CERN 99–05
19 August 1999

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
**CERN** EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

# CAS CERN ACCELERATOR SCHOOL

# VACUUM TECHNOLOGY

Scanticon Conference Centre, Snekersten, Denmark
28 May–3 June 1999

**PROCEEDINGS**

Editor: S. Turner

GENEVA
1999

CERN LIBRARIES, GENEVA

CERN LIBRARIES, GENEVA

B00013940

© Copyright CERN, Genève, 1999

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.

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.

ISSN 0007–8328
ISBN 92–9083–149–9

CERN 99–05
19 August 1999

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
**CERN** EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

# CERN ACCELERATOR SCHOOL

# VACUUM TECHNOLOGY

Scanticon Conference Centre, Snekersten, Denmark
28 May–3 June 1999

**PROCEEDINGS**

Editors: S. Turner

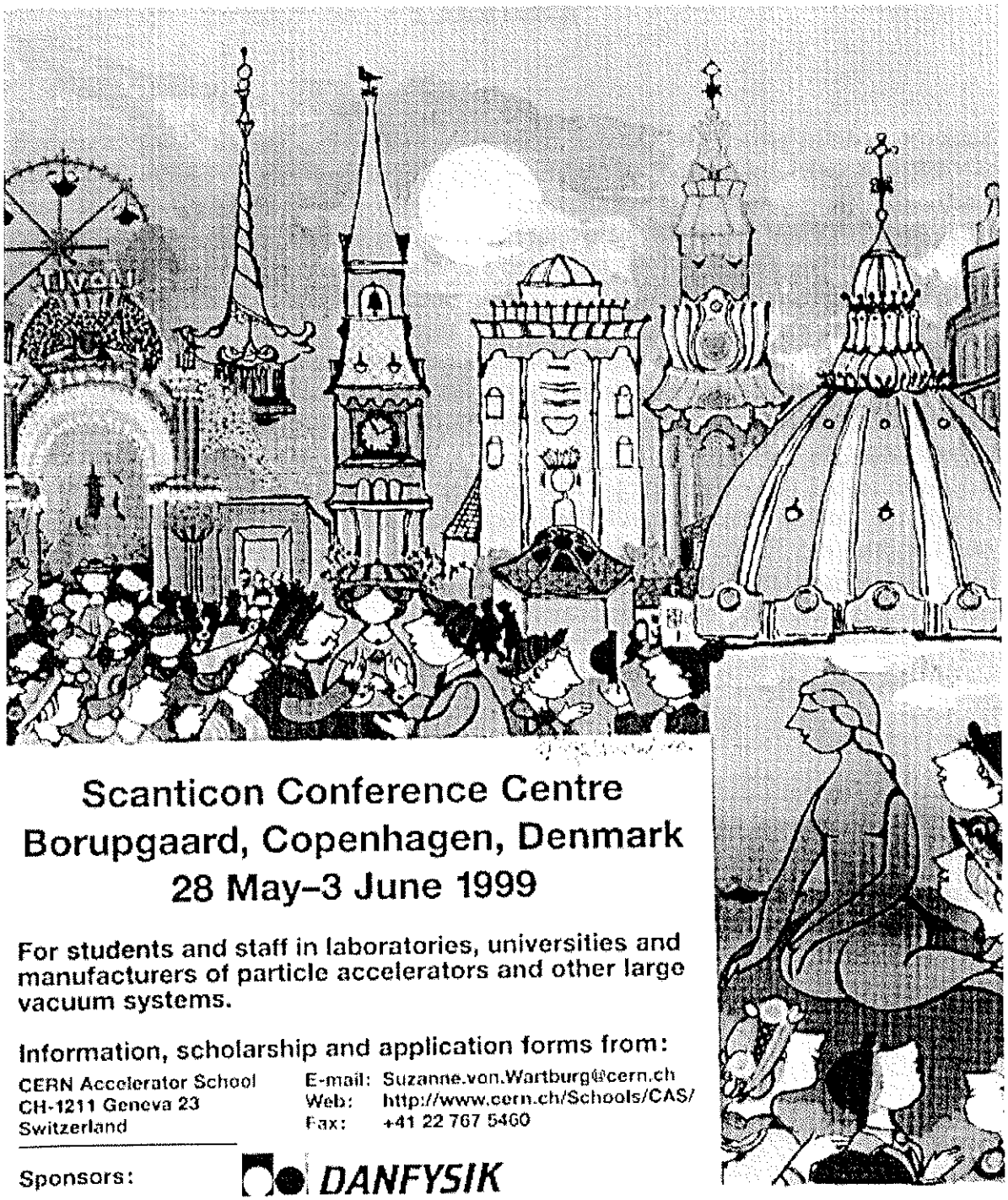
GENEVA
1999

CERN–Service d'information scientifique–RD/993–1800–août 1999

# ABSTRACT

These proceedings present the lectures given at the twelfth specialized course organized by the CERN Accelerator School (CAS), the topic this time being 'Vacuum Technology'. Despite the importance of vacuum technology in the design and operation of particle accelerators at CERN and at the many other accelerators already installed around the world, this was the first time that CAS has organized a course devoted entirely to this topic. Perhaps this reflects the facts that vacuum has become one of the more critical aspects of future accelerators, and that many of the pioneers in the accelerator field are being replaced by new, younger personnel. The lectures start with the basic concepts of the physics and technology of vacuum followed by detailed descriptions of the many different types of gas-pumping devices and methods to measure the pressures achieved. The outgassing characteristics of the different materials used in the construction of vacuum systems and the optimisation of cleaning methods to reduce this outgassing are then explained together with the effects of the residual gases on the particle beams. Then follow chapters on leak detection, materials and vacuum system engineering. Finally, seminars are presented on designing vacuum systems, the history of vacuum devices, the LHC (large hadron collider) vacuum system, vacuum systems for electron storage rings, and quality assurance for vacuum.

# Niels Bohr Institute
# CERN Accelerator School
## course on
# *Vacuum Technology*



**Scanticon Conference Centre**
**Borupgaard, Copenhagen, Denmark**
**28 May–3 June 1999**

For students and staff in laboratories, universities and manufacturers of particle accelerators and other large vacuum systems.

Information, scholarship and application forms from:

CERN Accelerator School
CH-1211 Geneva 23
Switzerland

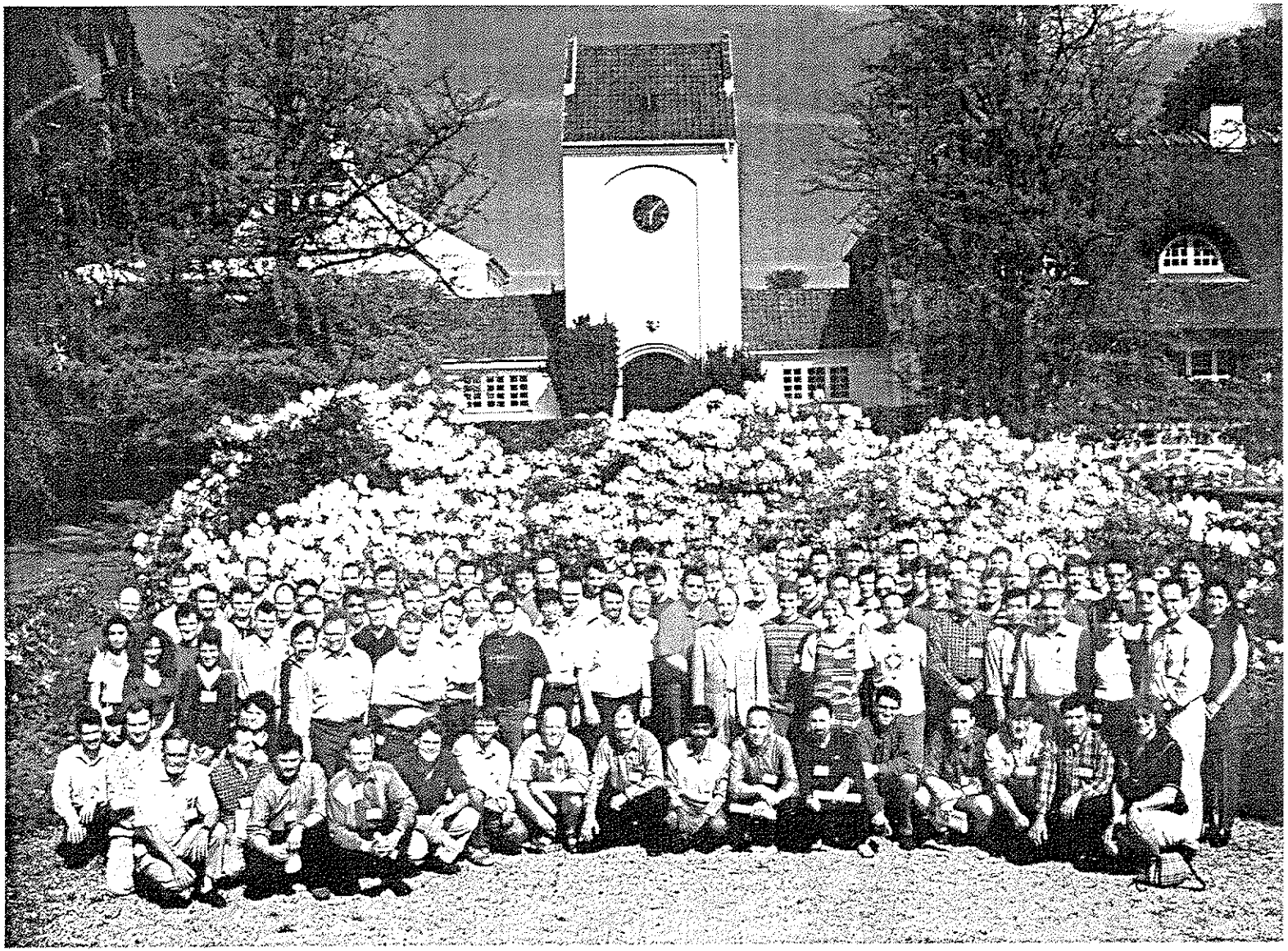
E-mail: Suzanne.von.Wartburg@cern.ch
Web: http://www.cern.ch/Schools/CAS/
Fax: +41 22 767 5460

Sponsors: DANFYSIK

# VACUUM TECHNOLOGY

Scanticon Borupgaard Conference Centre, 28 May - 3 June 1999

| Time | Friday 28 May | Saturday 29 May | Sunday 30 May | Monday 31 May | Tuesday 1 June | Wednesday 2 June | Thursday 3 June |
|---|---|---|---|---|---|---|---|
| | BREAKFAST | | | | | | |
| 09.00<br>10.00 | Introduction to vacuum physics<br>*N. Marquardt* | Outgassing I (physics)<br>*J. de Segovia* | Outgassing II (static outgassing)<br>*K. Jousten* | | Outgassing III (dynamic outgassing)<br>*O. Gröbner* | Engineering II<br>*L. Westerberg* | Leak detection<br>*N. Hilleret* |
| | COFFEE | | | E | BREAK | | |
| 10.10<br>11.10 | Pumps I (mechanical)<br>*N. Hilleret* | Materials<br>*E. Huttel* | Cleaning I<br>*R. Reid* | X | Engineering I<br>*L. Westerberg* | Vacuum controls and operation<br>*D. Schmied* | Extreme high vacuum<br>*P. Redhead* |
| | LUNCH | | | C | COFFEE | | |
| 11.30<br>12.30 | Pressure measurement I (total)<br>*Ch. Edelmann* | Pumps II (ion-pumps)<br>*L. Schulz* | Pumps III (getters)<br>*C. Benvenuti* | U<br>R | Pumps IV (cryo-pumps)<br>*C. Benvenuti* | Beam residual gas interactions I (lifetime effects)<br>*S. Pape Møller* | "The best laid schemes ..."<br>*L. Westerberg* |
| | TEA | | | S | LUNCH | | |
| 14.30<br>15.30 | **Seminar** Designing Accelerator vacuum systems<br>*P. Marin* | Pressure Measurement II (ionisation gauges)<br>*K. Jousten* | Pressure Measurement III (partial pressure)<br>*J.H. Leck* | I<br>O | Cleaning II<br>*R. Reid* | Beam residual gas interactions II (neutralisatio effects)<br>*A. Poncet* | |
| | Welcome Reception | | | N | TEA | | |
| 16.00<br>17.00 | | **Seminar** History of vacuum devices<br>*P. Redhead* | **Seminar** LHC Vacuum system<br>*O. Gröbner* | * | **Seminar** Vacuum system for electron storage rings<br>*D. Kraemer* | **Seminar** Quality Assurance for vacuum<br>*R. Reid* | |
| 20.00 | EVENING MEAL | | | | BANQUET | MEAL | |



Participants and setting of the Vacuum Technology course at Snekersten, Denmark, 1999

# FOREWORD

The aim of the CERN Accelerator School to collect, preserve and disseminate the knowledge accumulated in the world's accelerator laboratories applies not only to accelerators and storage rings, but also to their related sub-systems, equipment and technologies. This wider aim is being achieved by means of the specialized courses listed in the Table below. The latest of these was on the topic of Vacuum Technology, held at the Scanticon Conference Centre, Snekersten, Denmark, from 28 May to 3 June 1999, its proceedings forming the present volume.

Over the last few years vacuum technology has made great strides forward and has become of immeasurable importance. Industry currently uses ultra-high vacuum in many production processes while the more recently achievable extreme high vacuum is necessary for the simulation of interstellar space and in the processing of sophisticated semiconductor devices, and particularly in the field of particle accelerators and storage rings. Therefore it was felt to be the appropriate moment to design a course where the many facets of technology required to produce and measure the different degrees of vacuum could be brought together into a logical sequence. Naturally, with so much progress having been made in the accelerator field, it was clear that accelerator rerquirements should form the backbone of the course especially since the techniques used there are applicable in most other research institutes and in industry. The resulting series of lectures and seminars was presented at the course and written-up for the use of the present and future workers in this discipline.

Organization of this course and the production of its proceedings relied very heavily on the help and encouragement CAS received from many Institutes, Companies and individuals. In particular we wish to acknowledge the help of J.D. Hansen and H. Vedel of the Niels Bohr Institute, K.N. Henrichsen of CERN as well as Director Prof. Ove Poulsen for his lively and provocative opening talk. Financial support in the form of scholarships from UNESCO, NORFA, SNF/HEP and Danfysik was also most appreciated along with the generous budget made available by the CERN Management. The CAS Advisory, Programme and Local Organizing Committees also deserve sincere thanks for the guidance and attention to detail that ensured the great success of this course. The efficient and friendly service provided by the staff of the Scanticon Conference Centre made our stay there most enjoyable as well as educative. However the greatest effort was made by the course lecturers in preparing and presenting the various topics as well as writing the chapters of these proceedings. Not only the participants at the course but also all those who will use these proceedings in the future owe them a great mark of gratitude. Last, but certainly not least, the attendance and keenness of the participants from so many parts of the world made the efforts made for them so worthwhile and we thank them most warmly.

S. Turner
Editor

# LIST OF SPECIALISED CAS COURSES AND THEIR PROCEEDINGS

| Year | Course | Proceedings |
|---|---|---|
| 1983 | Antiprotons for colliding beam facilities | CERN 84-15 (1984) |
| 1986 | Applied Geodesy for particle accelerators | CERN 87-01 (1987) also Lecture Notes in Earth Sciences 12, (Springer Verlag, 1987) |
| 1988 | Superconductivity in particle accelerators | CERN 89-04 (1989) |
| 1989 | Synchrotron radiation and free-electron lasers | CERN 90-03 (1990) |
| 1990 | Power converters for particle accelerators | CERN 90-07 (1990) |
| 1991 | RF engineering for particle accelerators | CERN 92-03 (1992) |
| 1992 | Magnetic measurement and alignment | CERN 92-05 (1992) |
| 1993 | RF engineering for particle accelerators (repeat of the 1991 course) | – |
| 1994 | Cyclotrons, linacs and their applications | CERN 96-02 (1996) |
| 1995 | Superconductivity in particle accelerators | CERN 96-03 (1996) |
| 1996 | Synchrotron radiation and free-electron lasers | CERN 98-04 (1998) |
| 1997 | Measurement and alignment of accelerator and detector magnets | CERN 98-05 (1998) |
| 1999 | Vacuum Technology | Present volume |

# CONTENTS

**Foreword** ..... v

*N. Marquardt*
**Introduction to the principles of vacuum physics** ..... **1**
Introduction—definition, history and applications of vacuum 1
Conversion factors for pressure and pressure ranges of vacuum, vacuum gauges and pumps 10
Composition of air 13
Ideal and real gases 14
Kinetic theory of gases 15
Thermal conductivity and energy transport 17
Gas flow 18
Gas load and ultimate pressure 23
Conclusion 24

*N. Hilleret*
**Mechanical pumps** ..... **25**
Introduction 25
Mechanical pumps 25
Sorbtion pumps 28
Momentum transfer pumps 29
Pump characteristics 31
Layout of a roughing station 31
Conclusion 35

*L. Schulz*
**Sputter-ion pumps** ..... **37**
Introduction 37
Pumping mechanism 37
Operating principles 38
Pressure range 40
Special design for integrated linear pumps 40
Operational aspects 41
Conclusions 41

*C. Benvenuti*
**Molecular surface pumping: the getter pumps** ..... **43**
Generalities 43
Evaporable getters 45
Non-evaporable getters (NEG) 46
Thin-film getter coatings 48

*C. Benvenuti*
**Molecular surface pumping: cryopumping** ..... **51**
Generalities 51
Cryogenic considerations 54
Cryosorption pumps 55
Condensation cryopumps 55
The hydrogen "anomaly" 56

*Ch. Edelmann*
**Pressure measurement—total pressures** ..... **59**
Introduction 59
Gauges using force effects for pressure measurement 60
Gauges using the hydrostatic pressure of mercury or other non-volatile liquids 62
Modulation gauge 64
Gauges using the viscosity of a gas for pressure measurement 64
Gauges using the pressure dependence of the specific heat conductivity of the gas 66
Radiometer-effect gauges 69
Gauges using the interaction between electrons and gas particles 70
Survey of the operational ranges of vacuum gauges 73

*K. Jousten*
**Pressure measurement with ionization gauges** ..... **75**
Introduction 75
Brief historical review 75
The design of today's most common ionization gauges 76
Problems in using ionization gauges 82
Ionization gauge calibration 85
Ionization gauge calibration methods 86

*J.H. Leck*
**Partial pressure measurement** ..... **89**
Introduction 89
The quadrupole mass filter 91
The design of practical quadrupole filters 95
The four operating zones for the quadrupole filter 96

*J.L. de Segovia*
**Physics of outgassing** ..... **99**
Introduction 99
Permeation of gases through matter 99
Diffusion from a semi-infinite wall 101
Diffusion from a finite wall 102
Vapour pressure 104
Adsorption 105

*K. Jousten*
**Thermal outgassing** ..... **111**
Introduction 111
Pump-down behaviour and desorption 111
Vacuum systems after bake-out and out-diffusion 115
Pre-bake-outs to reduce outgassing 117
Outgassing-rate measurement methods and the problem of reabsorption 121
Conclusion 124

*O. Gröbner*
**Dynamic outgassing** ..... **127**
Synchrotron radiation induced desorption 127
Ion-induced desorption 132
Conclusions 137

*R.J. Reid*
**Cleaning for vacuum service** ..... **139**
Introduction 139
Choosing a cleaning process 140
Dependence of the base pressure required 140
Dependence on contaminants 141
How construction affects cleaning 141
Chemical cleaning agents 142
Chemical cleaning of specific materials 143
Physical cleaning techniques 145
Passivation 146
Special cleaning techniques 146
Safety 149
Cost 150
How do we know a surface is clean? 150
Surface analytical techniques 150
Phenomenological methods 152
Recontamination 152
Summary 153

*S. Pape Møller*
**Beam-residual gas interactions** ..... **155**
Introduction 155
The concepts of cross sections and lifetimes 155
The interaction between charged particles and residual-gas atoms 156
Local versus average pressure 163
Conclusions 163

*A. Poncet*
**Ion trapping, clearing, beam-ion interactions** ..... **165**
Introduction 165
Neutralisation of a beam: a simple description 165
The ionisation process 167
The ion or electron motion 171
A few examples of ion or electron motion 177
Bunched beams 178
Clearing means 179
The limit of accumulation 184
The effects of neutralisation 184
Diagnostics and phenomenology 192

*N. Hilleret*
**Leak detection** ..... **203**
Introduction 203
Methods 203
Helium lead detectors 204
Practical cases 208
Conclusion 211

*P.A. Redhead*
**Extreme high vacuum** ..... **213**
Introduction 213
General considerations 214
Pumps for XHV 215
Pressure measurement at XHV 218
Leak detection 221
Reduction of outgassing 222
Conclusions 224

*D. Schmied*
**Vacuum control and operation** ..... **227**
Vacuum control 227
Interlocks 231
Vacuum operation 232

*E. Huttel*
**Materials for accelerator vacuum systems** ..... **237**
Introduction 237
Physical restrictions 238
Selected materials 243
Bonding 250
Corrosion 250

*L. Westerberg*
**Engineering** ..... **255**
Design principles of a UHV system and its building blocks 255
Typical building blocks in a UHV system 258
Joining methods 260
Combined action of flanges, gaskets and bolts 261
Brazing and brazing preparations 261
Welding and weld preparation 262
Valves 262
Expansion bellows 263
Bakeable vacuum systems 263

*L. Westerberg*
**The best laid schemes ...** ..... **267**
Introduction 267
Recommendations 267
References for further information 270

*P. Marin*
**Designing accelerator vacuum systems** ..... **271**
Introduction 271
The physical and technical basis for the specifications of accelerator vacuum systems 271
The design of vacuum systems 275
Conclusions 279

*P.A. Redhead*
**History of vacuum devices** ..... **281**
Introduction 281
Vacuum for fun and science, the first two centuries, 1650 to 1850 282
Vacuum for science, the great leap forward, 1850 to 1900 283
Vacuum for industry, 1900 to 1950 285

*O. Grobner*
**LHC vacuum system** ..... **291**
Introduction 291
The vacuum system 293
Primary desorption by synchrotron radiation 296
Vapour pressure and recycling of physisorbed gas 297
Dynamic model for $H_2$ 298
Ion stimulated desorption and pressure instability 301
Beam-induced electron multipacting 302
Mechanical design aspects 303
Summary 305

*D. Kraemer*
**Vacuum systems of electron storage rings** ..... **307**
Introduction 307
Synchrotron radiation aspects in electron storage rings 307
Linear power density 309
Thermal- and photon-induced desorption 310
Pumping scheme & pressure distribution 312
Single-particle loss mechanisms 313
Implicit requirements for the vacuum system 316
Accelerator vacuum systems in routine operation 319
Conclusion 319

*R. Reid*
**Quality issues in vacuum** ..... **321**
What is "quality"? 321
Quality systems 321
Vacuum system quality 321
Specification of a vacuum system 322
Vacuum system performance and its measurement 323
Real-world vacuum gauging 323
Summary 324

**List of participants** ..... **325**

# INTRODUCTION TO THE PRINCIPLES OF VACUUM PHYSICS

*Niels Marquardt*
Institute for Accelerator Physics and Synchrotron Radiation,
University of Dortmund, 44221 Dortmund, Germany

**Abstract**
Vacuum physics is the necessary condition for scientific research and modern high technology. In this introduction to the physics and technology of vacuum the basic concepts of a gas composed of atoms and molecules are presented. These gas particles are contained in a partially empty volume forming the vacuum. The fundamentals of vacuum, molecular density, pressure, velocity distribution, mean free path, particle velocity, conductivity, temperature and gas flow are discussed.

