations); the angle measurements made externally exert no influence on the structure and can be carried out on an object in use. Nevertheless, it is prudent to employ a combination of angle and distance measurements whenever possible.

4.1.2 Forced centring systems

To obtain high accuracy, and because of the short distances involved, it is obvious that the entire geodetic system must use forced centring systems. The standard socket used is machined to better than a few microns and the top of the socket can receive either a positioning cylinder (with which every instrument is equipped) or Taylor Hobson spheres. The bottom part of the system can be adjusted precisely in a given theoretical position, the reference being the centre of the 3.5 inch Taylor Hobson sphere set in the
cup. The top part can be adjusted to the vertical so that all measurements are processed in the horizontal plane.

4.1.3 Object point marking

The type of reference is crucial for the accuracy of the determination. If there are no well defined points on the object, such as an edge, a corner or a punch mark, artificial marks must be added prior to measurement. In many cases the Taylor Hobson sphere is too big and heavy, and requires special additional equipment which renders it unsuitable for object point marking.

Precise reference holes of different diameters designed to receive plug-in target marks adapted to the measuring device (angles, distances) and easily visible from many directions are of more universal use. This system offers the most flexibility in the choice of measuring pattern because the forced centring disposition allows us to change the targets without affecting the repeatability of the measurement. Other possibilities are self-adhesive sighting marks (crosses, concentric rings, symbols taken from transfer films), while various target patterns are available for different conditions of sight distances and illumination. They do not require any special machining but only angle measurements can be made with them.

Sometimes it is impossible to make marks or attach references to the object. A possible solution is to use a laser whose beam is projected via the optical axis of the theodolite so that the target point thus created can be intersected by other theodolites.

Forced centring systems as developed at CERN appear to be the essential condition in high-precision measurement because of their reproducibility and flexibility of usage. They have helped us to perform measurements whose quality is similar to those achieved in laboratory conditions.

4.2 Technical surveying milestones of an experiment

4.2.1 Reference network

Setting up a micro reference network seems to be one of the more convenient ways to determine three-dimensional patterns. From the design stage up to normal running of the experiment, it ensures that the required accuracies will be attained. It is successively used for provisional installation, alignment of detectors within the tolerances, measurement of detectors in the data taking position. It also allows survey of the stability of these components; successive measurements, based on the network, attempt to maintain a correlation between the evolution of the components and of the geometry itself.

The network is really the fundamental frame of all survey activities. It acts as a stable and precise large calibration bench which surrounds the volumes to be measured so that the critical points are visible and accessible physically. This bench is defined by
pillars, brackets and/or rivets, whose positions must comply with a certain number of criteria:

- reference points must be easily accessible. The density of points has to be great enough to cover the objects to be surveyed.

- structure must allow simple, precise, easy and quick measurements of whatever kind on the points to be determined and between the reference points themselves.

- network structure must from the very beginning take account of the different steps of installation and be evolutionary enough for it to grow relevant to any future needs.

- key positions must be occupied in respect to the layout and be adapted to the point’s projected usage (angles, distances, direct levelling).

- creation must be made as soon as possible and in one operation for the homogeneity to be maintained; completion in hindsight must be avoided.

- installation avoids unstable areas while allowing rapid measurements to be performed to detect and evaluate casual movements or accidents.

The assembly and positioning tolerances guide the homogeneity and the precision to be obtained when measuring and controlling the network, the threshold being given by the most sensitive elements of the experiment (which are generally the most difficult to reach practically). It is useful to attain the best possible accuracy under current conditions since even if this is not strictly necessary, it can be of great help in the future as the experiments evolve. Homogeneity means that the precision of the bench must be identical at any location of the network. Redundant measurements must be performed to ensure quality and uniformity of accuracy. These points must be considered when setting up the network since good homogeneity and density provide us a greater flexibility of methodology and control.

The form of the network depends on the type of the experiment. For fixed target experiments, the network consists of two lines parallel to each other and to the beam. The method of defining these lines depends on the precision required, for example, the line can be defined by survey sockets inset into the floor or, in a less demanding case, be simply drawn on the floor with several rivets placed along its length. For collider physics the network consists of several levels of references commanding both the maintenance and the on-line position of the experiment.

In the case of the UA1 and UA2 experiments, references are situated on three distinct levels (+ 6.0 m, 0.0 m, - 3.0 m with respect to the beam). The geometrical links between the floors were achieved using high precision verticality measurements. These were carried out between brackets on the same vertical (within several centimetres) for the case of UA1, whereas in UA2 sloping invar wires were used. The following results were obtained:

- UA1 : 38 brackets – 160 measurements – range of accuracy .08 -.17 mm
- UA2 : 56 brackets – 270 measurements – range of accuracy .08 -.18 mm
These results were only possible because all the requirements quoted above (length measurements preferable to angle measurements, forced centring) were met. Unfortunately, the network cannot always be established in its ideal form; it must be compatible with the other services such as ventilation conduits, cabling and cooling systems which greatly reduce the options available.

4.2.2 Detector assembly

Preparation phases.--

A detector is rarely a single unit. It is often divided into component parts which we can term sub-detectors which geometrically are considered as single, rigid and non-deformable units. Each assembly of sub-detectors is considered to produce another rigid object.

To plan the measurement procedures, it is of overriding importance that the steps of assembly are clearly defined, this normally being determined by the structure of the detector. Specific discussions can then start on any remaining ambiguities, and the following questions answered:

- is the detector a single unit or composed of several independent modules?
- is it rigidly mounted on another detector? How is the relationship between the two units to be made (mechanically or by means of surveying)?
- is its determination to be made sequentially (in a laboratory, in the experimental area)?
- have external reference marks to be created? If so how many and where?
- is there a supporting structure or a mechanical adjustment?
- has the detector (as a unit or modules) to be placed at a theoretical position, in respect to what (beam, surrounding detectors) and with what precision?
- what precision is needed for the different detectors relative to each other?

The surveying procedure can then be decided upon; the best place for reference marks, the redundancy required, the most suitable moment and place for the work during the assembly of the detectors, etc.

Steps of the geometrical determination.--

To obtain the final version of the survey database, a cascade of operations is needed, from the first availability of individual objects up to the experiment's data-taking stage. The steps of geometry result from principles whose sequence follows closely the following assembly phases.
Internal geometry.

Each individual detector element (chamber or counter module) is equipped with fiducial marks defined in its own coordinate system (for example: the main axis of symmetry). They are deemed to be internal because they are part of the object. These marks can be the accessible sensing part of the detector, for example the cells of a calorimeter or wires directly visible from the exterior of the chamber.

Another possibility is reference marks directly connected to the sensing part of the detector; for example the pin holes used to centre planes of wires. Each plane is held within a frame and at each corner there is a reference hole whose diameter is precisely machined and whose position is known with respect to a wire (measurements carried out in the laboratory). If the chamber contains several planes of wires the relationship between each plane is assured by calibrated cylinders which fit precisely in the reference holes. From the measurement of the position of these cylinders, the position of each plane of wires and the position of each wire can be deduced.

Features of internal geometry are that it is carried out as an integral part of the manufacturing process and is used for geometry when easily, directly and permanently accessible. It is the preferable way of defining the geometry of the detector as a direct method of determination.

Link geometry.

Measurements in this category are carried out for one of two reasons. Firstly, if the internal marks are not convenient for surveying, special survey references must be installed. These have then to be situated with respect to the internal geometry by geometrical, optical or mechanical methods, carried out in the laboratory, workshops or assembly halls before the definitive installation of the detector. Secondly, when a detector consists of several individual modules, then for reasons of economy (many sub-detectors), accessibility of the sub-detectors or to derive parameters for the physics data banks, link measurements between adjacent modules become necessary. These are related to marks (at least three) which will be accessible when the detector is in the data taking position and can be either visible internal reference marks or specially installed auxiliary references. The position of every element of the detector can then be deduced from the measurements of several references.

Features of link geometry are:

- the intermediate step required for the geometrical knowledge of a detector
- the difficulty to foresee whether it is really necessary since it is dependant on the environment (encumbrance of zone etc.)
- that auxiliary marks are not an integral part of the detector (link geometry has to be repeated or plug-in system foreseen)
the use of mixed step link geometry, for example a mechanical plug-in system to extend internal reference marks

that, if possible, it should be carried out with a precision greater than, or equal to, that of the original references; this is often not possible hence loss of precision.

Transformation geometry.-

Before being moved onto the beam line, the detectors are installed one by one in their definitive relative position. Due to the planning requirements, detectors can be assembled at different places in the mounting area before they are fitted together. This allows us to verify that the theoretical parameters of the experiment are respected, and to determine the relationship between different detectors which will later be hidden. An element unmeasured at this stage means that the position of one of the detectors will not be known when the experiment is on the beamline. Equally annoying is a relationship known with insufficient precision since all detectors internal to the one measured cannot then be located with sufficient accuracy.

Features of transformation geometry are:

- the establishment of the relationship between successive layers of detectors

- that, by this stage, the internal and link geometry of the individual detectors must be known (internal data base)

- the measurements are related to an assembly network during system assembly

- the relationships must be recorded at each step.

Definitive geometry.-

Once on the beam line, the visible references are measured from a network whose zero is the theoretical beam interaction point of the accelerator. Consequently, the position of all the detectors and sub-detectors can be calculated using transformation programs, whose truth depends on the care with which the measurement procedures were planned and the rigidity of the detectors. These results constitute the external survey data base required by the physicists.

The main feature of definitive geometry is that the position of all detectors is deduced from the measurement of the visible references of the experiment in the system of the machine.

The geometrical phases described above require much work and it is important that the number of cascaded operations should be reduced as much as possible. To this end non-essential movements of the experiment should be avoided as this requires new measurements and spatial transformation programs to determine the new shape and dimensions of the ensemble.
The final result, if it does not come from direct measurement, will be affected by the successive adjustments introducing slight differences due to the computations. To reduce this loss of accuracy, a high precision (homogeneity of network) in measurement is needed so that intermediate survey steps are not biased by successive determinations.

Possible geometrical method.

Methods must be comprehensive enough to give a true 3-D picture of the object. The techniques used in most cases will be a mixture of the following: angles, distances (horizontal, vertical and spatial), offset measurements (optical or physical, in the horizontal or vertical plane). To the measurements made in the field can be added internal distances. That is, high precision values known from the manufacturing of the object or from calibration measurements performed in the laboratory. These values can be used in several ways:

- verification: to check that there has been no distortion of the object after movement (transportation, handling) by a comparison of the deduced internal distances and the known internal values. They can serve equally to verify the scale factor of the network (with an appropriate weighting). It could be envisaged to measure an object using auxiliary theodolite stations where the precise scale of the base is derived from the internal distances of the object.

- determination: these redundant distances are often known to a higher degree of precision than we can achieve in the field. They are also direct measurements made on the object itself. Thus, through their precision and their redundancy, they optimise the 3-D determination. In the case where points are unattainable or too numerous, internal distances between unobserved and observed points are a mean of deducing the unmeasured points.

Methods depend on the following considerations:

Characteristics of the object.— Its shape, dimension and behaviour (a priori instability in movement or a light structure).

Immediacy.— Because of fragility or limited availability of the object, rapid results may be required for verification.

Reference system.— When lab or preparatory works are necessary a local reference system is required. The choice depends on the spatial situation of the object and whether a direct knowledge of the parameters of the object are needed. Three cases can be distinguished for an object in any position:

- direct knowledge: the referential is directly linked to the object itself. The parameters are measured by trilateration or by triangulation but in the latter case the vertical axis of the theodolite must be parallel to the axis of the object;

- object can be adjusted and direct knowledge: it is obviously much more convenient to adjust the object, for instance with jacks, to produce a truly vertical axis;
- indirect knowledge: the object is measured from a reference system unrelated to the object itself. To obtain dimensions and relations in the object reference system, a transformation is required.

**Environment.** For example not enough space around object for comprehensive coverage; the use of artificial lights with the risk of phase error; temperature gradient rendering triangulation unreliable.

**Number, location, accessibility and type of survey marks used.** Visibility and required accuracy determine the most suitable method. Standard multi-purpose targets appropriate to any kind of measuring are very useful in this type of work, especially when access is difficult or installation of the point has to be permanent. Pattern and dimensions must be chosen in respect of the length of sight and thus have important considerations for the network. Insufficient room around the detector means too many stations (time-consuming) or the need to use special lenses to allow focusing on the objects with consequent loss of precision.

**Non rigidity of the object.** Unfortunately sub-detectors are not always rigid. This can be due to the environment of the object (local vibration) or just a weakness in the structure of the detector (the more sensitive detectors in particular are built with a minimum of material). The non-rigidity can be controlled either by using instruments like gauges, clinometers or by using geometrical methods. The former give punctual information of a high precision, can be placed at critical points, and are independent of the geometrical results. However, this information cannot be directly included in a 3-D determination; Moreover, that needs special equipment support and many instruments to give a complete picture. The geometrical method can provide a relationship between adjacent modules but requires a certain preparation and is relatively time consuming. It requires marks on the object which are not necessarily linked to the internal references. Adding marks in hindsight should obviously be avoided but because of the complexity of the detectors is sometimes the only solution.

The question of whether the detector or the network has moved is obviously a problem, in particular in underground experiment halls where movements of the order of mm’s continue for many years after the completion of the civil engineering work. This requires periodic verification of the reference network, often in conditions less favourable than when the network was established (experiment and services in place, lines of sight blocked). Redundancy becomes more and more essential but less easy to achieve. In any case with the geometrical method it is impossible to determine the precise moment of the movement, the amount and direction at any given time to be allowed for in the calculations and the rotation/translation centre. To avoid some of these problems it is a good idea to use instruments giving punctual information, especially if used to furnish on-line information to monitor the situation. Once the deformation exceeds a certain limit a general geometrical survey can be made. All in all, it emphasises the need for a good preparation of the object, to detect at an early stage any possible movements and allow the possibility to install special equipment if necessary.
Special tools.- Precise extension plugged into reference socket, various shaped "jigs" or pins, plug-in brackets, adjustable supports, targets with scales, self-illuminating marks, ball bearing fitting system to measure a distance in any direction are all of considerable help. Precision of the machining, a rigidity of structure, adaptability, adjustable if necessary and simplicity of use are generally required.

4.2.3 Survey and assembly sequences

During installation, it is essential that coordination is established between surveyor and the person in charge of the installation to ensure timely interventions and that allowance will be made for the need for space and time for survey measurements. A communication system has to be arranged to allow a fast access to the survey data at all steps of the geometrical process. Hence possible inconsistencies between different measurements can be detected during installation so that measurements can be repeated if necessary. This survey file should be used to test off-line alignment programs between the modules already in place (e.g. by Monte Carlo simulations).

4.2.4 Data-taking period: diffusion of results

Survey measurements have to be repeated often because of position instabilities and movements on replacement of modules. The survey results must be entered as soon as possible in the experiment's data-base. It is updated following the progress of the installation and must reflect any last minute changes or modifications. It can be considered as the definitive document between the surveyor and his customer. At the request of the latter, further information such as the precision of the results must be provided. Publication of the results must be meticulous, in particular when designations of the measured points are concerned. Graphics plotting in 3-D can help in the understanding of the results.

In fact, we also need intermediate data banks - specially constituted by the link measurements - and we have to create an adapted computing chain that will be described later.

5. MEANS AVAILABLE IN THE FIELD FOR THE SURVEYING OF LARGE PHYSICS EXPERIMENTS

5.1 Angle measurement

Theodolites (Wild T2 and T3, Kern E2) are used over short distances (a few metres) in conjunction with forced centring systems for detector assembly of medium accuracy, in particular when references are non-accessible or unreliable. Extensive use of vertical angles in conjunction with direct levelling improves the accuracy perpendicular to the base. Angles are also used in network observation when required to improve the strength of the geodetic figure.
5.2 Distance measurement

5.2.1 High precision (accuracy better than .1 mm)

The three main instruments for providing high precision distance measurements are listed below together with their relative merits and drawbacks.

Distinvar.—

This instrument can measure from .02 m up to 50 m with a rms error of .05 mm; due to the instability of the invar alloy, the elongation is directly proportional to usage, hence for work of the highest order the wire must be calibrated before and after measurement on a precise bench. The invar exerts a force of 15 kgf on an object being measured so the structure has to be rigid and requires a standard forced centring socket. It can be utilized on slopes up to 11%. It is rarely used on detectors unless prior provision has been made to accommodate it. Its main use in metrology for experiments is in the measuring of reference networks.

Self aligning interferometer.—

It is used for distances of .1 m to 60 m with an accuracy of .01 mm and does not apply traction to the detector but requires a free path and accessibility between reference sockets. It cannot be used for distances on slopes. More time is necessary but the instrument has the advantage of requiring straight line access rather than the catenary curve which may prohibit use of the distinvar.

The self aligning interferometer system is used for in-field auxiliary bases or for measuring high-precision networks when the reference points are all at the same height and freely accessible (linear network).

High precision bar.—

This new device (A digital micrometer based on differential condensers) will be used for distances up to 2 m and the accuracy expected is .05mm. Any mechanical adaptation can be fitted to make measurements between any kind of reference holes and in any plane, so that is is possible to increase length using precise extensions. Because of its light structure and easy handling, it does not deform the object.

This device is still under development and is expected to be of great use for precise micro-trilateration, in particular for link geometry in the metrology of experiments.

5.2.2 Medium precision (several tenths of a mm)

A stadiometric method has been developed which, although not of the accuracy of the devices mentioned above can cheaply produce good accuracy (.04 mm/m for distances up to
15 m and slopes of up to 20%). It consists of a 3 m invar tape mounted on an H bar. Four accurately aligned marks act as references on the tape, the distance between these points being known from a laboratory calibration. A staff is held vertically in a reference socket, vertical angles being observed to the marks. From these observations (vertical angle resection) can be calculated the distance between the sockets, difference in the height (bottom of staff and theodolite) and the non-verticality of the staff. Corrections can be calculated from the latter. The main attractions of this method are the versatility (especially when working in bad conditions) and the possibility of calculating the distance in the field using a pocket calculator.

5.2.3 Millimetric precision

This is usually achieved by electronic distance measurement (EDM) using, for example, the Kern DM 502/4 instrument. Its main features are an accuracy of 1 mm rms for distances up to 15 m and that it can be used to determine approximate coordinates or 1-2 mm metrology, while Kern mirrors can be adapted to fit any reference holes.

To benefit fully from the EDM, frequent calibration is required (bench or an invar network) in the range of the measured lengths. The instruments used are adjusted at the factory to reduce phase errors over short distances, analysis of calibration measurements having confirmed this for distances up to 30 m. Thus, using linear regression, the zero constant and the scale error can be calculated. Calculation of the standard errors of the zero constant and the scale error allow estimation of degree of confidence. Important points are:

- the calibration constants can be calculated using a pocket calculator (though attention must be paid to rounding errors)
- additional information about the zero error can be gathered from residuals computed and plotted with respect to distance
- it is important to calibrate often since the zero error changes with time and rough handling
- the possibility to use self-adhesive "scotchite" tape (up to 15 m) is very advantageous
- the absolute rotation angles of a large object (verticality, perpendicularity) can be obtained by means of differences in distances.

5.3 Levelling

Here we use an automatic high-precision levelling instrument such as Zeiss Ni2 with the limitations that long traverses cannot be made and intensive use is required of redundant measurements (nodal points) while frequent calibration is necessary since sights rarely of equal length. For convenience a vertical translation stage is used to measure directly on the point without using a staff though for experiments the large height differences often encountered require the use of 3/4 m staffs or vertical invar tapes.
5.4 Special instrumentation

5.4.1 Offset measurement

The offset is designed to replace angular observations by the precise measurement of the shortest distance (offset) from one point to a straight line defined by a nylon wire. Offsets of up to 500 mm can be measured in this way with an accuracy better than .1 mm. The method can be applied over large distances (up to 100 m) but requires stable air conditions, i.e. no draughts.

This instrument is used in conjunction with distinvar and/or interferometer measurements for the determination of linear networks and we intend to use it over short range and in any plane, in particular for redundancy in link measurements when they are combined with spatial distances done on the object itself. However, this special application will demand adaptation of the forced centring device.

For optical offsets a vertical plane is described by a theodolite telescope and an offset reading is made on a ruler which is approximately perpendicular to the plane. These measurements are used to give the verticality of an object and/or redundant measurements. The precision achieved is of the order of .3 mm over short distances.

An active system using a laser has also been used when turbulent air conditions prevent the use of a nylon wire. A rms error of .1 mm over 100 m is possible but this precision decreases rapidly over shorter distances.

5.4.2 Clinometers and hydrostatic levels

Clinometers give an accurate measurement of tilt angles in one direction and the value can be read by remote control. Accuracy is .01 mm/m and tilts of up to 2 mm/m can be measured.

The hydrostatic level is designed to measure the difference in height between several stations. The instrument installed on each station can be operated and read either directly or remotely and can be controlled either manually or by computer. The accuracy is of the order of .05 mm over distances of several tens of metres.

These instruments give periodic or permanent survey of altimetry, deformations or micro-movements. However, they require auxiliary equipment for their installation while generally their size and weight limit their application.

5.4.3 Vertical lines and plumbing methods

When a network is composed of several distinct levels it is necessary to find a way to transfer coordinates between floors. This can be achieved using the distinvar if the
slopes do not exceed 11%. In other cases the transfer is done using vertical lines between stations which are on approximately the same vertical using for example:

- a plumb-bob damped in an oil bath

- a specialized optical instrument: precision nadir plummet with mercury horizon (Wild GLQ)

- a nadiro-zenithal telescope

- optical methods using special adapters such as:

  - the diagonal eye piece for measuring zenithal distances in several planes between the vertical line of the theodolite and the point but only possible if measuring upwards.

  - the pentaprism which when fitted to the telescope of a theodolite describes a plane perpendicular to the optical axis. Observations (horizontal and vertical angles) describe planes whose intersection give the observed point coordinates. For good accuracy, the height difference between the points must be known precisely and the planes described must be well distributed.

Both the pentaprism and diagonal eye-piece methods use redundancy in the determination which can be calculated with a pocket calculator. Accuracies of better than .2 mm have been achieved over distances of up to 20 m.

6. COMPUTER PROCESSING

6.1 Importance to metrology

The variety of measurements and the redundancy, the number of points to be handled mean that least-squares adjustment programs are an essential instrument for calculating definitive coordinates. All surveying operations in metrology depend on this since the calculation gives the observed object as a truly geometrical figure. The definitive form is the one which fits most closely to the observations. Some differences will remain because of the inaccuracy of the methods used but the resultant form is the most probable (statistically speaking).

The following features are essential for a least-squares program adapted to the metrological method:

- capable of handling the wide variety of observations such as horizontal and vertical angles, horizontal, vertical and spatial distances, offsets in any plane, punctual information given by non-geometric instruments (autocolimation, clinometer etc.)

- of sufficient size to adjust the largest data set in one go

- do not influence the method in the field

- give statistical analysis of the results including redundancy factor, standard error a posteriori of the measurements, error ellipses, residuals after compensation, criterions of homogeneity (relative error ellipses)
- independent weighting of variables
- good, clear presentation of the results (histograms, plotting etc.).

6.2. In-field programs for metrology

A chain of programs has been developed to aid the work in the field; their simplicity and versatility of usage being a considerable asset. These programs were first developed on HP41 calculators but the increasing availability of portable computers such as the Epson HX20 has allowed the development of more comprehensive and versatile complementary programs. For example, the determination of one point in X and Y by least squares (resection, intersection, distances and angles) can be programmed easily. This possibility of field computing offers a flexible setting out method which can be adapted to the working conditions.

6.3 Data capture and computation programs

When a large number of points have to be measured in a limited time, automatic recording programs become almost obligatory. They reduce operator strain and the possibility of error both in data taking and in subsequent manipulation of the data, and ensure data is taken in a consistent way. The connection of a portable computer to an electronic theodolite means that angles (and distances) pass directly to the computer, the data is then checked (verification, between two pointings, collimation of the instrument) automatically and directly. In the case of error the operator has the possibility to verify and correct the observation.

An extension of this idea is to calculate automatically (intelligent recording program) the results in the field (distance/bearing, intersection, resection, levelling), which are then stored in the memory of the computer. This is used when there is little or no redundancy, and means that no computing is required in the office. If the measurements are periodic the previous values can be stored in the memory thus allowing a direct check before replacement by the new value. Observations or coordinates are then transferred to a larger computer for distribution and eventual further calculation.

6.4. Main core compensation programs, specific concepts

6.4.1 XYZ determination

Comprehensive planimetric, altimetric and spatial compensation programs have been developed over many years and are being continuously updated to take account of future needs. All are based on the variation of coordinates algorithm and specially cater for:

- offset measurements in any plane defined by the three connected points and offsets taken from a theodolite set on a given bearing
- gyroscopic bearings
- plumb-bob observations
- computation of any network as a free network (one fixed point and one specified direction), thus a minimum of constraints
- correction of distances by a scale factor or by a systematic constant after calibration
- the a posteriori transformation of angle residuals into radial vectors which are much more appropriate to short distance determination
- tracing of remaining mistakes by entering approximate coordinates as fixed points, big discrepancies between observed and calculated values highlight problems.

6.4.2. Simulations

The possibility of simulating future survey operations is intensively used (especially for the preparation of the LEP experiments). The methodology is established (set of measurements, choice of points), the program then calculates the theoretical measurement and adds an appropriate error calculated from a random number table and the a priori weight of the observation. The "observations" are then adjusted and the results calculated. The program repeats this several times (normally 10 is sufficient) using different random numbers. The precision of the definitive network can be seen plus the range of values for the definitive coordinates. If the program is used on several possible data sets, the most favourable pattern can be found and it can be seen if the required accuracies can be achieved. This allows planning and equipment requirements to be established in a rigorous way and well in advance.

6.4.3 Spatial-adjustment program

Spatial-adjustment programs are imperative in metrology since they give the complete set of coordinates even if only some points are remeasured from a previously established relationship and the possibility to change the reference system following needs (providing the relationship between the main axes is known). To define a coordinate system without ambiguity seven coordinates distributed among three points are needed (i.e. X1, Y1, Z1; X2, Y2, Z2 and Z3). Alternatively, six coordinates and a distance constraint have the same effect. In all of these cases, the relative geometry is not distorted and observation residuals are always the same. In an over abundant situation the network will be adjusted after the principle of least squares; this will produce inevitable inconsistencies in the residuals of the observations though these will be small or negligible. A set of three points known in X, Y and Z constitutes a redundant control set.

The decision to use minimal or redundant control should be based upon confidence in the control coordinates versus that in the measuring system and has important implications for the definition of methodology. If an item (network, detector) is periodically remeasured, then two cases are possible. First, all the points are used to fix the parameters of adjustment and the residuals dx, dy and dz prove the item to be stable or not. Second, due to changes in the environment the quality of the measuring method can not be maintained so that the relationships established from a previous measurement (in better conditions) have to be imposed to improve the quality and reliability of the
results. However, in an over abundant situation the distortion of the results by nonconform points is always a danger (micro-movements, instability). An a priori limit has to be defined to allow the rejection of suspect values which would otherwise bias the adjustment parameters (translation, rotation in the three planes and scale). Thus stable points can be selected and new coordinates computed for passive points.

6.5. Possible computable steps in metrology

6.5.1 Applications to periodic survey

In frequently repeated operations a methodology has to be developed to explain discrepancies in position. A comparison with the previous data set allows rapid detection of gross errors and gives an idea of possible discrepancies in position. Then computing a network as a free system (one fixed point one fixed direction) and adjusting the whole set of coordinates of the new network in relation to the old network is a possible way to detect stable points. The values of the residuals give a guide for introducing rejection criteria and give information on the accuracy; so a new adjustment (using so-called fix points) provides a direct comparison (absolute and relative positions).

Often the items to be measured are more stable than the network itself. This emphasises the need to use non-geometrical devices to control stability and the fact that a reference system related to the object itself (directly or through the use of internal distances) is often the most suitable way of obtaining true positions. When internal distances on the observed object impose unacceptable constraints the stability of the item or the so-called fixed points must be open to doubt (particularly if the network was not remeasured). Then, solutions have to be found to resolve the dilemma, such as computing network and object together (free calculation, one point fixed and one direction fixed) and processing successive spatial adjustment of the ensemble, taking different values of rejection criteria into account. The object may thus be used to relocate the network. After final adjustment the magnitude of the residuals dx, dy and dz gives an idea of the "exactness" of the final solution but, to envisage a solution along these lines, there must be redundant information otherwise casual deformations can not be brought into evidence. In any case these can only be detected if they are greater than the rejection value.

Example of link-transformation geometry adjustment.-

Let us consider the detector D3 which is part of the experiment UA2. The detector is composed of four identical modules each composed of four boxes containing three detector chambers: the physicists require the position of each of these 48 units. To achieve this, four distinct operations were required:

- laboratory measurements (carried out on a calibration bench) to an accuracy of a few hundredths of a mm (internal geometry),
- two separate operations of triangulation to establish link and transformation geometry,
- final triangulation to find position of the detector D3 in the experimental network (definitive geometry).
This entailed three successive adjustments, using the intermediate sets of coordinates, to provide the position of the detector chambers in the experimental network, the whole operation requiring more than 2700 angle measurements. However, in the following table, we can see that despite the number of operations, a sub-millimetric accuracy was still achieved.

Residuals after each of the adjustments
(values in mm)

<table>
<thead>
<tr>
<th>Measure 1</th>
<th>Measure 2</th>
<th>Measure 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>Y</td>
<td>0.05</td>
<td>0.17</td>
</tr>
<tr>
<td>Z</td>
<td>0.06</td>
<td>0.10</td>
</tr>
</tbody>
</table>

7. Specific Features for the LEP Experiments

7.1 Computerisation

The size and complexity of the LEP detectors and the constraints imposed by the assembly and installation schedules require an extension of the geometrical methods. All relevant theoretical data and results files have to be available in the field, allowing a direct comparison of theoretical and measured parameters. Thus, the problem is one of data management rather than field technique and requires the use of a comprehensive system of data capture and calculation.

7.1.2 Features of the system

An internal disc drive allows permanent storage of theoretical parameters of an experiment (detectors, sub-detectors), coordinates of network, measurements and results. Use of the Unix system of exploitation enables logical subdivision into as many parts as there are units, while least-squares 3-D compensation and adjustment programs can be introduced into the computer.

The system can be interfaced with different configuration such as main frame computers, (IBM, VAX), electronic theodolites (master and/or receiver), and micro-computers (HP-IPC, Epson HX20 and PX4) used as data loggers. Multi-task functions allow the system to be used in a versatile way (acquisition, control of instruments, calculations, communication lines) to reduce data handling, while special routines and facilities help in producing comprehensive description and analysis of the results.

7.1.3 Expected benefits

Compensation and adjustment programs facilitate on-line comparisons either with internal fiducial-mark coordinates (theoretical values) or with previous results if the
current measurement reveals displacement. Since verification is made in the field, any
discrepancy (residuals given by the on-line capture and computations) can be checked, thus
giving greater confidence in the final results. This produces considerable time saving
since assembly, measurement, verification and correction are carried out simultaneously.

7.2 Instrumentation

The complete system consists essentially of two parts, namely electronic theodolites
and computer. The former is the Kern E2 (with DM502/504). It is used singly or in
combination (several E2s can be independently controlled through one interface RS232 and
via ASCII SINGLE BUS System). Readings are taken and the circle can be set remotely.
Computer equipment includes the data loggers and the central unit of management. The HX20
and FX4 will continue in their role as individual data loggers either remotely or through
keyboards. Their portability, robustness and practicality plus the battery of existing
programs make them ideal for data capture and simple calculations, both in the field and
in the office. The management unit will use the HP-IPC as its master unit. Its compact
design (internal disc drive, screen, printer), its capacity (1 Mbyte), its versatility
(programmable in BASIC, FORTRAN and C, system of exploitation UNIX, easily interfaced with
other apparatus) make it well suited for this role.