## 1. INTRODUCTION — DEFINITION, HISTORY AND APPLICATIONS OF VACUUM

The word "vacuum" comes from the Latin "vacua", which means "empty". However, there does not exist a totally empty space in nature, there is no "ideal vacuum". Vacuum is only a partially empty space, where some of the air and other gases have been removed from a gas containing volume ("gas" comes from the Greek word "chaos" = infinite, empty space). In other words, vacuum means any volume containing less gas particles, atoms and molecules (a lower particle density and gas pressure), than there are in the surrounding outside atmosphere. Accordingly, vacuum is the gaseous environment at pressures below atmosphere.

Since the times of the famous Greek philosophers, Demokritos (460-370 B.C.) and his teacher Leukippos (5$^{th}$ century B.C.), one is discussing the concept of vacuum and is speculating whether there might exist an absolutely empty space, in contrast to the matter of countless numbers of indivisible atoms forming the universe. It was Aristotle (384-322 B.C.), who claimed that nature is afraid of total emptiness and that there is an insurmountable "*horror vacui*". Therefore, he doubted and even rejected an absolute vacuum. He assumed, for example, that the idea of empty space would invite the concept of motion without resistance, i.e. a motion at infinite velocity. This opinion became a paradigma for almost 2000 years. It was believed by famous writers, like Roger Bacon (1214-1299) and René Descartes (1596-1650) and was strongly supported also by the church.

Only in the 17$^{th}$ century were vacuum physics and technology born. Galileo (1564-1642) was among the first to conduct experiments attempting to measure forces required to produce vacuum with a piston in a cylinder. Torricelli (1608-1647), an associate of Galileo´s, succeeded in 1644 to produce vacuum experimentally by submerging a glass tube, which was filled with mercury and closed at one end, with its open end in a pool of mercury. By using mercury instead of water, he was able to reduce the size of the apparatus to convenient dimensions. He demonstrated that the mercury column was always 760 mm above the level in the pool, regardless of size, length, shape or degree of tilt of the tube (see Fig. 1) and, in this way, he measured for the first time the pressure of atmospheric air. Also, by other experiments, performed by the French scientist and philosopher Blaise Pascal (1623 - 1662), the idea of the "*horror vacui*" was definitely proved to be wrong (see Fig. 2). Pascal also measured, for example, the altitude with the Hg barometer (invented by Torricelli), contributed significantly to our understanding of vacuum physics and made many other important discoveries in physics. The unit of measure of the degree of vacuum of the International Standards Organization, i.e. the SI-unit of pressure, was called in honour of Pascal :

$$1\ \text{Pa} = 1\ \text{N/m}^2 = 7.501 \times 10^{-3}\ \text{Torr} = 10^{-2}\ \text{mbar}\ .$$

At the same time, the mayor of the city of Magdeburg, Otto von Guericke (1602-1686; see Fig. 3), modified the water pumps, invented the manometer (1661), constructed the first air pump (1650) and made many other discoveries in physics and astronomy. For instance, from the astronomical observation of the constancy of the time of orbit revolution of the planets, he concluded

that there is no friction in space, but vacuum instead. On the occasion of the Reichstag at Regensburg/Germany he performed publicly in 1654 his famous and most spectacular experiment with the two empty, gasketed copper "Magdeburg Hemispheres" of about 50 cm in diameter (see Fig. 4). By this experiment he demonstrated that it is impossible to pull the two halves apart against the air pressure, even by using 2 × 8 horses (the counter-pressure by air in the interior of the sphere is missing). During this time, it became clear that we are living on the bottom of a huge ocean of air and that the mass of the atmosphere corresponds to a pressure of about 1 kg per $cm^2$ or 10 tons on an area of 1 $m^2$. This is the reason, why the 16 horses of von Guericke were unable to pull the hemispheres apart. The reason, why we don´t feel anything of this tremendous pressure is simply that there is the same pressure inside our body. To determine the mass of air, Otto von Guericke performed his spectacular gallows experiment, shown in Fig. 5.

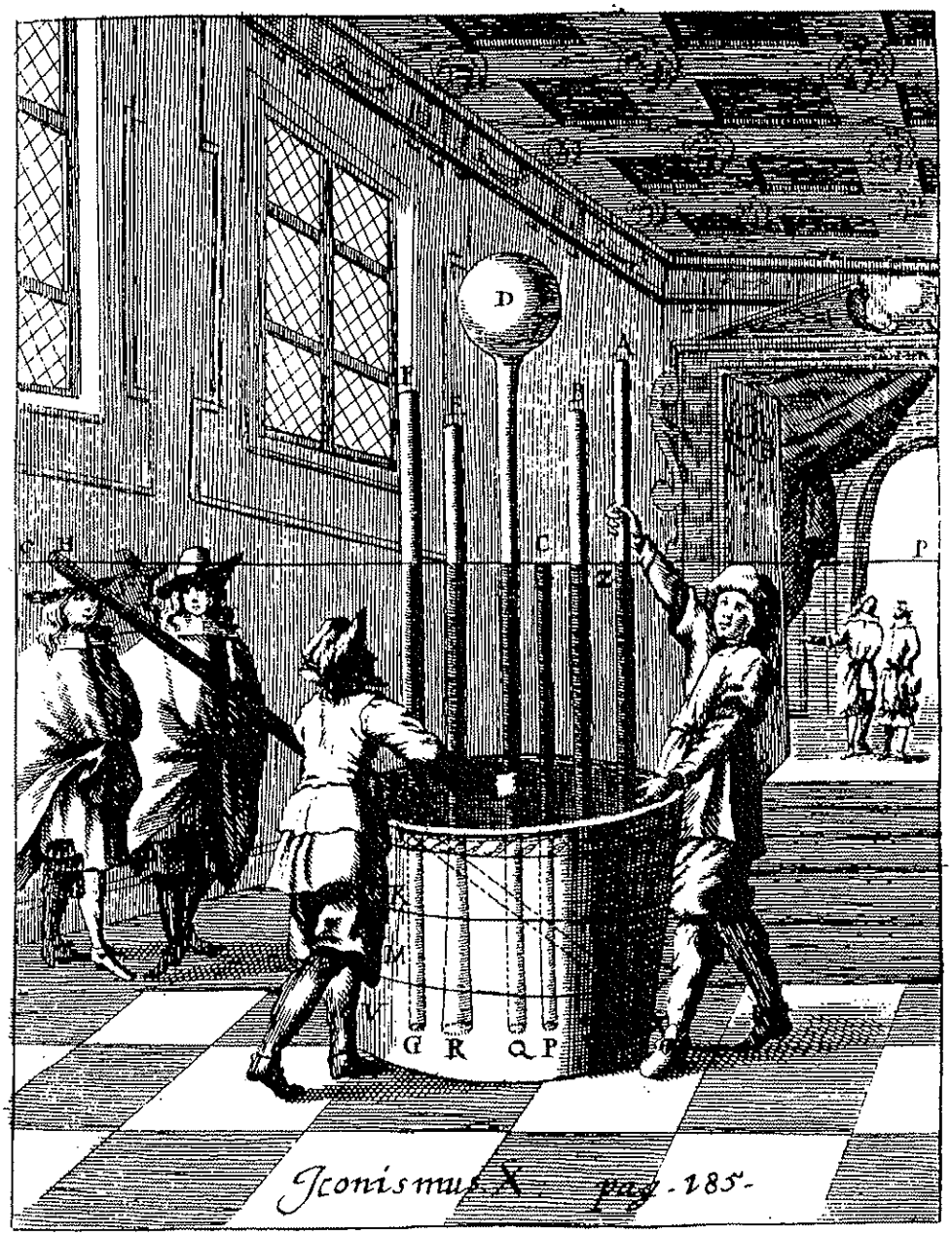



Fig. 1 Experiments of Torricelli (1608-1647), observing the height of a mercury column in one-sided glass tubes to be independent of the shape and inclination of the tubes. In the volume above the mercury meniscus was the first experimentally generated vacuum, the "Torricelli Vacuum" (taken from an engraving of P. Gasparis Schotti "*Technica Curiosa*", Nürnberg, 1664. Courtesy C. Edelmann [11]).

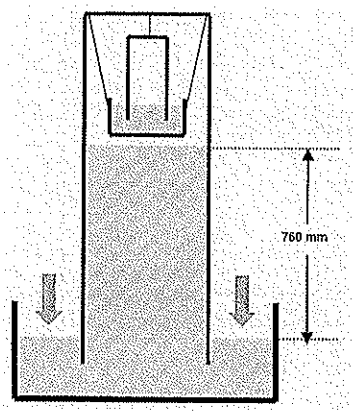



Fig. 2 Experiment by Blaise Pascal proving the action of atmospheric pressure and the existence of vacuum.

Fig. 3 Portrait of Otto von Guericke (1602-1686), (Courtesy "*Otto von Guericke Society*" in Magdeburg).

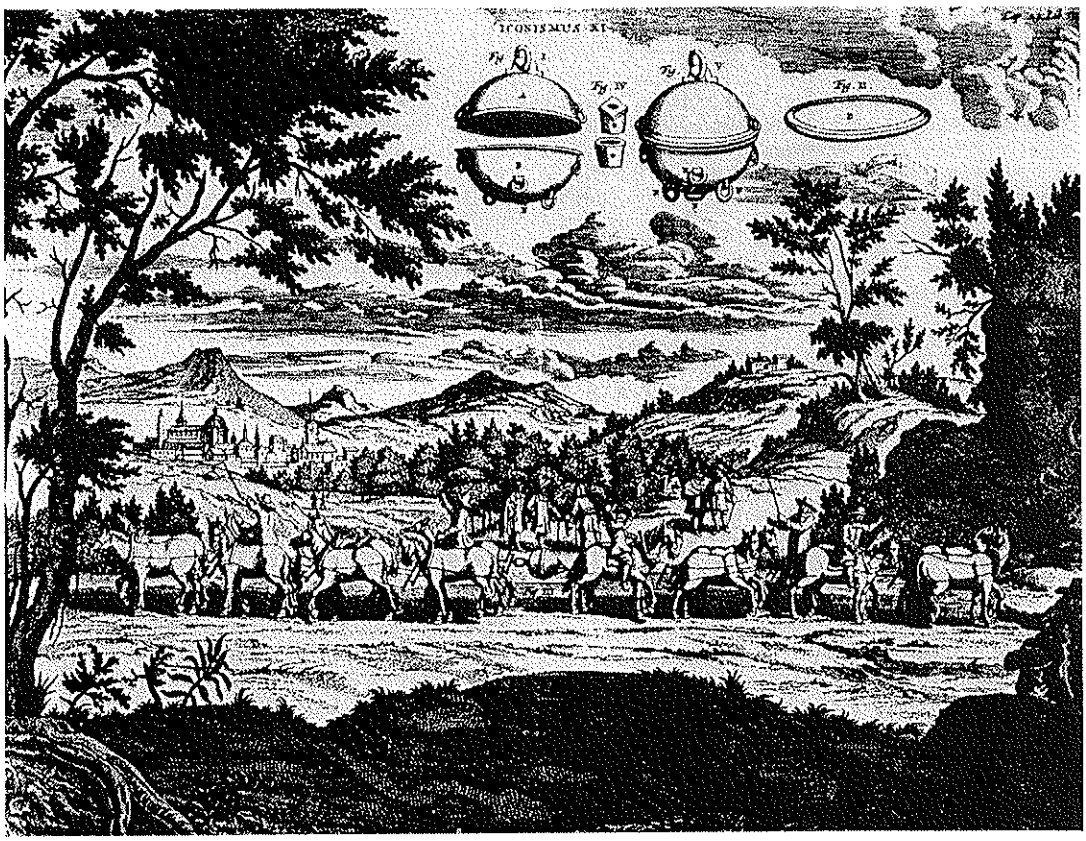


Fig. 4 Experiment with the "*Magdeburger Hemispheres*" (1654) by Otto von Guericke (taken from an engraving of P. Gasparis Schotti "*Technica Curiosa*", Nürnberg, 1664. Courtesy Otto von Guericke Society in Magdeburg).

Another important discovery was made by Robert Boyle (1627-1691, who built an improved vacuum pump, and Edme Mariotte (~1620-1684). They derived *the law of Boyle-Mariotte*, the fundamental equation of gas laws, valid for so-called "ideal gases" being in thermodynamical equilibrium :

$$\boxed{p\,V\,/\,T\;=\;\text{constant}}$$ (with $p$ = pressure, $V$ = volume and $T$ = absolute temperature of the gas).

Many technical inventions, concerning various pumps for vacuum generation, were made in the last, and until the middle, of our century, leading for instance to the discovery of the electron in 1897 by J.J. Thomson (1856-1940) and the invention of the X-ray tube by Wilhelm Conrad Röntgen (1845-1923) with his discovery of short-wavelength X-ray radiation on November 8, 1895. Finally, in the thirties of our century, the invention of the wide variety of particle accelerators started and went on till today with the construction of huge machines. With accelerators, nuclear and high-energy elementary-particle physics began. All the inventions and sophisticated technological developments of today would have been unthinkable without high vacuum and its technology. Therefore, vacuum physics and techniques paved the way for the development of modern high-technology and our industrial society.

What characterises vacuum? The particles in a volume are in constant motion. They hit the walls of the container and exert a force on its surface area, which is called "pressure". One can measure this force per unit area of the vessel, called pressure, by comparing it to the atmospheric pressure or by determining it in absolute units. Therefore, by a pressure measurement one obtains the number and intensity of particle impacts on a unit of surface area.

The highest pressures ever obtained in a laboratory happened in the European Laboratory for High Energy Nuclear Physics CERN during head-on collisions of two fast lead nuclei. This is the unimaginably large pressure of $10^{30}$ bar, which is estimated to be about 5-times larger than before a supernova explosion. Compared with this value, the maximum static pressure of about $3.5 \times 10^{6}$ bar, reached in experiments with diamond stamps, is very small. On the other hand, the lowest pressure or

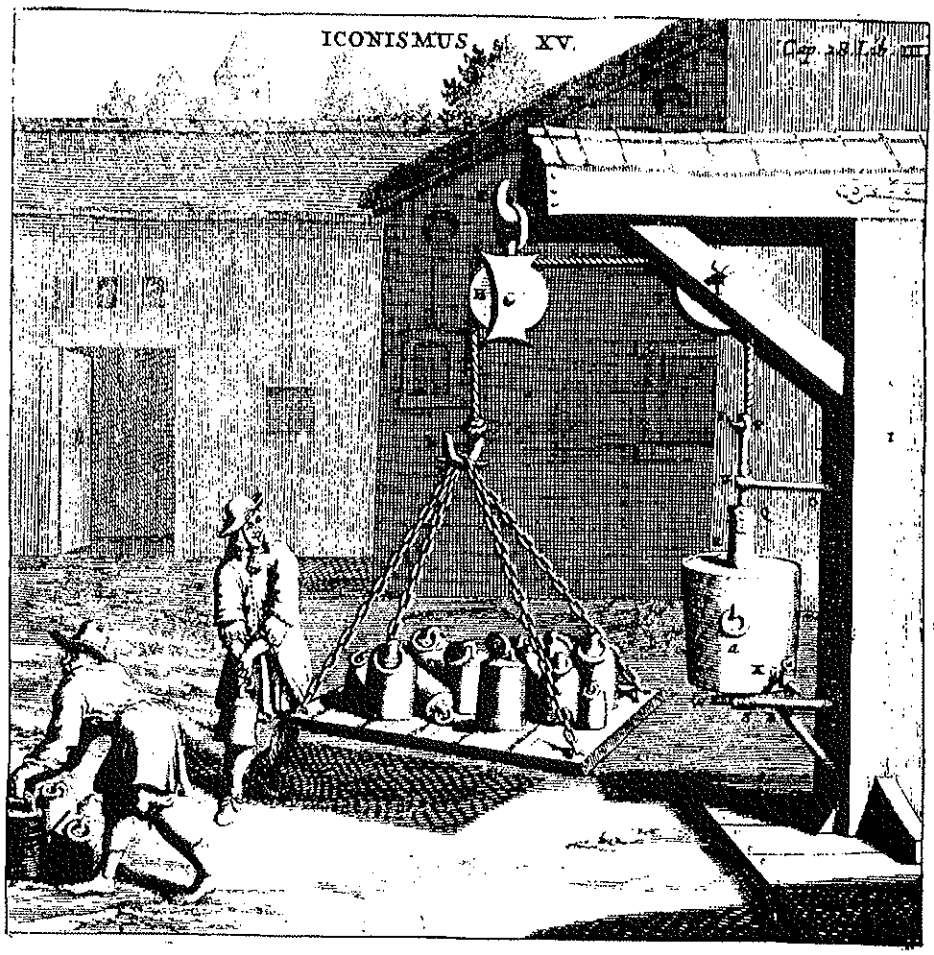



Fig. 5 Gallows experiment of Otto von Guericke (1602-1686), in order to determine the mass of air (taken from an engraving of *Ottonis de Guericke "Experimenta nova (ut vocantur) Magdeburgica de vacuo spatio"*. Courtesy C. Edelmann [11]). In a copper cylinder **a** , fixed at the gallows post on the right side of the picture, a piston **Q** can move vertically. The piston is connected with a rope running over two rolls **H** and **T** to a board loaded with several weights. At first, the piston is pulled up with the tap **X** at the bottom of the cylinder open. When evacuating the cylinder with an air pump, tap closed, the piston slides downwards due to the outside air pressure, lifting the weights.

highest vacuum generated by pumps in the laboratory amounts to $10^{-17}$ bar, which corresponds to just a few hundred particles in one $cm^3$ and which was also obtained at CERN. One should compare this with a number of ~ $3 \times 10^{19}$ particles per $cm^3$ at atmospheric pressure and room temperature. Therefore, between the highest and lowest pressure there are 47 orders of magnitude in between. But even the best vacuum obtained on earth is a high-pressure area compared with the almost total emptiness between stars in space. Accordingly, besides pressure vacuum is characterised by the density of particles. The interstellar particle density in the Milky Way, for instance, consisting of gas, plasma and dust, is only ~ $5 \times 10^4$ particles per $m^3$. Between Galaxies one has only one particle or at most a few of them per $m^3$. If one would distribute homogeneously the total matter of the universe in space, one would still have an extremely low particle density of only 3 particles per $m^3$. The universal density scale, presented in Table 1, exhibits the enormous range of about 60 orders of magnitude between the maximum density calculated for a black hole and almost totally empty regions in space with the earth somewhere in the middle. Table 2 shows examples of particle densities of various objects. However, even the almost perfect vacuum of interstellar space is not empty at all. There is electromagnetic radiation everywhere or superstrong fields, which are even able to generate new particles. Even if we would succeed to construct an empty volume, totally shielded against radiation from outside and cooled to absolute zero temperature to reduce radiation from its walls, there would still be radiation emitted from the particles of the walls which are never totally at rest (due to Heisenberg´s uncertainty principle).

Table 1
Universal density scale

| Universal density scale | g/cm$^3$ |
|---|---|
| Big Bang | $\infty$ |
| Black hole | $10^{27}$ |
| Neutron Star | $10^{14}$ |
| White Dwarf | $10^{6}$ |
| Solar Center | 158 |
| Sun (average density) | 1.4 |
| Water on Earth | 1 |
| Air | $10^{-3}$ |
| Vacuum in Laboratory | $10^{-18}$ |
| Interstellar Space | $10^{-24}$ |
| Universe | $10^{-30}$ |
| Empty Regions in Space | $10^{-33}$ |

$\sim 10^{60}$

Table 2
Examples of particle densities of various objects.

| Object | Number of Particles | Length of Cubic Edge | Mean Free Path |
|---|---|---|---|
| Gold (one of the most dense elements) | 60 000 000 | 0.0001 mm | - |
| Air (at normal pressure) | 30 000 000 | 0.001 mm | 0.000066 mm |
| Air (on Mt. Everest) | 9 000 000 | 0.001 mm | 0.0002 mm |
| Thermos Flask | 5 000 000 | 0.1 mm | 33.6 cm |
| TV Cathode-Ray Tube | 13 000 | 0.1 mm | 134 m |
| Atmosphere (at the orbit of space shuttle) | 1 000 | 0.1 mm | 1.68 km |
| Interstellar Space | 3 | 1 cm | 670 000 000 km |
| Intergalactic Space | 1 | 10 m | $1.7 \times 10^{18}$ km (=180 000 light years) |

In association with the interest in space research, the development of techniques to achieve ultra-high vacuum (UHV) is of utmost importance. The pressure of a conventional vacuum system with diffusion pumps, today normally replaced by turbo-molecular pumps, and elastomer gaskets is normally about $10^{-4}$ to $10^{-5}$ Pa, which corresponds to about $10^{14}$ to $10^{13}$ molecules of nitrogen arriving each second on each cm$^2$ of a surface at room temperature. Since an atomic monolayer corresponds to about $10^{15}$ atoms per cm$^2$ such conditions result in nitrogen arrival rates of a monolayer every 10 s (assuming that every molecule sticks to the surface). This represents an unacceptable level of contamination for most high-tech applications of today. For surface studies e.g., vacua below $10^{-7}$ Pa, i.e. UHV, became indispensable, corresponding to the formation of about one monolayer in about a few hours or to a mean free path between collisions of molecules of about 50 000 km. An UHV-system typical for this kind of research is shown in Fig. 6.

Vacuum vessels are fabricated of materials like glass, aluminium, (oxygen-free) copper, stainless steel, titanium etc., which corrode very slowly and have low rates of outgassing of adsorbed gas. Vessels are carefully cleaned by mechanical, chemical or thermal treatments and vacuum sealed by materials of different vapour pressure from rubber, synthetics to metals, depending on the quality of vacuum to be maintained. To obtain UHV conditions, free of organic contaminants, water vapour and hydrocarbons, metal gaskets are necessary, allowing baking of the whole chamber to about 520 K while the vacuum pumps operate. By baking vacuum systems for several hours an accelerated desorption of water vapour and other gases from all internal surfaces is obtained, the necessary condition for UHV. To avoid high vapour pressure, also the materials used inside vacuum chambers have to be chosen very carefully.

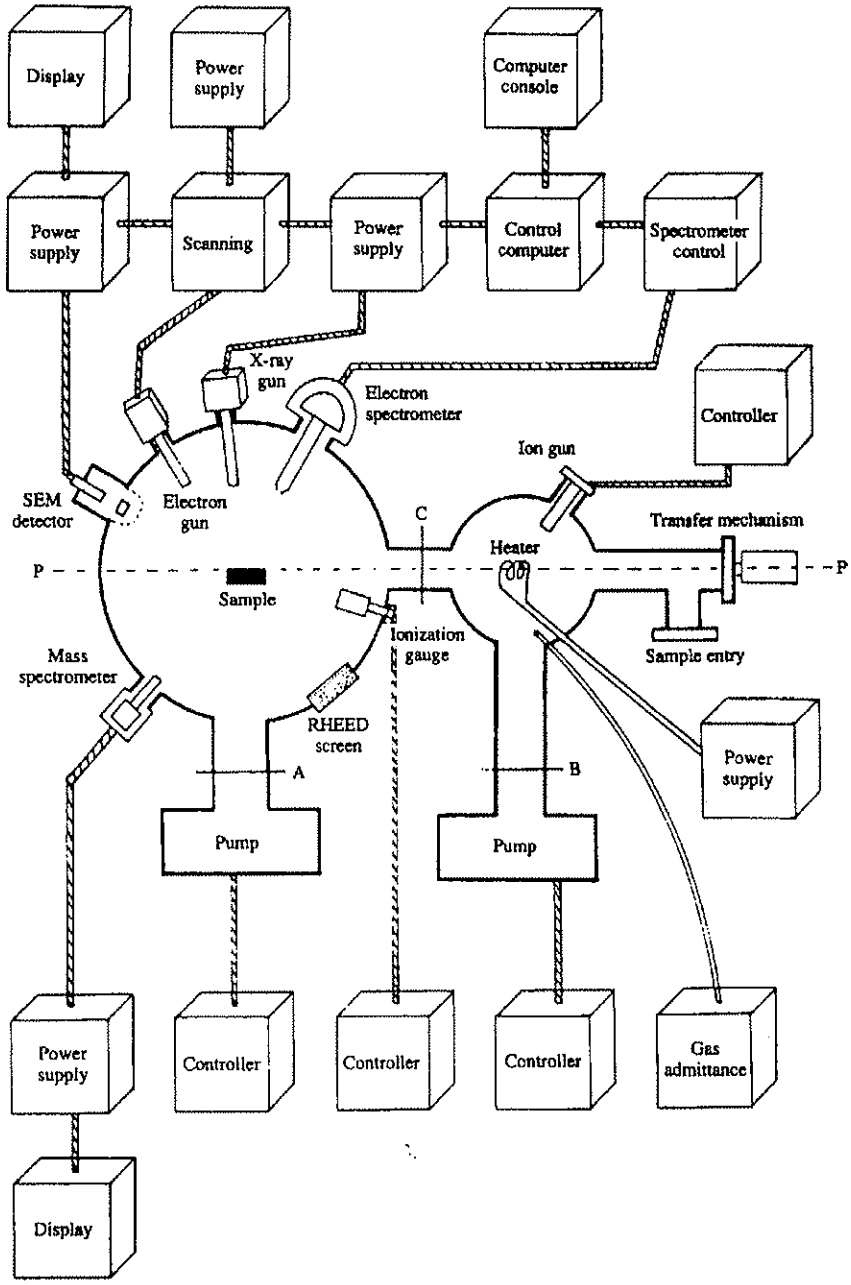



Fig. 6 Arrangement of vacuum components in a typical UHV-system for surface studies [9].

Vacuum is made by removing the various gases from the contained volume. This is done by many different kinds of vacuum pumps, which either remove the particles from the volume by real pumping or trap them by binding them via physical or chemical forces, changing their form and keeping them trapped in the bulk of the pumping material, a so-called getter. This is done by various mechanical and sorption pumps, diffusion pumps, cryo-pumps, ion-sputter pumps, non-evaporable-getter(NEG) pumps or titanium sublimation pumps. As an example for the tremendous development of mechanical pumps during the second half of the 20$^{th}$ century, in particular of fast rotating turbo-molecular pumps, Fig. 7 shows the so-called dry (oilfree) wide-range type of such a pump, where the rotor is kept in position by magnetic bearings. Not only pumps for vacuum generation, but also a wide variety of devices for pressure measurements, rest-gas analysis and mass spectrometry have been developed for the various ranges of vacuum. All these highly sophisticated devices for vacuum generation and measurement and the corresponding materials and cleaning techniques form the basis for modern vacuum technology and will be discussed in detail in the following chapters.

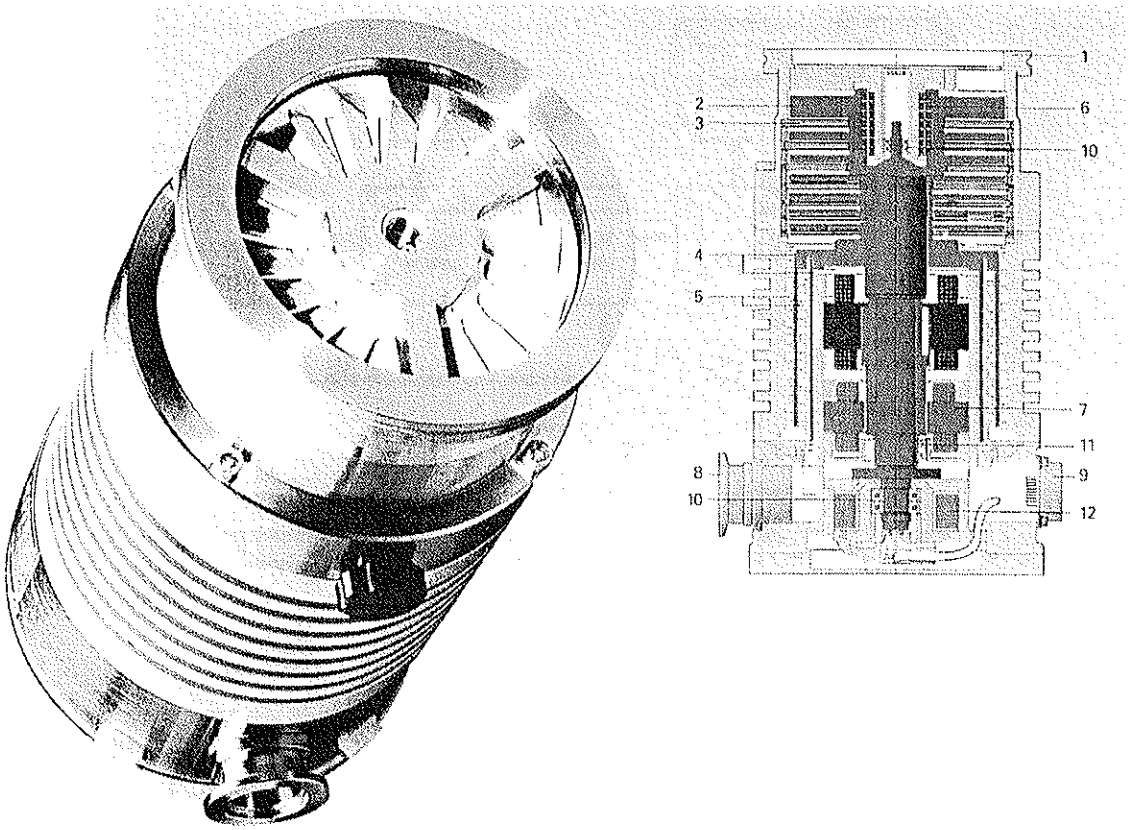



Fig. 7 Magnetic-bearing wide-range turbomolecular pump (from Pfeiffer-Balzers Co.).

There are many people who believe that vacuum technology is just dealing with valves, flanges and other vacuum components. But the science and technology that all this hardware enables are the keys to our high-tech society. Vacuum is used for a huge number of technical processes and products, like vacuum-packed coffee, thermos flasks, light bulbs, microwave and cathode-ray tubes for TV screens. One uses it always when one needs a "clean" space, which is free of gases and contaminants. It led to the development of ion-getter pumps and ultra-high-vacuum (UHV) techniques, to electron and ion microscopes, ion implanters, sensitive leak detectors and space simulation technology. Because its conductivity is lower than any other material, vacuum is best suited for thermal insulation, i.e. to reduce energy transfer or to prevent heat loss. It is used for thermal insulation at low temperatures (below 100 K), in particular. On the other hand, low temperatures are used to generate ultra-high vacuum by cryo-pumping. Contamination by gas particles can disturb or totally prevent processes or spoil products. Therefore, vacuum technology and equipment was essential to the development of electron and microwave tubes, the production of high-purity metals, the separation and storage of gases, any plant producing computer chips and semi-conductor components, displays, thin films, nanostructures, optical and microelectronic coatings, vacuum metallurgy, data storage and hard disk drives. Vacuum technology is growing in importance as more industries introduce advanced materials and devices and as technology goes more and more into the direction of components of micro- and nanometer scale.

Ultra-high vacuum is also of particular importance for chemical and biological techniques, for atomic-physics research and for the operation of modern particle accelerators, used for high-energy elementary-particle physics or synchrotron-radiation research. The development of the so-called 3$^{rd}$ generation synchrotron light sources with their tremendous fluxes of photons of all wavelength, used nowadays in many areas of natural science, would have been impossible without UHV-technology. A comparison of the brightness, i.e. the number of emitted photons in appropriate units, of a simple candle with the brightness of these modern machines, using magnetic undulators for its amplification, is presented in Fig. 8 . The graph shows that synchrotron radiation is about 10 orders of magnitude brighter than the radiation of the sun, which emits the huge amount of energy of ~41 ergs per year. A

supernova, however, is radiating in much shorter times more than one order of magnitude more energy. Figure 9 shows how this enormous amount of radiation energy is transferred into the kinetic energy of high-speed acceleration of gas, expanding in all directions of the extreme-high-vacuum of empty space.

For a more detailed study of vacuum physics, a list of a few general textbooks on high-vacuum physics is given at the end of this article.

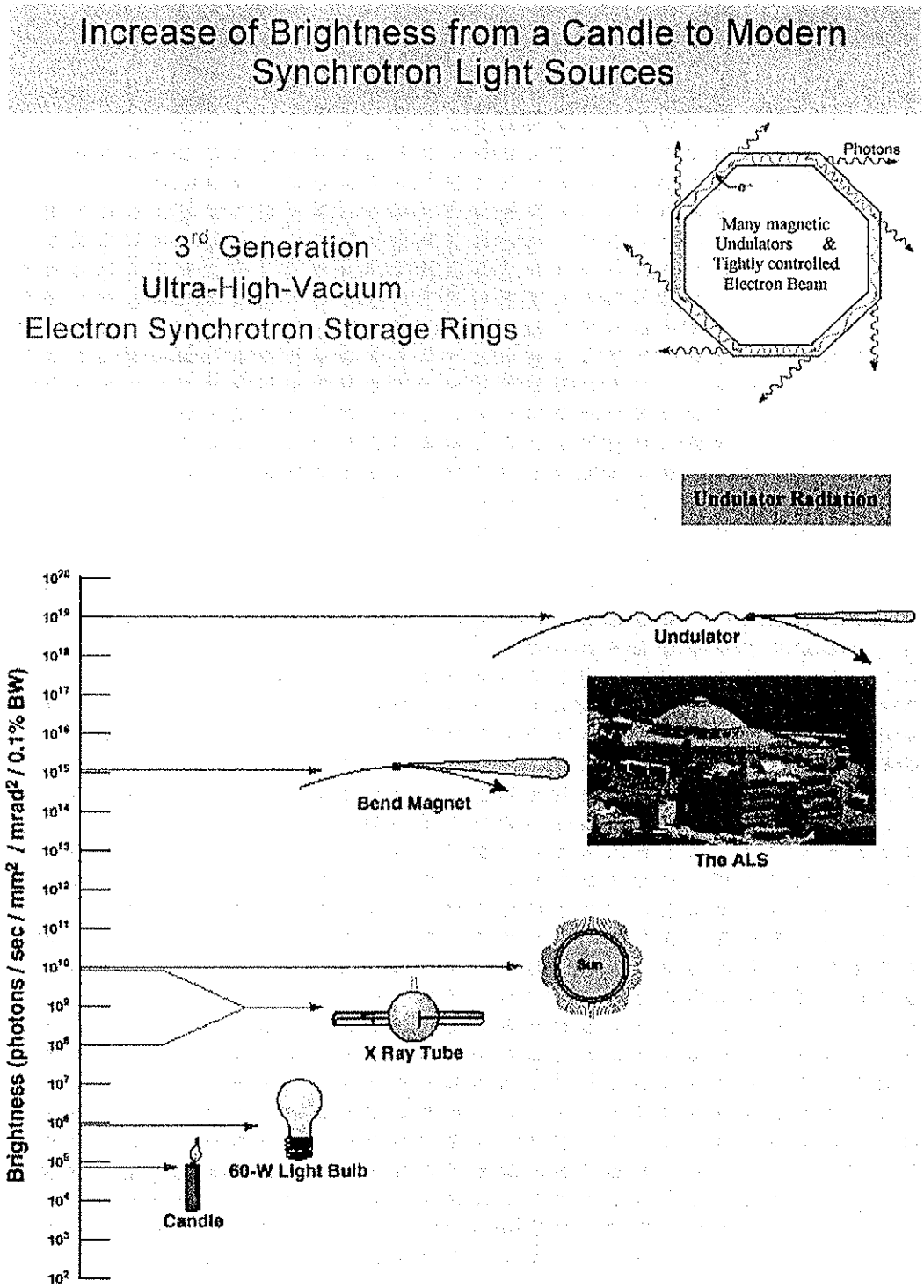



Fig. 8 Comparison of photon intensities emitted by various artificial light sources with the sun. (Courtesy Advanced Light Source, Lawrence Berkeley National Laboratory.)

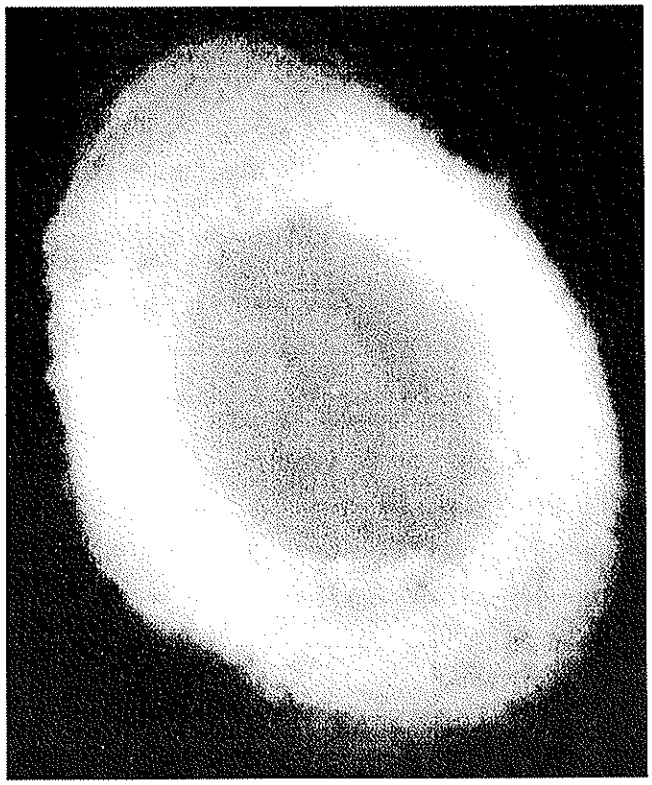


Fig. 9 A ring nebula observed by the Hubble Space Telescope (NASA/ESA, taken from CERN Courier, Vol.39, No.4, May 1999, p.10). The hot gas thrown into space by a supernova explosion travels at more than 100 km/s.

## 2. CONVERSION FACTORS FOR PRESSURE AND PRESSURE RANGES OF VACUUM, VACUUM GAUGES AND PUMPS

By *pressure* p one means the average force $F_\perp$ in Newton (or the rate of normal momentum transfer) exerted by gas molecules impacting on a surface of unit area $A$ (in $m^2$). Therefore, pressure can be expressed (in so-called "coherent" pressure units) as:

$$p = F_\perp / A \quad \text{(Pa)} \quad \text{(see Table 1).}$$

If there is a flow of gas, one has to distinguish between dynamic and static pressure, depending whether the gauge reading the pressure is stationary or moves with the same velocity as the flow. Similarly, one speaks of a steady-state pressure, if the pressure at different locations in a vacuum systems remains constant with time. Derived from the definition above, pressure can be expressed also (in "non-coherent" pressure units) as the height $h$ of a liquid with density $\rho$ (which is mercury for the unit Torr) of a manometer column under the acceleration $g = 9.806$ m $s^{-2}$ due to gravity:

$$p = h\, g\, \rho \text{ (Torr).}$$

It can also be expressed by the number density of molecules $n$, the Boltzmann constant $k$ and the temperature $T$ :

$$p = n\, k\, T \text{ (Pa)}$$

or by $n$, the mass per molecule m, the mean-square velocity $v_{ms}^2$ of molecules and the gas density $\rho$ :

$$p = (1/3)\, n\, m\, v_{ms}^2 \text{ (µbar)}$$

$$p = (1/3)\, \rho\, v_{ms}^2 \quad \text{(Torr).}$$

All the quantities used here are further explained in the following sections. As one can see, the gas pressure does not depend on the gas species. The conversion factors for pressure in various systems of units are given in Table 3.