8. CONCLUSION

Progress at CERN is not limited to the domain of physics and the different methods and
instruments presented in these proceedings underline this. The rapid evolution of high
energy physics constantly poses new problems to the surveyor and in meeting this challenge
prosper is made. Metrology as described in this paper has benefited greatly from recent
advances in electronics, data handling and computing. The new automatic instruments
(theodolites and short distance devices) have greatly aided us. This however, has not
blinded us to less "glamorous" but sometimes more effective solutions. Organisation of
field and office work becomes more and more important in order to meet the needs of our
customers in a more efficient way. Full benefit from the possibilities of computing
requires considerable forethought, organisation and self discipline.

It is indispensable in our case to have a realistic idea of the precision (standard
errors, confidence limits). This requires redundancy of measurement and often the only
possibility of achieving this is by a mixture of "unusual" observations requiring
comprehensive and flexible compensation programs. To handle any wide ranging alignment,
portable computers are essential in order to manage the flow of data, calculate the
results and to stock the results in a data bank.

ACKNOWLEDGMENTS

The author would like to thank D. VEAL for his help and his large contribution to this
paper.
GEODETIC NETWORKS FOR CRUSTAL MOVEMENTS STUDIES

P. Baldi and M. Unguendoli
University of Bologna, Italy.

ABSTRACT
Starting from a short review of the geophysical phenomena representing the major causes of crustal movements, we analyze some problems related to the geodetic networks used for the detection of the movement them selves. Some examples of real networks set up in Italy are also shown and discussed.

1. INTRODUCTION

The application of Geodesy to the observation of crustal deformation can be divided into two main categories: the continental scale monitoring of the relative movement of tectonic plates, and the regional or local scale observation of deformation connected with seismicity, volcanic activity and subsidence. On an even smaller scale one might study the phenomena of landslides and land settlement etc.

Since the relative movement of continental blocks is estimated to be within the order of a few centimetres a year, the relative positions of the points need to be defined with an accuracy of $10^{-8}$ to $10^{-9}$. This is now made possible by the use of Satellite Laser Ranging and Extragalactic Source Radio Interferometry (Very Long Base Interferometry).

On the other hand, the use of artificial satellites, particularly those using the GPS (Global Positioning System), will in the next few years furnish us with an exceptionally versatile instrument for small and medium scale geodetic applications; the use of a constellation of 18 satellites permitting one to determine the relative positions of points on the Earth's surface with an accuracy in the order of 1 cm.

However, classical geodetic techniques, i.e. the repeated measurements of distances, angular height differences, direction and intensity of the gravity field, remains, at least for the time being, the most effective way of observing and checking deformations associated with seismic activity and/or other local or regional scale geophysical phenomena. Here crustal movement is generally subdivided into horizontal and vertical deformations. The techniques commonly employed consist of triangulation and trilateration of reasonably complicated networks for the former, whilst vertical deformations are studied by means of geometric or trigonometric levelling integrated with gravimetric, with clinometric, extensimetric and mareographic measurements. In some cases, such as mountainous areas, plano-altimetric networks can also be used, integrating the various types of measurement, and including astronomic observations in order to obtain three-dimensional networks.

There is no doubt about the importance of studying surface deformations that may be connected with geological and geophysical phenomena taking place beneath the Earth's crust; one need only think, for example, about the possibility of monitoring in detail perhaps the most fascinating, though still not well-defined, phase of the seismic cycle taking place before an earthquake and characterised by the presence of the phenomena known as precursors. This phase, observed prior to several earthquakes, can be connected with changes in the physical properties of rock subjected to tectonic forces, according to models amply discussed by various authors. Without doubt, anomalous deformations of the terrain seem to be the most easily recorded signals of such activity among the various other effects observable, (namely the velocity variations of seismic waves, changes in rock resistivity, etc.).

Monitoring of the deformations in volcanic areas using geodetic techniques in association with seismic and geochemical measurements etc., is particularly important and useful when it facilitates the making of "deterministic" previsions.

Let us finally remember the possibility of monitoring and thus modelling subsidence phenomena, which, if not forecast or controlled, may result in damage to or destruction of surface facilities, the irrigation system and ground water regime, etc.
2. SOME EXAMPLES OF GROUND DEFORMATIONS

2.1 Seismic areas

The pattern of crustal deformation in a seismic area can provide essential information on the state of stress, and thus furnish further details concerning the danger level of the area. Generally speaking, a stationary deformation may be correlated with the seismic risk of the zone, the level of which depends upon the speed of deformation, and the time elapsed since the preceding earthquake; in those cases where the entire deformation characteristics over the whole area are known, the crustal movements measured can be used to estimate the period in which such events might recur. Clearly an affirmation of this nature would be based upon the hypothesis that the deformation cycle of large seismic events is not complicated by complex phenomena, and is repetitive. In the schematisation of Fig. 1 this is shown in the presence of a post-seismic transient and permanent deformations modify the simple scheme proposed by Reid. At the present moment the data available for various seismic areas in the world indicate a post-seismic transient in the overall coseismic deformation, notwithstanding the fact that a part of the period concerned shows constant deformations. Many cases seem to show that a large number of seismic deformations take place in a short time (several years), and are then followed by a slow and gradual transition to pre-seismic deformation levels.

![Diagram of seismic cycle](image)

**Fig. 1 Diagrams of seismic cycle**

Many authors have developed models to explain time-dependant crustal deformation observed in regions of ongoing tectonic activity. They found evidence for the existence of visco-elastic relaxation in the asthenosphere. In particular, visco-elastic stress relaxation was found to be the cause of the long-term inter-seismic/aseismic deformation occurring between major earthquakes in several regions of Japan. In this case the deformation takes the form of subsidence in the source region, with a compensating uplift on the flanks of the subsiding region. The width of the subsidence profile seems to be roughly proportional to the elastic plate thickness, and its time development was found to depend on the characteristic relaxation time (\( \tau = 2\eta/\mu \)), where \( \eta \) is the asthenospheric viscosity and \( \mu \) is the asthenospheric shear modulus).
Information concerning structural deformation on a geological scale allows one to point out permanent deformations. It is, however, necessary to bear in mind the fact that some of the deformation over the centuries may not be strictly connected with the storing of elastic energy, and that anomalous crustal movements, defined as precursor phenomena, may be adding their presence to the deformation cycle; a presence which in many instances shows different characteristics, overall behaviour and duration from that of the characteristic seismotectonic changes in the area.

As an example of the possible complexity of deformation see Fig. 2 in which are laid out the altimetric variations in the region of Honshu (Japan). Three different fields of deformation following three different processes can be identified: the dominant process produces a constant subsidence related to the westerly subsidence of the Pacific Plate along the Japanese trench. Local effects are added to this movement, namely uplift of the Quaternary structure, and sinking in the locality of the fault associated with the 1996 earthquake (M=7.3). In Fig. 2 the heights over time variations are repeated in relation to two different zones A and B. Whilst zone B, which is far from the fault, shows a linear process, in zone A the deformations agree with those forecast by the theory of stress relaxation in a visco-elastic asthenosphere following a faulting 6.

Important features of the earthquake mechanism include the size and shape of the rupture surface, its orientation, the faulting motion on this surface, and the time history of the process. Seismic data, in general, provide information about seismic source mechanisms; geodetic surveys could produce important advances in our knowledge of earthquake source models.

Theoretical expressions for surface and subsurface deformations accompanying faulting, obtained on the basis of the elastic theory of dislocation 7,8 have been given by Press 9, Mansinha and Smylie 10 and others. For example, the horizontal displacement components are derived for a finite rectangular strike-slip fault in a uniform elastic half-space with the geometry shown in Fig. 3. In particular the horizontal displacement component U along a median line perpendicular to the fault strike is shown in Fig. 4.

Fig. 2  Level changes, 1900-1975, in Northern Honshu, Japan, and region of surface faulting in the 1996 M=7.3 Riku-u earthquake (Thatker, 1981)4}
Fig. 3  Horizontal components of displacement, for a finite rectangular strike-slip fault

Fig. 4  U component of displacement for different parameters of a strike-slip fault. The displacement is given as a fraction of the slip of the fault.
In general, the asymmetry of the deformations at the two sides of the fault is strictly correlated to the slip and inclination of the fault plane, whereas the displacement fields near to and distant from the fault are indicative respectively of the upper and lower limits of the fault.

Obviously, if we consider a more realistic model, in which the uniform elastic half space is substituted with a variable shear modulus, we obtain different results. In fact assuming that rigidity \( z \) is a continuous and smooth function of depth:

\[
\mu(z) = C (1 - Te^{-z/z_0})
\]

where the value of \( T \) is derived from seismic data \( (T=0.9) \) as well as from laboratory measurements \( (T=0.5) \); Mahrer and Nur 13) obtain theoretical displacements almost identical for different values of \( T \) in the case of surface faults; in contrast, in the case of a buried fault, a notable dependance is present (Fig. 5).

![Diagram](image)

**Fig. 5** Normalized shear modulus as a function of depth (a); surface displacement as a function of distance from the fault: (b) surface fault case; (c) buried case. (Mahrer and Nur, 1979) 13)
It is evident that the quantitative information obtained by geodetic measurement should be interpreted in the light of one's total knowledge regarding the phenomenon being studied, and of the various factors that might strikingly influence the surface effects associated with a state of stress in the crust, or more simply a slip along a fault plane. One thinks, for example, of the possible correlations between morphology and disturbances in surface deformations using a simple elastic model. Starting from a model with topography of small slope, McTigue and Stein [14] analyse those deformations associated with a variation in the uniform tectonic force (Fig. 6), establishing that topography gives rise to local departures from regional strain of the order of the characteristic slope \( \frac{H}{L} \). Considering instead structural discontinuities caused by surface structures (alluvial deposits), one sees that these can introduce notable disturbances (dependant upon rigidity contrasts) into the general field of surface deformations, even if, in general, they are limited to areas of similar dimensions to those of the disturbed tract Fig. 6.

To these sources of disturbance one should add the local, changing effects produced by the weather, and related to rainfall, changes in atmospheric pressure and temperature, etc., which can generate a background noise detectable using modern geodetic instrumentation.

\[
\frac{U}{\varepsilon(1-\nu^2)}
\]

\[
\frac{V}{\varepsilon(1-\nu^2)(1+2\nu)}
\]

Fig. 6 Example of dimensionless perturbation displacement field, due to a far-field tectonic stress (compression); the surface relief is assumed to be characterized by a horizontal length scale, \( L \), and a vertical length scale \( H \). (from McTigue and Stein, 1984) [14]
2.2 Volcanic areas

A classic case where the integration of a multitude of different techniques is indispensable, is undoubtedly the control of deformations related to volcanic activity, due to their speed of change, and the real possibility of making predictions. Subsurface movements of magma generally cause rapid movements of the surface, with uplifting and subsidence that alternate and vary the magma level and pressures involved.

We have laid out the case of Campi Flegrei (in the region of Naples, Italy) as an example where periods of rapid uplift, accompanied by intense seismic activity, alternate with periods of calm. The main features of the expansion can be summarised as follows: the expanding area of quasi-circular shape, and radius of about 7 km, shows maximum uplift near the town of Pozzuoli (Fig. 7). The uplift practically vanishes at the edge of the caldera. The speed of uplift (as high as 4-5 mm/day) and its change over time constrained the operators following the phenomenon both to set up mareographic and clinometric networks in the area, which permit continual monitoring of the phenomenon, and to carry out the periodic checks by geometric levelling using a high number of working teams.

Horizontal measurements provided additional information on the strain field. The displacement vector referred to a point located in the area of maximum uplift, appearing to be roughly perpendicular to the contour lines of ground uplift, with an average horizontal strain value of about $10^{-5}$.

To this, microgravimetric measurements were added because of the magnitude and speed of the deformations; these being extremely useful for monitoring, as well as for supplying an important quantitative element in the definition of the source model.

Fig. 7 (a) Vertical ground deformation at Pozzuoli (Italy) during 1970-83 period; (b) area distribution of vertical ground deformation, and horizontal displacement for the period 1980-83 (from Berrino et al., 1985)
2.3 Subsidence

It is a well known fact that extensive withdrawals of fluid from subsurface resources is the major cause of land subsidence. This is due to a subsurface compaction which represents the response of a compressible, porous, medium to changes in the fluid flow field operating within it. The artificial withdrawal of fluid modifies the natural balanced condition by causing a decrease in pore pressure and a consequent increase in intergranular stress. Under the influence of this increase in the effective stress, the formations compact and the land surface subsides. Subsidence may in general be connected with vertical and horizontal surface displacements, cracks and fissures in soft ground, changes in the hydrological regime, contamination of fresh water aquifer, changes in local tectonic regimes due to altered stress and strain conditions with a consequential risk of increased seismicity, etc.

The qualitative and quantitative determination of the parameters governing the subsidence process involve measurements and analyses of geological, geotechnical and geophysical quantities, as well as surface and near surface measurements of vertical and horizontal movements, such as levelling and trilateration surveys, tilt-meter and extensometer surveys.

If a complete set of the above-mentioned information is available, the complex natural system will be a mean of those mathematical and numerical prediction models formulated from the equation of subsurface flow combined with the elastic equilibrium equations.

Major subsidence in oil and gas fields, or due to ground water pumping, has been extensively covered in the literature. For example, a surface lowering of about 8m from 1937 to 1962 has been observed in the harbour area of Los Angeles and Long Beach. A more recent case of particular interest is the subsidence of the Po Plain (Italy) and Venice, where the lowering, though modest, is nevertheless a matter of concern because it compounds the effect of the exceptionally high tides in the city.

3. GEODETIC MEASUREMENTS

Having started with this short review of the geophysical phenomena representing the major causes of crustal movement, we will now move on to consider the geodetic method used for the detection of such movement, basing our lecture more on practical than on theoretical considerations. We will thus describe classical networks, notwithstanding our awareness that soon much will change with the extensive use of spatial techniques (GPS), since at the present moment we do not have sufficient experience with such techniques to be able to supply exhaustive answers to all the problems related to their application. As we have mentioned, we will also subdivide our exposition into three different subjects: planimetric, altimetric and plano-altimetric networks.

3.1 Planimetric networks

Horizontal crustal deformation may be measured by means of a survey involving angles and distances; electromagnetic distance-measuring instruments are able to determine distances with an accuracy of 1 ppm or less, according to the extent to which one is able to estimate the average propagation speed of light in the atmosphere, using either meteorological measurements, or by simultaneously measuring optical path length at two or three different wavelengths. With a view to developing adequate geodetic networks, various criteria, such as accuracy, reliability and expense, need to be balanced in order to optimise the network. Such optimisation is usually classified in four different orders, starting with a parametric adjustment based on the observation equation:

$$Ax - d = v$$

which, by means of a least-squares solution and following the free adjustment scheme, gives:
\[ X = (A^T PA)^{-1} A^T P d \]

or following the free adjustment scheme

\[ X = (A^T PA)^+ A^T P d \]

and with the co-variance matrix:

\[ K_{xx} = \sigma^2 (A^T PA)^+ = \sigma^2 Q_{xx} \]

where \( A \) is the configuration matrix, \( P \) the weight matrix, \( \sigma^2 \) is the variance of the unit weight and \( ( \cdot )^+ \) denotes the Moore-Penrose inverse.

In the case of planimetric networks, for example, disregarding the "Zero-order design" relating to the optimal reference system, the optimization process is carried out by means of 'First, Second and Third-order design', which relate respectively to the configuration of the network and observation plan, to the search for an optimal distribution of observation weighting, and to the improvement of the network by including additional points and/or observations.

One of the criteria used to evaluate the reliability of a network, is based on the analysis of the diagonal elements \( r \) of the matrix \( R \), where:

\[ R = A Q_{xx} A^T P - I \]

These elements may vary from 0 to 1; good reliability is obtained for high values of \( z \), as well as with a homogeneous distribution of such values throughout the entire network.

As regards "First-order design", the theoretical approach is severely limited by topography and the available information relating to the geophysical phenomenon in question.

As an example of theoretical trilateration networks, we show you those in Fig. 8 which correspond to differing needs: the first is used when it is impossible to localise the direction of expected movements, and the others to concentrate research on the deformation along a traverse of the strike fault.

Fig. 8 Different schemes of trilateration networks.
On the basis of point distribution, measurements should be planned seeking the greatest possible reliability and precision, without forgetting the question of cost. The most practical approach is the interactive simulation.

Once the network has been drawn up and the theoretical aspect of the measurements established, one can take steps to improve the accuracy of single measurements, with a view to obtaining a better overall network design. This may be done by seeking a method for weighting the measurements themselves which respond more satisfactorily to the needs for which the measurements are taken (second-order design).

In effect, the analytic approach is founded on the solution of the following equation, as it relates to $P$:

$$(A^T PA) = Q_{xx}$$

in which the criterion matrix $Q_{xx}$ must be defined in advance.

In general, a geodetic network should tend towards homogeneity and isotropy. In this case, $Q_{xx}$ is normally constructed using the Tailor-Karman structure solution [23], the simplest case of which may be reduced to the equation:

$$Q_{xx} = I$$

where $I$ is the unit matrix, involving the drawback that the co-variance between the coordinates is assumed to be zero.

A different approach, which appears to yield results more in line with real needs, consists in the use of the singular value decomposition of the matrix $Q$ of the network:

$$Q_{xx} = \lambda \Lambda V^T$$

where the diagonal matrix $\Lambda$ contains the eigenvalues of $Q_{xx}$, and $V$ is the matrix of the corresponding orthonormalised eigenvectors. A possible derivation of the criterion matrix might consist in contracting the eigenvalue spectrum, without changing the eigenvectors, with a consequent diminution in the semi-axes of the error ellipses. A different approach, of particular interest in the study of deformations, might be to obtain a variation in the orientations of the error ellipses semi-axes by rotating the eigenvectors matrix $V$ [24].

In any case, these methods are severely limited by the shape of the network and also, especially in distance measurements, by instrument limitations. For these reasons practical results are generally negligible, but second-order design nevertheless supplies us with much useful information for the planning and execution of networks.

We have extensively used second-order design and reliability criteria in an undogmatic way to reduce the measurements in a rational manner involving only a slight loss in accuracy, with a consequent reduction in costs and execution time. In trilateration networks, for example, starting with the maximum number of measurements possible, we eliminate, on the basis of second-order design, the least weighted sides one after another, even if they are characterised by high reliability. This method, which may be easily programmed for automatic processing, gives good results and can be stopped when the points error becomes greater than a pre-fixed value and/or the minimum reliability is less than an acceptable value (Fig. 9).

Second-order design can also be useful for solving some of the problems related to the third order, such as adding measurements of different kinds.

As an example of this, see trilateration network $E$ of Fig. 8 in which the error ellipses are very flat, and may therefore be improved by the addition of angular measurements, in this example the angles around points 4, 5 and 6 improve the isotropy of the net. They were found by applying an empirical trial and error procedure, taking into account only those angles with maximum weighting.

When a network has different types of observation, (angles, distances, Laplace azimuth...), correct weighting of the observations presents a problem. This is a very difficult matter and at present we think that the best method is an empirical approach based
on experience, also if we look with interest to the method proposed by Prof. Kubik 25), for an a posteriori estimate of the various measurements' weighting. This method has been tested in my institute with good results 26).

Fig. 9 Effect of the reduction of measured sides on reliability and precision (net B, Fig. 8)

We shall describe at length the measurements themselves because they are so strictly related to the instruments used. For instance, atmosphere related problems can be overcome using a two-wavelength instrument such as the terrameter, while they represent the most limiting factor if a one wavelength model is used over long distances. Although much research has been devoted to this problem 27), we think that any atmospheric model used to correct measurements is likely to be far from perfect.

We have obtained good results using the method proposed by Robertson, based on the relation between two lengths measured from the same point in two directions with a similar topographical profile in the same period of time 28,29). For short or medium sides one tends towards limitation of the instrument's base error by using a graduated slide for the reflectors 30,31).

As regards the adjustment and analysis of data we have tended towards procedural rationalisation using a series of well organised programmes:

1) Handling and organisation of the data.

2) Adjustment by indirect observation. For the free net we use the bordering technique or the expression \( N^* = N (NN)^{-1} \) or the singular value decomposition \( N = USV^T \), \( N^* = U^* V^T \) SU. For the inversion of the regular matrices we use the modified Cholesky method.

3) The detection of the outliers can be performed using data snooping and considering the weighting matrix to be a diagonal using the variable

\[
W = \frac{v_i}{\sigma v_i}
\]

if \( |W_i| > F_{1-g,1,0.05} \) (normally 3.29) the observation is eliminated and the adjustment is repeated. It is possible to have an interactive version of this procedure. One can also use an alternative approach based on the minimisation of the sum of the absolute value of residuals \( r \), which in some situations seems to be more efficient 33).
4) The analysis of the network is carried out by considering the absolute and relative error ellipses, distance error, and that of the bearings between pairs of points. The ellipses can be plotted automatically.

5) Other programs for similarity transformation, for testing the significance of supposed movements, for studying the local strain field etc. are needed.

From amongst the various networks installed in Italy to study crustal movement, let us show you as an example the one we set up in the Cassino Area (south of Rome) in 1984.

Starting from the geophysical need to connect three mountainous blocks separated by two supposed fault zones, the work sequence has been as follows:

1) to search, using suitable cartography, for those intervisible points which might be used to form a very compact network (within a circle).

2) to mount a field campaign in order to verify the supposed possible connections by means of powerful lamps; to verify the possible routes and transportation needs (special cars, mules, helicopters...); to obtain permission from the owners of the land; to study the geological and lithological features of those places selected in order to choose the foundations of the bench-marks (reinforced concrete pillars); and so on.

3) After that we planned various possible networks on the basis of optimisation criteria, taking into account all the information collected during on-the-spot investigations. In this manner we chose the network in Fig. 10 consisting of three points on every three blocks.

4) Another field campaign was necessary to construct the pillars, built upon rock foundations, and with a self-centring device on the top.

On a geologically stable block we installed two pillars in order to have a baseline for the calibration of the instruments before, during and after every measurement campaign.

5) During the first measurements campaign we experienced difficulty in measuring some of the long sides, and thus decided to add another three points in order to eliminate such long connections. The final network is that shown in Fig. 11 in which the error ellipses are laid out. As you can see the network is characterised by very good homogeneity and isotropy, which is most important since the direction of the expected movements is not known.

3.2 Altimetric networks

A series of high-precision levelling campaigns is the most fruitful method of investigating vertical crustal movements; using this technique it is possible to examine in detail and with great accuracy displacements extending over even the largest areas. Vertical control networks should be planned bearing in mind a considerable number of parameters, such as the area affected by deformation, the characteristics of the displacement and its supposed magnitude and velocity, the type and location of the bench marks and their density, the possible level routes and the degree of accuracy expected, also taking into account the high cost of such measurements. It is not possible here to give a detailed examination of all the problems associated with precise levelling, such as instrumentation, operation methods, refraction, magnetism, rod calibration, bench-mark stability and so on. We will only emphasise the fact that interest in this technique has increased to such an extent that entire conferences are dedicated to it, many interesting suggestions and ideas are given in their proceedings [34,35,36]. As far as accuracy is concerned, we have no problems in the detection of coseismic deformations, large scale subsidence or bradiseism. However, we
Fig. 10 Scheme of the trilateration network of Cassino and comparison between the error ellipses obtained in the first campaign and the expected ones (dashed lines).

Fig. 11 Scheme of a modified network in Cassino.
encounter some difficulties in other cases, such as pre-seismic or post-seismic deformations, or with a slow subsidence or uplift where the small degree of movement concerns a large area.

Amongst the various sources of error we shall consider the following two in particular: refraction, and movement occurring during the measurement phase.

Regarding refraction, the greatest limiting factor for lines in hilly areas, there are essentially two ways in which it may be taken into account: the first is based on the well-known Kukumaki’s formula

\[
R = \left( \frac{L}{50} \right) \delta h \delta t
\]

\[
L = \text{length of sight in metres}
\]

\[
\delta h = \text{difference in level in metres}
\]

\[
\delta t = \text{difference in temperature in degrees at two sights, normally 0.5 m and 2.5 m.}
\]

\[
A = \frac{1190}{C^2} \left( \frac{z_1^{C+1} - z_2^{C+1}}{z_1^{C+1} - z_2^{C+1}} \right) - z_0 \left( z_1 - z_2 \right)
\]

where \(z_0\) is the height of the instruments, \(z_1\) and \(z_2\) are the sight heights, and \(C\) is a coefficient the value of which is normally assumed to be equal to -1/3. The second way is based upon the creation of atmospheric models which provide a means of adjusting the measurements using one temperature and general data regarding insolation, wind, etc.

Different techniques can be applied to check on any possible movements taking place during the measuring campaign, or to obtain a periodic indication of whether or not it is necessary to repeat the measurements over the entire network. One of these is based on the continuous recording of tilt components by tiltmeter; another method we have used extensively is based on the use of double levels at night. This technique makes it possible to ascertain with accuracy, and within a few hours, the difference in height between points at almost the same height, but located at a considerable distance (3-6 km) from each other.

By way of example we have set out the closing error of a control triangle used in various campaigns.

<table>
<thead>
<tr>
<th>(\Delta)</th>
<th>1977</th>
<th>1978</th>
<th>1980</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 km</td>
<td>2.9</td>
<td>-0.9</td>
<td>1.4</td>
</tr>
</tbody>
</table>

It must be remembered that the length of the precise levelling lines necessary to join the three points is over 20 km, with routes characterised by average differences in height of 50 m/km.

Having an idea of the movements occurring during the measurement period, it is possible for one to take them into account during the data processing phase. Nevertheless, we think that the best method is to limit the entire network’s measurement period to as short a time as possible. This can be achieved by contracting small areas of the network to a variety of companies working simultaneously.

The need to employ several firms is typical of levelling networks, and this is fact should be taken into account when preparing the specifications for the contracting firms concerning instrumentation, rod calibration, observation procedures, refraction, tolerances for a single section, for a line, for a loop etc.

In addition we must emphasise that when all the blocks, already analysed separately, are put together, other errors arise; it is no means being a simple matter to complete a serious data analysis or the adjustment of a whole network consisting of hundreds or thousands of measured sections.

As an example of this kind of work we present in Table I a flow diagram of the program prepared in my institute by Prof. Barbarella.
Table I
Flow diagram for network adjustment:

One can also apply other statistical tests for the study of error behaviour in order to obtain information about trends due to systematic errors. The tests are based on the random variable
\[ x_{ij} = \frac{D_{ij}}{R_{ij}} \]

in which is the discrepancy between the direct and reverse measurements of the relative height of two consecutive bench-marks, and \( R \) is the distance between them. The method is based on tests concerning each line, and cumulative tests regarding the entire network, preferably using non-parametric tests because of their generality. Many tests can be used, such as, for example: randomness tests, normality tests, the Wilcoxon test, the Person P. Test, the Kurskal Walis Test, the Bertlett Test 40).

At the end of this short description of altimetric networks we must briefly make one or two comments concerning gravity. Gravity remeasurements may to some extent be used as a substitute for a levelling survey. The accuracy obtainable is less satisfactory, but costs are lower and the method is quicker.

Simultaneous monitoring of elevation and gravity change provides us with further elements for the interpretation of the phenomenon in question 41). For example, generally speaking the gravity variation corresponding to a vertical displacement caused by an accumulating stress field in a seismic area, can be expressed as a sum of the free air and Bauguer effects, plus the gravity contribution made by the subsequent density change in relation to the volumetric strain 42).

With the use of modern microgravimeters, (e.g.: 1a Coste-Ronberg mod.D), standard errors smaller than \( 5 \times 10^{-5} \) m/sec² may be obtained 43). Improvements in accuracy depend mainly on the elimination of effects caused by earth tides, tidal loading, atmospheric mass movements, and so on.

Fig. 12 Scheme of the altimetric network in Bologna (Italy)
As an example of an altimetric net we shall show you the one set up in the Bologna area, (our home city), where over the last ten years we have had a differential subsidence velocity of at least about 11 cm/year, causing many problems regarding the stability of buildings, and water management. The network planned in my institute by Prof. Pieri and Dr. Russo was conceived in order to give different levels of information. In fact it consists of a large network, (460 km²), in which the bench-marks are spaced from between 1 to 0.5 km along the lines, and the distance between the nodal points ranges from 7 to 3 km, the density increasing towards the town; (areas 1, 2 and 3 of Fig. 12). Another network, (Fig. 13), covers the old town, with the bench-marks spaced at intervals of 0.25 km. along the lines, and with a distance between nodal points of about 0.5 km. Finally there are some local networks set up for the purpose of studying the movement of buildings or monuments, often in conjunction with other techniques, such as pendulums, tiltmeters and photogrammetry.

An example of these local networks and of the application of a variety of techniques is provided in the monitoring of two important monuments in Bologna: the two towers (Asinelli and Garisenda) and St. James’ Church in Via Zamboni, where the differential subsidence is very strong (Fig. 14).

Fig. 13 Scheme of the altimetric network in the town of Bologna
3.3 Plano-altimetric networks

In order to give an idea of some of the problems related to plano-altimetric networks we shall present an example of a network in which we have applied many different techniques in order to obtain the greatest possible information about movements in the seismic area of Ancona, and to create a test network for the study of these various techniques. As you can see in Fig. 15, it consists of a planimetric network of 10 stations, signalized by means of reinforced concrete pillars using 30 optical connections which provide a very good redundancy (1.76). Furthermore using length measurements only performed by the ADA geodimeter mod. 8. Seven of these stations are also connected by a precise levelling network, consisting of over 80 km of levelling lines with more than 100 bench-marks, and connected with the national first-order levelling network and with a tide gauge located at Ancona. A gravimetric network was set up consisting of nine base stations situated on levelling bench-marks. Another two stations were set up on a bell tower for the time monitoring of the gravity gradient. The measurements are taken using a La Coste-Romberg mod. D gravimeter.

In addition to these base networks we have also set up two altimetric control triangles using double levels (Zeiss Ni2).

Astrogeodetic measurements have also been taken at five planimetric stations for the direct measurement of vertical deflection.

Consideration of all the measurements together in order to make a tridimensional adjustment in a general or local framework presents many problems. However, this question will be treated at length in a later lecture in these proceedings. Instead we should now like to draw to your attention two problems that need to be solved when dealing with trigonometric levelling. The problem is to improve trigonometric levelling so that this technique may be more widely used. In some cases trigonometric levelling may be combined with, or even partially substitute, spirit levelling, with a considerable reduction in costs, and with an increased rapidity of data acquisition. To accomplish this we have to reduce the effects of refraction, and take into account the influence of vertical deflection without astrogeodetic measurements.
To carry out reciprocal and simultaneous zenithal angles we always make use of first-order theodolites, (Wild T3, DA13), equipped with a small but powerful lamp for collimation at day or night. The lamp is mounted directly on the telescope and allows it to rotate around its horizontal axis. To link points that are at approximately the same height, (less than 1.5 m difference), we use, as we have said, double levels with special rods designed for collimation at night. Atmospheric data, (temperature, pressure, vapour pressure, wind speed), were collected during the measurement period by means of small meteorological stations.
As is well known, the accuracy of trigonometric levelling is severely limited by the effects of refraction. Essentially there are two possible ways of reducing these effects. The first is based on two wavelength angular measurements; and the second approach uses atmospheric models in order to find the variation of refractivity. Since the first method does not seem to be feasible for the near future, we have devoted our attention to the extensive use of the second.

The theory is well known. Starting with the formula

\[ h_A - h_B = D (1 + \frac{h_A - h_B}{2R}) \tan \frac{1}{2} (Z_{AB} - Z_{BA} + \delta Z_{AB} - \delta Z_{BA}) \]

following Prof. Moritz, we have

\[ \delta Z_{AB} = \frac{10^{-6}}{S} \sin Z_{AB} \int_0^S \frac{dN}{dz} (S - x) \, dx \]

in which the refractivity gradient may be expressed:

\[ \frac{dN}{dz} = -0.2696 N \frac{P}{T^2} (0.0342 + \frac{dT}{dz}) + 11.25 \frac{e}{T^2} \frac{dT}{dz} + \frac{11.25}{T} \frac{de}{dz} \]

where the temperature gradient is the most important factor.