Table 3
Conversion factors for pressure in various systems of units [8].

| | Pa ($N\ m^{-2}$) | mbar | Torr (mm Hg at 0 °C) | Technical Atmospheres (at) | Physical Atmospheres (atm) |
|---|---|---|---|---|---|
| Pa ($N\ m^{-2}$) | 1 | $1.0 \times 10^{-2}$ | $7.5 \times 10^{-3}$ | $1.02 \times 10^{-5}$ | $9.87 \times 10^{-6}$ |
| mbar | $1.0 \times 10^{2}$ | 1 | $7.5 \times 10^{-1}$ | $1.02 \times 10^{-3}$ | $9.87 \times 10^{-4}$ |
| Torr (mm Hg at 0 °C) | $1.33 \times 10^{2}$ | 1.33 | 1 | $1.36 \times 10^{-3}$ | $1.32 \times 10^{-3}$ |
| Technical Atmospheres (at) | $9.80 \times 10^{4}$ | $9.80 \times 10^{2}$ | $7.36 \times 10^{2}$ | 1 | $9.68 \times 10^{-1}$ |
| Physical Atm.(atm) | $1.01 \times 10^{5}$ | $1.01 \times 10^{2}$ | $7.60 \times 10^{2}$ | 1.03 | 1 |

The ultimate pressure of a system designates the limiting pressure approached in a vacuum system after pumping long enough to demonstrate that a further reduction in pressure is negligible. Similarly, the ultimate pressure of a vacuum pump is the limiting low pressure of a pump due to leaks or gases trapped in the various components of the pump. *Pumping speed* means the ratio of the throughput of a given gas to the partial pressure of that gas near the pump intake. The throughput or the amount of gas in pressure-volume units flowing per unit time at a fixed temperature across a specified cross section near the pump intake is often referred to as *pumping capacity*.

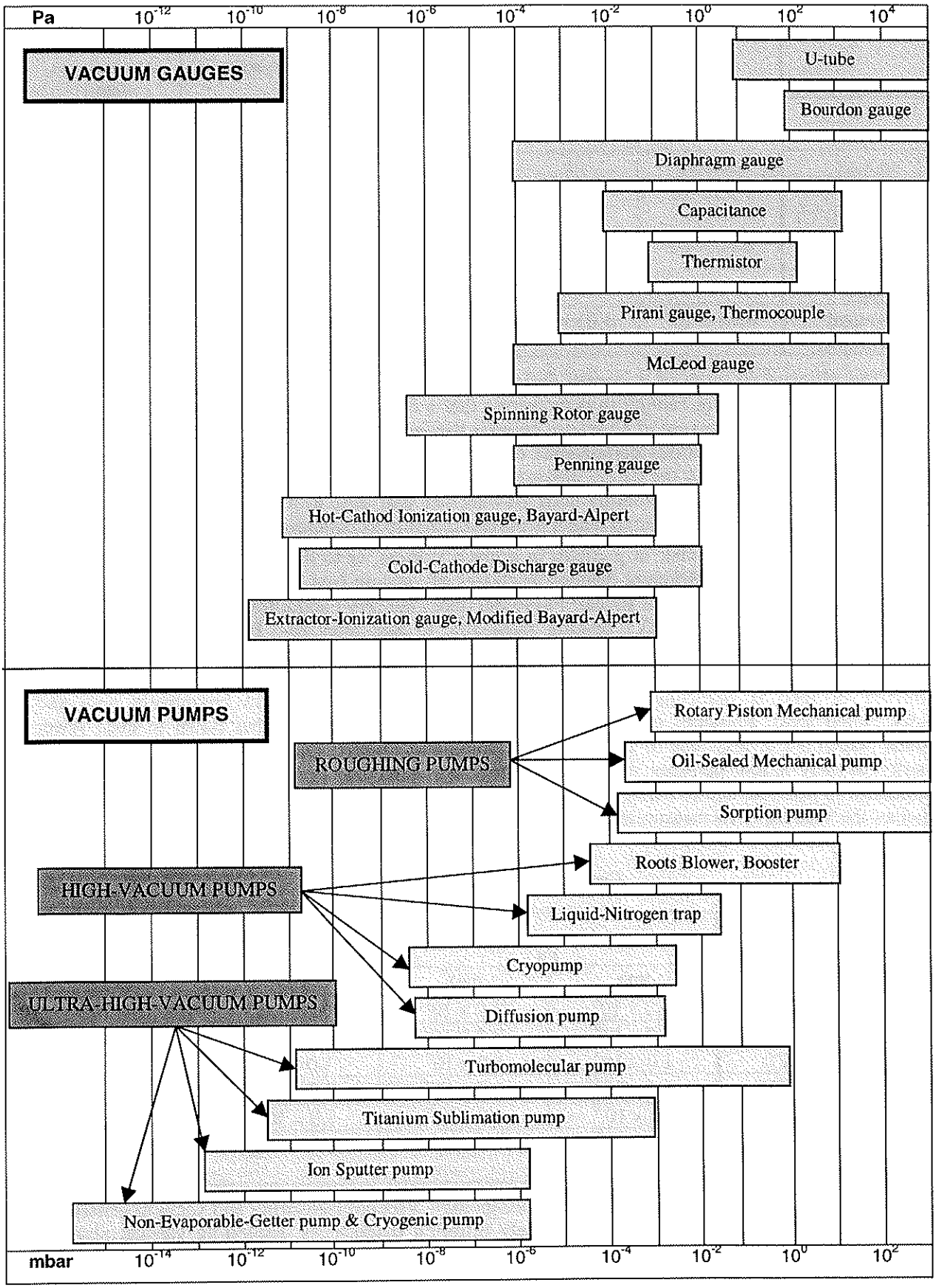
According to the number of particles in a given volume, or to the various regions of pressure correspondingly, there are different degrees or classes of vacuum (see Table 4), ranging from rough vacuum of some ten thousands of Pascal (Pa) to the fine, high (HV) and ultra-high vacuum (UHV) to the extreme high vacuum (XHV) ranging from $10^{-10}$ to less than $10^{-12}$ Pa (almost to $10^{-14}$ Pa). Accordingly, vacuum covers a range of almost 19 orders of magnitude. As always in physics, to reach the extreme values of vacuum, either the generation or measurement of pressures of about 1/100 of a femtobar or less, is particularly attractive, because it pushes science to completely new and undiscovered areas. Pressures measured in space so far, are reaching further down to even lower values, which are at least one or two orders of magnitude smaller. Such pressures correspond to a density of only a few particles per cubic centimetre. Clearly, it is a real challenge for modern techniques to generate and to measure absolutely such low pressures, and it is not obvious at present whether the vacuum or its measurement is better.

Table 4
Classification of vacuum ranges [8].

| Vacuum Ranges | Pressure Units | | | |
|---|---|---|---|---|
| | Pa | | mbar | |
| | min | max | min | max |
| Low (LV) | $3.3 \times 10^{3}$ | $1.0 \times 10^{5}$ | $3.3 \times 10$ | $1.0 \times 10^{3}$ |
| Medium (MV) | $1.0 \times 10^{-1}$ | $3.3 \times 10^{3}$ | $1.0 \times 10^{-3}$ | $3.3 \times 10$ |
| High (HV) | $1.0 \times 10^{-4}$ | $1.0 \times 10^{-1}$ | $1.0 \times 10^{-6}$ | $1.0 \times 10^{-3}$ |
| Very High (VHV) | $1.0 \times 10^{-7}$ | $1.0 \times 10^{-4}$ | $1.0 \times 10^{-9}$ | $1.0 \times 10^{-6}$ |
| Ultra-High (UHV) | $1.0 \times 10^{-10}$ | $1.0 \times 10^{-7}$ | $1.0 \times 10^{-12}$ | $1.0 \times 10^{-9}$ |
| Extreme Ultra-High (XHV) | $\leq 1.0 \times 10^{-10}$ | | $\leq 1.0 \times 10^{-12}$ | |

According to this wide range of vacuum technology, vacuum measuring and vacuum generating techniques have had to be developed for widely differing magnitudes, exploiting various physical properties of gases. Table 5 shows the pressure ranges of the huge variety of vacuum gauges and vacuum pumps, to be discussed in the following lectures.

Table 5
Pressure ranges of vacuum gauges and vacuum pumps.



## 3. COMPOSITION OF AIR

Whereas atmospheric air is a mixture of gases with over 99% nitrogen and oxygen (see Table 6, for instance) the rest gas of high and ultrahigh vacuum consists mainly of the lightest gas hydrogen. The reason is that hydrogen easily penetrates walls, is adsorbed by all surfaces surrounding the volume and is less effectively pumped than gases of larger active molecules.

Table 6
Composition of atmosphere (at sea level) [8].

| Gas | Percent by Volume | Partial Pressure (mbar) |
|---|---|---|
| $N_2$ | 78.08 | $7.93 \times 10^{2}$ |
| $O_2$ | 20.95 | $2.12 \times 10^{2}$ |
| Ar | 0.93 | 9.39 |
| $CO_2$ | $3.3 \times 10^{-2}$ | $3.33 \times 10^{-1}$ |
| Ne | $1.8 \times 10^{-3}$ | $1.87 \times 10^{-2}$ |
| He | $5.24 \times 10^{-4}$ | $5.33 \times 10^{-3}$ |
| Kr | $1.1 \times 10^{-4}$ | $1.11 \times 10^{-3}$ |
| $H_2$ | $5.0 \times 10^{-5}$ | $5.06 \times 10^{-4}$ |
| Xe | $8.7 \times 10^{-6}$ | $8.79 \times 10^{-5}$ |
| $H_2O$ | 1.57 | $1.72 \times 10$ |
| $CH_4$ | $2.0 \times 10^{-4}$ | $1.99 \times 10^{-3}$ |
| $O_3$ | $7.0 \times 10^{-6}$ | $7.06 \times 10^{-5}$ |
| $N_2O$ | $5.0 \times 10^{-5}$ | $5.06 \times 10^{-4}$ |

## 4. IDEAL AND REAL GASES

### 4.1 Ideal gases

The theoretical concept of an ideal gas is very useful for the description of vacuum-physics problems. It assumes that (i) the molecules are small spheres, (ii) the gas density is relatively small, i.e. the volume of molecules is small compared to the total volume occupied by the gas, (iii) the molecules do not exert forces on each other, i.e. the temperature of the gas is not too low, (iv) the path travelled by molecules is linear and random and (v) collisions between molecules are purely elastic. Under these conditions, which are the normal ones, the transition from one state of a gas, specified by its pressure $p_1$, volume $V_1$ and thermodynamic temperature $T_1$ (measured in degrees Kelvin) to another state with quantities $p_2$, $V_2$, $T_2$ can be described by *the Boyle-Mariotte law for an ideal gas :*

$$\boxed{p_1V_1/T_1 = p_2V_2/T_2 = \text{const}}$$

This means that for a given gas $pV/T$ is a constant value. Under normal conditions this is best fulfilled for $H_2$ and He. The constant depends for a given temperature only on the number of molecules contained in the volume, i.e. from the amount of gas enclosed.

For a given mass of a gas, three fundamental relations can be derived from Boyle-Mariotte´s law, namely:

| *Boyle´s law* | *Charle´s law* | *Gay-Lussac´s law* |
|---|---|---|
| $T = \text{const}$ | $p = \text{const}$ | $V = \text{const}$ |
| $p_1V_1 = p_2V_2 = \text{const}$ | $V_1/T_1 = V_2/T_2 = \text{const}$ | $p_1/T_1 = p_2/T_2 = \text{const}$ |
| (isothermal phase transition) | (isobar phase transition) | (isochore phase transition). |

For the case of thermal insulation one has $Q$ = const , with $dQ = 0$ and the $(p,V)$-relation : $pV^{\kappa}$ = const , corresponding to an adiabatic phase transition ($dQ = dU + pdV$ with the specific heat capacities at $p,V$ = const).

It was observed that by relating $pV/T$ to the concept of mole, its value is the same for any ideal gas. A mole is the weight in grams equal numerically to the molecular weight of a substance. For instance: 1 mole $O_2$, $H_2$, $H_2O$ = 32 g, 2.016 g, 18.016 g respectively. Avogadro found that the mass $m$ (of a standard volume of gas at the same temperature and pressure) is proportional to its molar mass (molecular weight) $M$. In other words the amount of moles $n_M$ in a given mass $m$ of a substance having a molar mass $M$ is: $n_M = m/M$. It was shown by experiments that under standard conditions of temperature (0 $^0$C or 273.16 K) and pressure ($1.013250 \times 10^5$ Pa or 760 Torr) one mole (or gram molecular weight) of any gas occupies a volume of 22.4 liter (= 22 415 cm$^3$). It is: $pV = 22.4\ m/M$, with $m$: mass of gas, $M$: molar mass, $V$ in liters and $p$ in Pa, where the pressure $p$ is proportional to the gas density $\rho = m/V$.

For a gas mixture, the total pressure is the sum of all partial pressures: $p = p_1+p_2+...+p_i = \Sigma\, p_i$ (Dalton´s law) and the mean molar mass is: $\langle M\rangle = (m_1+m_2+...+m_i) / (n_{m1}+n_{m2}+...+n_{mi})$ (in g/mol).

The number density n of particles (atoms, molecules etc.) is given by:

$$n = N/V = (N_A/R)\ (p/T) = n_M\ (N_A/V) = (m/M)\ (N_A/V) \quad \text{(particles/m}^3\text{)},$$

where $N$ is the total number of particles (molecules) and $V$ the volume. Under the same conditions of pressure and temperature, equal volumes of all gases have the same number of particles (Avogadro´s law, L.R.A.C. Avogadro : 1776 - 1856). This volume is called a mole, and the number of particles in one mol of a substance is given by Avogadro´s number $N_A$. By independent measurements of the mass of the hydrogen atom (particle mass) of $m_p = 1.67 \times 10^{-24}$ g and the molecular weight of $H_2$: $M$ = 2.016 g, one can calculate Avogadro´s number to be :

$$N_A = M/m_p = 2.016 / (2 \times 1.67 \times 10^{-24}) = F/e = 6.0221367 \times 10^{23} \text{ (molecules/mol)}.$$

One gets the number $N_A$ independently also from the quotient of the value of the Faraday F (defined as the electrical charge necessary to deposit electrolytically one mole of a substance) and the charge $e$ of an electron. One mole of any gas has $N_A$ particles and under standard conditions occupies 22.4 liters. From these values one can calculate that in a standard liter there are:

$$(6.022 \times 10^{23} \text{ particles}) / 22.4 \text{ liters} = 2.69 \times 10^{22} \text{ particles/liter}\ .$$

R is the *universal (or molar) gas constant*:

$$R = 8.314 \text{ J mol}^{-1} \text{ K}^{-1} \text{ (with } p \text{ in Pa, } V \text{ in m}^3\text{, } T \text{ in K)}.$$

Whereas the universal gas constant R is related to moles, *Boltzmann´s constant* k is calculated on a molecular basis:

$$k = R/N_A = 1.3806 \times 10^{-16} \text{ erg/K} = 1.3806 \times 10^{-23} \text{ J/K}, \text{ (with 1 J} = 1.0 \times 10^7 \text{ erg} = 2.388 \times 10^{-1} \text{ cal)}.$$

From this the equation of state of an ideal gas is derived, which describes how the measurable quantities of a gas depend on each other:

$$\boxed{pV = n_M RT = (m/M)\ RT}$$

Again, from Dalton´s law one gets for a mixture of gases: $pV_M = (\Sigma n_{Mi})RT$, with total volume $V_M = \Sigma V_i$ and total pressure $p = \Sigma p_i$ , if all components have the same pressure.

With $n = N/V$ (molecules/m$^{-3}$) and $m = M/N_A$ one obtains for the gas density: $\rho = mn = pM/RT = pm/kT = M/V_M$. The density of gas depends on the two parameters temperature and pressure, which both must be given, if the density is quoted.

The pressure $p$ exerted by a gas depends only on the number density of molecules n = $N/V$ and temperature $T$ of the gas, not on the particular gas. It can be expressed as the force F it exerts on a unit area $A$ or as the number density of molecules n = $N/V$, Boltzmann´s constant $k = R/N_A$ and the temperature $T$ (see: section 2).

### 4.2 Real gases

At low temperatures or high pressures a gas approaches the points of condensation or sublimation, where it changes its phase from the gaseous to the liquid or even solid one. Such a gas is called a real gas (or vapour). Due to the action of interatomar or intermolecular forces under these conditions, the $p$-$V$ curve of a real gas deviates from that of an ideal gas, exhibiting a flat plateau, which corresponds to the liquid-gas phase transition. Near the critical point (critical temperature and pressure) the behaviour of real gases can be described very satisfactorily by a modified equation of state, *the Van der Waals equation* (J.D. van der Waals: 1837 -1923):

$$(p + n_M^2\, a/V^2)\,(V - n_M b) = n_M\, RT\,, \quad (\text{for } n_M = m/M > 1 \text{ mol})$$

The correction term $n_M^2\, a/V^2 \sim (m^2/V^2) = \rho^2$ with $a$ = const takes the forces of intermolecular attraction into account, which brings molecules closer together, acting like an additional pressure. The volume-correction term $n_M b$ is a measure of the actual volume of all molecules of their own, which is four times that of the molecule itself and increases with the molecular diameter. Whereas the isotherms of an ideal gas are hyperbolas, those of a real gas at sufficiently low temperatures exhibit first a maximum and then a minimum with decreasing volume (see Fig. 10). Normally, at critical values of temperature $T_c$, volume $V_c$ and pressure $p_c$ these extrema merge at a critical point into a curve with horizontal turning tangent. The values of $a$ and $b$ are determined for most gases as a function of the critical point ($T_c$, $V_c$, $p_c$) of the corresponding gas ($a = 3\, p_c\, (V_c/n_M)^2 \sim \rho_c^{-2}$ ; $b = 3\, V_c/n_M$) and are tabulated. The limiting temperature $T_B$ (Boyle's point) up to which a gas can still be described like an ideal one, even up to high pressures, is given by $T_B = a/(bR)$.

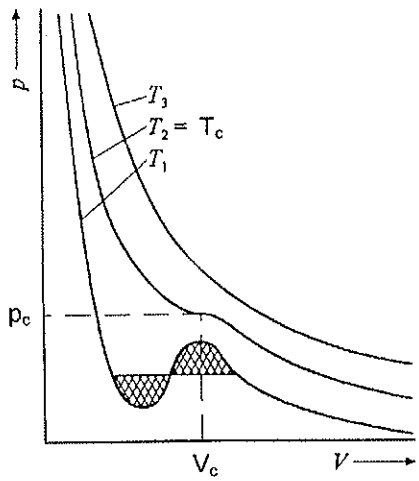



Fig. 10 Isotherms of a real gas according to van der Waals.

## 5. KINETIC THEORY OF GASES

### 5.1 Molecular velocities and temperature

Gas molecules are in constant, irregular thermal motion. The velocity of this motion depends on the molar mass $M$ and temperature $T$ only. The higher the molar mass, the lower is the velocity, for a given temperature. Concerning molecular velocities, one distinguishes between the most probable velocity $v_p$, the arithmetic average velocity $v_a$ (e.g. of a flow of gas), the mean-square velocity $v_{ms}$ and the root-mean-square velocity $v_{rms}$ (normally used, when the kinetic energy of the molecule influences the process), which are simply related in the SI-system of units by :

$$v_{rms} = (v_{ms})^{1/2} = (3RT/M)^{1/2} = (3kT/m)^{1/2}\,, \quad (\text{m s}^{-1})\ , \text{ with}$$

$$v_p = 129\,(T/M)^{1/2} \quad < \quad v_a = 146\,(T/M)^{1/2} \quad < \quad v_{rms} = 158\,(T/M)^{1/2}\ , \quad (\text{m s}^{-1})\ .$$

These velocities are independent of gas pressure and, as can be seen from Table 7, they are greater than the speed of sound of ~330 m/s in air.

Table 7
The root-mean-square velocity $v_{rms}$ of different gas molecules at 0 and 23 °C [8].

| Gas / T (K) | $v_{rms}$ (m s$^{-1}$) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | monoatomic | | | | | diatomic | | | | | polyatomic | | |
| | He | Ne | Ar | Kr | Xe | $H_2$ | $N_2$ | CO | $O_2$ | Air | $NH_3$ | $C_2H_6$ | $CO_2$ |
| 273 | 1305 | 581 | 413 | 285 | 228 | 1838 | 512 | 493 | 461 | 485 | 632 | 511 | 393 |
| 296 | 1358 | 605 | 430 | 297 | 237 | 1914 | 533 | 514 | 480 | 505 | 659 | 533 | 410 |

In contrast to the velocity, the concept of temperature of a gas cannot easily be associated with a single molecule. Temperature is understood in terms of the kinetic energy and velocity of motion. If the air, surrounding us, has a molecular velocity below 400 m/s, it appears to us too cold, if it is above 500 m/s it is too hot.

## 5.2 Velocity distribution and mean free path

The constant motion of molecules and their collisions produce a wide distribution of velocities, from about zero to very large values, which is expressed by the *Maxwell-Boltzmann law* (J.C. Maxwell: 1831 - 1879; L. Boltzmann: 1844 - 1906):

$$(1/n)\, dn/dv = f_v = (4/\pi)^{1/2}\, (m/2kT)^{3/2}\, v^2 \exp(-mv^2/2kT)$$

where $dn$ is the number of particles in the velocity range between $v$ and $v + dv$ (per unit of velocity range) and $f_v$ is the fractional number of particles between $v$ and $v + dv$, which is zero for $v = 0$ and $v = \infty$ and has its maximum at the most probable velocity: $v_p = (2kT/m)^{1/2} = (2RT/M)^{1/2} = 0.816\ v_{rms}$. In Fig. 11 $f_v$ is plotted versus $v/v_p$

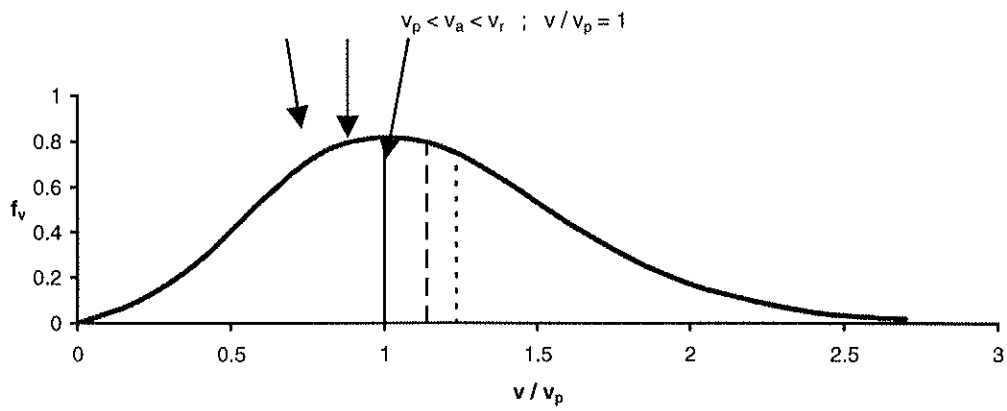



Fig. 11 Maxwell-Boltzmann molecular velocity distribution for a gas at a particular temperature $T$.