As regards the evaluation of the temperature gradient, one can make use of the heat flux estimate, a parameter that can be deduced from the radiation budget estimate. This approach involves the consideration of a large number of additional meteorological and ground parameters, such as heat flux into and out of the ground, evaporation, wind spread, cloudiness, sun height, terrain structure, vegetation and so on. We have applied this technique extensively following the suggestions of many authors all known for their interest in refraction problems, such as Angus-Leppan and Brunner, with good results.

### Table II

Example of result of computation of refraction angles for reciprocal and simultaneous trigonometric leveling measurements (from Achilli, Baldi and Unguendoli, 1985)

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Height differences (cm)</th>
<th>Time (hr)</th>
<th>Ground</th>
<th>Temperature (°C)</th>
<th>Refraction correction</th>
<th>Corrected height difference (cm)</th>
<th>Mean of the</th>
<th>Mean of the corrected values (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theodolite</td>
<td>822.9</td>
<td>13.0</td>
<td>30</td>
<td>-10.64</td>
<td>-8.61</td>
<td>-3.19</td>
<td>819.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>821.2</td>
<td>16.0</td>
<td>27</td>
<td>-10.64</td>
<td>-8.78</td>
<td>-3.27</td>
<td>817.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>812.0</td>
<td>20.0</td>
<td>25</td>
<td>-19.97</td>
<td>-22.03</td>
<td>-4.65</td>
<td>816.7</td>
<td>816.0</td>
</tr>
<tr>
<td></td>
<td>810.9</td>
<td>2.0</td>
<td>20</td>
<td>-19.26</td>
<td>-35.84</td>
<td>10.33</td>
<td>821.2</td>
<td></td>
</tr>
<tr>
<td>Double levels</td>
<td>813.2</td>
<td>21.0</td>
<td>13</td>
<td>-24.51</td>
<td>-28.43</td>
<td>6.16</td>
<td>819.4</td>
<td></td>
</tr>
</tbody>
</table>
As an example of how the method may be applied, Table II sets out the results of five series of reciprocal and simultaneous measurements of zenithal angles relating to a link 6 km in length belonging to a triangle, taken in a variety of atmospheric conditions. The corrected values listed in Table II show that the deviations from the mean are much smaller than in the previous case. Furthermore, correcting the other two sides of the triangle in the same way, it should be pointed out that the mean value of the corrected determinations shows an appreciable improvement in the disclosure error.

Fig. 16 Scheme of the local anomalies of vertical deflection (from Achilli, Baldi and Unguendoli, 1985)

Strictly speaking, if we want to refer the trigonometric height to the same surface, (the ellipsoid), we have to know the vertical deflection, obtainable by astronomic measurement, which are a little troublesome. In control networks where the area involved is always small enough, the problem of the reference surface may be solved in a local way, taking into account only the local anomalies of vertical deflection, and obtaining a smoothed reference surface that we call a "smoothed geoid" (Fig. 16). These anomalies are mainly due to the distortion of the gravitational field caused by the distribution of the surrounding masses. For the computation of these anomalies we make the "mass model" with a program prepared by Achilli and Baldi [33], and consisting of an evaluation of the disturbing gravitational attraction exerted by the surrounding topography, and by density anomalies. The area is subdivided into a regular quadratic grid, with the side lengths constant or diminishing towards the point considered. The area is therefore represented by a set of parallelepipeds with a density determined from geological information, the gravitational components of which can be calculated by means of the formula:

\[ T_\xi = \sum_{i=1}^{n} t_{xi}(P); \quad T_\eta = \sum_{i=1}^{n} t_{yi}(P) \]

in which:

\[ t_x(P) = \Theta G \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} \frac{x}{[(x-a)^2+(y-b)^2+(z-c)^2]^{3/2}} \, dx \, dy \, dz \]
The components of the local anomalies in vertical deflection are thus:

$$
\xi = \arctan \frac{T}{g} \quad ; \quad \eta = \arctan \frac{T}{g}.
$$

To verify the correction methods for refraction and local vertical deflection, we carried out, on a quadrilateral of the Ancona network, a series of trigonometric height measurements in different atmospheric conditions, paying particular attention during the measurement stage, and collecting atmospheric data. The sides range from 5 to 11.5 km, with a maximum difference in height of 400 m. The mean results of three series of measurements are listed in the table, and as you can see the closing error of the independent triangles improves markedly with the application of the corrections described above.

<table>
<thead>
<tr>
<th>side</th>
<th>$\Delta h$ (m) measured</th>
<th>correction (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>386.89</td>
<td>-0.01</td>
</tr>
<tr>
<td>1-3</td>
<td>54.57</td>
<td>0</td>
</tr>
<tr>
<td>1-4</td>
<td>115.79</td>
<td>0.01</td>
</tr>
<tr>
<td>2-3</td>
<td>-332.41</td>
<td>0.06</td>
</tr>
<tr>
<td>2-4</td>
<td>-271.18</td>
<td>0.09</td>
</tr>
<tr>
<td>3-4</td>
<td>61.19</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**Table III**

Differences in height obtained by means of trigonometric levelling and correction to take account of refraction and local anomalies in the deflection of the vertical (from Achilli, Baldi and Unguendoli, 1985) 54.

**MISCLOSURE ERRORS**

<table>
<thead>
<tr>
<th>triangle</th>
<th>measured (m)</th>
<th>corrected (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2-3</td>
<td>-0.09</td>
<td>-0.04</td>
</tr>
<tr>
<td>1-3-4</td>
<td>0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>1-2-4</td>
<td>-0.08</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

As we have said, seven planimetric stations of the Ancona network belong to a precise levelling network, and their heights were obtained both by trigonometric and spirit levelling, so that one could make an interesting comparison between the two means of measurement. This kind of comparison is only possible if the two types of measurement refer to the same surface, which, given the small size of the area involved, can be identified with the smoothed geoid. We have used the mass model, which permits calculation of the local anomalies of the geoid (Fig. 17). By means of such geoidal undulation the heights obtained by spirit levelling were corrected, while the trigonometric measurements were corrected for refraction and for the local anomalies of vertical deflection. The results set out in the Table IV show a high level of congruence between the two height measurements, with a maximum difference of 2 cm.
Fig. 17 Local anomalies of geoid in the Ancona area (mm)

Table IV
Comparison between height measurements obtained by means of trigonometric levelling and spirit levelling (From Achilli, Baldi and Unguendoli, 1985)\(^5\)

<table>
<thead>
<tr>
<th>Station</th>
<th>Adjusted Height (m)</th>
<th>Corrected and Adjusted Height (m)</th>
<th>Corrected Height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>195.81</td>
<td>195.81</td>
<td>195.81</td>
</tr>
<tr>
<td>2</td>
<td>230.22</td>
<td>230.19</td>
<td>230.18</td>
</tr>
<tr>
<td>3</td>
<td>502.77</td>
<td>502.74</td>
<td>502.73</td>
</tr>
<tr>
<td>4</td>
<td>121.57</td>
<td>121.55</td>
<td>121.55</td>
</tr>
<tr>
<td>5</td>
<td>251.15</td>
<td>251.12</td>
<td>251.13</td>
</tr>
<tr>
<td>6</td>
<td>311.61</td>
<td>311.61</td>
<td>311.61</td>
</tr>
<tr>
<td>7</td>
<td>250.40</td>
<td>250.41</td>
<td>250.39</td>
</tr>
</tbody>
</table>
4. CONCLUSIONS

As a conclusion to this short lecture on deformation measurement, we should like to consider the following:

Firstly, concerning methodology, the philosophy behind the mode in which we take our measurements is itself conceptually different from the classical determination of points location. In fact we measure large quantities to obtain small ones, and those problems related to precision, measurement conditions, instrumentation, and data analysis are more difficult and more delicate.

Secondly, considering the type of measurements to carry out, we think that the best way is to move towards an integrated application of temporally spaced measurements, (plano-altimetric networks, gravimetric networks, astrogeodetic measurements...) and continuous recording, (tiltimeters, strain meters, hydrostatic levelling), in order to acquire a better knowledge of the phenomenon studied.

Finally, we think that in this field firm cooperation between different disciplines is absolutely essential. Geodesists, Geophysicists and Geologists have to cooperate strictly in the various steps of the work; i.e. in the planning and execution of measurements, and in the analysis and interpretation of data.

* * *

REFERENCES

1) Goad, C.C., Remondi, B.W., Initial relative positioning results using the global positioning system, Bull. Geod. 58,(1984).


4) Thatcher, W., Crustal deformation studies and earthquake prediction research, Earthquake Prediction, American Geophysical Union, Washington,(1981).


19) Baldi P. and Unguendoli M., Some problems in monitoring crustal deformation by geodetic methods, Quaterniones Geodesiae, 3, (1982).


32) Fuchs H., Contributions to the adjustment by minimizing the sum of absolute residuals. Manuscripta Geodetica, 7, (1982).


THREE DIMENSIONAL ADJUSTMENTS IN A LOCAL REFERENCE SYSTEM

J. Iliffe *)
University College London, GB

ABSTRACT

This paper presents the way in which a general adjustment program was written for three dimensional networks at CERN. It describes the solutions which were adopted in developing an adjustment procedure based on the local geoid, and the constraints which existed regarding the generality of the program's application. The eleven types of observation currently permissible are given, together with a short description of some of the more "exotic" ones, which are peculiar to CERN. The input procedure, with its free-format convention and the system of key-words, is described, as is the information presented to the program user in the form of output.

1. INTRODUCTION

During the three decades and more of CERN's existence, the use of computers has played a central, and increasingly important, role in all areas. In the Applied Geodesy Group, as the techniques used to survey the larger and larger accelerators become increasingly sophisticated and complex, so the methods by which the surveying and geodetic networks are adjusted have had to develop apace. Each new accelerator has presented a new challenge, and has led to the development of calculating techniques and computer software to meet this challenge.

The latest project at CERN, the Large Electron Positron Collider (LEP), is almost four times as large as its immediate predecessor and has, in its turn, brought a new set of problems to be overcome. For the accuracy required in its construction, all computations relating to the adjustment must take account both of the ellipsoid and the geoid in the region of CERN. A spherical earth model is no longer sufficient.

In addition to the considerations of accuracy, the sheer volume of work, the profusion of data to be handled in such a project has entailed the writing of a computer program which would reduce the work done by the users to a minimum. This applies to the tedious process of preparing the data on computer input files; it also means that the program must wherever possible accept raw observations without the need for the prior reduction of data.

*) Former CERN technical student 1984-1985
Thus, the twin dictates of accuracy and ease of use already define the broad outlines of the program which is needed for the adjustment of the LEP geodetic network. Perhaps the next most important point was that the program, although designed specifically for the LEP, should also be applicable to all other surveying and geodetic networks at CERN. This follows naturally from a need for standardisation, and from a belief that the program could be made sufficiently attractive to the user for it to be preferred to all previously existing programs. So, although by default the calculations are referred to the ellipsoid and the geoid, the user may request that these be replaced by, for example, a spherical earth model with an unknown geoid. Eventually, it is intended that the program should incorporate functions which had previously been performed by complimentary programs (such as the calculation of approximate coordinates and two- or three-dimensional adaptation).

The program which has been written is called the Logiciel Général de Compensation (General Adjustment Program), or LGC for short. This paper describes the way in which it was written to conform to the principles outlined in this section. Other, more particular, constraints existed, but these will be described in context.

2. THE GEODESY OF LEP

2.1 The CERN coordinate system

The adjustment of the surveying and geodetic networks of the LEP is performed by LGC, the adjustment program, by the method of the variation of coordinates, with a least-squares minimisation of the weighted residuals. The coordinates which are varied in the adjustment are in the CERN XYZ cartesian system which, although not necessarily being linked to any physical quantity over the whole of the LEP area, nonetheless allows the easiest solutions mathematically. It is first necessary to define the coordinate system.

The origin, Po, is the pillar at the centre of the Proton Synchrotron, and the Z axis of the CERN system is deemed to be along the vertical at this point. The XY plane is thus tangential to the ellipsoid there, and the Y axis is defined as having an azimuth of 37.77864 gons. The system is a right-handed one, and hence the three axes are uniquely defined. For historical and practical reasons the coordinates of the origin, Po, are defined as follows:

\[X_0 = 2000.00000 \text{ m}\]
\[Y_0 = 2097.79265 \text{ m}\]
\[Z_0 = 2000.00079 \text{ m}\]
\[H_0 = 433.65921 \text{ m}\]

As has already been stated in the introduction to this paper, a spherical earth is no longer a sufficient model to use for the geodesy of the LEP. Such an approximation would have entailed an error of up to 22.5 mm in the transformation between measured altitudes
and Z coordinates \(^1\). The ellipsoid used instead was that chosen in 1980 by the Union Internationale de Géodésie et Géophysique, and of which the parameters are:

\[
\begin{align*}
    a &= 6378137 \text{ m} \\
    e^2 &= 0.0066943800229 \\
    e^{12} &= 0.0067394967755
\end{align*}
\]

It can be seen, then, that the Z axis is only parallel to the vertical at the origin, Po, and becomes less so as the distance from the origin increases. In general, a point's XYZ coordinates are only a convention without physical meaning, and yet there are rare occasions when it may be assumed that the XY plane represents the local horizon at any point. As this is not usually justifiable, it will be specifically stated when this assumption has been made.

There are several combinations of the coordinates which may be held fixed at any particular point in a network. Obviously, there is a need both for points which are completely free and for ones which are fixed in all three dimensions. In addition there are points which are fixed in X and Y but free to move in Z (in a levelling network for example) and ones which are fixed in Z but free to move in X and Y. An important variant on the latter is when it is not Z, but H which is held fixed; this cannot easily be expressed directly in the adjustment, but it is sufficient to recalculate Z from the new X and Y coordinates after each iteration (Fig. 1).

![Diagram of constraint onto the horizon](image)

**Fig. 1** Constraint onto the horizon

### 2.2 The Geoid

The geoid in the area of the LEP has been computed by calculating the contribution of mass elements over the whole of Switzerland and surrounding areas; the results obtained were checked by comparison with astro-geodetic measurements. The geoid in this area is defined as being tangential to the ellipsoid at the CERN origin, Po, and the deviation of the vertical and the geoidal heights are defined relative to the reference ellipsoid. It
should be noted here that the maximum deviation found was 15 seconds of arc, and the maximum separation was 200 mm$^2$.

Having computed the position of the geoid at 121 points on a 10 km x 10 km grid over the whole of the LEP area, the results were parameterised to obtain a simple algorithm by which the geoid (and its inclination) could be computed for any point in the area. This was done for two levels: at level zero and at the level of the LEP tunnel. For each one a "best fit" paraboloid was fitted to the known data, each one having certain dimensions and orientation. This information was included in LGC, the adjustment program, and hence the deviation of the vertical and the geoidal height could be calculated for any point at any stage in the program by calling a simple subroutine, XIETAN.

2.3 Heights and Z coordinates

The geoidal heights are used to convert a height relative to the geoid to one relative to the reference ellipsoid. This is done according to the formula:

\[ H_E = H_G + N \]  

(1)

where $H_E$ is height relative to the ellipsoid, $H_G$ is height relative to the geoid, and $N$ is the geoidal height.

A height relative to the geoid is, of course, the one which has a physical meaning; the geoid is an equipotential surface and $H_G$ can be measured. The ellipsoidal height, $H_E$, is a necessary quantity when converting from a measured altitude to the Z coordinate, as can be seen from Fig. 2.

![Diagram](image)

Fig. 2 Geometrical height above the reference ellipsoid
This conversion is performed (in both directions) several times in the course of the adjustment, especially in those cases where H is held fixed and Z is updated after each iteration. For this reason, there are two subroutines (TRANZH and TRANHZ) which perform the transformation from geoidal heights to Z coordinates, and vice versa.

2.4 The rotation matrix

When considering observations which are made with respect to a local vertical or in the local horizon (e.g. zenithal distances or theodolite observations) the solution process is assisted greatly if the observation equation is first derived with respect to a coordinate system based on the local vertical. Horizontal angles, for example, do not provide any information on the distance between two points in the direction of the vertical at the stationed point; but away from the origin, Po, it may have some bearing on the difference in Z.

A local reference system needs to be defined. The local z axis is, naturally, colinear to the vertical. The y axis is then defined as being in the plane of the z and Z (CERN) axes, pointing away from the origin, Po. The x axis forms a right handed system with the y and z axes (Fig. 3).

![Diagram of local parameters of an observation station](image)

**Fig. 3** Local parameters of an observation station

Having a standard definition of the local coordinate axes means that a general rotation matrix can be derived for converting from one coordinate system to the other. This rotation matrix expresses the compound action of three fundamental stages:

1. Rotation C about the Z axis,
2. Rotation \( \omega + dV_y \) (deviation of the vertical at M in the direction PoM) about the new X axis,
3. Rotation \( dV_x \) (deviation perpendicular to PoM) about the y axis.
The resultant rotation matrix, \( R \), is given by:

\[
R = \begin{pmatrix}
\cos G - dV_x \sin (w + dV_y) & -\sin G - dV_x \sin (w + dV_y) \cos G & -dV_x \cos (w + dV_y) \\
\cos (w + dV_y) & \cos (w + dV_y) \cos G & \sin (w + dV_y) \\
dV_x \cos G + \sin (w + dV_y) \sin G & -dV_x \sin G + \sin (w + dV_y) \cos G & \cos (w + dV_y)
\end{pmatrix}
\]

(2)

where it has been assumed that \( \sin dV_x + dV_x \) and \( \cos dV_x + 1 \).

In the first instance this matrix is used to derive the difference in xyz coordinates between two points from the difference between the tabulated approximate XYZ coordinates. These differences \( (\Delta x, \Delta y, \Delta z) \) are then used to derive the observation equation in terms of the incremental displacements in \( x, y, \) and \( z \) at each point. The rotation matrix is then used, in the opposite sense from the above, to express the equation in terms of \( dX, dY, \) and \( dZ \), the unknowns in the adjustment. That is, since:

\[
\begin{pmatrix}
\Delta x \\
\Delta y \\
\Delta z
\end{pmatrix} = R
\begin{pmatrix}
\Delta X \\
\Delta Y \\
\Delta Z
\end{pmatrix}
\]

(3)

and if the observation equation in the local horizon is expressed as:

\[
C \begin{pmatrix}
\frac{dx_N}{d\phi_H} \\
\frac{dy_N}{d\phi_H} \\
\frac{dz_N}{d\phi_H}
\end{pmatrix} - C \begin{pmatrix}
\frac{dx_N}{d\phi} \\
\frac{dy_N}{d\phi} \\
\frac{dz_N}{d\phi}
\end{pmatrix} = (\text{Obs. - Comp.}) + v_i
\]

(4)

where \( C \) is the 1x3 matrix of coefficients,

then,

\[
R^T C \begin{pmatrix}
\frac{dx_N}{d\phi} \\
\frac{dy_N}{d\phi} \\
\frac{dz_N}{d\phi}
\end{pmatrix} = (\text{Obs. - Comp.}) + v_i.
\]

(5)

Thus, the coefficients of the observation equation can be rotated as if they were a spatial vector.

This rotation procedure has been simplified by placing it in a subroutine, ROTVEC, where it can be called from any part of the program. It is necessary to supply, as dummy arguments to this routine, the origin of the local coordinate system (the name of the stationed point), the direction of the rotation (indicated by a 0 or a 1) and the three components of the vector. The components of the rotated vector are returned in the same locations. It is not necessary to supply \( w, G, dV_y, \) and \( dV_x \), the component elements of \( R; \) \( w \) and \( G \) are calculated by a subroutine \( \text{OMEGAS} \), called at the start of each iteration, and stored; \( dV_x \) and \( dV_y \) are called as necessary from \( \text{XIETAN} \) from within ROTVEC.

2.5 Special cases

Up to this point mention has only been made of the way in which LGC performs the adjustment for the LEP. Yet it was stated in the introduction that this was to be an
adjustment program for all surveys carried out at CERN. Two special cases have to be provided for: the spherical earth model and the limited dimension local origin survey.

The spherical earth model is applicable to those surveys related to the previous accelerators. This is the simplest case to deal with: instead of the radius of curvature of the ellipsoid depending on the azimuth from Po to a particular point, it is simply replaced by the mean value of $R = 6371$ km. The geoidal corrections are not applicable in this case, either, and so the subroutine XISTAN is modified to return a geoidal height and a deviation of the vertical equal to zero.

The other special case refers to small three-dimensional networks where the curvature of the earth is of no importance. In such cases the coordinates of the points are usually grouped about a local origin with coordinates $(0, 0, 0)^T$; the CERN origin is approximately at $(2000, 2000, 2000)^T$ and so there would be problems at the local origin, where the XY plane would not be parallel to the horizon. This is corrected by applying a coordinate shift within the program which is invisible to the user.

Both of these cases are selected by writing special key words (SPHERE, and OLOC, respectively) in the input list. The necessary modifications are carried out automatically.

3. **THE OBSERVATIONS**

3.1 **Formation of the equations**

The fact that LGC is an adjustment program for many different types of network has meant that a wide variety of observations have to be dealt with; at present there are around ten, including no less than four different sorts of off-set.

The observation equations are not stored as separate entities; they are computed from the approximate coordinates at each iteration, and one by one they contribute to the formation of the normal equations. This is done first of all by creating the coefficients and placing them in the vector $CO$. Each element in this vector has a corresponding rank in the normal matrix, and these are found by calling the subroutine IMAT, which calculates the ranks for each point involved in the observation. These ranks are placed in the vector $IR$, in exactly the same order as the corresponding coefficients in $CO$. The last element in $CO$ is the "observed minus computed" term, and this has no corresponding rank. Once all the coefficients have been found, the subroutine EQNORM calculates their contributions to the normal equations and the vectors $CO$ and $IR$ are cleared ready for the next observation.

Two types of observation have special characteristics which require further explanation. These are the horizontal angles and the zenithal distances. Rounds of
horizontal angles require a correction \((V_0)\) to be made to the horizontal circle reading; an approximation is made for each set-up before the first iteration by calling the routine \(WOAPP2\). The fine corrections are treated as unknowns and updated after each iteration. Since these are scalar, that is non-vectorial, values they are not rotated with the other terms of the observation equations.

It was decided, when considering the case of the zenithal distances, that the most satisfactory procedure would be to leave the heights of the instruments as unknown quantities. These are set to 0.200 m initially, and updated after each iteration.

3.2 Simulations

It is possible to run the adjustment program, LOC, in simulation mode to assess the strength of a proposed network. The observation equations will be exactly the same as for an observed network, except that the observations themselves will not exist and it is necessary to create them, in order to write the "observed minus computed" terms. This is done by obtaining a random number from a sample with mean zero and standard deviation equal to the standard error of the observation, and adding it to the computed "observation". This is done for each "observation", and the network is solved using these values.

This process is repeated several times (the actual number is specified by the user in the input file) and the result is several versions of the same network, which can be analysed to determine the spread, maximum movements of points, and so on.

3.3 Types of observation

The observations which can be handled by the adjustment program are as follows:

1) Spatial distances, completely independent of the geoid; Straightforward equation.

2) Horizontal distances, not horizontal distances in the usual sense; rather, they are the distances between points which have been reduced to level zero and then the \(XY\) plane (Fig. 4). This observation is for handling data which has already been reduced for other purposes. Observe, but occasionally useful.

3) Vertical distances, to take account of the geoidal heights at each point.

4) Horizontal angles, whose equation is determined in the local horizon and then rotated to the CERN \(XYZ\) system.

5) Orientations, as for horizontal angles, but without the horizontal circle correction, \(V_0\).

6) Zenithal distances, for which seven coefficients have to be calculated, including that for the height of the instrument. The program allows an extension to be specified on the targeted point. This can justifiably be assumed to be parallel to the vertical at the stationed point, and the equation is simplified considerably as a result.
7) Horizontal off-sets, perpendicular off-sets where the measurement is made in the horizontal plane regardless of the relative heights of the three points (Fig. 5). Since the three points A, B, and C are very often almost co-linear it is necessary to have the concept of "left" and "right" measurements. These are conventionally ascribed as negative and positive respectively. For this observation, the horizontal plane can, with negligible loss of accuracy, be assumed to be the same as the XY plane.

8) Spatial off-sets, being perpendicular off-sets where the measurement is made in the plane defined by the three points (Fig. 5). They can be made at any angle - on a wall, upside down, etc. - and for this reason the concept of positive and negative measurements had to be abandoned. It is impossible for the program to determine on which side of the plane the measurement was made.
9) Theodolite off-sets, made from a theodolite bearing, where there is no “third point”. This measurement cannot be made independently, but must be included in a round of angles for the horizontal circle correction to be calculated (Fig. 6).

![Fig. 6 Theodolite off-sets](image)

10. Vertical off-sets, made between one point and a plumb-bob hanging off another.

11. Radial observations, correctly described as constraints and not observations, but as they are treated in a similar way it is appropriate to describe them here. The “observation” is a direction in which a point is deemed to be free to move; movement perpendicular to this direction gives rise to a residual and is therefore restricted. This facility is useful when trying to avoid the “compression effects” obtained when both ends of a linear network are fully constrained.

4. THE NORMAL EQUATIONS

In considering the formation of the normal equations, the first question to be addressed is how they are to be stored. Obviously, since they are symmetric, it is only necessary to store the lower or upper half of the left hand sides. For larger networks, and especially the elongated variety typical in tunnels, the matrix will tend to be "banded". That is, the only non-zero elements lie in a narrow band on either side of the leading diagonal. In such a case, there will be advantages in only storing this central band, and not wasting space by occupying it with zero elements. Conversely, there are disadvantages in terms of the time taken to store and retrieve elements in such a fashion. So for LGC it was decided that it would be sufficient to store the whole upper half-matrix of the normal equations, and only introduce a banded storage pattern at a later date if it was required.

The matrix is stored in a vector of one dimension, which is composed of the successive columns of the upper triangular matrix (Fig. 7).

Hence the element in the \(i^{th}\) row and \(j^{th}\) column is stored in the \(p^{th}\) position, where:

\[
p = \frac{j^2 - j + 1}{2}.
\]
Figure 7  Columnwise storage

The elements of the normal matrix are formed by the subroutine EQNORM. This is called for each observation when the coefficients of the observation equation have been formed. The cross-product of each pair of coefficients is added to the corresponding element of the normal equations.

When the formation of the normal equations is complete, the matrix of the left hand sides, $N$, is decomposed as a first step to finding the solution vector. This is done in the subroutine EQSOLV. The pattern of decomposition chosen was:

$$N = U^T D U$$  \hspace{1cm} (7)

where $U$ is an upper triangular matrix with all elements on the leading diagonal equal to 1 and $D$ is a diagonal matrix. The advantage of this as opposed to a decomposition into $U^T U$ (Choleski) is that time is saved by not having to find square roots of the diagonal elements. The elements of $U$ and $D$ can be stored in the same locations as those of $N$, the latter being overwritten as the decomposition proceeds. It is not necessary to store the leading diagonal of $U$.

Having decomposed the matrix, the solution vector is found by forward and back substitution. If any of the unknowns are superior to the limit of 0.01 mm then another iteration is deemed to be necessary. The subroutine returns to the main program, the coordinates are augmented, and the solution re-starts with the formation of the observation and normal equations.

If, on the other hand, all the unknowns are less than the limit, then the vector solution is complete. Some elements of the inverse of the normal matrix will be needed for the analysis which follows the finding of the solution. Then, since

$$N^{-1} = U^{-1} D^{-1} U^{-1} = U^{-1} D^{-1} U^{-1}^T$$  \hspace{1cm} (8)

the first step is to find the inverses of $U$ and $D$. These are again stored in the same locations.
The subroutine EQSOLV stops at this point. Obviously, it would be wasteful to find all the elements of $W^{-1}$, since in general relatively few of them will be needed. It is more efficient to calculate individual elements from $U^{-1}$ and $D^{-1}$ as they are needed, even though there will often be cases where elements are calculated more than once.

So, having inverted $U$ and $D$, the subroutine EQSOLV returns to the main program. Here the coordinates are updated again, as for the previous iterations. But this time, instead of starting another iteration, the program calculates the residuals of the observations and from these the unit variance, $\sigma_o$, where

$$\sigma_o = \frac{\Sigma (v_i \sigma_i)^2}{n - k}$$  (9)

where $v_i$ are the residuals, $\sigma_i$ are the standard errors of observation, $n$ is the number of observations, and $k$ is the number of unknowns.

Individual elements of the inverse of the normal matrix can subsequently be found by calling the subroutine ELMINV. This multiplies out the relevant rows and columns of $U^{-1}$ and $D^{-1}$, and multiplies the result by $\sigma_o$. The result is returned as the inverse element corresponding to the rows and columns specified in the CALL statement.

5. THE INPUT PROCEDURE

5.1 Key-words

As was stated in the introduction, a principle feature of LGC is the ease with which it can be used. This is due in large part to the way in which the input routines have been structured, and an essential feature of this is the system of key-words.

A key-word is an asterisk, followed by a string of characters which define the meaning of the word. The meanings can be divided into two broad categories. Firstly there are those key-words which introduce a set, or a type, of data. For example, *ANGLES is followed by the list of horizontal angular observations, and this list is understood to terminate when the next key-word is encountered.

The second category of key-words is of those which do not precede a data set, but select particular options of the adjustment program. An example of this type is *SPHERE, which selects a spherical earth model, rather than an ellipsoidal one.

Almost any length of key-word may be written, but in general only the first four characters are checked. Hence, in the previous example, the key-word *SPHERICAL EARTH MODEL would have exactly the same effect as *SPHERE.
The complete list of key-words, together with a brief summary of their meanings, is given in the appendix.

5.2 Free format

Despite the fact that the program was written in FORTRAN, it has been done in such a way that the data on any one line does not need to correspond to any particular format. The important thing is the order of the data, and hence the lines:

\[
\begin{array}{ccc}
1.0 & 2.3 & 4.05 \\
1.0 & 2.3 & 4.05
\end{array}
\]

would both have the same effect. The only rule affecting the position of the data is that separate pieces should be separated by at least three blanks.

The way in which this is achieved is that each line is read as a character string of length 80, and then divided up into an array containing the 80 individual characters. Five subroutines may be called to "read" the data from this array, instead of having actual READ and FORMAT statements.

These five routines are:

- RINTEG - reads integers
- RREALN - reads real numbers
- RALPHA - reads character strings
- RMOT - reads words, curtailing after four characters, and
- RTITRE - reads entire lines.

These routines find the next piece of data, irrespective of its position on the line. Redundant blanks are ignored. The five routines are complemented by:

- RNEXT - called to fetch the next character
- RLINE - reads the next line when the current one ends
- RSTART - initializes all parameters
- RFAIL - called when a discrepancy is found
- REND - decides whether to continue or to print failure message and stop.

As can be seen from the presence of the routine RFAIL, various checks can be made on the data which is being read. These errors occur either half way through an item of data (e.g. finding a decimal point in an integer), or as a result of the unexpected presence or absence of an item. The general principle is that an error should be noted, and the user warned, but that action should be taken to jump over the suspect data and continue reading. The program is not halted until an attempt has been made to read all the data present.
5.3 Layout of data

Without giving detailed specifications of the ways in which the various classes of data are entered, it is nevertheless instructive to show an example of the flexibility which can be attained. For example, to enter measured distances the essential data is the key-word followed by a list of the pairs of points and the distances between them:

* DMES
  AA   BB   1234.5
  CC   BB2  896.12
  CDE  FGH  1066.88
  BB   CDE  1492.0

This is an irreducible minimum for the data, and clearly the next information required is the standard error of each measurement. If, as is often the case, these are the same for all observations, then the generic standard error can be put on the same line as the key-word, signifying that it applies to each observation. For example,

* DMES 1.0 1.0

implies that each observation has a standard error of 1.0 mm + 1.0 ppm. This value can be over-ruled for individual observations, however, by placing an alternative value next to the observation in question. So, in the previous example,

* DMES  1.0  1.0
  AA   BB   1234.5
  CC   BB2  896.12  2.0  5.0
  CDE  FGH  1066.88
  BB   CDE  1492.0

All observations have a standard error of 1.0 mm + 1.0 ppm, except CC-BB2 for which it is 2.0 mm + 5.0 ppm.

If no standard error is given for the group as a whole then the program will assume a value and warn the user that it has done so.

Another feature which can be introduced is the ability to specify a constant to be added to each observation. Constants are introduced with a "/", to distinguish them from standard errors, and apply to each observation after and including the one whose line they are on, until another constant is found. Thus, in the previous example, to add 100 mm to the first observation, 200 mm to the next two, and 0 to the final one:

* DMES  1.0  1.0
  AA   BB   1234.5 /0.100
  CC   BB2  896.12  2.0  5.0 /0.200
This assumes that the network is observed, rather than simulated. If the latter were the case then the observations would not be written. The previous example, with the same standard errors, but without the constants would become:

* DMES 1.0 1.0
AA BB
CC BB2 2.0 5.0
CDE FGH
BB CDE

6. RELATIVE ERRORS

One option available on LGC is that which allows a list of pairs of points to be specified, between which the relative errors are to be calculated. Since these relative errors are calculated in three dimensions, it is necessary to define the individual components which are displayed in the output list. They are: radial error, $\sigma_r$; bearing error, $\sigma_G$; vertical angle error, $\sigma_V$; height error, $\sigma_z$; and length error, $\sigma_L$. These are shown in Fig. 8.

![Fig. 8 Parameters for relative errors](image)

These errors are calculated from the characteristic equation for the observations which correspond to them. The first step is to form the $6 \times 6$ sub-matrix, $Q$, of the inverse of the normal matrix. This is composed of the six rows and columns corresponding to the three unknowns at each point (Fig. 9).
Note that if any coordinate at either point is fixed, then there will be no corresponding row and column in the inverse matrix. Since it is fixed, then it follows that its variance, and all its associated covariances, are zero. The relevant elements of Q are, accordingly, set to zero.

The relative height error is found directly from Q, and is given by the formula:

$$\sigma_Z^2 = \sigma_A^2 + \sigma_B^2 - 2\sigma_A \sigma_B \rho_{AB}$$

(10)

All the elements on the right hand side of this equation are found in Q. Having found $\sigma_Z$, the angular error $\sigma_\theta$ can be found from the relation:

$$\sigma_\theta = \frac{\sigma_Z}{L} \sin \theta$$

(11)

To find the relative length error, $\sigma_L$, it is necessary to perform the matrix equation:

$$\sigma_L^2 = C_L^T Q C_L$$

(12)

where $C_L$ is a $1 \times 6$ vector containing the theoretical coefficients of what would be the observation equation for a measured distance between the two points. In a similar way, the standard error of the bearing, $\sigma_\theta$, is found from the equation:

$$\sigma_\theta^2 = C_\theta^T Q C_\theta$$

(13)

where $C_\theta$ contains the coefficients of the observation equation for the bearing from one point to the other. The relative radial error, $\sigma_r$, is found from the formula:

$$\sigma_r = \sigma_\theta \cdot L \sin \theta$$

(14)

The pairs of points between which the relative errors are required are introduced by the key-word *EREL.*
7. **THE RESULTS OBTAINED**

7.1 **Final adjusted coordinates**

Each adjustment of an observed, rather than a simulated, network leads to a set of final adjusted coordinates in the X, Y and Z dimensions. These are printed for each point, with the points grouped as fixed points, free points, and so on. In general there is also a geoidal height, or measured altitude, corresponding to these coordinates, but if this is in a network with a local origin, in which case the geoidal height is meaningless, it is printed as zero.

The coordinates are usually printed up to the fifth decimal place, or hundredths of millimetres. If such precision is unnecessary, however, it may be clearer to specify that only the tenths of millimetres need to be printed. The user can do this by writing the key-word *DXI* in the input list. (The etymology of this may be obscure; it is from the French " dixièmes ", or " tenths ").

Together with the list of coordinates are printed the movements of the points in each dimension; DX, DY and DZ. These are the adjusted minus the approximate values. The standard errors of the positions of the points are also printed. These three values, in the X, Y and Z directions, are extracted from the inverse of the normal matrix.

7.2 **The observations**

After the list of adjusted coordinates comes the list of observations. These are grouped into class (measured distances, horizontal angles, etc.) and for each observation the following information is printed:

- the points stationed and observed
- the observation
- its adjusted value
- the standard error (sigma)
- the residual
- the ratio residual / sigma

In addition, certain types of observation have items of information printed which are peculiar to them. For the measured (spatial) distances, the altimetric sensitivity is printed. This value, which is useful in assessing the contribution to determining the Z coordinate at a point of a distance bearing upon it, is the variation caused to the distance by a change of 1 cm in the height of the point. For each zenithal distance observed, any extension specified on the target point, and the height of instrument which has been calculated, are printed.
The residuals of all angular observations are expressed both in centessimal seconds (cc), and in terms of the radial displacement at the observed point (Fig. 10).

![Diagram showing angular and radial residuals of an angle measurement](image)

**Fig. 10** Angular and radial residuals of an angle measurement

For each class of observation, the mean and standard deviation of the residuals are calculated. Denoting these as \( \bar{X} \) and \( S' \) respectively, \( S'^2 \) is taken as an unbiased estimate of the true variance of the population, and the statistic:

\[
\frac{\bar{X} - \mu}{S'/\sqrt{n}}
\]

is distributed as a Student \( t \)-distribution with \((n-1)\) degrees of freedom, where \( \mu \) is the population mean and \( n \) is the sample size.

So, taking a 5% level of significance, the corresponding 95% confidence interval for \( \mu \) is defined as:

\[
\left[ \bar{X} - t_{0.05} \cdot \frac{S'}{\sqrt{n}}, \quad \bar{X} + t_{0.05} \cdot \frac{S'}{\sqrt{n}} \right]
\]

(16)

where \( \text{Prob}\left[ \left| \frac{\bar{X} - \mu}{S'/\sqrt{n}} \right| > t_{0.05} \right] = 0.05. \)

The confidence limits for the standard deviation of the residuals can be calculated as well, since the statistic:

\[
\frac{(n-1) \cdot S'^2}{\sigma_0^2}
\]

is distributed as \( \chi^2 \) with \((n-1)\) degrees of freedom, where \( \sigma_0^2 \) is the variance of the population from which the sample is drawn. Again with a significance level of 5%, the limits \( \chi_0^2 \) and \( \chi_1^2 \) are defined as:

\[
\text{Prob.}\left[ \chi^2 < \chi_0^2 \right] = 0.05/2,
\]

\[
\text{Prob.}\left[ \chi^2 < \chi_1^2 \right] = 0.05/2.
\]
The 95% confidence interval for the variance is then given by:

\[
\left[ \frac{(n-1) S^2}{X_1^2}, \frac{(n-1) S^2}{X_0^2} \right]
\]

There are library routines in the CERN computing system which enable the Student t-distribution and the $\chi^2$ distribution to be used. Having the confidence limits for the mean and standard deviation of the residuals means that the user is better placed to interpret the effect of individual, or groups of, observations. Ultimately it is hoped that an option will be available where the residuals are printed as a histogram. A visual inspection can identify irregularities in the distribution which figures cannot.

7.3 Output for simulations

When LGC is running in simulation mode, several simulations are carried out on the network. It is not necessary to print the results of each simulation in full. Instead, the adjusted coordinates of all points are printed as before, but instead of a complete list of the observations a summary of the number in each class, together with the mean and standard deviation of the residuals, is printed. On the other hand, since it would be useful to have a list of the observations (but not have it repeated for each simulation) this is printed for the last simulation only.

After all the simulations have been completed, a final list of points is printed, showing for each one:

- the standard errors (1σ) of the position in the X, Y and Z directions,
- the maximum positive and negative movements in each direction over all the simulations.

7.4 Other information printed

The next item to be printed after the observations (or the summary of the simulations) is the list of relative errors. The five components defined in section 6 are printed here.

For all networks, simulated or observed, a final page is written which gives for each point:

- the X and Y coordinates at level zero
- the ellipsoidal height (non-measurable "geometric" altitude)
- the value of the geoid-ellipsoid separation at the LEP and zero levels
- the deviations of the vertical and n (North and East components)
- the deviations of the vertical in the CERN X and Y directions.
ACKNOWLEDGEMENT

I gratefully acknowledge the help and encouragement given to me by Michel Mayoud and many others during my stay at CERN.

* * *

REFERENCES


2) B. Bell, A simulation of the gravity field around LEP, Private communication, (1985).
APPENDIX

BREVIARY OF KEY-WORDS

*ANGL  List of horizontal angles
*CALA  List of fixed points
*DHOR  List of distances reduced to level zero and to the X,Y plane
*DIXI  Specify tenths of a millimeter rather than hundredths
*DMES  List of measured distances
*DVER  List of vertical distances
*ECHO  List of offsets in the X, Y plane
*ECSP  List of offsets in a three-dimensional plane
*ECHD  List of offsets from a theodolite direction
*ECVE  List of offsets from a plumb-bob line
*END   Last key-word
*EREI  List of pairs of points for relative errors
*FIN   Last key-word
*LEP   Work on the LEP level
*LOC   Local origin
*ORIE  List of orientation
*POIN  List of points variable in X, Y, z
*PDOE  Introduce the orientation point
*PUNC  Specify list of adjusted coordinates at end of output
*RADI  List of radial observations
*SIMU  For a simulation
*SPHE  Spherical earth model
*TITR  Introduction of title
*VXYI  List of points variable in X and Y
*VZ    List of points variable in Z only
*ZENH  List of zenithal distances - instrument height known
*ZENI  List of zenithal distances
INVERSE REGIONAL REFERENCE SYSTEMS:
A POSSIBLE SYNTHESIS BETWEEN THREE- AND TWO-DIMENSIONAL GEODESY

H.M. Dufour
Institut Geographique National, Saint-Mandé, France

ABSTRACT
For reasons perhaps more practical than theoretical, the early geodesists tended to consider horizontal and vertical coordinates separately. This introduced complications when it was decided to transform a straight line, in space, to a geodesic on an arbitrary ellipsoid. The appearance of three-dimensional geodesy in the last 30 years has permitted a simplification of the formulae - but the interest in the two groups of parameters remains, at times, fundamental, notably for the comparison of two networks (ancient or modern). The introduction of regional three-dimensional reference systems, defined, in principle, in the neighbourhood of a point O, by an inversion, permits the conservation of the simple formulae of three-dimensional geodesy. It also allows the division of the parameters into two categories: stereographic coordinates (\( \lambda, \phi \)) on a regional projection, in the plane tangent to O, and Z coordinate close to an altitude. Some practical applications are presented (for example the representation of the local gravity field).

1. THE HISTORICAL PERSPECTIVE OF THE PROBLEM

For many reasons, both theoretical and practical, geodesy has for a long time remained divided into distinct domains: horizontal positions defined by two parameters (two dimensional geodesy), and the third dimension characterised by levelling, giving an altitude HN, the evaluation of which has developed almost independently to that of the horizontal parameters (longitude \( \lambda \), latitude \( \phi \)). This separation was clearly not total, since a connection was made in using a "best-fit" ellipsoid for a given nation or region, these surfaces being estimated to minimize the deviations of the vertical.

\[
\xi = \phi(\text{Astro}) - \phi(\text{Geod.}) \\
\eta = (\lambda(\text{Astro}) - \lambda(\text{Geod.})) \cos \phi
\]

which implies the existence of a geometrical height H. The difference \( DH = H - HN \) was considered negligible.

One may wonder why the geodesists of the past, including among them the greatest mathematicians of their time, did not directly set the problem in three dimensions, but obstinately replaced the straight line AB in Euclidian space, by the geodesic linking the projections of "a" and "b" - the extremities of the segment AB on an ellipsoid. This geodetic line was a particularly difficult tool, even though its study permitted the development of remarkable theories on curved surfaces, such as Legendre's theorem concerning the spherical excess of a geodetic triangle, or the famous Gaussian
integral on a closed contour. I believe that the reasoning of our ancestors was more pragmatic: they had practically no means of handling matrix calculations, even of such small dimensions as 3 x 3, however the two dimensional field permitted the use of formulae of plane and eventually spherical trigonometrical tables.

The advent of spatial satellite surveying and the need for an improved precision in œdetic networks has induced the precise definition of a third dimension, distinctly different from the levelling height which is in fact, only a transformed potential. The geometric height H associated with geographical coordinates (λ, φ) must, when integrated with the latter, supply the cartesian coordinates (U, V, W):

\[ \begin{align*}
U &= (N + H) \cos \phi \cos \lambda \\
V &= (N + H) \cos \phi \sin \lambda \\
W &= (N(1 - e^2) + H) \sin \phi
\end{align*} \]  

(1) 

(N being the maximum radius of curvature at latitude φ, and e the eccentricity of the ellipsoid).

The transformation from a network (1) to a network (2) supposes that both are correctly measured in order to provide the principle translation parameters T (TU, TV, TW), or, for more complex problems, the translation, scale change, and rotation. For example:

\[ \begin{align*}
U_2 &= U_1 + T_U + a \cdot DU - r \cdot DV + q \cdot DW \\
V_2 &= V_1 + T_V + r \cdot DU + a \cdot DV - p \cdot DW \\
W_2 &= W_1 + T_W - q \cdot DU + p \cdot DV + a \cdot DW
\end{align*} \]  

(2) 

with DU = (U1 - U0), DV = (V1 - V0), DW = (W1 - W0), (U0, V0, W0) being the coordinates in network (1) of a central point (barycentric for example) of the working region. The scale change is represented by the parameter (a) and the rotation by the coefficients (p, q, r). [This type of formulae has been introduced by a number of authors such as: WOLF, BURSA, MOLODENSKY, BADESCAS, WEIS, etc ... The combination of Cartesian coordinates with those of a local system has been especially brought to the forefront1-3)].

Cartesian coordinates are irreplaceable in their use for worldwide representation, or for large expanses. On the other hand they can also lead to difficulties in the cartographic representation of results, which remain two dimensional. Until now, the cartographic representation has not evolved logically, treating the horizontal (λ, φ) coordinates separately from the vertical (H) coordinates. Most representations are simple with respect to (λ, φ) for example the Mercator (X, Y) projection

\[ X = \lambda ; \quad Y = \text{function of } (\phi) \]

But few of them really have the necessary spatial characteristics, that is to say they do not allow an acceptable transformation to a three-dimensional reference system.

In order to make use of these cartographic representations at the time of calculation, and for the presentation of results, it is necessary that they display the following quality: the uniqueness
of classical coordinates \((\lambda, \phi, H)\) must be replaced by systems of multiple coordinates, whilst retaining interchangeability for spatial transformations. One can therefore talk of the "dynamics" of reference systems.

There are really only two systems which are capable of allowing such transformations: the gnomonic and stereographic representations.

There exists a set of three-dimensional space representations which defines this space using three positional axes and an arbitrary orientation. Together these representations form a group having for their conjugate relationship a linear transformation (translation, rotation).

\[
\begin{align*}
X' &= aX + bY + cZ + d \\
Y' &= a'X + b'Y + c'Z + d' \\
Z' &= a''X + b''Y + c''Z + d''
\end{align*}
\]  

(3)

In two dimensions one can consider all the projections, from the same centre \(I\), to be from space to different planes \(\pi\). Hence one obtains the gnomonic representations, the link between them being the homographic relationships of the real variables \((a, b, c, a', b', c', P, Q)\):

\[
\begin{align*}
D \cdot X' &= a + bX + cY \\
D \cdot Y' &= a' + b'X + c'Y \\
D &= 1 + Px + QY
\end{align*}
\]  

(4)

The gnomonic projection conserves straight lines in space but introduces important distortions proportional to the distance from the central point (orthogonal projection of point \(I\) onto the plane of projection \(\pi\)).

The fundamental characteristics of the three-dimensional stereographic representation will be considered as:

- Choosing a reference plane \(\pi\) and centre \(O\) (the origin of the image space situated in the plane \(\pi\)).
- Performing from point \(P\) (the inversion pole), situated on the normal to \(\pi\) at \(O\), an inversion \((P, 4R^2)\) of pole \(P\) and of power \(4R^2\) (with, in principle \(2R = OP\), but this is not obligatory).
- The coordinates \((X, Y, Z)\) in image space are defined from the two axes \((OX, OY)\) situated in \(\pi\); the third axis \(OZ\) being determined by \(OP\) (in principle \(OZ\) is of the same sign as \(OP\)).

The advantage of the method lies essentially with the efficiency of the representation of the space to the proximity of an approximately spherical surface around \(O\); with the following conditions:

- Plane \(\pi\) is tangent (semi-rigorously) to the level surface at the origin \(O\).
- The pole \(P\) is (semi-rigorously) at the antipode of \(O\) on the surface.
- \(2R = OP\).
- The coordinates \((X, Y)\) then have the character of planimetric stereographic map coordinates.
- The coordinate \(Z\) is approximately an altitude.
In their general form, the different inverse three-dimensional representations \((X, Y, Z)\) of the same object space form a group obtained by the following type of relationships:

\[
\begin{align*}
D\cdot X' &= a + bX + cY + dZ + e(X^2 + Y^2 + Z^2) \\
D\cdot Y' &= a' + b'X + c'Y + d'Z + e'(X^2 + Y^2 + Z^2) \\
D\cdot Z' &= a'' + b''X + c''Y + d''Z + e''(X^2 + Y^2 + Z^2) \\
D &= P + QX + RY + SZ + T(X^2 + Y^2 + Z^2).
\end{align*}
\] (5)

In considering that relationships of the same type permit a direct passage from general cartesian coordinates \((U, V, W)\) to any \(X, Y, Z\) system, one realises that the geographical coordinates \((\lambda, \phi, H)\) can never be used in the calculations. There remains a few possible uses, but their main use in calculations with trigonometrical tables is no longer necessary.

It is important to note that all the envisaged transformations [Eq. (5)] are three dimensionally conformal, and that the properties of the figure in the image space \((X, Y, Z)\) are simply deduced from the only knowledge of the scale factor \(K\) (at a point \(Q\) in the object space):

\[
K^2 = \frac{dx^2 + dy^2 + dz^2}{(du^2 + dv^2 + dw^2)}
\]

\[
K = \frac{4R^2}{(PQ)^2}.
\] (6)

An important subset of transformations by inversion are those where the planes \(z\) are tangent to the same sphere \((S)\) at the points \(O\), the corresponding poles \(P\) being found at the antipoles of \(O\) on the sphere \(S\). In Eq. (5), the important simplifications are characterized by the separation of the coordinate \(Z\) from the rest of the equation [see Eq. (23)] and the possibility to uniquely define the subgroup of variables \((X, Y)\) under the following form:

\[
\begin{align*}
D\cdot X' &= a + bX + cY + e(X^2 + Y^2) \\
D\cdot Y' &= a' + b'X + c'Y + e'(X^2 + Y^2) \\
D &= P + QX + RY + T(X^2 + Y^2)
\end{align*}
\] (7)

which takes a remarkably different form if rewritten as:

\[
X + iY = z; \quad X' + iY' = z'
\]

\[
z' = \frac{a + bz}{1 + Pz}
\] (8)

\([z, z', a, b, P \in C (\text{complex numbers})]\).

An extension of these formulae for the ellipsoid \((E)\) requires a definition (of a two-dimensional nature) of the stereographic representation of the ellipsoid.

This general appraisal of the stereographic type transformations, has indicated the principle properties involved, in going from the simple to the complex. The appraisal of two-dimensional problems, thus appears worthwhile before considering the most general cases.
2. TWO-DIMENSIONAL PROBLEMS ON THE SPHERE

At an origin O (λ, φ) of the sphere, one considers the local system (XS, YS, ZS) which is deduced by rotation and translation from the terrestrial cartesian axes (U, V, W):

\[
\begin{align*}
XS &= U - O \\
YS &= RT V - O \\
ZS &= W R
\end{align*}
\]  
(9)

with

\[
RT = \begin{bmatrix}
-\sin \lambda & \cos \lambda & 0 \\
-\sin \phi \cos \lambda & -\sin \phi \sin \lambda & \cos \phi \\
\cos \phi \cos \lambda & \cos \phi \sin \lambda & \sin \phi
\end{bmatrix}
\]  
(10)

The following formulae give the STEREO (X, Y) coordinates

\[
X = XS/E; \quad Y = YS/E;
\]

\[
I = E = 1 + ZS \\
K = 1 + \frac{X^2 + Y^2}{4R^2}.
\]  
(11)

We will let \( z = X + iY \).

The whole of the STEREO representations of the sphere form a group for which the complex \( z \) coordinates are interrelated by the homographic transformations in the set of complex numbers C.

A simple formula relates the two elements of the group \((z1, z2)\):

\[
\frac{z1}{z1(2)} + \frac{z2}{z2(1)} + t^2 \frac{z1z2}{z1(2)z2(1)} = 1
\]  
(12)

\( z1(2) \) = coordinate of the origin of \((2)\) in system \((1)\)
\( z2(1) \) = coordinate of the origin of \((1)\) in system \((2)\)
\( t = \tan \frac{\theta}{2} \); \( \theta \) = the angular distance between the normals to the sphere from origins \((1)\) and \((2)\)

\[
t^2 = \frac{z1(2)z1(2)}{4R^2} = \frac{z2(1)z2(1)}{4R^2} \quad (z = \text{conjugate of } z).
\]

3. EXPLOITATION OF THE STEREO REPRESENTATION OF THE SPHERE

All the corrections necessary to transform measurements made in the image space (on the projected plane), to the object space (on the sphere) have been obtained from formulae involving the expressions of the scale factor \( K \).
Let $\hat{\widehat{ab}}$ be an arc of the sphere, the image of $\hat{\widehat{ab}}$ (of chord $\widehat{ab}$) on the plane.

We will assume the general form $X/2R = x$, $Y/2R = y$. Then one has:

$$K = 1 + x^2 + y^2, \quad E = \frac{1}{K} \quad \text{(Scale factor)} . \quad (13)$$

**Length of Great Circle**

$$\hat{\widehat{ab}} = 2R \arcsin \left( \frac{\hat{\widehat{ab}}}{2RK(a)K(b)} \right) . \quad (14)$$

**Arc to chord correction** $[p_1 = p_2 = dV/2]$

$$D \cdot tg(dV) = x(a)y(b) - y(a)x(b)$$
$$D = 1 + x(a)x(b) + y(a)y(b) . \quad (15)$$

**Correction of a direction**

This apparently unknown formula is easily deduced from the preceding correction.

At point a, the tangent to the image $\widehat{ab}$ of the great circle is the straight line $\widehat{ab}'$, $b'$ given by the vectorial relationship:

$$\frac{\partial \hat{\widehat{ab}}'}{\partial \hat{\widehat{ab}}} = \frac{1}{1 - \alpha}, \quad \text{with} \quad \alpha = (ab)^2/(4RK(a)) . \quad (16)$$

**Bearing (C) from a meridian**

$$-D \cdot tg(C) = 2x(y + tg\phi)$$
$$D = 1 + x^2 - y^2 - 2y\ tg\phi . \quad (17)$$

**Division of the arc of a great circle**

The division of the arc into two elements ($p$, $q$) according to the length relationship ($p + q = 1$) is possible, but only ($p = q = 0.5$) is simple.

The mid point G of the arc has an image g:

$$D \cdot X(g) = E(a)X(a) + E(b)X(b)$$
$$D \cdot Y(g) = E(a)Y(a) + E(b)Y(b)$$
$$D = E(a) + E(b) - 2 \sin^2 \frac{\alpha}{4} \quad (18)$$

with $\alpha = \hat{\widehat{ab}}/R$.

Equations (14) to (18) are valid and rigorous for $\alpha < 180^\circ$. 

4. TWO-DIMENSIONAL PROBLEMS ON THE ELLIPSOID

One can define a number of representations of the ellipsoid which can be qualified as "stereographic". However, we will only discuss one possible group in terms of functions of three parameters: \( F(\phi_0; \lambda, \phi) \):

- First, one considers the conformal representation of the ellipsoid on the sphere bitangential at parallel \( \phi_0 \), preserving the longitudes and the latitude of the origin \( \phi_0 \).

- One then considers the different STEREO representations of the image sphere possible at the points 0 on this sphere, which are images of the coordinates \((\lambda, \phi)\) of the ellipsoid.

From the results one can deduce the following:

- All the \( F \) representations form a set linked by a homographic relationship of the complex variables.

- One passes from the polar form of subgroup \((\phi_0)\) to those of subgroup \((\phi')\) by a scale change.

- For each element \( F(\phi_0, \lambda, \phi) \) there is a corresponding element \( F(\phi', \lambda, \phi') \) such that the relationship between the two reduces to a linear transformation (translation and scale change).

- There is an important point here as far as cartography is concerned. It seems useful to define a cartographic system

\[
F(0; \lambda, \phi)
\]

allowing local maps of different \((\lambda, \phi)\) i.e. local maps;

and a regional cartography: \( F(\phi_0 \neq 0; \lambda, \phi_0) \)

This consideration is interesting in applications close to latitude \( \phi_0 \) but the practical realization of regional maps is not necessary, since any local map of the general cartographic system is also (to a translation and scale change approximation) a regional map.

- The exploitation of stereographic representations of the ellipsoid necessitates the addition to the spherical corrections of residual terms. The use of the third dimension gives much more elegant solutions.

5. THREE-DIMENSIONAL PROBLEMS ON THE SPHERE

Figure 1 illustrates the three-dimensional stereographic representation on the sphere, in general: the origin of the projection is the point 0 of the sphere and the plane of the figure is defined by the great circle containing 0 and the point Q of the space represented.

\( X_0, Y_0, Z_0 \) are the local cartesian coordinates derived from the general cartesian coordinates \((U, V, W)\) by the relationship:

\[
\begin{align*}
X_0 &= U - 0 \\
Y_0 &= RT V - 0 \\
Z_0 &= W - R
\end{align*}
\]

(19)
The normal to the sphere passing through a general point $Q$ cuts the surface of the sphere at $M$, and must become in the inversion $(P, 4R^2)$ the point $q_1$ and by symmetry with reference to the tangential plane $\pi$ the point $q_2$ (the 3D - STEREO image of the general point $Q$).

If $(XS, YS, ZS)$ are the coordinates of $Q$ in the regional system, one has the relationships (in the same reference system):

\[
\begin{align*}
D \cdot X &= 4R^2(XS) \\
D \cdot Y &= 4R^2(YS) \\
(2R - Z)D &= 4R^2(ZS + 2R) \\
D &= XS^2 + YS^2 + (ZS + 2R)^2 .
\end{align*}
\]

With these formulae one can associate the following expressions of the radial distance $x = X^2 + Y^2$ and of the variable $Z$ as a function of the angular distance $\nu$ and of the real altitude $h$:

\[
\begin{align*}
D \cdot X &= (R + h) \sin \nu \\
D \cdot Z &= h(1 + \frac{h}{2R}) \\
D &= \cos^2 \frac{\nu}{2} (1 + \frac{h}{R}) + \frac{h^2}{4R^2} .
\end{align*}
\]
The transformation from a reference origin $O(\lambda, \phi)$ (coordinates $X, Y, Z$) to a reference origin $O'(0, \phi')$ (coordinates $X', Y', Z'$) corresponds to the following formulae, in which $O$ represents the angular distance of the arc $O'\overline{O}$, $A'$ being the direct azimuth and $A$ the reverse azimuth:

\[
\begin{align*}
D\cdot X' &= a + bX + cY + e(X^2 + Y^2 + Z^2) \\
D\cdot Y' &= a' + b'X + c'Y + e'(X^2 + Y^2 + Z^2) \\
D\cdot Z' &= Z \\
D &= P + QX + RY + T(X^2 + Y^2 + Z^2)
\end{align*}
\]

(22)

with

\[
\begin{align*}
a &= R \cos \phi \sin \lambda = R \sin \theta \sin \Lambda' \\
b &= \cos \lambda \\
c &= -\sin \phi \sin \lambda \\
e &= -\cos \phi \sin \lambda/4R = -\sin \theta \sin \Lambda'/4R \\
a' &= R \sin \phi \cos \lambda \\
b' &= \sin \phi' \sin \lambda \\
c' &= \cos \phi \cos \phi' + \sin \phi \sin \phi' \cos \lambda = -\sin \theta \cos \Lambda'/4R \\
e' &= \cos^2 \theta/2 \\
P &= \\
Q &= -\cos \phi' \sin \lambda/2R = \sin \theta \sin \Lambda/2R \\
R &= \sin \phi \cos \Lambda/2R \\
T &= \sin^2 \theta/4R^2
\end{align*}
\]

Note the behaviour of the variables $(Z, Z')$ in these formulae, which on the sphere $(Z = Z' = 0)$ correspond to the complex variable formula, linking the parameters $(z_1, z_2)$ [Eq. (12)].

6. GENERAL SPATIAL PROBLEMS

One recalls that it is possible to create, at a point anywhere in space, a regional inverse, where the points will be defined by the planimetric elements $(X, Y)$ and a third dimension $Z$ (which can semi-rigorously represent an altitude).

- One applies a motion displacement to the Cartesian coordinates $(U, V, W)$ i.e. a translation and rotation in order to define regional coordinates $(X, Y, Z)$ centered at point $O$.

- On the $ZS$ axis one chooses an inversion pole $P (OP = 2R)$ and applies the inversion $(P, 4R^2)$ followed by a symmetry relative to the plane $(X, O, Y)$. Hence one obtains the regional inverse coordinates $(X, Y, Z)$.

If one passes from an inverse space $V(O, P)$ to another inverse space $V'(O', P')$, with the following notation:
\( \omega \) = Rotation matrix of the 1st system to the second: \( \omega(ij) = a_{ij} \)

\( P^i P = \) Vector relating the inverse poles

\[
P^i P = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \text{ in V} ; \quad \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \end{pmatrix} \text{ in V'}
\]

\[
(P^i P)^2 = a^2 ; \quad OP = 2R ; \quad O^i P' = 2R'
\]

then one has:

\[
D \cdot X' = \left( \frac{R}{R'} \right)^2 \left[ a_{11}X + a_{12}Y + a_{13}Z \right] + \frac{R'^2}{4R^2} a' \left( X^2 + Y^2 + Z^2 \right)
\]

\[
\vdots
\]

\[
D \cdot Z' = \left( \frac{R}{R'} \right)^2 \left[ a_{31}X + a_{32}Y + a_{33}Z \right] + \frac{R'^2}{4R^2} \gamma' \left( X^2 + Y^2 + Z^2 \right)
\]

\[
D = 1 + \frac{a^2}{16R^4} \left( X^2 + Y^2 + Z^2 \right) + \frac{1}{2R^2} \left( \omega X + \beta Y + \gamma Z \right).
\] (24)

These formulae have the inversion poles \( P \) (for \( X, Y, Z \)) and \( P' \) (for \( X', Y', Z' \)) for their respective origins. Normally they must be brought back to points \( 0 \) and \( 0' \) by the following translations:

\[
\begin{align*}
Z & \text{ replaced by } (2R - Z) \\
Z' & \text{ replaced by } (2R' - Z')
\end{align*}
\]

which gives more complicated expressions of type (5). In the particular case of the sphere \( (R = R', P \Rightarrow P' \text{ interconnected with rotation } \omega) \) one arrives at Eqs. (22), (23).

7. REGIONAL INVERSES OF THE ELLIPSOID (E)

In the proximity of point 0 of the ellipsoid, of geographical coordinates \( (\lambda_0, \phi_0) \), one can define a regional stereographic representation, having the point 0 for its centre (cf. Section 4), but the exterior region does not have a simply definable image. For many reasons it seems preferable to define a regional inversion of centre 0, and pole at point P, on the normal from 0 to the ellipsoid. The straight line PO cuts the axis of revolution at \( I_1 \), eccentric to the origin \( I \), quantified by

\[
\zeta = \frac{P}{I} = Ne^2 \sin \phi_0 \quad \text{(Fig. 2)}.
\]
Fig. 2 Regional inverses of the ellipsoid (E)

We have adopted the principle of making $P$ the antipole of $O$, on the bitangential sphere of radius $N_0 = a/\sqrt{1 - e^2 \sin^2 \phi_0}$, but this solution is perhaps not optimal in certain cases.