The average distance travelled by all gas molecules between successive collisions is called the *mean free path* $\lambda$. It depends on the size and the number of the molecules, i.e. on gas density and pressure. Applying the kinetic theory of gases by using the Maxwell-Boltzmann distribution and the relation $p = nkT$, the average mean free path is given in m by:

$$\lambda = 1/(2^{1/2}\, \pi\, d^2\, n) = kT/(2^{1/2}\, \pi\, d^2\, p) = 3.107 \times 10^{-24}\, T/(d^2\, p)\,,$$

where $d$(m) is the molecular diameter, $n = N/V$ ($m^{-3}$) is the number of molecules per volume, $T$ (K) is the temperature and $p$ (Pa) is the pressure. Accordingly, $\lambda$ is proportional to the absolute temperature and inversely proportional to the gas pressure.

For air (assuming that all molecules have the same $d$ at room temperature (23 $^0$C), one gets: $\lambda_{Air}$(m) = 6.5 × $10^{-3}/p$ (Pa) . For common gases, the range of variation of $\lambda$ is about a factor of 10. When the mean free path becomes greater than the dimensions of the vacuum vessel, the collisions become less frequent and the molecules mostly collide with the walls of the vessel.

The mean free path was derived for pure elastic collisions between the molecules. Taking into account the attractive forces between the molecules, one found empirically that at very high temperature $T_\infty$ the mean free path is: $\lambda_T = \lambda_\infty / (1 + C/T)$, (with n, C = const).

It is interesting to compare the mean free path of electrons $\lambda_e$ (cm) and ions $\lambda_{Ion}$ (cm) in a gas with that of neutral molecules $\lambda$ (cm). One finds approximately: $\lambda_{Ion} = 2^{1/2}\ \lambda = \lambda_e/4$. Values of mean free path for different pressures or particle densities, respectively, are given below in Table 8.

Table 8

Rough values for characteristic vacuum parameters at a few pressures and gas flow regimes.

| | Pressure | | | | |
|---|---|---|---|---|---|
| | $10^5$ Pa<br>$10^3$ mbar | $10^2$ Pa<br>$10^0$ mbar | $10^{-1}$ Pa<br>$10^{-3}$ mbar | $10^{-4}$ Pa<br>$10^{-6}$ mbar | $10^{-7}$ Pa<br>$10^{-9}$ mbar |
| Particle Density, n ($cm^{-3}$) | $10^{19}$ | $10^{16}$ | $10^{13}$ | $10^{10}$ | $10^7$ |
| Mean Free Path, $\lambda$ (cm) | $10^{-5}$ | $10^{-2}$ | 10 | $10^4$ | $10^7$<br>(= 100 km) |
| Impingement Rate, $Z_A$ ($s^{-1}$ $cm^{-2}$) | $10^{23}$ | $10^{20}$ | $10^{17}$ | $10^{14}$ | $10^{11}$ |
| Collision Rate, $Z_V$ ($s^{-1}$ $cm^{-3}$) | $10^{29}$ | $10^{23}$ | $10^{17}$ | $10^{11}$ | $10^5$ |
| Monolayer Time, $\tau$ | 10 ns | 10 µs | 10 ms | 10 s | 3 h |
| **Type of Gas Flow** ← | viscous | Knudsen | | molecular → | |

From the Maxwell-Boltzmann equation given above, one can derive the number of molecules striking an element of surface perpendicular to their direction of movement, per unit time, the so-called *impingement rate* $Z_A$ as:

$$Z_A = 8.333 \times 10^{22}\ p/(TM)^{1/2} \text{ (molecules m}^{-2}\text{ s}^{-1}\text{)}, \text{ with } p \text{ (Pa), } T \text{ (K) and } M \text{ (kg/mol),}$$

$$= 2.635 \times 10^{22}\ p/(TM)^{1/2} \text{ (molecules cm}^{-2}\text{ s}^{-1}\text{)}, \text{ with } p \text{ (mbar), } T \text{ (K) and } M \text{ (g/mol).}$$

Similarly one can calculate the *collision rate* ZV (in molecules $cm^{-3}$ $s^{-1}$), which is the number of collisions in a unit volume per unit time.

For ultra-high-vacuum systems there is another quantity of interest, namely the *monolayer time* $\tau$, which is defined as the time required for a monolayer to form on a clean, gas-free surface. To derive $\tau$, it is assumed that every impinging gas molecule finds and binds to a vacant site (sticking coefficient ≡ 1). The monolayer time $\tau$, in seconds, is inversely proportional to the pressure and can be estimated from:

$$\tau = 3.2 \times 10^{-6}/p \text{ , with } p \text{ in mbar.}$$

Approximate values for particle density n , mean free path $\lambda$, impingement and collision rate $Z_A$ and $Z_V$ and monolayer time $\tau$ are given for different pressures in Table 8, where also the relevant types of gas flow are indicated (see also section below).

## 6 THERMAL CONDUCTIVITY AND ENERGY TRANSPORT

So far, all systems discussed have been in thermodynamical equilibrium. If there is a temperature difference in the system, a heat transfer will be observed, from which the specific thermal conductivity of the gas can be determined. Similarly, if differences in pressure or particle density exist, one has a flow of particles from positions of high pressure to low-pressure regions. Pressure differences lead to velocity distributions, which correspond to the transport of particles and are connected to the viscosity of a gas or its coefficient of internal friction of a gas. The *coefficient of viscosity* is given as:

$$\eta = 0.499\ n\ m\ v_a\ \lambda\ ,$$

with the gas density $\rho = n\ m$ , $v_a$ = average molecular velocity and $\lambda$ = mean free path.

The specific heat conductivity of a gas is proportional to its coefficient of viscosity. Therefore, they have the same dependence on pressure. Differences in particle density in a system lead to mass transport, called diffusion.

The mean free path of gas particles is related to both, the thermal conductivity and the viscosity of gases. Concerning heat conductivity and other phenomena of energy transport, one must distinguish between low and high pressures and, correspondingly, between different states of a gas or types of gas flow, which will be discussed in the following section. Depending on the state of the gas, transport phenomena behave differently. Above approximately 1.3 kPa (= 13.3 mbar = 10 Torr) the heat transfer through a gas inside a small chamber is dominated by convection due to the bulk motion of the gas. In other words, in this so-called *viscous state* of a gas the totality of molecules is responsible for the heat transfer. In this range of gas flow the heat conductivity of gases is independent of pressure, which is also true for the viscosity. At lower pressures, when the mean free path of molecules $\lambda$ becomes much greater than the dimensions of the container, which is called *molecular state*, the individual molecules carry the heat from wall to wall. Then the gas is no longer characterised by a viscosity. Under this low-pressure condition of so-called free molecular conduction the heat transfer depends only on the number of carriers. Thus the rate of energy transfer, the conductivity, is proportional to the pressure and the temperature difference.

Both, the variation of conductivity and of viscosity with pressure are utilised for pressure measurements, in the first case by monitoring the heat transfer rate from a hot filament to the surrounding walls (thermal conductivity gauges), in the second case by measuring the deceleration of a freely rotating body, a metallic sphere suspended in a permanent magnetic field, due to collisions with the rest gas (viscosity manometers, spinning rotor gauge). Heat transfer in high-vacuum is very low compared to conditions at atmospheric pressure. Therefore, there is very little heat flow for instance between two metal plates in high vacuum. Heat transfer by conduction takes place only through the small areas of contact, if not high clamping forces are used to enlarge the touching area.

# 7 GAS FLOW

## 7.1 Gas flow regimes and transport phenomena

When a system is pumped down from atmospheric pressure to high vacuum, the gas in the system goes through various regimes of gas flow. These regimes depend on the size of vacuum components, the gas species and the temperature. One distinguishes the *viscous state* at high pressures from the *molecular state* at low pressures and an *intermediate state* between these two. At atmospheric pressure up to about 100 Pa (= 1 mbar), the mean free path of the gas molecules is very small (see Table 8). Therefore, the gas flow is limited by viscosity and the type of flow is called *viscous*. This type of gas flow is characterised by intermolecular collisions that impart to the gas the appearance of a viscous fluid. At low pressures, however, where the mean free path is similar to the dimensions of the vacuum vessel containing the gas, its flow is governed by viscosity as well as by molecular phenomena. This type of gas flow is called intermediate or sometimes *Knudsen regime*. To specify the types of gas flow, the dimensionless *Knudsen number* (1910, M. Knudsen: 1871-1949) has been defined as

$$K = \lambda / d,$$

with the characteristic dimension of the vessel or the diameter $d$ of a cylindrical pipe (K is proportional to the ratio of Mach and Reynolds number, see below).

K<<1 holds for the viscous flow regime with the characteristics of a continuous fluid. At very low pressures, where the mean free path is much larger than the dimensions of the vacuum enclosure (K>>1) , the flow is molecular. In this state, the molecules interact mostly with the walls of the vessel not with themselves (rarefied "Knudsen gas").

In the viscous state of a gas, one distinguishes further between *turbulent* and *laminar* flow. When the velocity of the gas flow exceeds certain values, the flowing gas layers are not parallel and cavities are formed between them. Then the viscous flow is turbulent. On the other hand, at lower

velocities the flow is laminar, i.e. the layers are moving parallel, forming e.g. in a pipe a parabolic velocity distribution with increasing velocities from the wall towards the axis of the pipe. Whereas the limit between laminar, intermediate and molecular flow is defined by the value of the Knudsen number, the limit between turbulent and laminar flow with the characteristics of a continuous fluid is described by the value of another dimensionless quantity, the *Reynolds number* (O. Reynolds: 1842-1912), which is expressed by

$$R = d\, v\, \rho/\eta$$

the ratio of kinetic energy and frictional work, with $d$ = diameter of the pipe or characteristic size of the vacuum component, $v$ = gas velocity and $\rho$, $\eta$ the density and viscosity of the gas at the temperature of the flow.

By evaluating the Knudsen and Reynolds number, one can predict the various flow regimes, which can be distinguished approximately by the following relations :

| | | | |
|---|---|---|---|
| (1) $R < 1200$ | ( $K < 0.01$ ) | : laminar | (viscous gas state) |
| (2) $R > 2200$ | ( $K < 0.01$ ) | : turbulent | (viscous gas state) |
| (3) $1.0 > K > 0.01$ | | : intermediate | (transition gas state) |
| (4) $K > 1.0$ | | : molecular | (rarefied gas state). |

The exact value of R for which the flow changes from laminar to turbulent depends on the geometry of the component, its surface roughness and other experimental factors. During evacuation of a vessel, turbulent flow normally occurs only for a short period of time at the beginning. It has been found to be approximately proportional to the root of the pressure gradient.

Transport phenomena in the viscous state of a gas are characterised by its coefficient of internal friction or *viscosity* $\eta$, which is defined as the tangential force per unit area for unit rate of decrease of velocity with distance. The tangential force per unit area is given by the rate of momentum transfer between adjacent layers. From this one can derive the coefficient of viscosity (see section 6):

$$\eta = 0.499\, \rho\, v_a\, \lambda\,,$$

in (units of viscosity) poise (J.-L.M. Poiseuille: 1799 - 1869), 1 Pa s = 1 kg $m^{-1}$ $s^{-1}$ = 10 poise .

Contrary to liquids, for which the viscosity decreases as the temperature increases, the viscosity of gases at normal pressure increases with temperature. At very high pressures, however, the intermolecular forces are important and the momentum transfer is very different. Correspondingly, at very low pressures there are almost no collisions between molecules, and momentum transfer happens only between molecules and walls. Since the viscosity in the viscous state is not a function of pressure, thermal conductivity is in this case also independent of pressure.

Like viscosity, *diffusion* of gases into each other is also determined by the mean free path $\lambda$. In the case of diffusion of molecules in the same gas at constant temperature, the coefficient of self-diffusion has been found to vary inversely as the density $\rho$ of gas or directly with $\lambda$: $D_{11} = 1.342\ \eta/\rho = v_a\ \lambda/3$. Diffusion pumps function on the principle of interdiffusion (coefficient $D_{12}$) of a gas having a low number density of molecules $n = N/V$ into another gas that has a very high number density.

Transport phenomena in the molecular state, when $\lambda$ is very large, are no longer determined by molecular collisions or viscosity of the gas. In this case, molecules condense on a surface, rest on it a given time and then reevaporate in arbitrary directions. There is a certain time required to form a monolayer on a clean surface (see section 5.2 and Table 7). If the surface is in motion it can transfer a velocity component to the molecule. This is the principle exploited by (turbo) molecular pumps. Rotating molecular gauges (like the spinning rotor gauge) also use this principle of molecular drag.

Unlike the viscous state, in the molecular state heat conductivity or so-called free molecular conduction depends upon pressure. In this range the rate of energy transferred by molecules between two surfaces is proportional to the pressure and temperature difference between the surfaces. Thermal conductivity at low pressures is used for measuring the pressure by thermal conductivity gauges. Normally, with such gauges the pressure is determined by variation of temperature, keeping the energy input for heating a filament constant.

If two chambers A and B with gases at different temperatures $T_A$ and $T_B$ are separated by a narrow pipe or a porous plug, the following relation holds: $p_A\, T_B^{1/2} = p_B\, T_A^{1/2}$. In this case, <u>*thermal transpiration*</u> will occur, which means gas will flow between the two chambers until an equilibrium is established.

This is important in situations where a chamber A is at low temperatures, say at liquid air temperature $T_A = 90$ K, and the pressure is measured with a gauge at room temperature $T_B = 300$ K. Then the real pressure is: $P_A = (90/300)^{1/2}\, P_B = 0.55\, P_B$. At higher pressures ($\lambda$ smaller than tube diameter) molecular collisions become predominant and the condition of equilibrium is $P_A = P_B$ instead.

## 7.2 Conductance and gas flow for different vacuum components and systems

### *7.2.1 Conductance, pumping speed and throughput*

Since gases have much lower viscosities than liquids, they flow typically with velocities of tens to hundreds of meters per second at relatively small pressure differences. Under normal conditions the gas flow with the pressure difference as driving force can be compared with the flow of heat or electricity in a conductor, caused by a voltage difference. The flow rate $Q$ or electrical current $I$ is directly proportional to the potential difference and inversely proportional to resistance. In complete analogy to the electrical current $I = \Delta V/R$, one gets for a gas flow through a pipe with resistance, known as <u>*impedance*</u> Z:

$$Q = (p_1 - p_2)/Z = \Delta p/Z = C\,\Delta p\,.$$

The reciprocal of the impedance (resistance to the gas passage) is called <u>*conductance*</u> $C$ of a pipe:

$$C = 1/Z\ (\mathrm{m^3/h})\,.$$

For heat transfer one gets a similar expression for the thermal energy: $H = C_t\,(T_1 - T_2)$, with $C_t = kA/L$ (k = conductivity; $A$, $L$ = cross-section, length of conductor). Flow impedance Z and flow conductance $C$ of a vacuum component vary with gas species $M$ and temperature $T$ and except in molecular flow, with gas pressure. In other words, Z and $C$ depend on the gas-flow regime and the geometry of the component. For the total impedance of several vacuum components connected in series or in parallel hold the same relations as for the resistance of conductors for the electrical current, i.e.:

$$Z_{tot} = Z_1 + Z_2 + \ldots + Z_n \quad \text{or} \quad 1/C_{tot} = \Sigma_i\, 1/C_i \quad \text{(in series)}$$

$$1/Z_{tot} = 1/Z_1 + 1/Z_2 + \ldots + 1/Z_n \quad \text{or} \quad C_{tot} = \Sigma_i\, C_i \quad \text{(in parallel)}\,.$$

The conductance of a pipe $C = Q/\Delta p$ can be written as: $C = p\,v\,A/\Delta p$, with $p$ = pressure, $v$ = flow velocity of the gas and $A$ = cross-sectional area. For the energy associated with flow, namely the potential energy of pressure or the kinetic energy of motion, the principle of preservation of energy holds. This is known as the <u>*Bernoulli principle*</u> (D. Bernoulli: 1700 - 1792), which can be expressed as follows: in a fluid of negligible viscosity the sum of potential and kinetic energy, associated with $p$ and $v$ along a streamline, remains constant, which means acceleration (deceleration) of flow is accompanied by a drop (rise) in pressure.

Since $p$ is associated with the density $n = N/V$, the flow of gas can be interpreted as the number of molecules $N$, passing with flow velocity v per unit time through a cross-section $A$ of a pipe. This relation defines the rate of flow or the <u>*pumping speed*</u> $S$ as follows:

$$N = A\,v\,n = S\,n \ \text{ with the pumping speed } S \equiv A\,v\ (\mathrm{m^3/h})\,.$$

According to this definition, the rate at which gas is removed from a vacuum system by pumps is measured by the pumping speed $S_p$ of a vacuum pump. It is generally defined as the volume of gas per unit of time $dV/dt$ that is removed from the system at the inlet pressure p of the pump:

$$S_p = dV/dt = Q/p\,,$$

where $V$ = volume of gas flowing per unit time, $p$ = steady pressure at the plane through which the flow or <u>*throughput*</u> $Q$ of gas passes (not for transient pumping speeds). Therefore, the throughput $Q$ of a pump is given as the product of pumping speed $S_p$ and inlet pressure p:

$$Q = p\,S_p = p\,dV/dt = C\Delta p\ .$$

This equation holds only if $p$ is constant, it does not apply to the transient state of flow. It describes the throughput of a pump with pumping speed $S_p$ at the intake pressure $p$, if $p$ and $V$ are constant. Correspondingly, the throughput for a passive element can be expressed as the product of the conductance $C$ with $\Delta p$, where $\Delta p$ is the difference between the pressures at the entrance and exit of the element. By analogy with this expression, the pumping speed at any point of the vacuum system is:

$$S = Q/p\ .$$

This leads to the following expression for the effective pumping speed $S_{eff}$ of a pump with nominal pumping speed $S_p$ and the conductance of connections between the pump and the volume to be evacuated :

$$\boxed{1/S_{eff} = 1/S_p + 1/C_{tot}}$$

This is a basic formula in vacuum physics, because in most cases a pump is not connected directly to a vacuum vessel. Almost always an intermediate system of pipes and various vacuum components is necessary, which introduces a flow resistance ($C_{tot}$ = total flow conductance) resulting in an effective pumping speed $S_{eff}$ smaller than the pumping speed $S_p$ of the pump. In order to ensure a definite effective pumping speed at the container, a correspondingly higher speed of the pump must be chosen. Only in the case where $C_{tot} = \infty$ (or impedance $Z_{tot} = 0$) is $S_p = S_{eff}$.

From the equation given above one sees also, that it does not make sense to increase the pump, if the conductance of the connecting components (pipes etc.) is limiting the pumping speed. One should always make sure that in a certain pressure region only the corresponding relevant conductance value is used. In general, it must be emphasised that the conductance of a vacuum component has by no means a constant value, independent of the pressure. It depends greatly on the kind of flow, viscous or molecular (see below), and therefore, on the pressure.

### *7.2.2 Conductance in the viscous flow and transition regime*

When a gas at constant pressure $p_1$ escapes from a large volume through a small aperture into a region where the pressure is $p_2 < p_1$, the gas acquires a velocity and streamlines are formed in the flow direction towards the aperture. After passing to the low-pressure side $p_2$, the gas jet has a minimum cross-section and then expands and contracts again approximately 10 times until finally it diffuses in the mass of gas $p_2$. This is the characteristic of a *laminar* flow.

By decreasing $p_2$ ($p_1$ = constant), the quantity and velocity of gas are increasing up to the state where $p_2/p_1$ reaches a critical value, corresponding to a velocity equal to the sound velocity (*Mach number*: Ma = 1; E. Mach: 1838 - 1864). Further decrease of $p_2$ does not increase the flow or velocity further. The dimensionless Mach number is defined as Ma = $v_a/v_s$, where $v_a$ is the average gas-flow velocity and $v_s$ is the local acoustic velocity, which is a function of the gas species and temperature (for air at 23 $^0$C: $v_s$ = 329 m s$^{-1}$).

As has been discussed already, at higher pressures, the value of the conductance depends on pressure, whereas it is independent of pressure in the region of molecular flow. Resistance to mass flow for a given pressure difference is reduced with rising pressure, because fewer molecules interact with the stationary walls and the flow rate depends on the viscosity of the gas.