In the inversion $(P, 4\Omega^2)$ followed by the symmetry with reference to the tangent plane at $O$ on the ellipsoid, the image of $(E)$ is found entirely in the part of the coordinate system under the plane $(O, XS, YS, ZS)$ with contact all along the parallel of latitude $\phi_0$. The final coordinates are called $(X, Y, Z)$.

The residual depression of the horizon in the image space is positive towards south or the north, and of the form

$$a \frac{e^2}{2} \cos^2 \phi_0 (\phi - \phi_0)^2 :$$

zero at the north pole and a maximum when $O$ is equatorial, and being $650$ m for $|\phi - \phi_0| = 10^2$.

The formulae giving the correspondence between the altitude $H$ and the variable $Z$ must be corrected to take account of this depression.
8. THREE-DIMENSIONAL CORRECTIONS

Referring to Fig. 1 for the notation = the formulae for the sphere are valid in the general case because they only use the radius R of the inversion sphere, the spatial coordinates (X, Y; Z) and in particular the expression of the scale factor K. They are completely soluble using the three-dimensional inversion formulae.

\[
\text{Scale: } K = \frac{(Pq)^2}{4R^2} = 1 - \frac{Z}{R} + \frac{X^2 + Y^2 + Z^2}{4R^2}
\]

(25)

\[
\text{Inverse: } E = \frac{1}{K} = \frac{4R^2}{(Pq)^2} = 1 + \frac{ZS}{R} + \frac{XS^2 + YS^2 + ZS^2}{4R^2}.
\]

The real length \(\overline{AB}\) of a spatial segment

If \(\overline{ab}\) is the image of \(\overline{AB}\) in the inverse representation, one has the relationship

\[
\overline{AB} = \frac{ab}{K(a)K(b)}.
\]

(26)

Correction to a direction

In the image space it is necessary to introduce the point \(J\), symmetric of pole \(P\) with reference to \(O\) \((J = -P)\).

To reproduce the straight line \(\overline{AB}\), it is necessary in projection to place the point \(b'\), shifted along the line \(\overline{ab}\), in the following manner:

\[
\frac{\overline{ab}'}{\overline{ab}} = \frac{1}{1 - \alpha}, \quad \text{with } \alpha = \frac{(ab)^2}{(4R^2 K(a))}.
\]

(27)

In the projection onto \(OZ\), this relationship introduces an important depression of the horizon. On the axes \((OX, OY)\) it reduces to:

\[
\frac{\overline{ob}'}{\overline{ob}} = \frac{1}{1 - \alpha}
\]

(28)

which is none other than a generalization of expression (16).

Partition of the spatial segment \(\overline{AB}\)

The image "g" of \(G\), divisor of \(\overline{AB}\) in the relationship \((p, q)\), has for coordinates:
\[ D \cdot \mathbf{j} = pK(a) \mathbf{a} + qK(b) \mathbf{a} \]
\[ D = pK(a) + qK(b) - pq(ab)^2/4R^2. \]

The horizontal expression is not identical to Eq. (18), because the spherical arc \( \widehat{AB} \) and the rectilinear arc \( \overline{AB} \) are not compatible. (They are however tangent at \( a \) and \( b \) in the horizontal projection.)

**Azimuths**

The correction of direction is generalized when "b" approaches infinity. One can thus calculate, in the image space and in proximity to "a", the fundamental vector transformations, the physical vertical, parallel to the polar axis - and as a result, evaluate the azimuth \( AZ \) and the Zenithal distance \( DZ \) of the segment \( \overline{AB} \) from the analogous elements in the STEREO image space. These calculations, as well as the corresponding differential terms \( d(AB), d(AZ), d(DZ) \), make finite formulae for a length of segment \( ab \).

The fact that we have defined a rigorously conformal transformation is naturally essential for the rigour and the generality of the results.

9. **POSSIBLE USES**

9.1 **The three-dimensional stereographic representation**

The representation that one arrives at is the simplest of the conformal two dimensional representations - and at the same time, has a precise meaning for the third dimension.

We can say however, that there is a small difference between the stereographic representation of the ellipsoid as it has been defined in Section 4 and the image of the ellipsoid as it has been determined by the coordinates \((X, Y)\) of the inverse space. But these differences are inconceivably small with respect to the facilities offered by the third dimension: we really consider the problem of the straight line \((AB)\) in space without projection onto any surface.

Equations (26) and (27) are remarkable in their simplicity and diversity of use, a practical limitation issues from the variation of the scale and the difficulty of finding a working relationship between the variable \( Z \) and the altitude \( H \) (an extent of 3000 km can be proposed for practical purposes).

9.2 **The refined analysis of three-dimensional networks**

The regional inverse reference system allows one to maintain the advantage of the three-dimensional mode whilst carefully distinguishing the third dimension from the other two.

One has to be able to put formulae (2) into the following form:
\[ dX' = X' - X = T_X \left[ K - \frac{X^2}{2R^2} \right] - T_Y \left( \frac{XY}{2R^2} \right) - T_Z \left( \frac{X}{R} \right) \left( 1 - \frac{Z}{2R} \right) + \delta X \]
\[ dY' = Y' - Y = -T_X \left( \frac{XY}{2R^2} \right) + T_Y \left( K - \frac{Y^2}{2R^2} \right) - T_Z \left( \frac{Y}{R} \right) \left( 1 - \frac{Z}{2R} \right) + \delta Y \]
\[ dZ' = Z' - Z = T_X \left( \frac{X}{R} \right) \left( 1 - \frac{Z}{2R} \right) + T_Y \left( \frac{Y}{R} \right) \left( 1 - \frac{Z}{2R} \right) + T_Z \left( 2 - K \right) \left( 1 - \frac{2Z}{R(2-K)} \right) + \delta Z \]

\[ \delta X = aX(1 - \frac{Z}{R}) - rY + q(Z + \frac{X^2 - Y^2 - Z^2}{2R}) \]
\[ \delta Y = rX + aY \left( 1 - \frac{Z}{R} \right) - p(Z + \frac{Y^2 - X^2 - Z^2}{2R}) \]
\[ \delta Z = (-qX + pY)(1 - \frac{Z}{R}) + a(Z + \frac{X^2 + Y^2 - Z^2}{2R}) \]  

(30)

These formulae allow the use of the fundamental parameters: translation \((T_X, T_Y, T_Z)\), scale \((a)\), and rotation \((p, q, r)\) with the rigorous coefficients. These elements are expressed on the regional axis, and with reference to the origin \(O\) (which perhaps can be taken on the ellipsoid \((E)\)), the \(Z\) axis being the normal from \(O\) to \(E\) including the inversion pole \(P\), situated at the antipole position of \(O\) on the bitangent sphere at that point (see Fig. 2). One will note that the rotations \((p, q)\) about the horizontal axis are different from the rotation \(r\) about the vertical axis from \(O\).

These formulae can be completed without difficulty for the planimetric variations by the complementary terms which add to \((\delta X, \delta Y)\):

\[
\begin{align*}
\delta X' &= bX + cY + eX^2 + fXY + gY^2 \ldots \\
\delta Y' &= cX - bY + e'X^2 + f'XY + g'Y^2 \ldots 
\end{align*}
\]  

(32)

permitting a more comprehensive analysis of the two networks, incorporating in particular the linear deformation\(^4\):

\[ v = \sqrt{a^2 + c^2} \]

**Note:**

The formulae (30) and (31) are rigorously equivalent to (2), if one removes the principal part of the translations \((\sqrt{TX^2 + TY^2 + TZ^2 < 50}^3)\) - and if one takes account of the fact that the term \((Z^2/2R^2)dZ\) (quite negligible) has been neglected.

9.3 The possible reference systems for studying the local gravity field

We have already evoked time and time again the possible use that can be made of the local inverse coordinates \(X, Y, Z\) obtained by an inversion (and a reflection) from the cartesian points
(U, V, W). If T is the potential harmonic perturbation in the original space, then \( T' = T/\sqrt{K} \) is a potential harmonic image in the inverse space \((X, Y, Z)^5\).

An analysis of the problem, still in progress, tends to eliminate the representation of \( T \) by the Fourier double series in the plane \( O, X, Y \) or by the use of BESSEL functions - in favour of the use of translated functions rapidly decreasing in the plane \( Z = 0 \), their effect over the whole space would be simple to calculate = the possibility to create grid squares on the plane, with binary subdivisions, gives a real advantage for the process, since one can then greatly shorten the calculations by precalculating the number of influencial terms between the grid points at level 0.

10. **CONCLUSION**

It is not impossible to imagine a global system of regional inverse representations organized around an ellipsoid or a mean sphere, with centres positioned according to a system of origins which should be precisely given. They can also serve to support the exchange of information in three dimensions, in concurrence with the actual geographical coordinate system, and the altitude \((\lambda, \phi, H)\). The multiplicity of origins is perhaps an inconvenience, the possible advantage in large calculations being the absence of trigonometrical lines. The materialization of the \( X, Y \) coordinates so created, by stereographic maps can be an important advantage.

One will note that the system can be used at any point where the curvature of level lines is not far from being spherical (on the Earth, at a particular height, or on the planets) - and that it naturally incorporates all the local systems that one can create.

For example, a local CERN system could be defined, which covers an area of 100 km\(^2\) around a central point, and an altitude range between 0 and 1000 m. The system consists of \((X, Y, Z)\) coordinates that can be considered as follows:

- \(X, Y\) are the stereographic map coordinates of an ellipsoid, with a precision which is better than 0m,000 at altitude 0
  - 0m,050 for \(X = 0, Y = \pm100\) km, \(Z = 1\) km
- \(Z\) represents the altitude:
  - at level 0, better than 2m,45 at \(X = 0; Y = \pm100\) km
  - at level 1 km, better than 0m,36 at \(X = Y = 0\)
    - better than 2m,57 at \(X = 0; Y = \pm100\) km

The spatial scale coefficient \(K\), related to the origin, is a maximum when \(Z = 0\), at the limit of the zone \((K = 1,000062)\), and a minimum when \(Z = 1\) km at \(X = Y = 0\) \((K = 0.999843)\).

The corrections of the direction for a distance \(ab = 10\) km, correspond to a shift of the observed point \(b\):

- Radially, towards the exterior, a distance of 15 mm,4 at 25 km from the centre,
- Vertically (downwards) a distance of the order of 7 m,80 (which approximately corresponds to the classical depression of the horizon).
REFERENCES


ON CONTINUOUS THEORY FOR GEODETIC NETWORKS

Kai Borre
Aalborg University, Fibigerstræde 11, DK-9220 Aalborg Øst, Denmark.

Abstract
Geodetic levelling networks and networks with distance measurements are analysed in order to elucidate their special features. This analysis naturally leads one to look for continuous analogons of these networks. By means of tools known from the method of finite elements we derive the corresponding Green's functions for various boundary value problems. The Green's functions act as a formal covariance function for the corresponding least squares problem and provide information on the general error behaviour in the networks. The theory is illustrated with examples relevant to the CERN survey problems.

1. INTRODUCTION

During the last decade the theoretical analysis of geodetic networks has advanced tremendously. Not least, the geodetic literature about optimization problems of various kinds has left an overwhelming impression.

For a better and more comprehensive understanding of the special features of geodetic networks, I am still of the opinion that we need more basic knowledge of the real matter, before we can solve most optimisation problems in a professional manner.

The number of mathematical theories is tremendous, but the problem is to introduce into geodesy only the relevant ones or, in the words of Robert M. Pirsig\textsuperscript{1}), "The purpose of scientific method is to select a single truth from among many hypothetical truths. That, more than anything else, is what science is about. But historically science has done exactly the opposite".

The mathematical analysis of geodetic networks is mostly based on matrix algebra, theory of elliptic difference and differential equations, Fourier analysis, statistical estimation theory, theory of electrical networks, elasticity theory and above all the finite element method which has very much in common with the theory of geodetic networks.

As the main aim of this paper is to emphasize the interplay between discrete and continuous models I shall now examine this topic. The original problem can be traced back to ideas published by G. Förster in 1931. He wanted to establish a numerical model for combining neighbouring networks with some common boundary points. (In fact, F.R. Helmert considered that problem
already in 1893.) Förster proposed a solution which combines the two networks by means of a function of place which turns out mainly to be a conformal mapping. In later publications he uses concepts like "quadratic sum of residuals per unit area or the strength", homogeneity and isotropy, free networks, rotation of networks, scale-factor of net deformation, conformal mapping, etc. It is obvious that Förster was inspired by the theory of elastically deformable solids.

The finite element method combined with the ideas of Förster led to the publication by T. Krarup and myself2 of the paper "Foundation of a Theory of Elasticity for Geodetic Networks".

The mathematical abstract of the method is the following:

The usual adjustment problem of a single triangle is set up. Next, the original observations are transformed into so-called pseudo-observations. In levelling they represent the slope of terrain in the direction of the coordinate axes and the closing error of this triangle. In the case of distance and azimuth networks they are linear functions of the elements of the metric tensor. Henceforth we calculate the weighted square sum of residuals. The kernel of this symmetric form is analogous to what in elasticity theory is called the stiffness matrix.

Now we look upon the network as a whole consisting of such elementary triangles. Here the crucial point appears, namely how we prolongate the discrete scalar and vector fields to continuous functions. We solve this problem by following simple principles from the method of finite elements. The weighted square sum of the residuals can now be extended to an approximating Riemann-Stieltjes' sum over the entire network. The search for a minimum of this sum leads to a variational problem the solution of which is a boundary value problem of the Neumann type.

This boundary value problem is the continuous analogue to the normal equations. The inverse of the partial differential operator (including the boundary conditions) is an integral operator, the kernel of which is the Green's function of the operator concerned. The Green's function acts analogously to the usual covariance matrix of the original discrete adjustment problem.

So the method yields one Green's function which is an approximating substitute for an n-dimensional covariance matrix. Instead of trying to grasp an n-dimensional covariance matrix we are furnished with a single covariance function which depends on certain characteristic parameters for the network. Such a simplification of course has its costs, e.g. we lose information about local phenomena of the network.
Let us illustrate the whole procedure by the following simple but characteristic example:

Along a line of length L we have observed the length of n sub-intervals. As formal observation we introduce the logarithm of the length of the sub-intervals rather than the lengths themselves:

$$\ln(x^0_{i+1} + u_{i+1} - x^0 - u_i) = I_{i+1} + r_{i+1}$$  \hspace{1cm} (1)

where $x^0_i$ denotes the preliminary value of the abscissa of subpoint i, $I_{i+1}$ is the observational value, $r_{i+1}$ the residual and $u_i$ the correction to the preliminary abscissa $x^0_i$. (1) is expanded in a truncated Taylor series of order 1:

$$\ln(x^0_{i+1} - x^0_i) = I_{i+1} - I_i - r_{i+1} - x^0_{i+1} - x^0_i.$$

We introduce the following abbreviations

$$\epsilon_i = x^0_{i+1} - x^0_i$$

$$f_i = I_{i+1} - I_i - \ln(x^0_{i+1} - x^0_i)$$

and prescribe

$$u_{i+1} - u_i = \epsilon_i u'.$$

Then (2) reads

$$-f_i + u' = r_{i+1}.$$  \hspace{1cm} (3)

The weighted square sum of residuals for the whole line L is then

$$\Delta E = \sum_{i=1}^{n} \frac{P_i}{\epsilon_i} (u' - f_i)^2.$$  \hspace{1cm} (4)

At this point it is appropriate to make a remark. We are interested in making a transcription of our algebraic problem into the language of differential operators. So we are discussing what is going to happen as we bring our sub-points closer and closer together. We are obviously interested in the limit to which our construction tends as the number of sub-points increases to infinity by letting $\epsilon_i$ decrease to zero. In order to save the valuable feature of a metrical space in relation to the study of continuous operators.
we have to modify the usual concept of weight $p_i$ by dividing it by $c_i$ exactly as shown in (4). This is an important modification.

Next we look for the minimum for $E$. Assuming $p_i/c_i=1$, which leads to a simple result and is no serious restriction, we get the result

$$-u'' = f'.$$

(5)

For a so-called free network we have the boundary conditions

$$u'(0) = u'(L) = 0.$$  

(6)

The generalized Green's function for this problem is

$$G(u,v) = \begin{cases} \frac{u^2 + v^2}{2L} + \frac{L}{3} - u & \text{for } v < u \\ \frac{u^2 + v^2}{2L} + \frac{L}{3} - v & \text{for } v \geq u. \end{cases}$$

(7)

This Green's function acts as a formal covariance function for the adjustment problem. An example is the variance of the difference of abscissae $\phi_{pq}$ of any two points $p$ and $q$ as follows:

$$\phi_{pq} = x_p - x_q, \quad p < q$$

$$\sigma^2(\delta) = G(p,p) + G(q,q) - G(p,q) - G(q,p)$$

$$= \left(\frac{p^2}{L} + \frac{L}{3} - p\right) + \left(\frac{q^2}{L} + \frac{L}{3} - q\right) - \left(\frac{p^2 + q^2}{2L} + \frac{L}{3} - p\right)$$

$$- \left(\frac{p^2 + q^2}{2L} + \frac{L}{3} - q\right)$$

$$= q - p$$

(8)

i.e. the variance of a difference of length is proportional to itself. This result is in good agreement with the result known from the discrete case.

2. SYSTEMATIC ANALYSIS OF CERN-LIKE NETWORKS

What are the special features of the CERN survey problems?

The detail network is (nearly) 1-dimensional, but along the circumference of a circle, and mainly distance observations are performed. So we shall concentrate on a thorough analysis of the error behaviour of that type of network.

In the following investigations we make certain assumptions in order to obtain explicit results. All observations are taken with variance $\sigma^2$, and they are supposed to be uncorrelated, i.e. the a priori covariance matrix is $\sigma^2 I$. Furthermore, observations are only performed between neighbouring points which are equally spaced.
In an appendix we include a systematic investigation of 1-dimensional networks. Remember that levelling and distance observations give rise to similar normal equation matrices.

2.1 Distance observations along a straight line

For reasons of reference we shall briefly repeat the characteristics about distance observations along a straight line. The normals for a so-called free network are, with the above-mentioned assumptions, proportional to

\[ N = \begin{bmatrix} 1 & -1 \\ -1 & 2 & -1 \\ & & \ddots \\ 0 & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix} \]  \hspace{1cm} (9)

cf. \( N_3 = R_1(2,-1,-1) \) in the appendix. Obviously, \( N \) is singular, so for its inverse we use the uniquely determined pseudo-inverse \( N^+ \). The elements of \( N^+ \) are given as:

\[
\frac{1}{2n} \begin{bmatrix}
i^2 - i^+ (n+1) (2n+1) \\
i^2 - (2n+1) i^+ (n+1) (2n+1) + j(j-2n)
\end{bmatrix}
\]

for \( i \leq j \) \hspace{1cm} (10)

\[
\frac{1}{2n} \begin{bmatrix}
i^2 - i^+ (n+1) (2n+1) \\
i^2 - i^+ (n+1) (2n+1) + j(j-2n)
\end{bmatrix}
\]

for \( i \geq j \).

The variance of a difference of distance between points \( P \) and \( Q \), say with ordinates \( p \) and \( q \), \( p < q \), is

\[
\sigma^2_{PQ} = \frac{1}{2n} \left[ p^2 - p^+ (n+1) (2n+1) + p(p-2n) \right] + \frac{1}{2n} \left[ q^2 - q^+ (n+1) (2n+1) + q(q-2n) \right]
\]

\[
\times \frac{1}{n} \left[ p^2 - p^+ (n+1) (2n+1) + p(p-2n) \right] + \frac{1}{n} \left[ q^2 - q^+ (n+1) (2n+1) + q(q-2n) \right] = q - p.
\]  \hspace{1cm} (11)

We see, as we also expected, that the variance of a distance difference grows in proportion to the difference itself.

The continuous analogon to the problem just treated is the following boundary value problem:

\[-u''(\xi) = \delta \]

with \( u'(0) = u'(L) = 0 \).  \hspace{1cm} (12)

The (generalized) Green's function for the problem (12) is
\[ G(x, \xi) = \begin{cases} \frac{x^2 + \xi^2}{2L} + \frac{L}{3} - x & \text{for } \xi < x \\ \frac{x^2 + \xi^2}{2L} + \frac{L}{3} - \xi & \text{for } \xi > x, \end{cases} \] (13)

cf. the appendix.

For the same variance we now have

\[ \sigma^2_{PQ} = \left( \frac{p^2 + p^2}{2L} + \frac{q^2}{2L} + \frac{L}{3} - p \right) + \left( \frac{q^2 + q^2}{2L} + \frac{L}{3} - q \right) - 2 \left( \frac{p^2 + q^2}{2L} + \frac{L}{3} - q \right) \]

\[ = q - p. \] (14)

This result is exactly the same as that one obtained in the discrete case (11). The approximation error is zero.

### 2.2 Distance observations along the circumference of a circle

In this case the normals are modified slightly to

\[ N = \begin{bmatrix} 2 & -1 & 0 & \ldots & \ldots & -1 \\ -1 & 2 & -1 & 0 & \ldots & \ldots \\ 0 & -1 & 2 & -1 & \ldots & \ldots \\ & \ldots & \ldots & \ldots & \ldots & \ldots \\ -1 & 0 & \ldots & \ldots & 0 & -1 & 2 \end{bmatrix} \] (15)

But this modification has great consequences. By the way, in the case of observations along a circle the boundary conditions are of no interest, just because of the absence of a boundary!

This matrix - a circulant matrix - is also singular, so we look for its pseudo-inverse. It is explicitly given by

\[ N^+_{ij} = \frac{1}{2n} \left[ \frac{n^2 - 1}{6} n(i-j) + (i-j)^2 \right] \text{ for } i \geq j \] (16)

and is also symmetric.

The variance of distance difference along the circumference is

\[ \sigma^2_{PQ} = \frac{1}{2n} \left[ \frac{n^2 - 1}{6} n^2 - 2 \left( \frac{n^2 - 1}{6} n(p-q) + (p-q)^2 \right) \right] \]

\[ = \frac{(p-q)}{n} \frac{(p-q)^2}{n}. \] (17)

In comparing (14) and (17) we note that the variance on the circle has decreased as compared to the "free-line" case. This was also to be expected. The variance has a maximum, when the points P and Q are situated diametrically opposite!
The continuous formulation of the problem is

\[ -u''(\phi) = \delta \]
\[ u'(0) = u'(2\pi) \]  \hspace{1cm} (18)

where the solution now has to be periodic with the period \(2\pi\).

We obtain the Green's function formally as the solution B.3 in the appendix with \(x\) and \(\xi\) substituted by the angles \(\phi\) and \(\psi\), and \(L = 2\pi r\), where \(r\) denotes the radius of the circle:

\[
G(\phi, \psi) = \begin{cases} 
\frac{\phi^2 + \psi^2}{4\pi r} + \frac{2\pi r}{3} - \phi & \text{for } \psi \leq \phi \\
\frac{\phi^2 + \psi^2}{4\pi r} + \frac{2\pi r}{3} - \psi & \text{for } \psi > \phi.
\end{cases}
\]

Next we obtain

\[
\sigma^2_{PQ} = (\psi - \phi)\sigma^2. \hspace{1cm} (19)
\]

Seemingly, \(r\) has been eliminated, but observe the following fact. The distance \(PQ\) along the circumference of the circle with radius \(r\) is, of course, \(r\) times the corresponding one on the unit circle. If we want to preserve the same number of network points in the sector \(PQ\) along the circle with radius \(r\) as along the unit circle, there will be \(r\) times the distance on the unit circle between the points, and consequently the variance increases (possibly proportional to \(r\), but this depends on the specific circumstances of the observational method).

There exists an alternative method for obtaining \(N^+\), viz. through the spectral decomposition of (15).

We shall report the results in detail, as it may be of interest to geodesists. Those interested in the procedure used in deriving the results are advised to consult Polozhii.\(^3\)

The eigenvalues \(\lambda_j\) of \(N\) in (15) are given as

\[
\lambda_j = 4 \sin^2 \left( \frac{(j-1)\pi}{n} \right), \hspace{0.5cm} j = 1, 2, 3, \ldots, n. \hspace{1cm} (20)
\]

Immediately, we observe that \(0 < \lambda_j < 4\), but also it turns out to be essential to distinguish between whether \(n\) is odd or even. In both cases we have \(\lambda_1 = 0\), while for \(n\) even
In order to describe the set of eigenvectors in an easy manner we first reorder the eigenvalues determined by (20) according to magnitude. Then the corresponding eigenvectors are as follows — in the right succession, and with \(\alpha = \frac{2}{n}(i-1)\pi\):

**n odd**

\[
\phi = \sqrt{\frac{2}{n}} \begin{bmatrix}
\frac{1}{\sqrt{2}} \cos \alpha & \sin \alpha & \cos 2\alpha & \sin 2\alpha, & \ldots, & \cos \frac{(n-1)\alpha}{2} & \sin \frac{(n-1)\alpha}{2} \\
\frac{1}{\sqrt{2}} \cos 2\alpha & \sin 2\alpha & \cos 4\alpha & \sin 4\alpha, & \ldots, & \cos \frac{2(n-1)\alpha}{2} & \sin \frac{2(n-1)\alpha}{2} \\
\frac{1}{\sqrt{2}} \cos 3\alpha & \sin 3\alpha & \cos 6\alpha & \sin 6\alpha, & \ldots, & \cos \frac{3(n-1)\alpha}{2} & \sin \frac{3(n-1)\alpha}{2} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{\sqrt{2}} \cos n\alpha & \sin n\alpha & \cos 2n\alpha & \sin 2n\alpha, & \ldots, & \cos \frac{n(n-1)\alpha}{2} & \sin \frac{n(n-1)\alpha}{2}
\end{bmatrix}
\]

**n even**

\[
\phi = \sqrt{\frac{2}{n}} \begin{bmatrix}
\frac{1}{\sqrt{2}} \cos \alpha & \sin \alpha & \cos 2\alpha & \sin 2\alpha, & \ldots, & \cos \frac{(n-2)\alpha}{2} & \sin \frac{(n-2)\alpha}{2} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} \cos 2\alpha & \sin 2\alpha & \cos 4\alpha & \sin 4\alpha, & \ldots, & \cos \frac{2(n-2)\alpha}{2} & \sin \frac{2(n-2)\alpha}{2} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} \cos 3\alpha & \sin 3\alpha & \cos 6\alpha & \sin 6\alpha, & \ldots, & \cos \frac{3(n-2)\alpha}{2} & \sin \frac{3(n-2)\alpha}{2} & \frac{1}{\sqrt{2}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{\sqrt{2}} \cos n\alpha & \sin n\alpha & \cos 2n\alpha & \sin 2n\alpha, & \ldots, & \cos \frac{n(n-2)\alpha}{2} & \sin \frac{n(n-2)\alpha}{2} & \frac{1}{\sqrt{2}}
\end{bmatrix}
\]

**Examples**

n=3 \(N^+ = \phi \Lambda^+ \phi^T\)

where \(\Lambda^+\) is a diagonal matrix with the inverse eigenvalues — in the reordered succession. If \(\lambda_k = 0\), then \(\lambda_k^+ = 0\).

\[
\phi = \sqrt{\frac{2}{3}} \begin{bmatrix}
\frac{1}{\sqrt{2}} \cos \frac{2\pi}{3} & \sin \frac{2\pi}{3} \\
\frac{1}{\sqrt{2}} \cos \frac{4\pi}{3} & \sin \frac{4\pi}{3} \\
\frac{1}{\sqrt{2}} \cos \frac{6\pi}{3} & \sin \frac{6\pi}{3}
\end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{\sqrt{3}}{2} \\
\frac{1}{\sqrt{2}} & \frac{1}{2} & -\frac{\sqrt{3}}{2} \\
\frac{1}{\sqrt{2}} & \frac{1}{2} & 0
\end{bmatrix}
\]

\[
\Lambda^+ = \text{diag}(0, \frac{1}{3}, \frac{1}{3})
\]
\[ N^+ = \frac{1}{9} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \]

\[ \Phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad , \quad \Lambda^+ = \text{diag}(0, \frac{1}{2}, \frac{1}{2}, 4) \]

\[ N^+ = \frac{1}{16} \begin{bmatrix} 5 & -1 & -3 & -1 \\ -1 & 5 & -1 & -3 \\ -3 & -1 & 5 & -1 \\ -1 & -3 & -1 & 5 \end{bmatrix} \]

2.3 Distance observations in an annulus

Now one can argue that the CERN networks are not of the simple line nature discussed so far. In order to deal with this objection we shall report some results covering this case.

The fundamental figure consists of a quadrilateral with measured diagonals. This case has just been investigated by Meissl\(^4\)) who also includes an additional observation, but this does not change the results qualitatively.

The variance in the tangential direction is asymptotically proportional to \( \frac{n}{12} \) and in the normal direction asymptotically proportional to \( \frac{n^3}{720} \), where \( n \) denotes the number of quadrilaterals around the circle. The reader is referred to Meissl\(^4\) for further details.

In the continuous case we shall derive the Green's function for an annulus. The problem has, with Neumann boundary conditions, the following formulation:

\[-\Delta u(z) = g(z) \quad \text{in } \Omega \quad (23)\]
\[ \frac{\partial u(z)}{\partial r} = 0 \quad \text{on } \omega. \quad (24) \]

The corresponding Green's function \( G(z, \zeta) \) shall satisfy the following conditions, cf. Meschkowski:\(^5\):

(i) For a fixed \( \zeta \in \Omega \) and \( z \in \Omega^+\omega \) \( G \) must be harmonic for any \( z \neq \zeta \),

(ii) The function \( G + i \Re |z - \zeta| \) must be harmonic in the neighbourhood of \( z = \zeta \),

(iii) At the boundary \( z = \omega, z \in \Omega, \frac{\partial G(z, \zeta)}{\partial z} = \text{constant} \) for a fixed \( z \).

The index \( z \) indicates that the differentiation has to be performed with respect to the variable \( z \).

Furthermore, at the boundary \( G \) has to be normalized through the condition

\[ \int_{\omega} G(z, \zeta) d\omega = 0. \quad (25) \]

Then in \( \Omega^+\omega \) the harmonic function \( u(z) \) may be obtained in the following manner

\[ u(\zeta) = \frac{1}{2\pi} \int_{r_i}^{r_e} \int_0^{2\pi} G(z, \zeta) \frac{\partial G(z, \zeta)}{\partial r} d\Omega. \quad (26) \]

By standard methods we obtain for \( G \)

\[
G(r, \Phi; \rho, \Psi) = \frac{r_i^n r_\rho^n - r_e^n r_\rho^n}{r_i^n + r_e^n} - \frac{r_i^n r_\rho^n - r_e^n r_\rho^n}{(r_i^n + r_e^n)^2} \]

\[
+ \sum_{n=1}^{\infty} \left[ \left( \frac{r_i}{r} \right)^n + \left( \frac{r}{r_i} \right)^n \right] \left[ \left( \frac{\rho}{r} \right)^n + \left( \frac{r}{\rho} \right)^n \right] \cos n(\Phi - \Psi) \left[ \frac{r_i^n}{r_i^n} - \frac{r_e^n}{r_e^n} \right], \quad \text{for } r < \rho. \quad (27)
\]
3. CONCLUSION

By a simple example it was demonstrated how we can proceed from a discrete network to a continuous analogous one. The Green's function for the continuous case acts as a formal covariance function which describes the propagation of random errors. In the appendix the influence of various boundary conditions is studied in more detail.

As specific examples relevant to the CERN networks, we investigated circulant matrices and the pertaining error propagation. In nearly all cases we are faced with error functions which are described by means of various powers of $n$. Only in the case of the annulus do these functions change character to be functions involving the logarithm of $n$. This feature reflects the general behaviour. 1-dimensional networks are described by powers of the distance between points, 2-dimensional networks with the logarithm of the distance between points, and 3-dimensional networks with the inverse of the distance between points.

The theory is described in a more elaborate publication which will be published in the near future. The title is: Untersuchung geodätischer Netze mittels partieller Differentialoperatoren.
APPENDIX

COVARIANCE MATRICES/FUNCTIONS FOR VARIOUS 1-D GEODETIC PROBLEMS

A. DISCRETE CASES

Consider a straight line with \( n \) equidistant points. Any height or distance difference between neighbouring points is measured with variance \( \sigma^2 \). Furthermore, all observations are considered to be uncorrelated.

Thus the weight matrix is \( \sigma^2 I \). Using the method of least squares we obtain the following coefficient matrix of the normal equations (adjustment by variation of parameters).

(i) Control is given at points 1 and \( n \):

\[
N_1 = \begin{bmatrix}
2 & -1 & & \\
-1 & 2 & -1 & \\
& -1 & 2 & -1 \\
& & \ddots & \\
& & & -1 & 2 & -1 \\
& & & & 0 & -1 & 2
\end{bmatrix}_{n,n}
\]

(ii) Control is given only at point 1:

\[
N_2 = \begin{bmatrix}
2 & -1 & & \\
-1 & 2 & -1 & \\
& -1 & 2 & -1 \\
& & \ddots & \\
& & & -1 & 2 & -1 \\
& & & & 0 & -1 & 2
\end{bmatrix}_{n,n}
\]

(iii) No control available at all. Free network:

\[
N_3 = \begin{bmatrix}
1 & -1 & & \\
-1 & 2 & -1 & \\
& -1 & 2 & -1 \\
& & \ddots & \\
& & & -1 & 2 & -1 \\
& & & & 0 & -1 & 1
\end{bmatrix}_{n,n}
\]
In the sequel we shall use a notation (apart from the sign) introduced by D.E. Rutherford\textsuperscript{6}:

\[
R_n(x,a,b) = \begin{bmatrix}
x+b & -1 & & 0 \\
-1 & x & -1 & \\
& -1 & x & -1 \\
& & & & \ddots \\
0 & & -1 & x & -1 \\
& & & & & -1 & x+a
\end{bmatrix}.
\]

The subscript \( n \) denotes the dimension of the tri-diagonal symmetric matrix. Hence

\[
N_1 = R_n(2,0,0)\\
N_2 = R_n(2,-1,0)\\
N_3 = R_n(2,-1,-1).
\]

Next we give a useful summary of characteristics of these matrices and their inverses.

(i) \( R_n(2,0,0) \)

The eigenvalues are \( \lambda_i(R_n(2,0,0)) = 4 \sin^2 \frac{i\pi}{2(n+1)} \), \( i = 1,2,\ldots,n \), and the normalized eigenvectors are

\[
\psi_i(R_n(2,0,0)) = \sqrt{\frac{2}{n+1}} \sin \frac{i\pi}{n+1}, \quad i,j = 1,2,\ldots,n.
\]

The matrix composed of the eigenvectors is not positive definite.

Let \( \text{cond}(A) \) denote the spectral condition number of a symmetric, real matrix \( A \), defined by

\[
\text{cond}(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}
\]

then

\[
\text{cond}(R_n(2,0,0)) \sim \frac{4(n+1)^2}{\pi^2}.
\]

Further,

\[
\det(R_n(2,0,0)) = n+1.
\]
Finally

\[
R_n^{-1}(2,0,0) = \frac{1}{n+1} \begin{bmatrix}
    n & 2n & 3n & \ldots & n^2 \\
    2n & 2(n+1) & 3(n-1) & \ldots & n(n-1) \\
    3n & 3(n-1) & 3(n-2) & \ldots & n(n-2) \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    n^2 & n(n-1) & n(n-2) & \ldots & n \\
\end{bmatrix} = \begin{cases} i(n+1-i) & \text{for } i\leq j \\
\frac{i(n+1-j)}{n+1} & \text{for } i> j
\end{cases}
\]  

(A.1)

(ii) \( R_n(2,-1,0) \)

\[
\det(R_n(2,-1,0)) = 1
\]

\[
R_n^{-1}(2,-1,0) = \begin{bmatrix}
    1 & 1 & \ldots & 1 \\
    1 & 2 & \ldots & 2 \\
    1 & 2 & \vdots & \vdots \\
    1 & 2 & \ldots & n
\end{bmatrix} = (\min(i,j)).
\]  

(A.2)

\( R_n^{-1}(2,-1,0) \) is a so-called Stieltjes matrix. Such a matrix has very nice features, especially in connection with SOR-methods. The optimal convergence rate can be predicted (see Varga7).

The largest eigenvalues of \( R_n^{-1}(2,-1,0) \) are very well-conditioned and the smallest very ill-conditioned. For \( n=12 \) the first few \( \lambda \)'s are of order unity, while the last three are of order \( 10^{-7} \). With increasing values of \( n \) the smallest eigenvalues become progressively worse conditioned. As stated, \( \det(R_n(2,-1,0)) = 1 \), but if the \( (1,n) \)-element is changed to \( (1+\varepsilon) \) the determinant becomes \( 1+(n-1)!\varepsilon \). If \( \varepsilon=10^{-10} \) and \( n=20 \) the determinant is changed from 1 to \( (1-19!10^{-10}) = -1.2 \cdot 10^7 \). Now, as the determinant equals the product of its eigenvalues, at least one eigenvalue of the perturbed matrix must be very different from that of the original matrix (see Wilkinson8).

(iii) \( R_n(2,-1,-1) \)

The eigenvalues are \( \lambda_i(R_n(2,-1,-1)) = 4 \sin^2 \frac{(i-1)\pi}{2n}, \ i = 1,2,\ldots,n \) and the normalized eigenvectors

\[
\psi_i(R_n(2,-1,-1)) = \begin{cases} \frac{1}{\sqrt{n}} & i = 1 \\
\sqrt{\frac{2}{n}} \cos \frac{(i-1)(2j-1)\pi}{2n} & i = 2,3,\ldots,n \\
\end{cases}
\]

\( j = 1,2,\ldots,n \)
\[ \text{cond}(R_n(2,-1,-1)) = \infty \]
\[ \text{det}(R_n(2,-1,-1)) = 0. \]

For the inverse we use the pseudo-inverse given by

\[ R_n^T(2,-1,-1) = \frac{1}{2n} \left[ \begin{array}{c} 3(n-1)(2n+5-6j)+(j-1)(j-2)+i(i-1) \\ \end{array} \right] \quad \text{for } i \leq j \quad \text{(A.3)} \]

else symmetric.

### B. CONTINUOUS CASES

(i) The continuous analogon is given through the problem:

\[ -u''(\xi) = \delta \]

with boundary conditions \( u(0) = u(L) = 0 \). The corresponding Green’s function is

\[ G(x,\xi) = \begin{cases} 
\xi - \frac{x^2}{L} & \text{for } \xi \leq x \\
\frac{x^2}{L} - \xi & \text{for } \xi \geq x 
\end{cases} \quad \text{(B.1)} \]

cf. Lanczos\(^9\), Problem 203.

(ii) The problem is

\[ -u''(\xi) = \delta \]

with boundary conditions \( u(0) = u'(L) = 0 \). Solution

\[ G(x,\xi) = \begin{cases} 
\xi & \text{for } \xi \leq x \\
x & \text{for } \xi \geq x 
\end{cases} \quad \text{(B.2)} \]


(iii) The problem is

\[ -u''(\xi) = \delta \]

with boundary conditions \( u'(0) = u'(L) = 0 \). Solution

\[ G(x,\xi) = \begin{cases} 
\frac{x^2 + \xi^2}{2L} + \frac{L}{3} - x & \text{for } \xi \leq x \\
\frac{x^2 + \xi^2}{2L} + \frac{L}{3} - \xi & \text{for } \xi \geq x 
\end{cases} \quad \text{(B.3)} \]

cf. Lanczos\(^9\), Problem 228.
C. MEAN VARIANCE OF ALL POINTS ALONG THE LINE

We define

\[ \sigma^2_{\text{mean}} = \frac{1}{n} \text{tr} \left( R_{-1}(2,\ldots) \right) \sigma^2, \]

cf. e.g. Meissl\textsuperscript{4}, eq. (6). Consequently

(i) According to (A.1) we get

\[ \sigma^2_{\text{mean}} = \left\{ \frac{1}{n} \sum_{i=1}^{n} i(n-1+i) \right\} \sigma^2 = \left( \frac{n(n+1)}{6} \right) \sigma^2. \]

(ii) According to (A.2) we get

\[ \sigma^2_{\text{mean}} = \left\{ \frac{1}{n} \sum_{i=1}^{n} i \right\} \sigma^2 = \left( \frac{n}{2} \right) \sigma^2. \]

(iii) According to (A.3) we get

\[ \sigma^2_{\text{mean}} = \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left[ (n-1)(2n+5-6i) + (i-1)(i-2) + i(i-1) \right] \right\} \sigma^2 \]

\[ = \left( \frac{n}{6} - \frac{1}{6n} \right) \sigma^2. \]

Notice that we obtain minimum variance in the case of the free line!

Next, in the continuous case we define

\[ \sigma^2_{\text{mean}} = \left( \frac{1}{L} \int_{0}^{L} G(x,x) \, dx \right) \sigma^2. \]

So

(i) According to (B.1) we get

\[ \sigma^2_{\text{mean}} = \left( \frac{1}{L} \int_{0}^{L} \left( x - \frac{x^2}{2} \right) \, dx \right) \sigma^2 = \frac{L}{6} \sigma^2. \]

(ii) According to (B.2) we get

\[ \sigma^2_{\text{mean}} = \left\{ \frac{1}{L} \int_{0}^{L} x \, dx \right\} \sigma^2 = \frac{L}{2} \sigma^2. \]
(iii) According to (B.3) we get

\[ \sigma^2_{\text{mean}} = \left\{ \frac{1}{L} \int_0^L \left( \frac{x^2}{L} - x + \frac{3}{3} \right) dx \right\} \sigma^2 = \frac{L}{6} \sigma^2. \]

Notice that in the continuous case we miss the constant terms. In general the fixed-free-case (ii) is the poorest as far as the overall determination of points is concerned.

D. VARIANCE OF POINT DIFFERENCES. APPROXIMATION ERRORS

We shall consider the variance of a point difference \( H_Q - H_P \) between any two points \( P \) and \( Q \) with nodal numbers \( p \) and \( q, \ p < q \). The discrete formulations are considered first:

(i) \[ \sigma^2_{PQ} = \left\{ R^{-1}_n (2,0,0)_{QQ} + R^{-1}_n (2,0,0)_{PP} - 2 R^{-1}_n (2,0,0)_{PQ} \right\} \sigma^2 = \left\{ (q-p) - \frac{(q-p)^2}{n+1} \right\} \sigma^2, \quad \text{cf. eq. (A.1)} \]

(ii) \[ \sigma^2_{PQ} = R^{-1}_n \left\{ (2,1,0)_{QQ} + R^{-1}_n (2,-1,0)_{PP} - 2 R^{-1}_n (2,-1,0)_{PQ} \right\} \sigma^2 = (q-p) \sigma^2, \quad \text{cf. eq. (A.2)} \]

(iii) \[ \sigma^2_{PQ} = \left\{ R^{-1}_n (2,-1,1)_{QQ} + R^{-1}_n (2,-1,1)_{PP} - 2 R^{-1}_n (2,-1,1)_{PQ} \right\} \sigma^2 = (q-p) \sigma^2, \quad \text{cf. eq. (A.3)} \]

The corresponding expressions for the continuous formulations follow from

\[ \sigma^2_{PQ} = \left\{ G(q,q) + G(p,p) - 2 G(p,q) \right\} \sigma^2 \]

(i) \[ \sigma^2_{PQ} = \left\{ (q-p) - \frac{(q-p)^2}{L} \right\} \sigma^2, \quad \text{cf. eq. (B.1)} \]

(ii) \[ \sigma^2_{PQ} = (q-p) \sigma^2, \quad \text{cf. eq. (B.2)} \]

(iii) \[ \sigma^2_{PQ} = (q-p) \sigma^2, \quad \text{cf. eq. (B.3)}. \]
There is an astonishingly good agreement between the exact discrete formulation and the corresponding results of the continuous approach. The latter is often easier to perform than the discrete one, thus we have demonstrated how the continuous approach is superior to the discrete analysis which is also the more difficult one. On the other hand, we normally lose the constant terms in the continuous solution.

*   *   *

REFERENCES


NEW TRENDS IN MATHEMATICAL GEODESY

H. Moritz
Technical University, Graz, Austria

ABSTRACT
Modern mathematical geodesy is faced with increasingly large data
sets. This requires the use of large computers as well as an
adequate theoretical understanding of the underlying problems. For
this purpose, linear functional analysis, a generalization of linear
algebra to infinite-dimensional spaces, furnishes powerful mathema-
tical tools, especially Hilbert space methods and Fourier transform
techniques. A general review of such methods and their applications
is given.

1. HANDLING LARGE DATA SETS

Geodetic analysis is called upon to handle data sets which rapidly increase in size.
Sets of the order of a million data are no longer unrealistic. There are several reasons for
this:

- Modern computers are able to perform rigorously extensive calculations which formerly
  had to be broken up, approximately, into a number of smaller tasks. An example is the con-
temporary readjustment of the North American triangulation, which involves several hundred
thousand stations.

- Another reason lies in the nature of modern geodetic measuring techniques which often
  furnish a data stream which flows almost continuously in time. Satellite techniques such as
  Doppler, GPS or satellite altimetry serve as examples for this phenomenon.

Thus \( n \), the number of data or the number of unknowns, may be very large. To get rid
of a definite number \( n \), it may well be useful to let \( n \to \infty \). Surprisingly enough, this
may frequently simplify matters and facilitate understanding the problem, with positive
effects on the actual numerical treatment of the data. It is like considering a definite
forest, rather than a set of 713 692 trees.

To give a more appropriate illustration: difference equations are known to be math-
ematically very difficult. For this reason, one tries to approximate them by differential
equations for which both a highly developed theory and efficient numerical techniques exist.
The truth of this is strengthened, rather than impaired, by the dialectical fact that these
numerical techniques again use finite-difference methods: their development has become
possible only on the basis of the theory of differential equations.
Going to the infinitely large is conceptually almost the same as going to the infinitely small as in the case of differential equations. Now we have, for instance, a data set which forms a vector of \( n \) elements, \( n \) being a very large number. If we formally let \( n \to \infty \), we get a vector in an infinite-dimensional space which is called Hilbert space (provided certain mathematical conditions are satisfied). Hilbert and similar infinite-dimensional spaces (e.g. Banach spaces) are very fashionable in contemporary mathematics. They are the subject of functional analysis which is almost as well developed nowadays as classical analysis (differential and integral calculus). In fact, some abstract modern treatments ("Bourbaki", Dieudonné) introduce differential and integral calculus right away on such general spaces ...

If the data are, in some way, regularly distributed, for instance with equal sampling intervals, then a branch of linear functional analysis, namely spectral analysis or Fourier transform techniques, provides very powerful theoretical and computational tools. Fourier transforms are an analog of diagonal matrices and can therefore be handled very well. You have certainly heard of Fast Fourier Transforms which are increasingly used in geodetic applications as well as in numerical mathematics.

Letting \( n \to \infty \) is not only a mathematical artifice but comes natural in many instances. Consider the earth’s gravity field. Its complete mathematical description requires an infinite number of parameters, e.g., the coefficients in a series of spherical harmonics. Other examples will be given later.

2. THE POWER OF SYMBOLISM

Consider the matrix equation

\[
Ax = y
\]  
(1)

and its solution

\[
x = A^{-1}y.
\]  
(2)

If \( A \) is a regular square matrix (of nonvanishing determinant) of dimension \( n \times n \), and if \( x \) and \( y \) are \( n \)-vectors (more precisely, column vectors \( n \times 1 \)), the (2) is the standard solution of a regular system of \( n \) linear equations for \( n \) unknowns, \( A^{-1} \) being the ordinary inverse matrix of \( A \).

But the formalism (1) and (2) also works if \( A \) is a rectangular matrix \( m \times n \), whether \( m \) is smaller or larger than \( n \). Then the corresponding system of linear equations has the form
\[ a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n = y_1, \]
\[ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2n}x_n = y_2, \]
\[ \vdots \]
\[ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \ldots + a_{mn}x_n = y_m. \]

In the sequel, we shall always consider the case \( m < n \). Then the system \( (3) \) admits an infinite number of possible solution vectors \( x = [x_1, x_2, x_3, \ldots, x_n]^{T} \), a column vector being considered as the transpose \( T \) of a row vector.

Equation \( (2) \) can still be interpreted to hold, but in the sense of a generalized inverse matrix; (cf. Bjerhammer, 1973\(^1\)) From the theory of generalized inverses it is known that \( A^{-1} \) has the form

\[ A^{-1} = K(A^T A)^{-1} A^T, \]

where \( K \) is an arbitrary \( n \times n \) square matrix such that the \( m \times m \) square matrix \( A A^T \) has a regular inverse (verify by substitution!). Various matrices \( K \) give various admissible solutions.

An "underdetermined" system of type \( (3) \) is furnished by the linearized condition equations in adjustment by conditions: there are less conditions than observations. The least-squares solution in classical:

\[ x = K(A^T A)^{-1} y, \]

\( K \) being the inverse weight matrix and \( x \) being the residuals (usually denoted by \( v \)). It has the form \( (2) \) with \( (4) \).

Let us go one step further and let \( n \rightarrow \infty \). Then the matrix \( A \), originally

\[
\begin{array}{c}
\text{n} \\
m \\
A
\end{array}
\]

becomes

\[
\begin{array}{c}
m \\
A
\end{array}
\]

extending to infinity towards the right. Then obviously \( x \) must be a Hilbert space vector, whereas \( y \) remains a \( m \)-vector.

Then \( (3) \), on \( n \rightarrow \infty \), becomes a system of \( m \) equations for infinitely many unknowns, and now comes a curious fact. The solution can still be written in the form \( (2) \), with the generalized inverse again given by \( (4) \)! Of course, now \( K \) will be an infinite square matrix, but it can be arbitrary as long as all sums, which are now infinite series, will converge.
Infinite systems of such type do occur in geodetic practice, namely in the determination of spherical-harmonic coefficients from satellite observations; (cf. Moritz, 1980, sec. 21)\textsuperscript{2}). In fact, there are infinitely many harmonic coefficients, to be determined from a large but finite number of equations. This is a special case of least-squares collocation, to be mentioned in the next section.

So the simple and innocent equations (1) and (2) are of much wider validity and applicability than intended by the inventors of regular square matrix inverses. A good symbolism has a peculiar power which goes far beyond the original intention, almost autonomously tending towards generalization. Sometimes children outgrow their parents beyond their boldest expections.

3. HILBERT SPACE TECHNIQUES

3.1 Error propagation. Consider a linear function $F$ of a vector $f$:

$$F = h^Tf = \sum_{i=1}^{n} h_i f_i \quad (6)$$

the real number $F$ may be regarded as the inner product of $f$ with an auxiliary vector $h$ or as the matrix product of the row vector $h^T$ (transpose of the column vector $h$) with the column vector $f$.

If the variance-covariance matrix of $f$ is denoted by $\Sigma$, of elements $\sigma_{ij}$, then the standard error $m_F$ of $F$ is well known to be

$$m_F^2 = h^T \Sigma h = \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \sigma_{ij} \quad (7)$$

Consider next a \textit{linear functional} $F$ of a function $f(u)$:

$$F = \int_{u=a}^{b} h(u) f(u) du \quad (8)$$

Obviously there is a close analogy between (6) and (8): there corresponds:

<table>
<thead>
<tr>
<th>function $f(u)$</th>
<th>to vector $f_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable $u$</td>
<td>to subscript $i$</td>
</tr>
<tr>
<td>integral $\int_{u=a}^{b}$</td>
<td>to sum $\sum_{i=1}^{n}$</td>
</tr>
</tbody>
</table>

By analogy with (7) we expect the standard error of the functional $F$ to be given by

$$m_F^2 = \int_{u=a}^{b} \int_{v=a}^{b} h(u)h(v) \sigma(u,v) du dv \quad (10)$$
\( \sigma(u,v) \) denoting the error covariance function of \( f(u) \), which is the continuous analogue of the matrix \( \sigma_{ij} \).

So far we have exploited the analogy between the continuous function \( f(u) \) and the discrete vector \( f_1 \). An even closer analogy can be found in the following way.

Expand \( f(u) \) into a series of orthonormal functions \( \phi_i(u) \):

\[
 f(u) = \sum_{i=1}^{\infty} f_i \phi_i(u) \ .
\]  

(11)

For instance, let \( a = 0 \) and \( b = 2\pi \) and

\[
\phi_1(u) = \frac{1}{\sqrt{2\pi}}, \\
\phi_2(u) = \frac{1}{\sqrt{\pi}} \cos u, \\
\phi_3(u) = \frac{1}{\sqrt{\pi}} \sin u, \\
\phi_4(u) = \frac{1}{\sqrt{\pi}} \cos 2u, \\
\phi_5(u) = \frac{1}{\sqrt{\pi}} \sin 2u, \\
\phi_6(u) = \frac{1}{\sqrt{\pi}} \cos 3u, \\
\phi_7(u) = \frac{1}{\sqrt{\pi}} \sin 2u, \\
\ldots
\]

(12)

Then (11) is nothing else than a Fourier series, and \( f_i \) are the Fourier coefficients of the function \( f(u) \).

If we similarly expand

\[
 h(u) = \sum_{i=1}^{\infty} h_i \phi_i(u) 
\]

(13)

\[
 \sigma(u,v) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sigma_{ij} \phi_i(u) \phi_j(v), 
\]

then (8) and (10) may be readily seen to become

\[
 F = \sum_{i=1}^{\infty} h_i f_i, 
\]

(14)

\[
 \sigma_F^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} h_i h_j \sigma_{ij}. 
\]

(15)

Now the analogy to (6) and (7) is perfect; the only difference is \( n \to \infty \).

Thus we see that we have two isomorphic kinds of Hilbert spaces: the space of (square integrable) functions \( f(u) \) and the space of infinite vectors \( f_1 \), the isomorphy (one-to-one equivalence) being established by (11). The theoretical physicist, of course, will smile: this is the equivalence of Schrödinger's wave formulation and Heisenberg's matrix formulation of quantum mechanics!
So far, so good, but what is the use of all this in geodesy? When I introduced these ideas in my doctoral thesis in 1958, I was working in the cadastral service where areas of pieces of land were still determined by an old-fashioned apparatus called a polar planimeter, which essentially is an analog integrator. So I started from the error theory of integration (Moritz, 1961)\textsuperscript{3}).

The most natural application of error theory in function spaces is, of course, in physical geodesy, since the gravitational potential, gravity anomalies, etc., are such continuous functions, the corresponding infinite vectors being formed by the coefficients of spherical-harmonic series, as we have already mentioned in the preceding section. Since these applications are now standard, I shall not dwell on them here too much.

However, other applications present themselves readily. Integrations are involved, \textit{e.g.}, in integrated Doppler measurements or in inertial surveying where coordinate differences are obtained by twice integrating measured accelerations.

In the terminology of mathematical statistics, the transition to infinite-dimensional space corresponds, of course, to the transition from random vectors to \textit{stochastic processes}.

3.2 \textit{Adjustment Computations}. Least-squares collocation, now very popular in physical geodesy, may be regarded as an adjustment in Hilbert space, quite similar to an adjustment by conditions (Kraup, 1969\textsuperscript{4}), Moritz, 1980, sec. 25\textsuperscript{2}). This is understandable since it involves the gravitational potential, which is a continuous function; see also the simple example given in sec. 2.

In fact, least-squares adjustment in Hilbert spaces can be developed in an abstract and rigorous way, both adjustment by parameters and adjustment by conditions, cf. (Sieg, 1983)\textsuperscript{5}). I need not give you the formulas: they look exactly the same as in usual adjustment computations, only that vectors may now be continuous functions or infinite vectors, and that matrices may be infinite or, more generally, will be linear operators. Don't be afraid: linear operators are in Hilbert space what matrices are in finite-dimensional spaces.

The geodesist who may have to apply these formulas will be gratified with their familiar look: this is another example of the power of symbolism. The mathematician, however, will pay attention to the fact that transition to the infinite involves nontrivial questions of convergence, existence, etc.

Let me finally mention an example where continuous adjustment by parameters comes natural. Consider an "integrated navigation system" for a land vehicle, a ship, or an airplane, consisting of an inertial system combined with a Global Positioning System (GPS) receiver. In principle, both systems can provide continuous recordings of position; in both cases, the observations could be considered continuous functions. Combining them by a suitable adjustment can provide elimination of systematic effects of either system and even information on the gravity field.
3.3 Nonlinear Functional Analysis. Linear functional analysis is the mathematical domain that treats Hilbert and Banach spaces, linear operators and functionals, etc. Nonlinear functional analysis, of course, is much more difficult mathematically, but formally sometimes surprisingly simple, because nonlinear operators are formally quite analogous to nonlinear functions; note again the power of symbolism ... So I cannot refrain from mentioning that a very simple first conceptual understanding of Molodensky's boundary value problem, the dread and delight of physical geodesists, can be obtained in this way, (cf. Moritz, 1980, sec. 40)²).

4. SPECTRAL METHOD

4.1 Diagonal matrices. Consider again the linear equation system (1), \( A x = y \), or if \( A \) is a square 3 × 3 matrix:

\[
\begin{align*}
  a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= y_1, \\
  a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= y_2, \\
  a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= y_3, \\
\end{align*}
\]

(16)

or briefly

\[
y_i = \frac{1}{3} \sum_{j=1}^{3} a_{ij}x_j;
\]

(17)

\( y_i \) is obtained from \( x_j \) by a "discrete convolution", that is, multiplication followed by summation.

Matters are much simpler if \( A \) is a diagonal matrix: (16) reduces to

\[
\begin{align*}
  a_{11}x_1 &= y_1, \\
  a_{22}x_2 &= y_2, \\
  a_{33}x_3 &= y_3,
\end{align*}
\]

(18)

or if we write \( x_i = x(i) \), \( y_i = y(i) \) —this may be motivated by the analogy (9)—and also \( a_{ii} = a(i) \), then instead of (17) we simply have

\[
y(i) = a(i)x(i),
\]

(19)

that is, convolution is replaced by a simple multiplication.

The simplification is particularly striking if we consider the inverse of a diagonal matrix:

\[
x = A^{-1}y
\]

(20)

reduces to
\[ x(i) = \frac{y(i)}{a(i)} \quad (21) \]

that is, inversion reduces to a simple division.

We could, of course, bring every square matrix to diagonal form by an appropriate linear transformation, so that the simple relations (18) through (21) hold—clearly, \( n = 3 \) was chosen only for simplicity and the procedure is valid for any \( n \), but finding such a linear transformation is in general certainly no less difficult than solving the original system \( Ax = y \).

In certain cases, however, such a diagonalization is rather simple, and this is the basic idea of spectral analysis. Not only matrices, but also linear operators, their equivalent in Hilbert space, can be diagonalized in this way.

4.2 Fourier Series and Integrals. We have met with Fourier series already in eqs. (11) and (12). Using the well-known Euler-Moivre identity

\[ \cos \alpha + i \sin \alpha = e^{i \alpha} \quad (i = \sqrt{-1}) \quad (22) \]

we can instead write

\[ f(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt} \quad (23) \]

where the Fourier coefficients \( c_k \) are given by

\[ c_k = \frac{1}{2\pi} \int_{0}^{2\pi} f(t)e^{-ikt} dt \quad (24) \]

Here we have denoted the independent variable by \( t \) instead of \( u \); in fact, \( t \) is often, but by no means always, time; it may also be distance along a spatial profile. Accordingly we say that the function \( f \) is defined in a "time domain" or in a onedimensional "space domain".

We note a considerable analogy between (23) and (24), similar to (9): both formulas have a similar structure, an integral corresponding to a sum. This analogy may even be emphasized by putting

\[ c_k = \overline{f(k)} \quad (25) \]

the set of these Fourier coefficients forms the spectrum of the function \( f(t) \).

It is well known that the expansion (23) holds for periodic functions with period \( 2\pi \). Similar Fourier series can be obtained for functions with an arbitrary period \( T \). By letting \( T \to \infty \) we get functions that are not periodic at all. For them, Fourier series are replaced by Fourier integrals; instead of (23) and (24) we then have
\[ f(t) = \int_{-\infty}^{\infty} \tilde{f}(\lambda)e^{i\lambda t} d\lambda \quad , \quad (26) \]

\[ \tilde{f}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)e^{-i\lambda t} dt \quad . \quad (27) \]

Here \( \tilde{f}(\lambda) \) denotes the spectrum, or Fourier transform, of the function \( f(t) \); note that we now have a continuous spectrum instead of the discrete spectrum (25).

Now the symmetry between (26) and (27) is almost perfect; it is disturbed only by the inessential factor \( 2\pi \); both \( i \) and \( -i \) have essentially the same nature, both being square roots of \( -1 \).

The continuous analogue of (17) is the formula

\[ g(t) = \int_{-\infty}^{\infty} h(t,\tau)f(\tau)d\tau \quad ; \quad (28) \]

especially important is the case

\[ h(t,\tau) = h(t-\tau) \quad , \quad (29) \]

which is a function of the difference \( t-\tau \) only.

Then

\[ g(t) = \int_{-\infty}^{\infty} h(t-\tau)f(\tau)d\tau \quad (30) \]

denotes a convolution in the proper sense: multiplication followed by integration.

We are purposely disregarding all questions of convergence, etc., which justly are a major concern to mathematicians but can be omitted in our heuristic introduction.

Now we have a beautiful and fundamental fact, which reminds us of the diagonal matrices mentioned at the beginning of this section. Consider the Fourier transforms, or spectra, (27) and

\[ \tilde{g}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)e^{-i\lambda t} dt \quad , \quad (31) \]

\[ \tilde{h}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(t)e^{-i\lambda t} dt \quad . \]

Then the convolution in the "time domain" is equivalent to a simple multiplication

\[ \tilde{g}(\lambda) = \tilde{h}(\lambda)\tilde{f}(\lambda) \quad (32) \]

in the "frequency domain" (\( \lambda \) has the character of a frequency).
The equivalence of (30) and (32) constitutes the basic theorem of spectral analysis. Spectral analysis, or harmonic analysis, not only holds for functions on the line, i.e., of a variable \( t \)--this leads to Fourier series and Fourier integrals—but also for functions defined on the plane or on the sphere. This leads to two-dimensional Fourier series and Fourier integrals for the plane and to spherical harmonics for the sphere. They represent transformations from a "space domain" (plane or sphere) to a frequency domain.

The basic integral formulas of physical geodesy, such as Stokes' formula, are convolutions on the sphere or in the plane. This makes the usefulness of spectral methods in physical geodesy immediately obvious; the evaluation of integral formulas by the "Fast Fourier Transform" (cf. Brigham, 1974)\(^6\) is now well established (cf. Sideris, 1985)\(^7\).

Similarly, matrix inversions are replaced by division in the frequency domain, which makes "frequency domain collocation" very attractive, although it is restricted to rather regular data distributions (cf. Colombo, 1979)\(^8\).

However, in line with my endeavor to conceal (not very successfully) my bias towards physical geodesy, I shall finally consider an example from the theory of

4.3 Geodetic Networks. Sufficiently regular networks can be successfully studied by spectral methods (cf. Meissl, 1976)\(^9\). This presupposes a regular network structure, but gives extraordinary mathematical insight which can be used for understanding networks of a more irregular structure.

I shall try to illustrate this by means of a very simple example (taken from Sünkel, 1985)\(^10\). Consider a regular levelling line consisting of equidistant points:

```
... o---o---o---o---o...
```

\( k-2 \quad k-1 \quad k \quad k+1 \quad k+2 \)

It is very long, so that it can be regarded as extending to infinity in both directions.