The viscous flow rate of gas through a long ($L \geq 20\ d$, $L$ = length in cm, $d$ = diameter in cm) straight pipe of uniform circular cross section is proportional to the average pressure $p_a = (p_1 + p_2)/2$ (in dyn cm$^{-2}$), the pressure difference at the ends of the pipe ($p_1 - p_2$) and the diameter d of the pipe to the 4$^{th}$ power and inversely proportional to the length $L$ of the pipe and the gas viscosity $\eta$ (in poise). This is *Poiseuille´s law* (Poiseuille: see above):

$$\boxed{Q_{vis} = (\pi\, d^4/128\ \eta\ L)\ p_a\ (p_1 - p_2)\ ,\quad (\text{erg s}^{-1})}$$

For air at 23 $^0$C this gives: $Q_{vis}(\text{air}) = 4.97\ (d^4/L)\ p_a\ (p_1 - p_2)$ , (Pa m$^3$ h$^{-1}$) .

The laminar conductance of straight pipes of uniform, circular cross section is :

$C_{vis} = (\pi\, d^4/128\, \eta\, L)\, p_a$ (cm$^3$ s$^{-1}$) and for air at 23 $^0$C: $C_{vis}$(air) = 4.97 $(d^4/L)\, p_a$ (m$^3$ h$^{-1}$).

The conductance for an aperture of surface area $A$ (in cm$^2$) in the case of viscous laminar flow of air is given by the *formula of Prandtl* (L. Prandtl: 1875 - 1953) :

$$C_{vis} = 76.6\, \delta^{0.712}\, (1 - \delta^{0.258})^{1/2}\, A/(1 - \delta)\ (\mathrm{l\ s^{-1}})\,, \quad \text{with } \delta = p_2/p_1 < 1\,.$$

At the critical pressure ratio (see above): $\delta_{crit} = (p_2/p_1)_{crit} = 0.52$ the conductance for an aperture $C_{vis}$ amounts to 40 $A$ liters s$^{-1}$ and approaches for $\delta \rightarrow 0$ the limiting value $C_{vis} = 20\, A$ (l s$^{-1}$).

Only a few examples for conductance have been given here. Formulae for conductance values of vacuum components of different geometries and numerous nomograms for determining conductances can be found in the literature.

The steady flow of gas in the *transition regime* is also called *Knudsen flow*. This happens when the mean free path of molecules is of the same order of magnitude as the transverse dimensions of the pipe through which the gas flows. Approximate formulae for this condition have been derived for example by combining the laminar flow conductance with the molecular conductance of a long pipe.

### *7.2.3 Conductance in the molecular flow regime*

At very low pressures, molecules will pass through an aperture from one chamber to the other in both directions without making any collisions near the aperture. In contrast with laminar flow, there is no mass motion of the molecules and the gas is practically unaffected by the presence of the aperture, the reason being that for the molecular flow regime (high vacuum) the mean free path $\lambda$ is very large compared to the diameter $d$ of the aperture ($\lambda >> d$, aperture with length $L \approx 0$).

The molecular *throughput* and the molecular *conductance* of an **aperture** (for $p_1 > 10\, p_2$) are given by:

$$Q_{mol} = (1/4)\, (8\, \mathrm{k}\, T/\pi\, \mathrm{m})^{1/2}\, (p_1 - p_2)\, A = (1/4)\, (8\, \mathrm{R}\, T/\pi\, M)^{1/2}\, (p_1 - p_2)\, A \text{ (erg/s)},$$

$$C_{mol} = Q_{mol}/(p_1 - p_2) =$$

$$C_{mol} = (1/4)\, (8\, \mathrm{k}\, T/\pi\, \mathrm{m})^{1/2}\, A = (1/4)\, (8\, \mathrm{R}\, T/\pi\, M)^{1/2}\, A \text{ (cm}^3\text{/s)} = 3.64\, (T/M)^{1/2}\, A \text{ (l/s)},$$

with $T$ (K), $M$ (g/mol), $A$ (cm$^2$) and k = $R/NA$ = 1.3806 erg/K (see above). For air at $T$ = 23 $^0$C:

$$C_{mol} = 11.6\, A \text{ (l/s)}.$$

One notices again that the molecular flow conductance of an aperture depends only on the kind ($M$) and temperature ($T$) of gas and not on the pressure. Accordingly, for constant temperature, $C_{mol}$ varies with the molecular mass $M$ only.

From these equations one can also derive the molecular *pumping speed* of an aperture:

$$S_{mol} = C_{mol}\, [(p_1 - p_2)/p_1] = 3.64\, (T/M)^{1/2}\, (1 - p_2/p_1)\, A \text{ (l/s)}$$

and for air at $T$ = 23 $^0$C:

$$S_{mol} = 11.6\, (1 - p_2/p_1)\, A \approx 11.6\, A \text{ (l/s)}\,, \quad \text{for } p_2 \leq p_1\,.$$

Another example is the molecular conductance of a **long cylindrical pipe** of uniform cross-section with perimeter $o$, inner diameter $d$ and length $L$ in cm:

$$C_{mol} = (16/3)\, (\mathrm{R}\, T/2\, \pi\, M)^{1/2}\, A^2/o\, L = 19.4\, (T/M)^{1/2}\, A^2/o\, L = 3.81\, (T/M)^{1/2}\, d^3/L \text{ (l/s)}$$

and for air at T = 23 $^0$C:

$$C_{mol} = 12.1\, d^3/L \text{ (l/s)}\,.$$

From these equations for short apertures and long pipes one can also derive equations under certain assumptions for the molecular flow through short pipes and channels of various cross-sections and through vacuum components of simple and complex geometries, which all can be found in the literature.

If the vacuum system contains bends these can be considered on the basis that a greater effective length is assumed, which can be estimated in the following way:

$$L_{tot} < L_{eff} < L_{tot} + 2.66\, n\, r\,,$$

with $L_{tot}$, $L_{eff}$: total and effective length (in cm), $n$ the number of bends and $r$ the tube radius.

Normally, most components are constructed such that the pumping speed is not significantly throttled. As a general rule, one should always construct a vacuum line as short and wide as possible and with at least the same cross-section as the inlet port of the pump.

The values for the laminar and molecular conductance of air, given above, should be multiplied by the factors given in Table 9 for working with other gases.

Table 9
Conversion factors of conductance for air to other gases.

| Gas at 23 °C | Molecular flow | Laminar flow |
|---|---|---|
| Air | 1.00 | 1.00 |
| $O_2$ | 0.947 | 0.91 |
| $N_2$ | 1.013 | 1.05 |
| He | 2.64 | 0.92 |
| $H_2$ | 3.77 | 2.07 |
| $CO_2$ | 0.808 | 1.26 |
| $H_2O$ vapour | 1.263 | 1.73 |

## 8 GAS LOAD AND ULTIMATE PRESSURE

In a vacuum system there are mainly three sources of gas, the so-called *gas load*, namely:

(a) the residual gas in the system,
(b) the vapour in equilibrium with the materials present,
(c) the gases produced or introduced by
- leakage (also "virtual leaks" by captured gas, without penetration through the walls)
- outgassing (adsorption),
- permeation (transfer of gas through a solid, through porous materials, glass etc.).

The *ultimate pressure* $p_u$ of high-vacuum systems normally depends only on the gas load $Q_G$ mentioned in (c) and on the effective pumping speed $S_{eff}$. It is given by:

$$p_u = Q_G/S_{eff}\,.$$

In the case of a leak, $Q_G$ is constant and, therefore also $p_u$ is constant, whereas $p_u$ is a function of time when one has $Q_G = f(t)$, as it is when outgassing dominates. There exist extensive outgassing data and various tables and nomograms in the literature relating gas load and ultimate pressure and other system parameters. Table 10 presents outgassing rates $K_1$ for several vacuum materials.

In cases where the pumping process is dominated by residual gas, pumpdown in a high-vacuum region can be described by:

$$p = p_{t=0} \exp[-(S_{eff}/V_{tot})/t]\,,$$

where $p$ is the pressure after time $t$, $p_{t=0}$ is the pressure at $t = 0$, $S_{eff}$ is the effective pumping speed and $V_{tot}$ is the total volume of the system.

However, by far the most important uncertainty, associated with pump performances, pressure and flow measurements, and external leakage, is due to outgassing. The outgassing rates can easily vary by many orders of magnitude, depending on the history and the material of a surface, its treatment, humidity, temperature, and the period of exposure to vacuum. Since one usually approaches the ultimate pressure of a system asymptotically, even small changes in gas loads result in large differences of evacuation times.

Table 10
Approximate outgassing rate $K_1$ for several vacuum materials, after one hour in vacuum at room temperature.

| Material | $K_1$ ( mbar l $s^{-1}$ $cm^{-2}$ ) |
|---|---|
| Aluminium (fresh) | $9 \times 10^{-9}$ |
| Aluminium (20 h at 100 $^0$C) | $5 \times 10^{-14}$ |
| Stainless steel (304) | $2 \times 10^{-8}$ |
| Stainless steel (304, electropolished) | $6 \times 10^{-9}$ |
| Stainless steel (304, mechanically polished) | $2 \times 10^{-9}$ |
| Stainless steel (304, electropolished, 30 h at 250 $^0$C) | $4 \times 10^{-12}$ |
| Perbunan | $5 \times 10^{-6}$ |
| Pyrex | $1 \times 10^{-8}$ |
| Teflon | $8 \times 10^{-8}$ |
| Viton A (fresh) | $2 \times 10^{-6}$ |

## 9. CONCLUSION

This introduction to a special branch of classical physics attempts to cover the fundamental principles and most important phenomena of vacuum physics and technology described in much more detail in a series of textbooks. It was written with the intention of providing a short summary and useful background for the students of the CERN Accelerator School 1999 on Vacuum Technology, and to help them better understand the subsequent lectures on current areas of research in vacuum physics.

## BIBLIOGRAPHY


[1] P.A. Redhead, The Physical Basis of Ultra-High Vacua, Chapman & Hall, London (1968).

[2] G.L. Weissler and R.W. Carlson (editors), Vacuum Physics and Technology, Methods of Experimental Physics, Vol. 14, Academic Press (1979).

[3] M. Wutz, H. Adam, and W. Walcher, Theorie und Praxis der Vakuumtechnik, F. Vieweg & Sohn, Braunschweig/Wiesbaden (1986).

[4] H.J. Halama, J.C. Schuchmann and P.M. Stefan (editors), Vacuum Design of Advanced and Compact Synchrotron Light Sources, Am. Inst. Phys., Am. Vacuum Soc. Series 5, Conf. Proc. No. 171, BNL, Upton, New York (1988).

[5] A. Chambers, R.K. Fitch and B.S. Halliday, Basic Vacuum Technology, Adam Hilger, Bristol (1989).

[6] A. Roth, Vacuum Technology, North-Holland, Elsevier Sci. Publ. B.V., 3rd updated ed. (1990).

[7] Y.G. Amer, S.D. Bader, A.R. Krauss and AR.C. Niemann (editors), Vacuum Design of Synchrotron Light Sources, Am. Inst. Phys., Am. Vacuum Soc. Series 12, Conf. Proc. No. 236, Argonne, Illinois (1990).

[8] A. Berman, Vacuum Engineering Calculations, Formulas, and Solved Exercises, Academic Press, Inc. (1992).

[9] M. Prutton, Introduction to Surface Physics, Oxford Sci. Publications, Clarendon Press (1994).

[10] M.H. Hablanian, High-Vacuum Technology, a Practical Guide, Marcel Dekker, Inc. (1997).

[11] C. Edelmann, Vakuumphysik - Grundlagen, Vakuumerzeugung und -messung, Anwendungen, Spektrum Akademischer Verlag (1998).

# MECHANICAL PUMPS

*N. Hilleret*
CERN, Geneva, Switzerland

**Abstract**
Various types of mechanical pumps are presented together with their pumping mechanisms and main characteristics. The combination of these elements to construct a roughing station is discussed in terms of reliability and efficiency. Various situations corresponding to real accelerators are considered to give guidelines for the design of roughing stations.

## 1. INTRODUCTION

The development of mechanical pumps was intimately linked with the progress of vacuum science up to the middle of this century when the introduction of both the sputter ion pump and the sublimation pump provided an alternate method for the production of low pressure. Nevertheless, the mechanical pumps are, with the exception of the sorbtion pumps, the unavoidable initial stage before any attempt can be made to activate capture pumps and to reach the ultra-high vacuum domain. This is due to their ability to evacuate efficiently from the vacuum vessel the large number of molecules present at atmospheric pressure. However the limited compression ratio of the pumps limits the achievable ultimate pressure. In this presentation an overview of the operating principles and of the characteristics of these pumps will be given. The rules for combining several mechanical pumps to build a roughing station will then be reviewed together with considerations on the control system and the operation of these types of pumps in accelerator vacuum systems.

## 2. MECHANICAL PUMPS

In mechanical pumps, the gas present in a vacuum system is transferred from a low pressure to a higher pressure region. In order to be efficient, the pump has to provide simultaneously a finite pumping speed and compression. To carry out this double task two main mechanisms have been used up to now:

- Isolation and subsequent compression of a gas volume defining a first category often named "Positive-displacement pumps".
- Transfer to the gas molecules of a preferential velocity direction introducing a mean drift of the gas towards the high pressure region. This category is known as: "Momentum transfer pumps".

### 2.1 Displacement pumps

#### *2.1.1 Principle*

The principle of these pumps is shown in the Fig. 1. During a cycle, a given volume of gas at low pressure is trapped mechanically and subsequently compressed to the exhaust pressure. During this cyclic process two difficulties arise:

- The swept volume has to be sealed in order to ensure a good compression ratio hence an efficient dynamic seal has to be provided.
- The compression of the gas during the pumping cycle creates heat, which must be evacuated.

These two difficulties can be efficiently mastered by the use of a fluid, which produces a sealing film between the moving and fixed parts of the volume while efficiently evacuating the heat liberated during the gas compression. This role was played by water or mercury in the first mechanical pumps but oil now replaces it in a very efficient way.

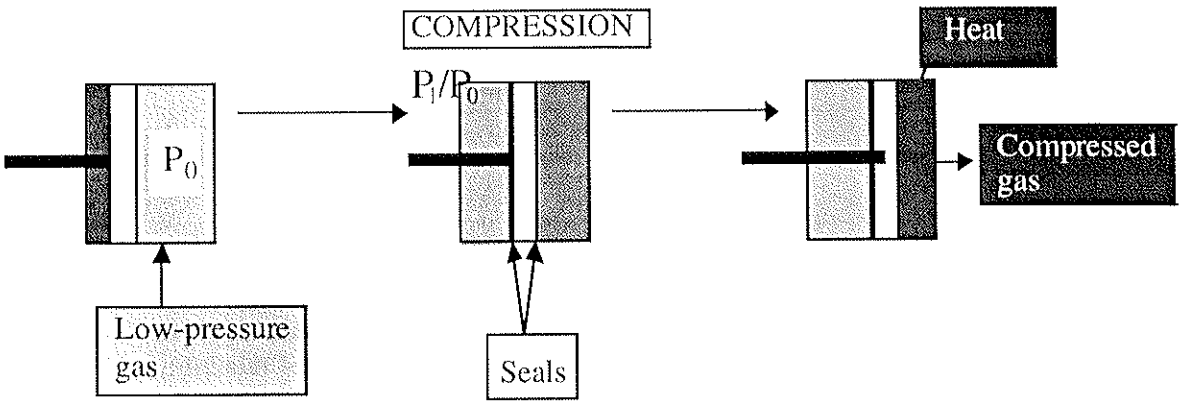



Fig. 1 Principle of the displacement pump

In order to avoid the presence of oil and the related maintenance and backstreaming problems, dry pumps have been developed. Here the absence of an efficient dynamic seal reduces the compression per stage of the pumps and the necessity to evacuate the heat implies the use of an external water cooling circuit. Several successive elementary stages are often combined in dry pumps to improve their compression ratio and to allow gas exhaust at atmospheric pressure.

### *2.1.2 Oil-sealed pumps*

In all these pumps, the compression is produced between one fixed cylinder and a second one rotating off-centred inside the fixed cylinder. The two main types are the piston pump and the rotary-vane pump. Both are able to exhaust gas at atmospheric pressure and their typical main characteristics are presented in Table 1.

Table 1
Main characteristics of oil-sealed pumps

| TYPE | ULTIMATE VACUUM (Total pressure) (Pa) | PUMPING SPEED ($m^3/h$) |
|---|---|---|
| Rotary-piston pump | 1 | 30–>1500 |
| Rotary-vane pump (single stage) | 1 | 1–>300 |
| Rotary-vane pump (double stage) | $2 \times 10^{-2}$ | 1–>300 |

The rotary-vane pump is mainly used for small to medium pumping speed, and is the most commonly used forevacuum pump in accelerators. The rotary-piston pump is often reserved for applications where high throughput is needed, in which case water cooling is necessary.

The compression ratio of double-stage rotary-vane pumps can be very high ($10^8$ for air) [1] and their ultimate pressure correspondingly low provided special care is taken [2]. When these pumps operate close to their ultimate pressure, i.e. in the molecular flow regime, the risk of oil-vapor back-streaming is important. Various trapping methods have been investigated [3, 4] and have shown the efficiency of zeolite traps, when they are properly reconditioned, to prevent the oil backstreaming.

### *2.1.3 Gas ballast*

The role of the oil is very important in these pumps and determines greatly their performance. The pumping of aggressive gases (seldom in the case of accelerators) or the condensation of gases such as water might perturb their performance. This condensation takes place usually in the high-pressure stage of the rotary-vane pump when pumping humid gases. It occurs when the saturation vapor pressure of water is reached in the compressed mixture. In this case, the formation of an oil/water

emulsion reduces the lubricating properties of the oil. Furthermore, the circulation of liquid water transported with the oil from the outlet stage to the inlet stage fixes the ultimate pressure of the pump at the corresponding water vapor pressure. To remedy this detrimental effect gas ballasting was proposed in 1935 by Gaede [5].

The aim of the gas ballast is to introduce in the pump, at the end of the compression cycle where condensation might occur, a certain volume of unsaturated vapor to lower the partial pressure of the condensable gas. This injection results also in a higher pressure in the compressed volume and hence causes the discharge valve to open at an earlier stage of the compression. The amount of gas to be injected during the ballasting process can be calculated in the following way:

If $P_e$ is the exhaust pressure of the pump and $P_i$ the inlet pressure then the compression $C$ can be written as:

$$C = \frac{P_e}{P_i} = \frac{P_e}{P_G + P_W}$$

$$P_s > C \times P_W$$

where $P_G$ is the uncondensable and $P_W$ the condensable gas pressure. In order to avoid condensation the partial pressure of condensable gas at the exhaust must be lower than the saturation pressure $P_s$ of the condensable gas. Hence:

$$P_W < \frac{P_G \times P_S}{P_E - P_S}$$

gives the highest condensable pressure in the pumped gas to avoid condensation in the pump without gas ballast.

If a gas ballast (pressure $P_B$, speed $S_B$) is introduced then $P_G$ becomes:

$$P_G \rightarrow P_G + P_B \times \frac{S_B}{S}$$

and the pressure $P_W$ is increased to:

$$P_W > \frac{\left(P_G + P_b \times \frac{S_b}{S}\right) \times P_S}{P_E - P_S}$$

From this equation it can be concluded that raising the ballast flow ($P_B$*$S_B$) avoids condensation. It also shows the great importance of the pump temperature during the evacuation of condensable gases. Raising the pump temperature very quickly increases the vapor pressure of the condensable gas. Part of the action of the gas ballast results from an increase of the operating temperature of the pump because of the enhanced gas flow.

In practice it is important before starting to evacuate a vessel in which a large quantity of condensable gas is present, to start the pump with the gas ballast open before the beginning of the roughing procedure. This has two beneficial effects: the oil is purged from possible traces of condensation and the pump reaches a higher temperature. This is not necessary if the pump is expected to run, during the roughing of the vessel, at high pressure for a long period and hence is expected to reach a sufficiently high temperature before attaining the pressure at which condensation can occur.

### *2.1.4 Dry pumps*

Despite the very important role played by the oil in the operation of displacement pumps, its presence has two major inconveniences:

- Interaction of the pumped gases, often aggressive in the case of the semiconductor industry, degrades the oil and necessitates frequent oil changes even in the case of the expensive fluorinated oils. This results in excessively high maintenance costs for oil pumps exposed to aggressive chemicals.
- Oil vapor present in the foreline pipe when the pump operates close to its ultimate pressure can contaminate sensitive equipment to which the pump is connected. Even worse, if the pump stops under uncontrolled conditions this can lead, if the internal safety valves of the pump do not operate, to an important oil migration towards the vacuum system.

For these two reasons the leading market for vacuum technology, the semiconductor industry, has forced the vacuum manufacturers to develop oil-free pumps.

The elimination of oil in the displacement pumps introduces major constraints on the design of the pump: the sealing and the lubrication effects of the oil are eliminated and heat transfer from the mobile parts is much reduced. This results in a reduced compression-per-stage of the pumps and the need for external cooling. In consequence, most of the dry pumps available on the market need water cooling and have more stages (about double) than an equivalent oil-sealed pump.

The various types of commercially available dry pumps are summarised in Table 2.

Table 2
Various types of dry pumps

| Type of pump | Compression per stage | Speed ($m^3/h$) | Ultimate pressure |
|---|---|---|---|
| Piston | 15-20 | 12 | 3-4 Stages, atm –> 4 Pa |
| Diaphragm | 200 | 0.8 –>5 | atm –>200 Pa |
| Scroll | $10^5$ | 15 –>600 | atm –>1 Pa |
| Roots | 30 | 25 –>1000 | 4 Stages, atm –> 0.1 Pa |
| Screw | 30-100 | 25 –>2500 | atm –>1 Pa |
| Claw | 20-50 | 25 –>500 | 3 Stages, atm –> 10 Pa |
| Tongue/groove | 10-20 | 50 | 4 Stages, atm –> 5 Pa |

Different types of pumps are often combined into a single system, e.g. a Roots pump is used to provide high pumping speed at the entrance stage after which claw stages are used to increase the compression at the high-pressure stages where the volume throughput is reduced. Two types of pump have neither oil nor joints between vacuum and oil: these are the scroll pumps and the diaphragm pumps.

The advantage of oil-free pumping is certainly more important for applications where the cost of maintenance is high. In the case of accelerators, the pumping of radioactive zones, where the disposal of contaminated oil and the maintenance of activated pumps is a concern, dry pumps offer an interesting alternative to the oil-sealed type.

## 3. SORBTION PUMPS

Although the sorbtion pump belongs to the family of cryopumps, it is also used for the roughing of vacuum systems and is perhaps the only absolutely dry pump. In this pump zeolites are used because of their porous crystal structure which is able to pump a large amount of gas when activated. This unique property is due to the size of the pores, comparable to the size of molecules, providing a high binding energy (low equilibrium pressure) and a large specific area (large pumping capacity). This

geometrical effect is enhanced by lowering the surface temperature of the zeolite to liquid-nitrogen temperature in order to lower the equilibrium pressure for a given gas quantity. The interaction of the molecules with the adsorbent depends on the nature of the pumped molecule and is of course much greater for nitrogen and water than for noble gases. When adsorbed, these latter gases will be very quickly displaced by other air molecules and will finally be desorbed when the pump is saturated. For this reason several pumps (two or three cartridges) are used to evacuate a volume from atmospheric pressure down to 1 Pa. The first pump reduces the pressure from atmospheric to some $10^4$ Pa. A second pump brings the pressure down to some tens of Pa, the third pump providing a final pressure in the 0.1 Pa region, mainly limited by noble gases. The size of the pumps depends directly on the volume of the vacuum vessel: 1 kg of molecular sieve is needed to evacuate $8 \times 10^3$ Pa.m$^3$ of nitrogen.

Sorption pumps are very simple and robust and consist of a metallic can containing the sieves and providing a good thermal contact between the sieve and the liquid nitrogen surrounding the pump. Between two pumping cycles, the pump has to be regenerated by allowing the sieve to warm to room temperature. During this regeneration, the pump has to be properly vented to let the pumped gas escape and avoid dangerous pressure build up. Water has a great affinity for the sieve so, after several pumping cycles, the pump has to be reactivated by baking to 300 °C under vacuum. The main limitation of these pumps is their limited capacity, which restricts their use to small unbaked systems. They are nevertheless very interesting for their simplicity, reliability and absolute cleanliness.

## 4. MOMENTUM TRANSFER PUMPS

### 4.1 Principle

In displacement pumps, a volume of gas is isolated in a moving volume in order to transmit a preferential direction to the molecules despite the viscous type of flow corresponding to the pressure in this volume during most of the compression cycle. In molecular flow, it is possible to alter their uniform velocity distribution by repeatedly giving to the molecules a preferential velocity direction. This can be achieved when molecules hit a fast-moving surface. To be efficient the velocity transmitted to the molecules must be comparable to their thermal velocity (460 m/s for nitrogen at room temperature); hence the pumping effect requires fast-moving surfaces. Furthermore, this pumping mechanism is only efficient in the molecular flow regime and hence this type of pump cannot pump gas against atmospheric pressure.

To illustrate the effect, the following simplified calculation shows the pumping performance obtained in a thin rectangular duct with one moving surface [6]. The geometry of the system is shown in Fig. 2 .

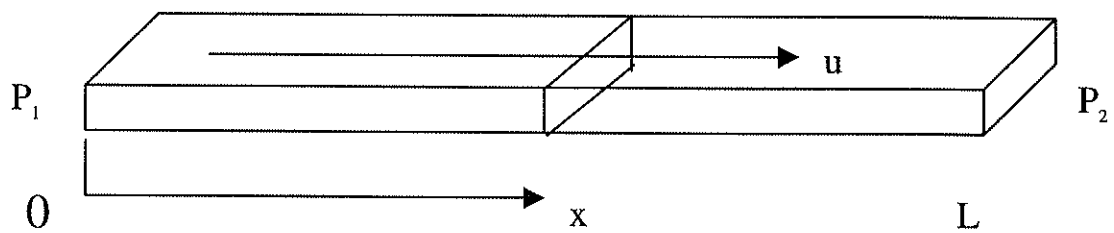



Fig. 2 Geometry of momentum-transfer pump

The molecular conductance $c$ of a rectangular ($W*h$) duct, with the perimeter:

$$H = 2(w+h)$$

and the section $A$ is given by:

$$c = \frac{4\bar{v}}{3\int_0^x \frac{H}{A^2}\,dl} = \frac{2}{3}\bar{v}\frac{h^2 w^2}{x(h+w)} .$$

In molecular flow the relation between pressure gradient and flow is:

$$\frac{dP}{dx} = \frac{Q}{c}$$

Assuming that one surface is moving with the speed $u$, the gas flow, which traverses a given cross section, corresponds to:

$$\frac{dP}{dx} = \frac{u}{2c}AP .$$

Thus

$$\int_{P_1}^{P_2} \frac{dP}{P} = \frac{u}{2c}A\int_0^L dx .$$

Integrating over the length of the duct one obtains:

$$K = \frac{P_2}{P_1} = e^{\frac{u}{2c}AL} .$$

These calculations show the exponential dependence of the compression on the velocity of the moving surface, and through $c$, its inverse exponential dependence on the square root of the mass of the gas. Accurate calculations can be found in Refs. [7–9] and again demonstrate the drastic influence of the surface speed on the performance of the pumps. This speed and the narrowness of the gap between the surfaces needed to ensure an efficient pumping has restricted the use of this type of pump until recently.

### 4.2 Molecular and turbomolecular pumps

The first molecular pump was proposed by Gaede [10] and later types by Holweck [11] and Siegbahn [12]. Up to the late 1970's this type of pump was not used because of its poor reliability and low pumping speed. Recently with the advent of dry pumping, these pumps were revised and integrated as the first stage of turbomolecular pumps in order to improve their compression and hence make them compatible with the high backing pressure of dry roughing pumps.

In 1957, Becker [13] proposed the first turbomolecular pump. A development using rapidly rotating blades was initially made to reduce the backstreaming from diffusion pumps. This type of pump with its high compression for hydrocarbons replaced the diffusion pump as the workhorse in laboratories. The success of the turbomolecular pump is based on its high compression (between 300 and $3 \times 10^4$ for hydrogen, $10^8$ to $10^{12}$ for nitrogen) and its high pumping speed (50 to 5000 l/s). In modern turbomolecular pumps, the rotation speed of the blades (up to 1500 Hz) is such that the linear speed is close to 500 m/s.

The geometry of the blades influences the pumping and the compression performance of the pump. Usually the blades are more open on the low pressure side providing a higher pumping speed. On the high pressure side, the volume flow being smaller, the blades can be more closely spaced (lower pumping speed) to provide more compression. The tips of the blades are submitted to very high mechanical stresses which limit the rotational speed of the turbine and hence the performance of the pumps. Because the pumping effect requires a molecular flow between rotor and stator, these

pumps reach their full performance for compression and pumping speed below 1 Pa exhaust pressure. In the 1980's the demand for oil-free pumps able to evacuate vacuum systems in the pressure range used for coating processes led to the introduction of combined molecular-drag pumps providing a higher compression and a possible backing pressure between $10^2$ and $10^3$ Pa. To suppress completely the presence of hydrocarbons in the pumping systems, magnetically suspended pumps have been developed.

Turbomolecular pumps operating in a metallic and baked environment can reach a very low pressure ($10^{-9}$ Pa) [14]. They are especially interesting for vacuum measurements as they have no memory effect nor selectivity compared to ion pumps or sublimation pumps, and a stable pumping speed across a large pressure range ($10^{-1}$–$10^{-7}$ Pa). Because of these features these pumps are widely used in accelerators for the initial pumping and during bakeout.

## 5. PUMP CHARACTERISTICS

It is often useful to check the performance of commercially available pumps in order to detect possible deviations from the expected values. Apart from the pumping speed, the main characteristics of mechanical pumps are the ultimate pressure and the compression. The methods and procedures used to measure these quantities are defined by, for example, the International Standards Organisation (ISO), the American Vacuum Society and the Pneurop Association.

### 5.1 Displacement pumps

The usual reception tests consist of measuring the ultimate pressure, the pumping speed and the water-vapor pumping capacity. In the case of pumps using oil, a distinction is made between condensable and non-condensable gases, depending on the type of gauge used to measure the pressure. For the measurement of the pumping speed $S$ a fixed and known flow $Q$ is injected in a dome where the resulting pressure $P$ is measured. The pumping speed is given by the simple equation:

$$Q = P \times S .$$

Procedures also exist for the measurement of water-vapor pumping and are described in Refs. [15–17].

### 5.2 Turbomolecular pumps

The control of turbomolecular pump performance could be important especially for the compression ratio, which could be subject to variation. The other characteristics, which are usually measured, are: the ultimate pressure and the pumping speed. For turbomolecular pumps, these measurements are often made on a baked measuring bench. Procedures and methods are described in Ref. [18].

## 6. LAYOUT OF A ROUGHING STATION

The roughing stations are used in accelerators to lower the pressure from atmospheric to the pressure required either to switch on the sputter ion pumps, the most widely used holding pumps in such machines, or to start a bakeout. The design of these stations should provide safe and efficient pumping to reach these targets. It must be safe in order to avoid the contamination of the beam pipe, often exposed to the enhanced desorption induced by the bombardment of energetic particles or to protect the vacuum system against accidents (e.g. power cut or human errors) leading to uncontrolled venting. It must be efficient in fulfilling its role: the achievement of a given pressure and gas composition in a given time with a minimum investment.

To fulfil these constraints one should take into account some general rules for the construction and the operation of a safe and clean pumping station and adapt the size of the station's elements to the particular case envisaged. The following paragraphs will deal first with the construction and the operating procedure of a roughing station and then consider some particular aspects depending on the characteristics of the accelerator.

## 6.1 Layout of a roughing station

Apart from breakdowns, which, with modern pumps, are very rarely due to misuse of the station, the main accidents which can take place with a roughing station are undesired venting of the accelerator vacuum system through the roughing station, or contamination with oil vapors. Both these events can be avoided by using an appropriate layout of the station and proper operational procedures controlled by an automatic controller system.

Some of the common reasons for accidental venting of an accelerator vacuum system are: the connection of a leak detector without isolating the pumping station or, during a power cut, an uncontrolled opening of a venting valve. To avoid these incidents, normally-closed valves interlocked by an appropriate control system are used. The sequence used in this controller must take into account various characteristics ensuring a clean operation of turbomolecular pumps.

The question of oil contamination through roughing stations is only relevant during the periods when the turbomolecular pump is not running at full speed. At its nominal speed, the compression ratio is such that contamination by oil vapor is not visible and certainly irrelevant in a normal accelerator vacuum system. On the contrary, if a turbomolecular pump is left idle in contact with a vacuum system, an important contamination can occur. This is illustrated in Fig. 3 showing the evolution of the residual gas composition in a system in contact with an unpowered turbomolecular pump. Even after a pump restart the system remains permanently polluted with traces of heavy hydrocarbons. To avoid this, a valve must be positioned to isolate the pump from the system (system isolation valve, bakeable if the accelerator will be baked) when the rotational speed of the pump falls below 70% of its nominal value. The reason for this can be seen in Fig. 4, which shows the variation of the partial pressure of various gases as a function of the speed of a turbomolecular pump. Traces of light hydrocarbons become visible for a rotational speed lower than 60% of the nominal value. To avoid contamination of the pump itself, it is a well-established and sound practice to vent the pump before it is stopped. For this a venting valve must be connected to the orifice generally provided by the manufacturer in the high-pressure stages of the pump. A pump must never be vented through its foreline.

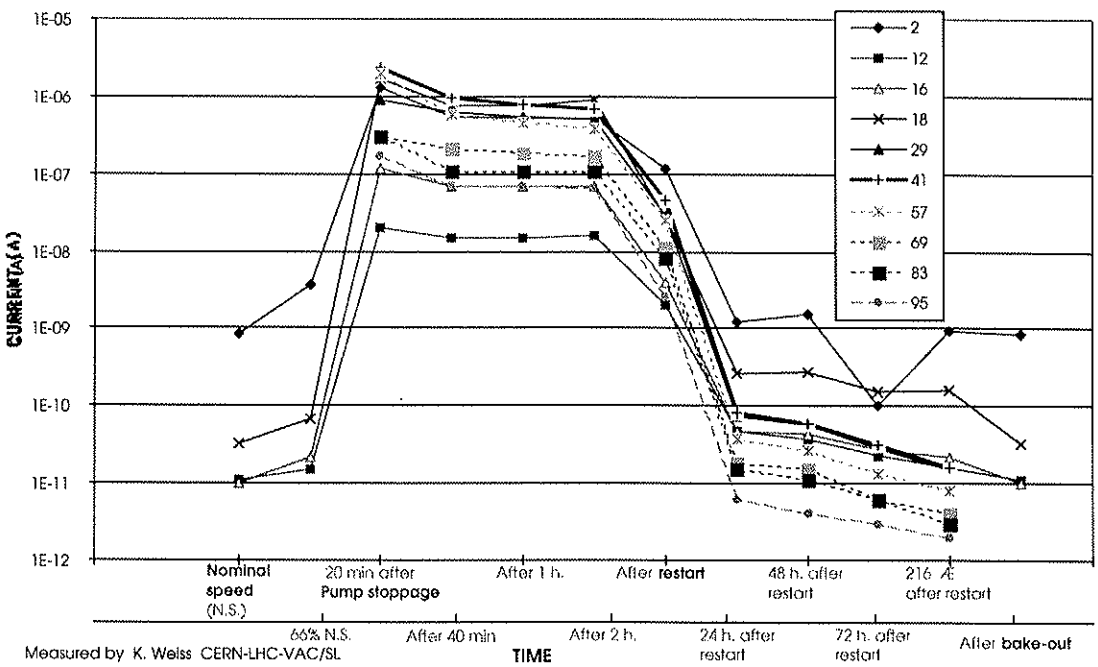



Fig. 3 Evolution of the residual gas composition after the stoppage of a turbomolecular pump

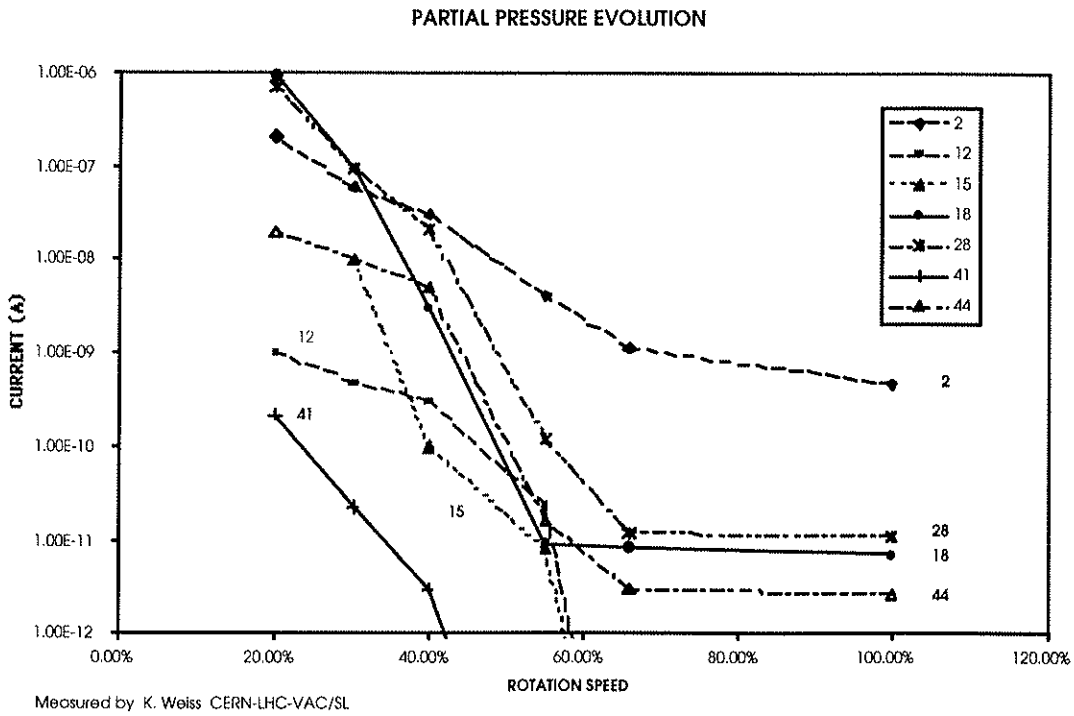



Fig. 4 Variation of the gas composition during the slowing down of a turbomolecular pump.

The contamination of a station or, worse, of the accelerator vacuum system through a massive flow of the rotary-pump oil is certainly the absolute nightmare against which the station must be perfectly immune. Although most of the modern roughing pumps are equipped with automatic valves closing in case of a rotary-pump stoppage, it is good practice, and also useful during leak detection, to have a valve isolating the two pumps (roughing pump isolation valve) when the group is stopped or in the case of some other anomaly. Lastly, to avoid contamination of the foreline with rotary-pump oil, this pump must also be automatically vented if the pump is stopped.

Most commonly, the roughing stations are also used during leak detection and a leak detection valve has to be installed on the high-pressure side of the turbomolecular pump. In order to vary the gas flow to the leak detector, this valve should be installed upstream from the roughing pump isolation valve.

Hence five valves as described above when properly activated can protect an accelerator vacuum system against contamination by oil backstreaming and ensure a flexible and reliable use of the station. A possible scheme of the roughing station corresponding to this description is given in Fig. 5. For these valves, electropneumatic types of DN16 or DN25 size are generally well suited with the exception of the system isolation valve which must be sized in relation to the turbomolecular pump. The use of normally-closed valves is a good protection against accidental venting in case of failure. Furthermore, a hard-wired safety system forbidding the actuation of venting valves when the system isolation valve is open has been found useful on some occasions. In the same way, protection can be installed on the leak-detection valve to forbid its opening when the system isolation valve is open and so avoid the roughing of the leak detector connecting line by the accelerator.

The control of the station must ensure the appropriate sequencing for the start up and the shutdown of the station and is a key element for obtaining clean pumping of the accelerator. The development of programmable controllers (PLC) gives cheap and versatile control systems fulfilling the most complicated requirements to be made.

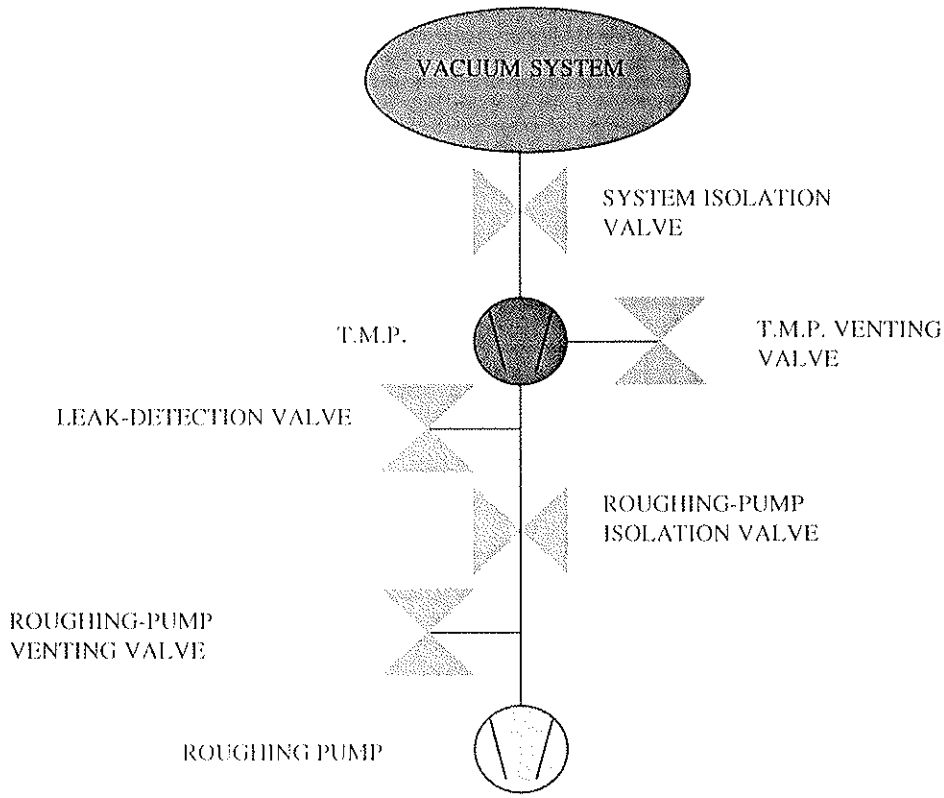



Fig. 5 Layout of a roughing station

## 6.2 Operational aspects

After these considerations on the layout of the roughing station, several other questions must be studied concerning their size, number and position. To address these points, the characteristics of the vacuum system must be considered i.e. the length and bakeability of the accelerator, the presence of radioactive or contaminated zones, of large tanks or of especially sensitive equipment (e.g. superconducting cavities). The vacuum system of accelerators is usually split in sectors. In large accelerators, the use of mobile stations reduces the investment, which could be quickly prohibitive since the spacing between stations should not exceed 100 m for conductance and leak detection reasons. Mobile stations are also useful in radioactive zones, as they are not exposed continuously to radiation that quickly damages their power supply. In fact, the trend in new turbomolecular pumps is to integrate the pump power supply with the pump. The use of fixed pumps in radioactive areas will hence require the purchase of specially built mechanical frequency converters at very high cost.