The observation equations are

\[
\begin{align*}
\hat{h}_{k+1} - \hat{h}_k &= \ell_{k,k+1} ,
\end{align*}
\]

\[
\begin{align*}
\hat{h}_{k-1} + 2h_k - \hat{h}_{k+1} &= -\ell_{k,k+1} + \ell_{k-1,k} = r_k.
\end{align*}
\]

For an infinite levelling line, \( k \) may assume all positive or negative integer values.
The normal equations (34) may be written in the matrix form

\[ \mathbf{A} \mathbf{h} = \mathbf{r} \]

or, more explicitly

\[
\begin{bmatrix}
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & -1 & 2 & -1 & 0 \\
0 & -1 & 2 & -1 & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots
\end{bmatrix}
\begin{bmatrix}
h_{-1} \\
h_0 \\
h_1 \\
\vdots
\end{bmatrix}
= 
\begin{bmatrix}
r_{-1} \\
r_0 \\
r_1 \\
\vdots
\end{bmatrix}
\]

(36)

The matrix \( \mathbf{A} \) is infinite but has a very regular band structure; it is a Toeplitz matrix. In fact, we have

\[ a_{ij} = a_{j-i} \]

(e.g. \( a_{\infty} = a_{11} = a_{22} = \ldots = 2, a_{10} = a_{21} = a_{32} = \ldots = -1 \), etc.), reminiscent of (29).

Because of this regularity, we may introduce the function

\[ \tilde{a}(\lambda) = e^{-i\lambda} + 2 - e^{-i\lambda} = 2(1 - \cos \lambda) \]

(38)

to represent the Fourier transform of the matrix \( \mathbf{A} \), the coefficients being the non-zero elements in any row (or column) of \( \mathbf{A} \), and similarly

\[ \tilde{h}(\lambda) = \sum_{-\infty}^{\infty} h_{\kappa} e^{-i\kappa \lambda}, \]

(39)

\[ \tilde{r}(\lambda) = \sum_{-\infty}^{\infty} r_{\kappa} e^{-i\kappa \lambda}. \]

(40)

Since (36) is a "discrete convolution" (cf. (17)), the Fourier transform of the infinite system (36) is the simple equation

\[ \tilde{a}(\lambda) \tilde{h}(\lambda) = \tilde{r}(\lambda), \]

(41)

corresponding to (32). Its solution is, of course,

\[ \tilde{h}(\lambda) = \frac{\tilde{r}(\lambda)}{\tilde{a}(\lambda)}, \]

so that the unknowns \( h_{\kappa} \) are simply given as the Fourier coefficients of \( \tilde{h}(\lambda) \)

\[ h_{\kappa} = \frac{1}{2\pi} \int_{0}^{2\pi} \tilde{h}(\lambda) e^{i\kappa \lambda} d\lambda = \frac{1}{2\pi} \int_{0}^{2\pi} \frac{\tilde{r}(\lambda)}{\tilde{a}(\lambda)} e^{i\kappa \lambda} d\lambda, \]

(42)

by (24); note that the roles of "time domain" and "frequency domain" are now interchanged.
The substitution of (40), written with \(j\) as summation index:
\[
\bar{r}(\lambda) = \sum_{-\infty}^{\infty} r_j e^{-ij\lambda}
\]
into (42) finally gives \(h_k\) as a linear combination of the observations \(r_j\) defined by (34):
\[
h_k = \sum_{j=-\infty}^{\infty} b_{k-j} r_j
\]
(43)
with
\[
b_k = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ik\lambda}}{a(\lambda)} d\lambda = \frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{\cos k\lambda}{1-\cos \lambda} d\lambda,
\]
(44)
by (38) and by the symmetry of \(\bar{a}(\lambda)\). Using a trick to eliminate a singularity, the integral can be evaluated to give
\[
b_k = -\frac{1}{2} |k|
\]
(45)
so that (43) becomes simply
\[
h_k = -\frac{1}{2} \sum_{j=-\infty}^{\infty} \sqrt{|k-j|} r_j
\]
(46)
Finally we note that the difference forming the left-hand side of (33), and the "second differences" appearing on the left-hand side of (34) are discrete analogues of first and second derivatives. Such analogies are useful if we make the formal transition from regular nets of high point density to "continuous nets", which can be treated by differential equations as pointed out by Krrup and others; (cf. Grafarend and Krumm, 1985)[11].

5. **CONCLUDING REMARKS**

Contemporary mathematical geodesy is characterized by an interplay between the finite and the infinite, between the discrete and the continuous, between time (or space) domain and frequency domain, and between deterministic and stochastic aspects. Continuous measurements are discretized for digital processing; on the other hand, discrete, regularly distributed quantities may be approximately considered continuous to get a deeper mathematical insight.

This interplay furnishes the basis of a fruitful cooperation between theoreticians and numerical analysts in geodesy and in other applied sciences as well.
REFERENCES


VERY LONG BASELINE INTERFEROMETRY

J. Campbell
Geodetic Institute University of Bonn, Fed. Rep. of Germany

ABSTRACT
This contribution reviews the technique of Very Long Baseline Interferometry as a tool for high precision measurements of relative point positions and spatial baseline orientation. The geodetic and geophysical applications of these measurements are discussed in relation to the objectives of global and regional programs of Earth dynamics research. Following a description of the method and the instrumentation, the systematic errors limiting the accuracy of VLBI baseline vector determinations are discussed and the different approaches of error elimination are indicated. Some of the most interesting results recently obtained by the different groups involved in geodetic VLBI are shown.

1. INTRODUCTION

Radio interferometry originated as an astronomical observation method with the aim of increasing as much as possible the low angular resolving power of single telescopes. In its simplest form a radio interferometer consists of two antennas separated by a given distance D, the baseline, and connected via the receivers and a phase stable electrical connection to a phase meter. The response of such a system to the observed radio source carries information on its angular size and structure, which are of a prime interest to astronomers. For the study of extremely compact sources, however, the resolution of the connected element radio interferometers (CERI) proved insufficient and some means to increase the baseline length, which is proportional to the resolving power, had to be found. A major breakthrough was made in 1967, when independent stations both in Canada and in the U.S.A. were able to produce interference fringes by using very precise oscillators and tape recorders, thus eliminating the need for a cable connection (BROTEK et al. 19671; BARE et al. 19672). First experiments that were explicitly aimed at achieving geodetic accuracy took place in 1969 on the 845-km baseline between the Haystack Observatory, Massachusetts and the 43-m antenna of the National Radio Astronomy Observatory in Greenbank, West Virginia (HINTEREGGER et al. 19723). In order to reach the high group delay resolution of ±1 ns attained in the above experiments, a bandwidth synthesis technique had been developed (ROGERS 19704). With this technique it became possible to
use tape recording equipment with a limited bandwidth and sample the broadband receiver window at a set of widely spaced frequencies. Both the introduction of independent high stability frequency standards and the bandwidth synthesis technique have contributed to transform radio interferometry into a geodetic measuring system of unprecedented accuracy, now commonly known as Very Long Baseline Interferometry (VLBI).

The following description of the VLBI technique, its potential and its achievements, is primarily oriented towards the geodetic applications with emphasis on the achievement of high group delay accuracy, while omitting the astronomical source imaging techniques.

2. FUNDAMENTALS OF THE VLBI-TECHNIQUE

The most striking difference between Very Long Baseline Interferometry and other space techniques is, that the interferometric observables are obtained a posteriori by the alignment in a processor of the two identical signal streams received at different times at the two telescopes. Thus the telescopes plus the processor could be seen to embody one instrument and the baseline separating the telescope sites may be (and often is) named an instrumental calibration constant.

Fig. 1
Functional diagram of a VLBI - system
The main elements of a VLBI-system are shown in Fig. 1: The radio signals coming from a given source are observed simultaneously at two or more stations at a preselected frequency in the GHz-region, converted to baseband (video frequency) and recorded (usually in a digitized form) or high data-rate tape. Before registration the signal streams are provided with precise time information derived from the local frequency standards. Later, when the tapes are brought together in the playback processor, this permits a coherent correlation of the approximately aligned signal streams for a certain interval. Because the actual time lag is still unknown, the correlation is done at a number of different delays separated by $\Delta t = 1/2B_s$ where $B_s$ is the synthesized bandwidth of the observed channels.

During correlation the data stream from one station is delayed quasi-continuously in such a way that the changing geometric delay $\tau_1$ is almost completely compensated for. This gives rise to a rather low residual fringe frequency, the phase of which slowly varies on the scale of a few turns per minute.

The output of the correlator is usually described by the complex cross-correlation function $R(\tau,t)$ which translates the response of an interferometer system to a radio source, see e.g. (MORAN 1976$^5$).

**RECEIVED SIGNALS**

Station 1: $x_1(t) = \int_{-\infty}^{\infty} X_1(\omega) e^{j(\omega - \omega_f)t} \, d\omega$

Station 2: $x_2(t-\tau) = \int_{-\infty}^{\infty} X_2(\omega) e^{j[(\omega - \omega_f)(t-\tau)]} \, d\omega$

**CROSSCORRELATION FUNCTION** $R(t, \tau_1) =$

a) computed: $\frac{1}{2\pi} \int_{-\tau}^{\tau} x_1(t) x_2(t-\tau_1) \, dt$

b) model: $\frac{\sin \pi B_s (\tau - \tau_1)}{\pi B_s (\tau - \tau_1)} e^{j\left[\omega_f t - \pi B_s (\tau - \tau_1)\right]}$

amplitude term phase term

$K = \text{constant proportional to SNR, } T = \text{integration time}$

$B_s = \text{maximum spanned bandwidth, } \omega_f = \text{fringe frequency,}$

$\tau = \text{true delay, } \tau_1 = \text{discrete correlation lags}$

The crosscorrelation function amplitude has its first zero at $\pi B_s (\tau - \tau_1) = 1$, hence $\Delta t = 1/B_s$, which is the half-width of the main peak.
In Fig. 2 the aspect of a typical fringe pattern as produced by a single 2 MHz-channel interferometer observing a point source is shown. Here the delay spacing is \(1/2 \cdot 2 \text{ MHz} = 0.25 \mu\text{s}\).

The analysis of the fringe pattern yields a wealth of astronomical and geodetic information. As many as four - albeit partially dependent - observables can be derived from the crosscorrelation function:

\[
R(\tau, t) \quad \text{General Purpose Computer Software Package} \quad \begin{cases} 
\phi(t) = \text{fringe phase} \\
A(t) = \text{fringe amplitude} \\
f(t) = \text{fringe rate} \\
\tau(t) = \text{time delay}
\end{cases}
\]

The actual travelling time \(\tau\) of a particular wavefront between the antennas at two sites can be expressed in two ways, i. e. as the delay of a wave group formed by the wide-band signals (group delay) and as the phase difference of a given monochromatic constituent of the signal stream (phase delay). The group delay is defined as the derivative of phase versus frequency in the band and can be estimated unambiguously by finding the maximum of the crosscorrelation function, which for an ideal square bandpass assumes the form of a \(\sin x/x\) function (see amplitude term above). The delay observable yields a full baseline solution and therefore plays the most important role in geodetic VLBI.
The fringe phase and the fringe rate (fringe frequency) are obtained from the sine and cosine parts of the crosscorrelation function. Due to the close relationship of $\phi$ and $f$, these observables are determined simultaneously, either from an ordinary sine wave adjustment or using the fourier transform into the frequency domain, where the maximum of $S(f)$ is estimated. These methods, which allow to establish the function $\phi(t)$ over a certain interval of time (usually the duration of an uninterrupted source scan) are often referred to as "phase tracking" methods (Thomas 1972). In local interferometry (CERI) this method is the common way of calibrating the short baselines and measuring source positions. Due to large phase fluctuations on longer baselines the phase observable can be used in VLBI only under special conditions.

The fringe rate is unambiguous, but it is insensitive to the $z$-component of the baseline. So, compared to the delay observable it plays a less important role in geodetic baseline determinations.

The fringe amplitude is of essential importance for source structure mapping.

Figure 3 shows a typical example of the output of the Mk III fringe analysis software package developed by A.R. Whitney (1976).

Fig. 3
Multichannel delay resolution function (Mk III fringe analysis)
The accuracy with which the time delay $\tau$ is estimated from the cross-correlation function depends chiefly on the halfwidth of the main peak, which is given by $\Delta \tau$ (see above).

There are different possible methods of delay estimation (WHITNEY et al. 1976$^7$), all of which yield a precision $\sigma_\tau$ of roughly 1% of $\Delta \tau$, depending of course on the SNR (signal to noise ratio) and the available integration time per observation.

The signal-to-noise ratio after averaging a total of $2B_sT$ correlated signal samples can be expressed by

$$\text{SNR} = \frac{S}{2k} \cdot \frac{\sqrt{A_1 \cdot A_2}}{T} \cdot \sqrt{2B_sT},$$

where $k$ is Boltzmann's constant.

This expression shows that, apart from the fundamental relationship $\Delta \tau = 1/B_s$, the delay accuracy is proportional to the flux density $S$ of a point source, the geometric mean of the antenna apertures $A_1$ and $A_2$, the square root of the recorded bandwidth $B_s$ and integration time $T$, and inversely proportional to the geometric mean of the system noise temperatures at both stations. If the ratio $\lambda/D$ becomes very small, i.e. at high observing frequencies and on long baselines, many of the compact sources are resolved. This results in a marked decrease of the correlated flux density with the effect that sources which show strong fringes on baselines of a few hundred km become very weak on intercontinental baselines (see e.g. KELLERMANN et al. 1971$^8$). Therefore in the latter case a high system sensitivity is of particular importance.

Efforts to improve the sensitivity of a Very-Long-Baseline interferometer have been concentrated on the sampling rate $2B_s$. With the newly developed Mk III system it is possible to record a data stream of 112 Megabit per second on 28 tracks of 2 MHz bandwidth each (CLARK 1979$^9$). This has produced an increase of the sensitivity by a factor of 5.3 over the commonly used Mk II system.

The available coherent integration time $T$ depends on the stability of the frequency standards and on the state of the atmosphere (troposphere and ionosphere). Due to the latter the ultimate stability that is achievable is limited to about $10^{-15}$ over time scales of $10^2 - 10^4$ seconds. Hydrogen maser frequency standards guarantee a stability of $10^{-14}$ over the same periods of time which is acceptable for geodetic applications. Using the above expression for the SNR the instrumental phase error

$$\sigma_\phi = (\text{SNR})^{-1}$$
and the group delay error

\[ \sigma_T = (2\pi \text{ SNR } B_s)^{-1} \]

can be computed. To illustrate these expressions, a typical example for a VLBI system consisting of two 20m antennas equipped with Mk III data acquisition terminals is given:

- **X-band** (8.4 GHz): 8 channels 2 MHz each; \( T = 300 \text{ sec} \)
- Correlated flux of radio sources: 1 Jansky
- \( \Theta_A = 0.057^\circ \) (two antennas of 20m diameter and 50 \% eff.)
- \( \Theta_R = 160^\circ \) K (uncooled paramps)

**SNR = 18.2** \( \sigma_T = \pm 3^\circ \)

With spanned bandwidth = 360 MHz:

**\( \sigma_T = 0.025 \text{ ns} \) (\( \approx 0.7 \text{ cm} \))**

This extremely high instrumental precision is, of course, curtailed to a certain degree by systematic instrumental and environmental error sources as will be discussed later in this section.

The fundamental observation equation relating the time delay with the baseline and source vectors may be written as

\[ \tau(t) = -\frac{1}{c} \vec{b} \cdot \vec{s}(t) \]

where

\[ \vec{b} \cdot \vec{s}(t) = b_x \cos \theta \cosh + b_y \cos \delta \sinh + b_z \sin \delta \]

with the baseline components \( b_x, b_y, b_z \)

the radio source position \( \alpha, \delta \)

and the Greenwich hour angle of the source

\[ h = \text{GST} - \alpha \]

The baseline vector components are referred to the instantaneous Earth rotation axis. The unit vector of the source points to the apparent position at the time of observation (see Fig. 4).

At this stage there are \( 3 + 2 \cdot n \) fundamental parameters to be determined in a least-squares fit: the three baseline components \( b_x, b_y, b_z \) and the coordinates \( \alpha, \delta \) of \( n \) observed sources.

In order to make theory consistent with observation, the above model has to be supplemented with terms allowing for a number of physical and instrumental effects listed below. Some of these effects can be predicted to a high degree of accuracy while others have to be parameterized or supple-
mented with additional measurements. In Fig. 5 an example of a geodetic VLBI software system is shown.

Fig. 4
Geometric VLBI Model

Model refinements:

a) Effects of precession and nutation

Precession and nutation are motions of the Earth's axis with respect to the celestial system represented by the positions of the observed compact radio sources. These motions are caused by external forces, the gravitational attraction of the members of the solar system, acting upon the non-spherical, inhomogeneous and visco-elastic body of the Earth. The main coefficients of the standard model have been determined empirically by long astronomical observing series and therefore are not error-free. In turn, any misalignment of the Earth's axis due to these errors will be detected by the highly sensitive VLBI observations. Significant corrections to the precession constant and to some of the nutation coefficients have been derived from the analysis of only a few years of VLBI data (HERRING et al. 198510).
b) Relativistic effects (space-time geometry)

- special relativity. Due to the relative motion of the VLBI antennas during observation the finite speed of light causes significant additional delays which can be precisely modelled (diurnal aberration etc.).

- general relativity. The effect of gravity on the propagation of electromagnetic waves is considerable (near the sun the delay may be many orders of magnitude larger than the measurement error). VLBI observations have been used to verify Einstein's theory to an accuracy of about 0.1%.
c) Instrumental effects

- local oscillator instabilities. Usually a clock model is introduced that accounts for the unknown epoch difference and a differential drift rate (two parameters). Unmodelled clock offsets and rates may lead to large systematic errors especially in the z-component of the baseline.
- change in the electronic circuitry and cable delays. While a constant delay is absorbed by the clock offset parameter, any delay changes have to be monitored by a phase and delay calibration system.
- deformations of the telescope structure, displacements of the reference points. These effects can be checked by models supplied with local geodetic measurements.

d) Environmental effects

The effect of the atmosphere on VLBI-observations is considered the most serious problem, because at widely separated stations the look-angles of the telescopes aimed at the same source differ greatly as well as the meteorological conditions themselves. The ionosphere, which is a highly dispersive medium for radiation in the radio frequency band, can be dealt with by using two different observing frequencies. Moreover, its influence diminishes considerably with increasing frequency. At 8 GHz the extra-zenital path due to the ionosphere reduces to about 2 dm. The remaining effect, however, is very unstable; therefore the dual frequency method using a second frequency at 2.3 GHz is applied. The neutral atmosphere, essentially the troposphere, presents the same problems in VLBI as in Doppler and radar satellite observations. Its influence on radio signals adds up to an extra-zenital path of 2.0 - 2.5 m. The contribution of the dry air is rather stable and can be described by suitable models. The wet component, although much smaller, changes rapidly and has to be monitored during the observations. The most promising method appears to be the radiometer technique, which consists of measuring the micro-wave thermal emission from water vapor near 22 GHz in the line-of-sight.

e) Geophysical effects

- polar motion and UT1-variations. These are effects that change the components of the baseline but not its length. To a certain degree of accuracy (~1 m) these variations can be obtained from the regularly published bulletins of the BIH to apply "corrections" to the observables (or to the model). In recent years the observations have proven to be more accurate than the corrections. Therefore the problem has been inverted, the changes showing in
the baseline components being used as precise polar motion and Earth rotation data.

- Earth tides. The tides of the solid Earth are mainly semidiurnal and diurnal effects of a few decimeters in amplitude that change the baseline components as well as the length. The solid Earth tides are predictable to a few cm, but it is also possible to derive significant estimates of Love's numbers from VLBI observations. For stations close to the shores an additional correction due to the ocean tides has recently been included in the model.

- crustal motions. The theory of plate tectonics which stipulates that the Earth's crust is formed by a mosaic of separate plates that are in motion relative to each other has now been universally accepted. Predictions derived from geophysical evidence yield motions of a few centimeters per year. These small displacements constitute one of the main interests of a continued series of VLBI experiments aimed at the detection of the present rates of motion.

f) Radio-astrometric effects

- Precise source positions are usually determined simultaneously with the baseline components. Even so, special astrometric campaigns are carried out in order to establish consistent radio-source catalogues for fundamental astronomy, optical counterpart identification and for the support of geodetic programs. The accuracy of some thirty compact radio sources, most of which are located in the northern hemisphere, is now at a level of 2 - 3 milliarcseconds as determined from VLBI-experiments. Work is in progress to improve the source coverage over the sky and to provide a firm connection to the existing optical reference system.

- Most of the observed compact sources tend to show structure at the level of a few milliarcseconds. This effect, in particular any changes in the structure, poses a limit on the accuracy of the radio-reference system. Permanent monitoring of the structure, which can be done by using the geodetic VLBI data, will help to reduce the contribution of these errors in the future.

The total geodetic error budget on baselines of about 1000 km sums up to about 1 cm in length and 3.5 cm in the individual components (see Fig. 6). Intercontinental baselines of about 6000 km show errors of 2 - 3 cm in length and 10 cm in the components. These extremely high accuracies (∼0.01 ppm) form the basis of the increasing efforts invested in the use of the VLBI-technique for the determination of such subtle effects as variations in the Earth's rotation and crustal motion.
2. SCIENTIFIC INTEREST; PROGRAMS AND RESULTS

A detailed description of the geophysical applications of VLBI has been presented as early as 1969 at a conference held in London, Canada on "Earthquake displacement fields and the rotation of the Earth" (SHAPIRO, KNIGHT 1970). In subsequent years virtually all of the goals could be realized or at least reach the stage of initial successful verification. In the following paragraph the scientific goals and projects are summarised and some of the recent results of the ongoing campaigns are shown.

Geodetic and geophysical interest in VLBI is based fundamentally on the use of an inertial reference frame of highly compact extragalactic radio sources. With the VLBI-technique it is possible to measure very accurately the baseline vectors and their changes in time between distant points on the Earth's crust. Therefore the primary objectives to use VLBI observations are:

1. the realization of a unified global reference system in order to satisfy the needs of global geodetic and navigational systems;
2. the monitoring of polar motion and changes in the Earth's rotation rate to enable a better understanding of the kinematics and dynamics of the Earth-Moon system and the structure of the Earth's interior;
3. the estimation of the Earth's elasticity parameters from directly measured tidal deformations;
4. the determination of plate motion and plate stability to improve the understanding of global plate tectonics;
5. the investigation of regional crustal movements in order to deduce the building up of strain and provide input to earthquake prediction programs.

Other important activities are:

6. the determination of improved values for the precession and nutation constants, and

7. the precise verification of the effects of general relativity.

The accuracy requirements necessary to attain these goals are ± 5 to ± 10 cm in each baseline component for items 1 and 2 (± 0.1 ms for UT1) and ± 1 to ± 3 cm in baseline length repeatability for items 3 through 5. By now these accuracies have been demonstrated in hundreds of VLBI campaigns on baselines connecting nearly all major continents of the globe.

Corresponding to the international nature of the anticipated goals several programs of multilateral cooperation have been initiated, among which the following most important projects should be mentioned:

1. NASA Crustal Dynamics Project (CDP).

This project is part of a US Federal program involving several government agencies for the application of space technology to crustal dynamics and earthquake research. Major cooperative arrangements have been made with European and other countries extending the project to a global research program (NASA 198312).

The VLBI part of the CDP comprises regular experiments (10 - 20 a year) of one to three days duration between the major geodetic VLBI facilities in the US, Europe and in and around the northern Pacific. In addition, so-called bursts of observations using the mobile VLBI units are carried out each year in the tectonically active zones of California and Alaska. These campaigns are aimed at the creation of a detailed picture of the local crustal motion pattern to assist the investigation of earthquake mechanisms.

The global station distribution is shown in Fig. 7. Some of the stations are still in the process of being fully equipped for Mk III VLBI.

The CDP in its present form is planned to extend through 1988. Follow-up programs will be set up in order to insure the repetition of the measurements at regular intervals of one or two years.

2. Project IRIS (International Radio Interferometric Surveying).

The goal of project IRIS is to conduct joint activities between the NGS-(US National Geodetic Survey) POLARIS-Network and other international stations that are dedicated to full-time geodetic work, such as the VLBI-station of Wettzell, for the regular monitoring of Polar Motion and UT1.

Fig. 7 Global VLBI network
The IRIS activities consist of a series of VLBI observing sessions of 24 hours duration at five-day intervals. The IRIS observations normally involve three stations in the United States (the Westford telescope in Massachusetts, the George R. Agassiz Station (Pt. Davis) in Texas and the Richmond Observatory in Florida) and the 20 m radiotelescope of the Wettzell Observatory in the Federal Republic of Germany. One session per month also includes the Onsala Space Observatory in Sweden (see Fig. 8).

Fig. 8
IRIS earth rotation network

The National Geodetic Survey (NGS)/Rockville, Md. regularly analyses all of the IRIS data to obtain polar motion and UT1 time series, which are published monthly in the IRIS Bulletins A. Thus, since Jan. 5, 1984, the date that the Wettzell Observatory began regular operations, the IRIS system has been routinely providing the x and y components of polar motion with an accuracy of 1 to 2 marcsec, and UT1 with an accuracy of 0.04 to 0.1 msecs (Carter, Robertson, Mackay, 198513)).

Figures 9 and 10 show the remarkably smooth trace of the instantaneous pole of rotation from Sept. 26, 1980 (the time of the MERIT preliminary campaign) until March 15, 1986. In particular, a significant improvement of accuracy can be seen from the moment that the full IRIS network with the station of Wettzell has become operational in January 1984 (Fig. 10).

In addition to the regular 5-day observations a special campaign of daily observations on the baseline Westford - Wettzell is being performed
Figs. 9 and 10
Merit and POLARIS/IRIS pole position
Sept. 1980 - April 1984
with the aim of looking at the short period UT1-vari-ations. A frequency ana-
lysis on 90 days of daily VLBI observations already permitted the detection
of the predicted 13.6 day and 9.1 day periods caused by tidal deformation of
the polar moment of inertia of the Earth (Fig. 11). Both the amplitudes (0.9
and 0.2 msec) and phases are in good agreement with the theoretical values.
These results represent a first purely empirical confirmation of theoretical-
ly derived short period tidal oscillations in the Earth's rotational speed,
which until now had been hidden in the noise and the poor temporal resolu-
tion of the BIH data (ROBERTSON et al. 1985\textsuperscript{14}), CAMPBELL, SCHUH 1986\textsuperscript{15}).

The NASA-CDP and the IRIS VLBI campaigns have been providing baseline
length results since the late seventies when the Mk III VLBI system became
fully operational. As the length of a baseline vector is independent from
changes in its orientation, the observed baseline length series are free from
errors in the Earth rotation parameters and can be treated as an independent
result of the campaign. However, as has been shown in the error budget (Fig.
6), other factors such as source positions (including the associated pre-
cession and nutation models) and atmospheric effects may still cause problems
when long series are to be interpreted. At present only two baselines across
the Atlantic (Onsala - Haystack and Onsala - Westford) cover a long enough
period (4 years) to be able to make first estimations of a meaningful rela-
tive motion. Both indepently observed series seem to indicate the same annual
drift rate of 1 - 2 cm per year between the North-American and the Eurasian
plate (Fig. 12). Before any firm conclusions can be drawn the further evolu-
tion of the much more frequently observed baseline Westford - Wettzell should
be awaited. The continental baseline Westford - GRAS (Ft. Davis, Texas) also
shows an appreciable change, which could confirm theories of an elastic
behaviour of the plates.

While the North-Atlantic section of the CDP and IRIS networks has been
observed on a routine basis, great efforts are made to include new stations,
in particular Shanghai, China, Hartebeeshoek, South Africa and Atibaia,
Brazil (Fig. 7). The stations of Kashima, Japan and Kauai, Hawaii have been
active since the beginning of 1984. On this baseline, where large motions
are expected, a contraction of 5 cm per year has been derived from the first
two epochs. The length variation on another Pacific baseline, Kashima -
Kwajalein, of -8 cm in one year is the largest so far observed. The agree-
ment of these observations with the predicted rates derived from geophysical
evidence is very good (KONDO, HEKI, TAKASHI, 1986\textsuperscript{16}).

Thus in recent years the VLBI technique has lived up to its expecta-
tions. By now large amounts of data have proven conclusively that the anti-
cipated accuracy levels can be effectively realized in routine observing
campaigns.
Fig. 11
Short period UT1 variations
(obtained from 90 days of daily VLBI observations)
Fig. 12
Transcontinental and intercontinental baseline evolution 1981 - 1985

SLOPE = 1.36 ± 0.51 cm/yr

Westford - Onsala +2.0 ± 0.3 cm/yr
Westford - Wettzell +1.2 ± 0.4 cm/yr
Westford - GRAS -1.3 ± 0.1 cm/yr
REFERENCES


FUTURE ACCELERATORS

K. Johnsen
CERN, Geneva, Switzerland.

ABSTRACT
Against the background of the well-known Livingston diagram, the current status and complementarity of the hadron collider projects SSC, LHC and Eloisatron, the lepton collider projects LEP and SLC, and the hybrid collider HERA are reviewed. Whereas future hadron colliders have no clear technological limit, lepton colliders are at an important turning point and the search for new techniques in the direction of linear colliders is in full flood. A natural approach is to extrapolate parameters for normal conducting structures to fit the special requirements of linear colliders. Superconducting accelerating structures, the use of pulsed rather than harmonic RF power with optoelectric switches, wake fields and the plasma beat-wave scheme are some of the new ideas which are reviewed in this paper. Much of what is included comes from the findings of the Long Range Planning Committee established by the CERN Council under the chairmanship of C. Rubbia.

1. INTRODUCTION

Particle accelerators became an important tool for experimental nuclear physics from around 1930. Since then there has been tremendous progress in the construction of such accelerators with an increase of about one and a half orders of magnitude in beam energy per decade, as illustrated by the Livingston plot shown in Fig. 1. In spite of this energy increase of eight orders of magnitude the cost of a typical accelerator installation has only gone up by one order of magnitude if inflation is corrected for.

It is interesting to look a little more closely at how this has happened. The progress of each type of accelerator has saturated fairly quickly, whereas new ideas have been proposed regularly and have been the main contributors to the rapid advance. Two startling examples can be pointed out: the invention of the alternating gradient focusing in the early fifties and the application of colliding beams in the sixties and onwards. For the future the technological development within such fields as superconductivity will be of great significance. And to satisfy the far-reaching visions of high-energy physics in the more distant future, we may have to rely upon new inventions and developments. This talk will be speculative on how this development will continue in the future in order to satisfy the desires of the high-energy physics community for ever higher energy and at the same time, increased luminosity. Another interesting observation can be made with reference to Fig. 1. The highest energies attainable at present are already far beyond what can ever be reached by fixed-target machines. In other words, when we talk about future facilities we are only talking about accelerators that can function as colliding beam machines.
Fig. 1 Livingston plot showing the maximum energy as a function of year for various accelerator techniques.

For hadron colliders there is no clear technological limit. For this reason future plans for such colliders are based on known principles, but extended in energy and luminosity as far as one can possibly hope to reach within assumed financial limits.

For electron colliders the situation is very different. The difficulties created by the radiation from electrons in circular machines severely limits the further development of circular electron colliders to such an extent that it is considered unlikely that such colliders will be constructed beyond the LEP energies. New ways must therefore be sought.
2. STUDIES OF FUTURE HADRON COLLIDERS

The highest energy hadron colliders are the proton-antiproton collider (SpS) at CERN of about $2 \times 300$ GeV energy and the proton-antiproton collider (Tevatron I) at Fermilab of about $2 \times 800$ GeV. The former has been operational since 1982, the latter had a trial run in October, 1985, and will operate for physics experimentation from late 1986. Both will always have inherently modest luminosities, although these are incredibly high when we take into account how they are achieved.

The next natural step is to put proton rings with the highest possible magnetic field into the LEP tunnel (see Fig. 2). Indeed, such considerations formed part of the arguments that determined the radius and size of the LEP tunnel. A range of possibilities exists for such a collider. The conceptually simplest option would be a pp ring with a single beam channel, preferably with the highest possible guide field. The luminosity would, however, be relatively low because antiproton sources are not very intense. Furthermore, in order to make provision for bunch separation at unwanted crossings, the aperture must be somewhat enlarged compared with a single-beam machine.

![Fig. 2 Large hadron collider in the LEP tunnel](image)

Another, and probably more interesting, option would be two beam-channels side by side. For space reasons, it is, in that case, an advantage to use the so-called "2 in 1" solution to the magnet design, i.e. the two channels within the same magnet yoke and the same cryostat. The two-channel approach will allow high-luminosity pp collisions with many bunches. For this reason this approach has formed the basis of some studies within ECFA and CERN of such a project. The main parameters that have come out of these studies are listed in Tables 1 and 2.
Table 1
General parameters

<table>
<thead>
<tr>
<th>Collider in LEP tunnel</th>
<th>Proton-proton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dipole magnetic field (T)</td>
<td>10</td>
</tr>
<tr>
<td>Operating beam energy (TeV)</td>
<td>8-9</td>
</tr>
<tr>
<td>Separation between orbits (mm)</td>
<td>165-180</td>
</tr>
<tr>
<td>Number of bunches</td>
<td>3564</td>
</tr>
<tr>
<td>Bunch spacing (ns)</td>
<td>25</td>
</tr>
<tr>
<td>Number of crossing points</td>
<td>8</td>
</tr>
<tr>
<td>Beta value at crossing point (m)</td>
<td>1</td>
</tr>
<tr>
<td>Full crossing angle (µrad)</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 2
Performance

<table>
<thead>
<tr>
<th>&lt;n&gt; at ∑ = 100 (mb)</th>
<th>1</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity (cm⁻² s⁻¹)</td>
<td>4 x 10^{32}</td>
<td>1.5 x 10^{32}</td>
</tr>
<tr>
<td>Beam-beam tune shift</td>
<td>0.0013</td>
<td>0.0025</td>
</tr>
<tr>
<td>Beam stored energy (MJ)</td>
<td>63</td>
<td>121</td>
</tr>
<tr>
<td>RMS beam radius (µm*)</td>
<td>12</td>
<td>* at interaction point for p*= 1 m</td>
</tr>
<tr>
<td>Beam life-time (h)**</td>
<td>42</td>
<td>21</td>
</tr>
</tbody>
</table>

** particle loss due to beam-beam collisions

A few clarifying comments to these tables may be useful.

Since the tunnel is given it is important to get as high a magnetic field as possible. It is hoped that an R & D programme on Nb₃Sn magnets will be successful, which should result in about 10 T bending field. This should then give 8-9 TeV beam energy, depending on the final lattice design, where a few options are still open. Another possibility is to follow the more conventional NbTi approach, but with 1.8 K operating temperature. This would result in about 8 T field.

The shortest acceptable bunch spacing (and therefore the maximum number of bunches) is determined by the response times of detector elements, which can hardly be assumed shorter than 25 ns.

Another uncertain element is the number of events per bunch crossing that a detector system can accept. It is believed to be of the order of one. In Table 2 one event per bunch crossing has been assumed for the first column. A second column where four events per bunch crossing have been assumed is also presented. It seems clear that for a machine to operate at 10^{33} cm⁻² s⁻¹ some detector development is needed, whereas the corresponding machine parameters seem alright. If higher luminosities are needed both machine performance and detector performance will have to be stretched.
This study has demonstrated that a hadron collider in the LEP tunnel would technically and scientifically be a very sound proposition. Centre-of-mass energies of 10-12 TeV would be attainable with present-day superconducting technology, with the prospect of 15-20 TeV with successful Nb₃Sn development. Luminosities of 10^{32}-10^{33} cm⁻² s⁻¹ seem entirely feasible in each of the eight interaction regions.

Whether or not such a machine will be built depends, however, on the development over the next few years of high-energy physics in the world in general and in Europe more specifically. It also depends on how new ideas develop during the rest of this century.

In the USA a vigorous effort goes into the study of a 2 × 20 TeV facility called the SSC (for Super Superconducting Collider). For this project there is no existing tunnel to put constraints on the magnitude of the guide field. They have therefore considered many field options between 3 T and 6.5 T, with little emphasis on the 10 T option since the time element is so important in their case. Otherwise the considerations are very similar to the ones described in the previous paragraph. A summary of the present results for the preliminary study is given in Table 3 and Fig. 3.

Table 3

| Primary SSC design objectives |  
|-----------------------------|---|
| Maximum beam energy (TeV)   | 20 |
| Injection energy (TeV)      | 1  |
| Maximum luminosity (cm⁻² s⁻¹)| 10^{33} |
| Maximum number of interactions per bunch crossing (at max. luminosity) | 10 |
| Number of interaction regions | 6 (4 initially developed) |
| Field (T)                   | 6.5 | 5 | 3 |
| Circumference (km)          | 90 | 113 | 164 |

The studies of this project are far enough advanced that technically the project is ripe for a decision. The fate of the project is therefore now largely in the hands of the politicians. With the options chosen by the American physicists the construction of this project, or an approximately equivalent one, has become very urgent as the HEP community will be very short of experimental facilities in some years time.

Another European hadron collider study should be mentioned: the Eloisatron project in Italy. Its aim is an energy of 2 × 50 TeV with 10 T magnets on a circumference of about 150 km. The Italian involvement in HERA, by the supply of 50% of the bending magnets for the superconducting proton ring of this facility, is considered as a kind of preparatory work for the Eloisatron project.
Before we turn to $e^+e^-$ colliders, a few words should be said about the possibility of electron-proton colliders, which falls somewhat in between the pure hadron colliders and the electron colliders. The first and only such facility is under construction at DESY, with the name of HERA, already mentioned above. The main parameters for HERA are listed in Table 4.

This project will be operational in 1989, and will then constitute an experimental facility very complementary to other facilities in the world.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Main parameters for HERA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-ring</td>
</tr>
<tr>
<td>Energy (GeV)</td>
<td>820</td>
</tr>
<tr>
<td>Circumference (km)</td>
<td>6.3</td>
</tr>
<tr>
<td>Dipole strength (T)</td>
<td>4.5</td>
</tr>
<tr>
<td>Luminosity (cm$^{-2}$ s$^{-1}$)</td>
<td>$3 \times 10^{31}$</td>
</tr>
<tr>
<td>Number of interaction points</td>
<td>4</td>
</tr>
</tbody>
</table>
Assuming this becomes as interesting for physics experimentation as hoped for, one would want to think of even higher energy e-p collisions in the future. Again LHC lends itself naturally to this possibility, as it is sitting in the LEP tunnel, and relatively minor modifications are needed to collide the electrons with the protons, giving an order of magnitude higher collision energy than HERA. To do similar things with the SSC would require the construction of a special electron ring.

3. PROSPECTS FOR ELECTRON-POSITRON COLLIDERS

Electron-positron colliders have a few clear advantages over hadron colliders. The main ones are: the e⁺e⁻ collisions are much cleaner than hadron collisions, and a much higher luminosity can be exploited, if achievable. In addition, an order of magnitude less energy is required since the colliding particles are themselves "constituents". Energy-wise an e⁺e⁻ collider with centre-of-mass energy of 2 TeV is roughly equivalent to a hadron collider with 20 TeV centre-of-mass energy.

However, the disadvantage is that the radiation makes it so much harder to make very high energy e⁺e⁻ colliders. It is therefore believed that LEP presently under construction at CERN, with a circumference of 27 km, will be the largest circular e⁺e⁻ collider ever to be built.

A layout of LEP is shown in Fig. 4. As presently being built, the beam energy of LEP will be 50 GeV. However, it is foreseen to extend the project to higher energies, as illustrated in Table 5. The luminosity is expected to be a little above $10^{31}$ cm$^{-2}$ s$^{-1}$ for Phase 1, increasing with increasing energy to perhaps $10^{32}$ cm$^{-2}$ s$^{-1}$ for the 'ultimate' phase. The project is under construction and is planned to have beams in the spring of 1989.

![Fig. 4 The LEP machine](image)

<table>
<thead>
<tr>
<th>LEP stages</th>
<th>Beam energy (GeV)</th>
<th>Power Dissipation (MW)</th>
<th>Installed RF power (MW)</th>
<th>Expt. areas at points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>50</td>
<td>75</td>
<td>16 (copper cav.)</td>
<td>2,4,6,8</td>
</tr>
<tr>
<td>Planning permission</td>
<td>100 a)</td>
<td>150</td>
<td>48 (supercond.cav.)</td>
<td>1,2,4,5,6,8</td>
</tr>
<tr>
<td>'Ultimate'</td>
<td>125 a)</td>
<td>(a)</td>
<td>96 (supercond.cav.)</td>
<td>all 8</td>
</tr>
</tbody>
</table>
The radiation problem in a circular $e^+e^-$ collider is fundamental and can only be counteracted or solved by increasing the circumference of the machine, hence the very large circumference of LEP. The ultimate limit of this approach is to go to infinity with the bending radius, in other words to consider colliding beams from linear accelerators. The first speculations on this kind of approach, with a host of new problems, started about 10 years ago and has recently gained considerable momentum. Because of the importance attached to this possibility, it has been made the subject of one of the Advisory Panels organised by C. Rubbia for the Long Range Planning Committee (LRPC) set up by the CERN Council.

The first attempt at colliding beams from a linear accelerator will be done with a facility now under construction at SLAC, the so-called SLAC Linear Collider (SLC), with a beam energy of 50 GeV and luminosity of $6 \times 10^{30}$ cm$^{-2}$ s$^{-1}$. The layout and main parameters are shown in Fig. 5. Contrary to circular colliders the beams pass through each other only once, which means that beam parameters must be optimised such as to make maximum use of this one crossing. One consequence of this is illustrated by the very small beam sizes that have to be achieved in the collision region with correspondingly difficult requirements on the quality of the system for the final focus, the surveying and the automatic beam-position controls, etc. The project will provide crucial information in such areas when it starts its operation in late 1986 or early 1987.

![SLAC Linear Collider (SLC)](image)

*Fig. 5 SLAC Linear Collider (SLC)*
At CERN we are considering the possibilities of a linear collider (CLIC) with energies in the TeV range, and luminosities up to $10^{34}$ cm$^{-2}$ s$^{-1}$, i.e. more than an order of magnitude higher energy than SLC, and about three orders of magnitude higher luminosity. This requires more than a simple extrapolation of present-day techniques.

Below we list the most important approaches that have been proposed for acceleration in linear colliders, with some comments on their main characteristics, and the need for further detailed studies before reliable comparisons can be made and firm conclusions drawn.

3.1 Normal-conducting accelerating structures

The classical travelling wave linac appears to be capable of accelerating gradients well above 100 MV/m, especially at high microwave frequencies. A high frequency is also desirable in order to reduce stored energy. Longitudinal and transverse wake fields as well as construction problems may impose a limit on the choice of frequency and this remains to be studied in detail.

The Q factor is too small to permit conservation of stored energy over a realistic repetition period. In spite of this, RF to beam efficiencies above 5% can be obtained. The price for this is a very short power pulse and an appreciable but probably correctible energy spread of the beam, owing to the need for extracting about 10% of stored electromagnetic energy from the accelerating structure at each beam pulse. The RF to beam efficiency could be raised above 10% if the electromagnetic energy remaining at the output of the travelling wave structure could be recovered. Two two-stage schemes resulting from the CERN study seem to permit that. Multibunch operation with bunch-to-bunch compensation of energy deviation may yield 30% efficiency provided higher-order wakefield problems are solved and a suitable final focus scheme is found.

The worst remaining problem seems to be the economic and efficient generation of peak RF power. Microwave power sources are being developed which may permit feeding individual linac sections with tens of megawatts peak power at wavelengths down to the order of 1 cm, but an economic problem remains. Alternatively a two-beam scheme employing microwave free electron lasers, induction linacs and a low energy drive beam has been proposed and is being developed (Sessler, 1982).

Two proposals of two-stage schemes, both employing a superconducting RF drive linac and a drive beam have emanated from our group (Schnell, 1986; U. Amaldi and C. Pellegrini, 1986).

A few tentative parameters of a 1 TeV main linac are given in Table 6. Energy recovery at the output of the travelling wave structure, would raise the luminosity (for given input power) by a factor of two. Compensated multibunch operation may raise the luminosity to $0.6 \times 10^{34}$ at only 25% extension of length and average power.
Table 6

Tentative parameters of a 1 TeV main linac

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>1 TeV</td>
</tr>
<tr>
<td>Frequency</td>
<td>29 GHz</td>
</tr>
<tr>
<td>Accelerating gradient</td>
<td>80 MV/m</td>
</tr>
<tr>
<td>Active length</td>
<td>12.5 km</td>
</tr>
<tr>
<td>Peak power per metre of section length</td>
<td>96 MV/m</td>
</tr>
<tr>
<td>Total average RF power</td>
<td>80 MW</td>
</tr>
<tr>
<td>Beam power</td>
<td>5 MW</td>
</tr>
<tr>
<td>Repetition frequency</td>
<td>5.8 kHz</td>
</tr>
<tr>
<td>Bunch population</td>
<td>$5.4 \times 10^9$</td>
</tr>
<tr>
<td>Beam radius at collision</td>
<td>77 nm</td>
</tr>
<tr>
<td>Luminosity (single bunch without energy recovery)</td>
<td>$10^{33}$ cm$^{-2}$ s$^{-1}$</td>
</tr>
</tbody>
</table>

3.2 Superconducting accelerating structures

Linear colliders based on superconducting cavities (SC) were proposed and studied at CERN more than ten years ago (U. Amaldi, 1976). In recent years studies of thermal breakdowns and electron loading in SC cavities have led to the construction of 350 MHz cavities having a gradient of 7 MV/m and a Q-value of $3 \times 10^9$. For small 3 GHz cavities, gradients of 18 MV/m and $Q = 10^{10}$ have been obtained. A few years of technological developments on the preparation of clean and defect-free surfaces may allow us to reach of 25 MV/m and $Q = 5 \times 10^{10}$. The development of type II SC, like Nb$_3$Sn and NbN, possibly sputtered on a copper substrate, offers the promise of reaching even higher gradients and quality factors. For linear colliders, economic fabrication and treatment will also become of paramount importance. Examples have been studied by U. Amaldi, H. Lengeler and A. Piel (1986) assuming the following parameters: $\nu = 1$ GHz, $G = 25$ MeV/m, $Q = 5 \times 10^{10}$, temperature = 1.8 K.

Table 7 lists the main parameters of two reference designs of a high luminosity ($10^{34}$ cm$^{-2}$ s$^{-1}$) (1+1) TeV collider for which some cost optimisations have been performed. The first design uses the principle of energy recovery to save on the total power and has parameters which are extrapolations by a factor of 10 only with respect to the SLC design. U. Amaldi (1986) and others have demonstrated that the damping ring system is a very important and expensive part of the complex. The second reference design has no energy recovery and requires an emittance 10 times smaller, at the limit of what can be achieved today with the best storage rings for synchrotron radiation.

The cost optimisation procedure, which takes into account klystron replacement and electricity consumption over 10 years, leads to the conclusion that the accelerator has to be run with a macroscopic duty factor $C = 0.15$, and that both designs need about 350 MW for running.
### Table 7

The two reference designs of a superconducting (1+1) TeV collider with $L = 10^{39} \text{ cm}^{-2} \text{s}^{-1}$ ($D = 2$)

<table>
<thead>
<tr>
<th>Accelerator</th>
<th>Energy recovery</th>
<th>No recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam power $P$(MW)</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>Recovery efficiency $\eta$</td>
<td>0.90</td>
<td>0.0</td>
</tr>
<tr>
<td>Invariant emittance $\varepsilon$ (m)</td>
<td>$2.5 \times 10^{-6}$</td>
<td>$3.1 \times 10^{-7}$</td>
</tr>
<tr>
<td>Bunch length $\sigma_z$ (mm)</td>
<td>3.6</td>
<td>0.36</td>
</tr>
<tr>
<td>$\beta$-value at crossing (mm)</td>
<td>10.7</td>
<td>1.07</td>
</tr>
<tr>
<td>Bunch radius $\sigma_r = \sigma_y$ ($\mu$m)</td>
<td>0.12</td>
<td>0.013</td>
</tr>
<tr>
<td>Particles/bunch $N$</td>
<td>$5.4 \times 10^9$</td>
<td>$6.5 \times 10^8$</td>
</tr>
<tr>
<td>Macroscopic duty factor $C$</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Bunch peak frequency $f$ (kHz)</td>
<td>670</td>
<td>820</td>
</tr>
<tr>
<td>Train peak frequency</td>
<td>67</td>
<td>82</td>
</tr>
<tr>
<td>Gradient (MeV/m)</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Q-value</td>
<td>$5 \times 10^{10}$</td>
<td>$5 \times 10^{10}$</td>
</tr>
</tbody>
</table>

### 3.3 Structures excited by optoelectric switches

#### 3.3.1 Conventional structures driven by optoelectric RF generators

The first such device to be suggested was the laser-klystron, where a train of laser pulses at 3 or 6 GHz striking a photocathode switches on bunches of electrons, which are accelerated by a static field and passed through a resonant cavity to extract the energy. It is hoped that the efficiency can be high and the cost lower than for conventional klystrons. Another approach uses the wire photocathodes discussed in the next section with laser pulse trains at 30 GHz, the "microlasertron" (Palmer, 1986).

#### 3.3.2 Radial pulse line structure driven by optoelectric switched power

This proposal (Willis, 1904) uses very short pulses instead of RF power, increasing the breakdown limit on the accelerating gradient. It may prove that the cost of switches may be relatively low and their efficiency high, despite the fact that the peak power is quite high in these non-resonant structures.

Tests are underway on two types of photodiodes, vacuum and semiconductor. For the former, measurements are being made on wire photocathodes, charges close to the field emission limit by nanosecond electrical pulses and then illuminated by a 20 ps pulse of 0.35 nm light. It is known that in this "field assisted photoemission" regime, the electron yield from rugged photocathodes is considerably increased. The efficiency of these photodiodes mounted in a radial line have been studied by the numerical simulation program "MASK". Using a sequence of four suitably timed laser pulses to discharge the wire, an overall efficiency of 15% has been predicted.

Studies have been started on the volume photoconductivity in GaAs (Cr) induced by 1.06 µm laser pulses. Results indicate that 1-5 mJ of light can switch one GW of electrical power.
The performance of the radial lines, consisting of copper discs spaced by a small gap and driven at the outer radius, has been studied in scaled up models, with measurements being made on the effect of misalignments.

A set of tentative parameters for a typical example with a beam energy of 1 TeV and a luminosity of $10^{33}$ cm$^{-2}$ s$^{-1}$ are given in Table 8.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer radius of radial line</td>
<td>120 mm</td>
</tr>
<tr>
<td>Inner radius (beam hole)</td>
<td>1 mm</td>
</tr>
<tr>
<td>Gap between disks</td>
<td>2 mm</td>
</tr>
<tr>
<td>Photodiode capacitance</td>
<td>16 pF</td>
</tr>
<tr>
<td>Pulse length</td>
<td>25 ps</td>
</tr>
<tr>
<td>Charging voltage</td>
<td>80 kV</td>
</tr>
<tr>
<td>Stored energy</td>
<td>48 MJ</td>
</tr>
<tr>
<td>Switched power</td>
<td>2 GW</td>
</tr>
<tr>
<td>Optical energy</td>
<td>3 mJ</td>
</tr>
<tr>
<td>Gradient on beam, in gap</td>
<td>0.6 GV/m</td>
</tr>
<tr>
<td>Average gradient, for 2 mm disc</td>
<td>0.3 GV/m</td>
</tr>
<tr>
<td>Stored energy, electrical</td>
<td>24 KJ/km</td>
</tr>
<tr>
<td>Beam energy, for 5% loading</td>
<td>1.2 KJ/km</td>
</tr>
<tr>
<td>Number of electrons per pulse</td>
<td>$10^{10}$</td>
</tr>
<tr>
<td>Optical energy</td>
<td>1.5 KJ/km</td>
</tr>
<tr>
<td>Bunch radius $\sigma_x$</td>
<td>0.1 $\mu$m</td>
</tr>
<tr>
<td>Bunch length $\sigma_z$</td>
<td>1 mm</td>
</tr>
<tr>
<td>Disruption parameter $D$</td>
<td>2</td>
</tr>
<tr>
<td>Beam radiation parameter $\delta$</td>
<td>0.3</td>
</tr>
<tr>
<td>Repetition frequency $f$</td>
<td>2 kHz</td>
</tr>
</tbody>
</table>

3.4 Wake-field acceleration

High accelerating gradients can be produced by the passage of bunches of charged particles through structures such as cavities or disk-loaded wave-guides. A number of accelerators based on this principle have been proposed and models have been or are being constructed in several laboratories around the world.

Very high transformer ratios, i.e. ratios of the maximum accelerating field behind to the decelerating field inside the driver, are hoped to be obtainable with a "wake-field transformer", in which the driver passes along a different trajectory from the beam to be accelerated (Voss and Weiland, 1982). In the coaxial type, the driver passes through an annular slot on the outside of a series of circular cylindrical cavities. The induced fields converge towards the beam to be accelerated on the axis. A model of such an accelerator is under construction at DESY and will be tested shortly. More limited transformer ratios can be obtained in elliptical cavities which have been tested recently in Japan.
3.5 Plasma wakefields

A bunch of electrons passing through a plasma will excite plasma oscillations travelling with a phase velocity equal to the bunch velocity. Gradients of several GeV/m seem possible. The fields behind the driving beam are similar to the wakefields in a linac structure, with the difference that only a single mode is present and the group velocity is zero so that the field pattern does not spread out laterally. A following bunch can be accelerated by this field.

The principle was proposed by Chen, Huff and Dawson (1984). It seemed at first that the driven beam cannot gain more than twice the energy than the driving beam loses on average. However, it was found by Chen & Dawson (1985) that a driving beam with a sawtooth-shaped longitudinal density profile will permit larger transformer ratios. Another way to obtain this is suggested by van der Meer (1986).

Some of the properties of the scheme important to accelerator application were described by Ruth, Chao, Morton and Wilson (1984). It appears that the transverse focusing by the wakefield tends to be very strong; with the low-emittance beams needed for a narrow final focus this results in beam diameters of the order of a micron. The driving beam, however, must have a diameter of the order of a plasma wavelength (0.1 – 1 mm for typical plasma densities) to prevent excessive radiation loss by particles with large transverse amplitudes (Zotter, 1986). It is not evident that the energy transfer between the beams will be efficient enough with parameters suitable for TeV accelerators. The inherent simplicity of the scheme seems, however, to justify further effort.

Theoretical work on the scheme is at present going on at the University of California, at SLAC and at some other places, including CERN. Much more of this will certainly be needed. Experimental tests on a small scale, to be performed at ANL, have been proposed by a Wisconsin group (Rosenzweig et al., 1985). An electron linac capable of producing dense, short bunches is essential for such tests. The LL electron linac at CERN has also been suggested as a possible source for tests (E.J.N. Wilson, 1985).

3.6 Plasma beat-wave acceleration

Two collinear laser beams propagating in a plasma with frequencies differing by the plasma frequency can drive a Langmuir wave whose phase velocity is close to the velocity of light. Resonant build-up of plasma-electron oscillations into the non-linear regime generates a ponderomotive force leading to charge separation and a travelling electrostatic wave. With plasma densities in the range of $10^{16} - 10^{17}$ cm$^{-3}$ accelerating fields of several GeV/m can be produced.

Initial experiments (UCLA and Quebec, CO$_2$ lasers), together with computer simulations, have shown that the basic principle is valid. A Rutherford Lab./Imperial College experiment, in which CERN is participating, is underway; it uses the RAL Nd-glass laser (1.06 micron) and is thus complementary to experiments with CO$_2$ lasers (10.6 microns). Other groups, both in the USA and in Europe, intend to start experimental work in this field.
The plasma beat-wave accelerator (BWA) and the plasma wakefield accelerator are unique in being the only accelerating methods so far proposed that could plausibly offer energy gains of several GeV/m. These high fields result from charge separation in a medium which, being fully ionised, cannot break down, and offer a particular attraction for the very long term. However, plasmas are notoriously difficult to control and are prone to many instabilities.

The plasma beat-wave principle could be applied to the generation of very strong focusing fields for the interaction region. In this application the problems of tolerances and stability over long distances would be much less severe than for a TeV accelerator. There is therefore a considerable incentive to pursue plasma studies in this regime for possible use with accelerators of a more conventional type in the not-so-far future.

4. **Concluding Remarks**

When demands for $e^+e^-$ collisions first moved into the TeV energy region, with corresponding increases in desired luminosity, conventional approaches to acceleration seemed unpromising and this gave a strong incentive to consider more exotic ideas. In particular it seemed desirable to create extremely high accelerating fields to reduce the physical size of the accelerator. A large number of ideas were born, much larger than those listed earlier. Some were rather short-lived, but quite a few are sufficiently promising that it is well worth putting effort into further studies and development. Nevertheless, no obvious snag-free solution has yet been found, and this is the main reason why new emphasis has recently been put on more conventional approaches. By being open-minded towards the possibilities of very different frequencies, very different power sources, very different beam structures etc., optimism has increased that quasi-conventional solutions maybe found, so that linear colliders need not wait for more exotic ideas to mature. However, several important problems still need attention before one can have sufficient confidence to start detailed design and preparation for construction. All linear colliders will require extremely small beam cross-sections in the interaction regions. It is unclear what challenges this will give to future surveying techniques. Automatic beam centring methods will be needed, but these will need to be based on very accurately surveyed accelerator components.

In short, for linear $e^+e^-$ colliders extensive accelerator R & D is needed, but they offer great promises for the somewhat far future. Hadron colliders, on the other hand, can be built for $E_{cm} > 10$ TeV any time, with requirements to surveying techniques not substantially more difficult than on present day projects.

The CERN Council established in June 1985, a Long Range Planning Committee with Rubbia as Chairman to evaluate the problems related to the many different options for future accelerator facilities and to propose a balanced long-term programme for Europe, also taking into account complementarity in the world. Much of what I have reported is a result of studies initiated by this Rubbia Committee, whose final report is due in June, 1987. Although it is perhaps hoped that this report will crystallise clearly which path to follow into the future, it will also have to contain flexibility and options to allow for desirable changes as programmes develop and become more concrete in other parts of the world, in particular in the USA, where considerable uncertainty reigns at the moment.
ACHILLI, V.  
Istituto Nazionale di Geofisica, Rome, Italy

ANDERSON, E.G.  
University of Calgary, Alberta, Canada

ANDERSSON, A.  
MAX-lab, University of Lund, Sweden

BAUER, M.  

BEYER, W.  

BOCCHIO, F.  
Istituto di Geodesia e Geofisica, Trieste, Italy

BUSCH, W.  

CAGNETTI, V.  
ENEA, S.M. Galeria (Rome), Italy

CAMPANA, H.  
Eurolep, Ferney-Voltaire, France

CERQUEIRA BASTOS, M.L.  
Observatório Astronómico, N.V. de Gaia, Portugal

CHABLAIS, H.  
Office Fédéral de Topographie, Wabern, Switzerland

CHAPERON, F.  
ETH-Hönggerberg, Zürich, Switzerland

CHAPMAN, D.  
University College London, UK

CHARRIER, D.  
SERCEL, Carquefou, France

COLLOMB, G.  
Groupe Dégau, Grenoble, France

CONROY, P.J.  
Harza Engineering Company, Chicago, IL, USA

DE JONG, G.  
Delft University of Technology, Netherlands

DEMOMIE, L.  
Institut Géographique National, Brussels, Belgium

DERWAEL, J-J.  
Omgving, Mortsel, Belgium

DORÉ, V.  
CERN, Geneva, Switzerland

DUPRAZ, H.  
Ecole Polytechnique Fédérale, Lausanne, Switzerland

EICHLHOLZ, K.  
Intitut für Markscheidewesen, Bochum, Fed. Rep. Germany

ELMAGHRABY, S.  
Geodätisches Institut, Hannover, Fed. Rep. Germany

FEKETE, S.  
MOM, Budapest, Hungary

GALVENIUS, G.  
Royal Institute of Technology, Stockholm, Sweden

GAO, B.  
I.H.E.P., Beijing, People's Republic of China

GAUBERT, P.  
Wild + Leitz France, Rueil-Malmaison, France

GOTTWALD, R.  
Kern & Co. AG., Aarau, Switzerland

GROOM, D.  
Super collider CGD, c/o Lawrence Berkeley Lab., CA, USA

GROSS, D.L.  
Illinois State Geological Survey, Champaign, IL, USA

GROSSE, H.  
Technische Hochschule Darmstadt, Fed. Rep. Germany

HERRERO, R.  
Instituto Geográfico Nacional, Madrid, Spain

HIERNAUX, M.  
Administration du Cadastre, Brussels, Belgium

HOEGGERL, N.  
Bundesamt f. Eich- und Vermessungswesen, Vienna, Austria

HULLEIN, F.  
SERCEL, Carquefou, France

ILLNER, M.  
Karlsruhe University (TH), Fed. Rep. Germany

INGENSAND, H.  
Wild Heerbrugg AG, Heerbrugg, Switzerland

KETCHAM, L.  
Fermi National Accelerator Laboratory, Batavia, IL, USA

KNIGHT, J.A.  
North East London Polytechnic, Dagenham, UK

KOHL, A.  
Technische Hochschule Darmstadt, Fed. Rep. Germany

KYLE, S.  
University College London, UK

LAUBERTIE, B.  
INSA, Villeurbanne, France
LAVRAUW, W.  Bureau d'Etudes Lavrauw, Wezembeek-Oppem, Belgium
LICHTENEGGER, H.  Inst. f. Angewandte Geodäsie, TU Graz, Austria
LOSER, R.  KERN & Co. AG, Aarau, Switzerland
MACKAY, W.  Texas Accelerator Centre, The Woodlands, TX, USA
MANZONI, G.  Ist. di Strade e Trasporti, University of Trieste, Italy
MARCHESINI, C.  Ist. di Geodesia e Geofisica, University of Trieste, Italy
MARCHETTA, A.  Italian Military Geographic Institute, Florence, Italy
METAILLIER, N.  CEA-CNRS, Gif-sur-Yvette, France
MOORE, C.  Fermi National Accelerator Laboratory, Batavia, IL, USA
MURPHY, C.T.  CERN, Geneva, Switzerland
NAPIER, M.  Nottingham University, UK
PERNICE, B.  Dipartimento di Fisica, Bari University, Italy
PICCA, D.  Dipartimento di Fisica, Bari University, Italy
PITTHAN, R.  SSC-SLAC, Stanford, CA, USA
PUSZTAI  MOM, Budapest, Hungary
REGGIOARDO, G.  Genoa University, Italy
RIEHL, J.  Technische Hochschule Darmstadt, Italy
RIEMENSPERGER, M.  Kern & Co. AG, Aarau, Switzerland
RINNER, K.  Institute of Space Research, Graz, Austria
SCHAUERTE, W.  Geodätisches Institut, Bonn University, Fed. Rep. Germany
SCHNEIDER, D.  Bundesamt für Landestopographie, Wabern, Switzerland
SPIEGEL, A.  Wild Heerbrugg AG, Heerbrugg, Switzerland
STAIGER, R.  Texas A & M University, College Station, TX, USA
STUDEMANN, B.  ETH-Hönggerberg, Zürich, Switzerland
SUDAU, A.  Geodätisches Institut, Bonn University, Fed. Rep. Germany
TIBERGHIES, V.  Soc. Franc. de Stereotopographie, Levallois Perret, France
TROIANI, M.  LNF-INFN, Frascati, Italy
VAN CRANENBROECK, J.  Institut Géographique National, Brussels, Belgium
VAN DE VEGT, H.J.W.  Meetkundiger Dienst, Rijkswaterstaat, Delft, Netherlands
VAN TWEMBEKE, U.  Ecole Royale Militaire, Brussels, Belgium
VEAL, D.  CERN, Geneva, Switzerland
VERRONE, G.  Dipartimento di Fisica, Bari University, Italy
WEI, F-Q.  LNF-INFN, Frascati, Italy
WILLIS, P.  Institut Géographique National, Saint Mandé, France
WITTE, B.  Geodätisches Institut, RWTH Aachen, Fed. Rep. Germany
WUNDERLICH, T.  Technische Universität Wien, Austria