The timing of an intervention is completely different in baked and unbaked systems. In baked systems, interventions are usually long due to the time needed to heat and cool the chambers. On the contrary, in unbaked systems the time needed for the roughing of a sector is an important part of the down time of a machine in case of failure. So the constrains on the roughing station are quite different.

### *6.2.1 Baked vacuum systems*

In this case the installation of a mobile roughing station does not change the intervention time. The function of the pumping groups is to evacuate the water vapor during the bakeout. The ultimate pressure after bakeout is not strongly dependent on the pumping speed during this operation. Furthermore accelerators have typically a specific conductance of about 100 l/s.m. For these reasons, small pumping stations equipped with pumps not exceeding 200 l/s, backed by primary pumps of about 15 $m^3/h$, are a good choice. The spacing between these stations is dictated by leak detection

considerations and should not exceed 100 m for reasons of ease of localisation of leaks and of the time before leak detection can commence.

*6.2.2 Unbaked vacuum systems*

In such machines, the relevant target during an intervention is to reopen the sector valves as quickly as possible. The various steps required are the roughing of the sector and the leak detection followed by the ignition of the ion pumps. In a small non-radioactive machine, fixed groups are better suited for this application. If mobile groups are preferred, they should be connected as closely as possible to an ion pump in order to ease the start up of the pump. Large primary pumps or auxiliary pumps connected during the pre-evacuation phase through the leak detection valve will help to reduce the non-negligible time of this roughing phase. Larger turbomolecular pumps (not exceeding 400 l/s) can be of help for a quick start of the ion pumps if their pumping speed is not throttled by the connections to the vacuum system. Note that the price of the station increases drastically with the size of the turbomolecular pump especially since the price of the connecting valve and of the necessary connection scales accordingly. Lastly in the case of a radioactive area, the use of mobile stations is highly recommended for the reasons explained above. In that case special care must be taken to choose for the connection of the group a position of reduced radioactivity, easily accessible and equipped with a quick and reliable connection system. This will reduce the radiation dose received by the operators during the installation and the dismounting of the roughing stations.

*6.2.3 Special devices*

Some devices require special care for their evacuation. This is, for example, the case for superconducting cavities that are very sensitive to pollution and, in particular, to dust particles which might be transported during the venting or roughing of these cavities. For this reason special care has to be taken in order to avoid their transportation during pumping and it is often considered that laminar flow conditions must be maintained throughout the evacuation. The upper limit of the flow can be calculated from the diameter of the sensitive areas. The calculated flux can be maintained by keeping the corresponding pressure at the entrance of the primary pump by means of an automated valve. Of course the cleanliness of the roughing station is in that case of utmost importance. A preliminary baking of the station before an intervention on the cavity vacuum allows to remove all residual contamination and to test the cleanliness of the station which is otherwise masked by the predominance of the water vapor in the residual-gas composition.

*6.2.4 Dry pumping*

With the availability of more-and-more sophisticated dry pumping systems, the vacuum specialist is increasingly confronted with users asking for dry-pumping stations. The real need of these users is clean pumping, and dry pumping is certainly one way to achieve this. There is no indication that dry pumping is the only way to produce a clean roughing but it is certainly the most expensive way. This expense is certainly justified by the reduced maintenance in the case of the semi-conductor industry where the stations are exposed routinely to aggressive gases degrading the most stable and expensive oils. This is perhaps not the case in regular accelerators where the most aggressive gas pumped is water vapor. On the other hand, in some very sensitive places, e.g. photocathodes for the production of electrons, or windows and mirrors submitted to a high flux of synchrotron light, the doubt could subsist and justify the choice of a dry system with the implied extra cost.

Nevertheless one should not forget that oil is not the only source of hydrocarbon degassing and that many components of dry pumps (e.g. seals, electrical isolation for coils, ...) in contact with the vacuum are also a possible source of hydrocarbons. Lastly, in most of the dry primary pumps, the "dry" property relies on the integrity of sliding gaskets which can also fail.

## 7. CONCLUSION

The vacuum expert has nowadays a wide variety of mechanical pumps covering the complete pressure scale from atmosphere to UHV. These pumps can satisfy the most stringent requirements in terms of ultimate pressure, cleanliness and resistance to aggressive media but, with the new trend towards active components, have the tendency to loose some of their resistance to radiation. This larger variety of pumps and their increasing price requires a careful analysis by the vacuum expert before choosing

the type and the size of a pumping element. The design of a roughing station must take into account a great variety of parameters such as the volume of the vessel and its length but also the radiation level, the presence of radioactive contaminants, the accessibility and, of course, the budget available. The large number of components developed by the European vacuum industry allows adequate answers to the most difficult problems to be found in most cases.

**REFERENCES**

[1] M.H. Hablanaian, J. Vac. Sci. Technol. A 12, 897 (1994).

[2] M.H. Hablanaian, J. Vac. Sci. Technol. A 19, 250 (1981).

[3] N.S. Harris, Vacuum 28, 261 (1978).

[4] M.H. Hablanaian, J. Vac. Sci. Technol. 18, 1158 (§1981).

[5] M. Pirani and J. Yarwood, Principles of Vacuum Engineering, Chapman and Hall, London (1961).

[6] O. Grobner, private communication (1999).

[7] C.H. Kruger and AH. Shapiro, Trans. 7th Nat. Vac. Symp, 6–12, Pergamon, New York (1961).

[8] C.H. Kruger, The Axial Flow Compressor in the Free-Molecular Range, Ph. D. Thesis, Department of Mechanical Engineering, M.I.T., Cambridge, MA (1960).

[9] K.H. Bernhardt, J. Vac. Sci. Technol., A12, 136 (1983).

[10] W. Gaede, Ann. Phys. (Leipzig) 4, 43, 337 (1913).

[11] M. Holweck, C. R. Acad. Sci., 177, 43 )1923).

[12] M. Siegbahn, Arch. Math. Astrom. Fys. 30B, 17 (1994).

[13] W. Becker, Vak. Tech. 16, 625 (1958).

[14] B. Cho, S. Lee and S. Chung, J. Vac. Sci. Technol., A 13, 4, 2228 (1995).

[15] International Standards Organisation, Geneva, Switzerland, http://www.iso.ch ISO 1607-2:1989, ISO 1607-1:1993.

[16] M.H. Hablanaian, J. Vac. Sci. Technol., A 5, 2552, 1987.

[17] Pneurop British compressed air society, London, UK, Vacuum Pumps Acceptance Specifications, Part 1 Positive Displacement Pumps.

[18] International Standards Organisation, Geneva, Switzerland, http://www.iso.ch ISO/CD 5302 1998.

# SPUTTER-ION PUMPS

*L. Schulz*
Paul Scherrer Institut, Villigen, Switzerland

**Abstract**
The following paper outlines the working principles of sputter-ion pumps. The pumping mechanisms for reactive and inert gas species, as well as argon instability are explained. The operating pressure range, special designs for integrated linear pumps in accelerators and operational aspects are discussed.

## 1. INTRODUCTION

In the ultrahigh vacuum technique applied for accelerator vacuum systems special pumping methods are required. Capture pumps, which trap the pumped gas molecules in the pump body, are the dominating pumps in the ultrahigh- and high-vacuum (UHV and HV) region of accelerator systems. Capture pumps have many advantages: they form a closed vacuum system with the vacuum chamber of the accelerator; no additional valves with complicated valve interlocks are necessary; there are no moving parts in the pump and no vibrations are transmitted to the magnet lattice.

The principal pumping mechanism employed is chemical transformation whereby gases are chemically combined into solid compounds with very low vapour pressure. At UHV and HV conditions a surface can hold large quantities of gases compared to the amount of gas present in space. A pumping action can be produced by physisorption or gettering, which refers to a chemical combination between the surface and the pumped gas.

Many chemically active materials can be used for gettering. In vacuum systems titanium is commonly used because it is chemically reactive with most gases when it is deposited on a surface as a pure metallic film, but it is rather inert in bulk form because of the tenacious oxide film covering its surface. Pumps using chemical and ionisation pumping effects can generally be called sputter-ion pumps. Early designs (after 1955) had a variety of arrangements for electron sources and for titanium evaporation. Today the most common designs are based on a Penning cell [1, 2] and are called sputter ion pumps because the supply of a fresh titanium film is produced by a process called sputtering.

## 2. PUMPING MECHANISM

The pumping effect of sputter-ion pumps is produced by sorption processes, which are released by ionised gas particles. The pumping speed is achieved by parallel connection of many individual Penning cells.

A sputter-ion pump consists basically of two electrodes, anode and cathode, and a magnet (Fig. 1). The anode is usually cylindrical and made of stainless steel. The cathode plates positioned on both sides of the anode tube are made of titanium, which serves as the gettering material. The magnetic field is orientated along the axis of the anode. Electrons are emitted from the cathode due to the action of an electric field and, due to the presence of the magnetic field, they move in long helical trajectories which improves the chances of collision with the gas molecules inside the Penning cell (Fig. 2). The usual result of a collision with the electron is the creation of a positive ion that is accelerated to some kV by the anode voltage and moves almost directly to the cathode. The influence of the magnetic field is small because of the ion's relatively large atomic mass compared to the electron mass.

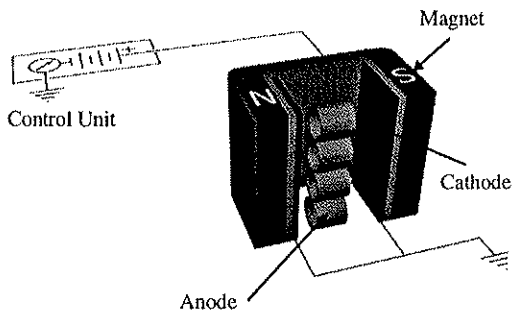



Fig. 1 Configuration of a sputter-ion pump.

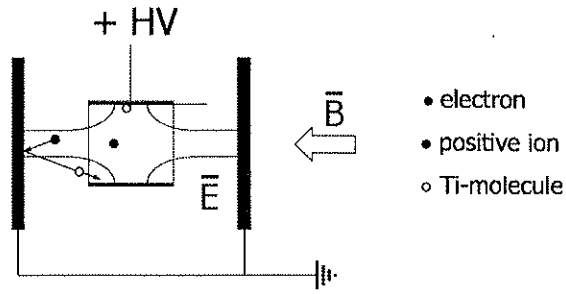



Fig. 2 Penning cell.

Ions impacting on the titanium cathode surface sputter titanium away from the cathode forming a getter film on the neighbouring surfaces and stable chemical compounds with the reactive or "getterable" gas particles (e.g. CO, $CO_2$, $H_2$, $N_2$, $O_2$). This pumping effect is very selective for the different types of gas and is the dominating effect with sputter ion pumps. The number of sputtered titanium molecules is proportional to the pressure inside the pump. The sputtering rate depends on the ratio of the mass of the bombarding molecules and the mass of the cathode material. The higher this ratio, the higher is the sputtering rate. For hydrogen, the lightest gas molecule, the sputtering rate of titanium is negligible.

In addition to the sputtering process a second important effect can be observed. The energy of the ionised gas particles allows some of the impacting ions to penetrate deeply (order of magnitude 10 atomic layers) into the cathode material. This sorption process pumps all kinds of ions, in particular ions of noble gases which do not react chemically with the titanium layer formed by sputtering. However, this pumping effect is not permanent since due to the erosion of the cathode material previously implanted molecules are released.

The cathode sorption process also works for hydrogen. Large amounts of hydrogen ions can diffuse deep into the bulk material beyond the range of the implanting ions and are permanently buried there.

## 3. OPERATING PRINCIPLES

### 3.1 Standard diode

The configuration described above is typically referred to as a standard diode pump. The anode cells are electrically isolated from the pump body and work with a positive voltage while the two cathode plates, made of titanium, are at ground potential. The electrodes are contained in the pump body, and the magnetic field is induced by external permanent magnets (Fig. 1).

A cell's pumping speed depends on several parameters i.e. its diameter, length, electrical and magnetic field and these parameters have been optimised in several theoretical studies and experimental tests [3–7]. Common designs for sputter-ion pumps use anode cell diameters between 15 and 25 mm, magnetic fields between 1 and 1.5 kGauss, and electrical fields between 3 and 7 kV. The ratio I/P (pump current/pressure) which is the main parameter of a Penning cell reaches values between 3 and 25 Ampere/mbar in such a configuration while the typical pumping speed for one cell is between 0.3 and 2 liters/second.

The diode pump has the highest pumping speed for all getterable gases but only low pumping speed for noble gases. For argon, the most common noble gas (1% in air) the pumping speed in a standard diode is only 2–5 % of the nominal pumping speed.

### 3.2 Noble gas pumping mechanism, argon instability

When pumping noble gases the following mechanism can be observed. Some of the noble gas ions that bombard the cathode surfaces become neutralised and rebound maintaining a part of their energy. They can then reach, unaffected by magnetic or electrical fields, the anode or the pump walls or other zones of the cathode. Energetic neutrals have far greater penetrating depth than ions that are quickly decelerated in the bulk materials by coulomb forces.

Argon instability occurs when the standard diode pump is pumping on an "air leak". The high rate of ion bombardment in the centre of the standard Penning cell erodes away cathode material. Previously buried molecules are released after the pump has removed a relatively large amount of argon. The pump dramatically and suddenly releases part of the buried molecules in sharp and repeated bursts of pressure.

### 3.3 Noble gas stabilised diode

A solution to argon instability is to increase the number of gas ions that "bounce back" as neutrals and sink into the anode or the pump body, where they are permanently buried. The reflection probability is a function of the mass ratio of the ion species and the cathode material and depends also on the incidence angle of the impinging ions on the cathode surface.

When an ion bombards the surface of a titanium cathode and is neutralised much of the kinetic energy of the ion is absorbed in the lattice structure of the bulk material. The penetrating power of the rebounded neutrals can be increased when for example a titanium cathode is replaced by a tantalum cathode. Tantalum is also a chemically active material but with a very heavy atomic mass (Ta = 181 amu, Ti = 48 amu). The number of elastic collisions is increased when an ion hits on a tantalum surface. The neutralised and rebounded ion now maintains much of its original kinetic energy and can be buried deeply in the anode or pump wall. For this type of pump, commercially referred to as a "noble diode" or "differential ion (D-I) pump", higher pumping speeds for noble gases have been observed, characteristically 20% of the pumping speed for air under stable conditions. However, with the same configuration of Penning cells the overall sputter rate for titanium and tantalum is reduced which leads to a reduction of the total pumping speed for getterable gases (typically by 15%) compared to the standard diode.

A different approach to increase the number of "rebounded" ions is the design of a diode pump with "slotted cathodes". In a standard diode the ions bombard the cathode more or less under normal incidence. With the slotted cathode (Fig. 3) more favourable incidence angles can be obtained. The reflection probability and the sputter yield of cathode material is much higher for glancing incidence than for normal incidence angles. A pumping speed for argon of 5–10% of the pumping speed for air can be achieved and is quite stable and leads to a higher pump capacity for getterable gases.

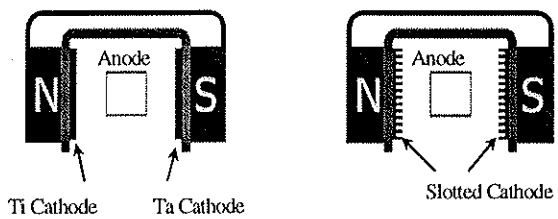



Fig. 3 Diode cell with tantalum and slotted cathode

### 3.4 Triode

In the triode configuration the effect of reflected neutrals due to glancing incidence on the cathode surface is increased. The cathode plates are replaced with a grid made of several titanium and is powered with negative high voltage (Fig. 4). The same relative voltage potentials as in the diode configuration are obtained when the anode and the auxiliary electrode are connected to ground potential.

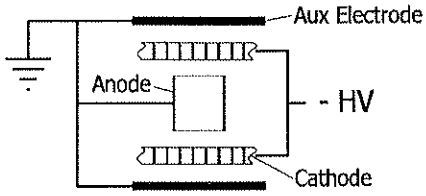



Fig. 4 Triode cell configuration.

In the triode the energetic ions bombard the cathode grid under glancing incidence with a high sputter yield of titanium. The energetic neutrals created by ion impingement on the cathode grids hit the auxiliary electrode or are reflected back and buried in the anode.

For energetic ions it is impossible to reach the pump body or anode. There is no sputtering of the body or anode and sorbed noble gases can remain buried on these surfaces. This leads to a very stable pumping of noble gases even after long-term pumping. The pumping speed for noble gases is increased up to 20–25% of the air pumping speed. Here, with the same number of anode cells of the same magnetic gap, the length of the anode cells is reduced and leads to a reduction of the total pumping speed for getterable gases (typically by 20%) compared to the standard diode. At high pressure (>$10^{-6}$ mbar) the triode reaches a higher pumping speed compared to a diode pump.

## 4. PRESSURE RANGE

The pumping speed of ion sputter pumps varies with pressure. The operating pressure is in the range of less than $10^{-4}$ mbar since at higher pressures the space charge in the Penning cells changes into a glow discharge and the sputtering process stops. The maximum pumping speed, also called the nominal pumping speed ($S_N$) is reached at about $10^{-6}$ mbar (Fig. 5).

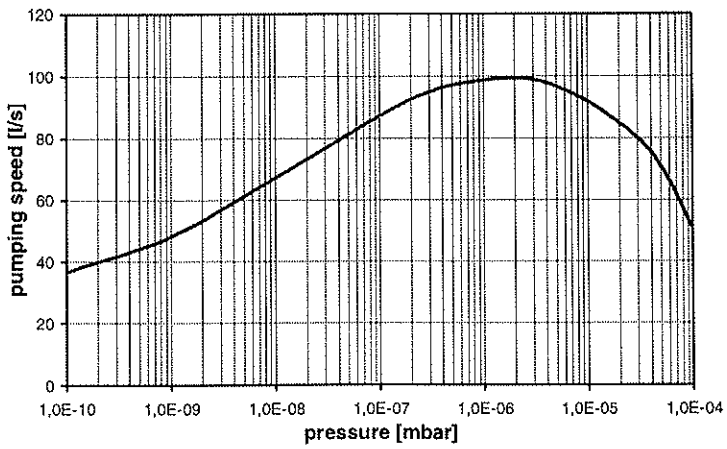



Fig. 5 Pumping speed vs pressure for a standard diode with $S_N$ = 100 l/s (for air after saturation).

Down to the base pressure the pumping speed decreases. The base pressure is limited by the equilibrium between pumped and desorbed molecules and not by the absence of discharge. The ultimate lowest pressure range of $10^{-11}$ and $10^{-10}$ mbar can only be reached after a bakeout of the ion sputter pump [8, 9].

## 5. SPECIAL DESIGN FOR INTEGRATED LINEAR PUMPS

For pumping systems in particle accelerators, especially in storage rings, distributed sputter-ion pumps (Fig. 6) can be used [5–7, 10]. The pump elements typically are integrated in the vacuum chamber of the bending magnets that provide the required magnetic field. In electron storage rings the beam induced gas desorption is proportional to the electron energy which is also proportional to the field of

the bending magnets with ranges typically between 0.5 and 15 kGauss. The resulting pumping speed of the distributed sputter-ion pumps is proportional to the magnetic field.

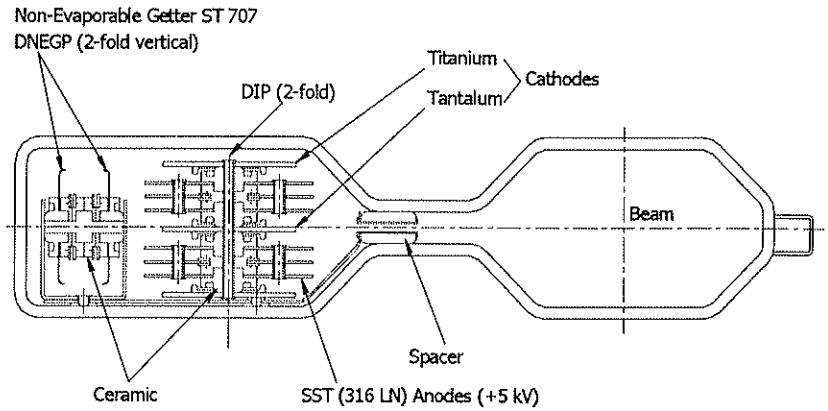



Fig. 6 Distributed sputter-ion pump in the DELTA storage ring.

Use of distributed sputter-ion pumps is ideal when long pump elements are installed in the vacuum chambers and when the desorption rate has a homogenous distribution in the longitudinal direction. The resulting pumping speed depends on the magnetic field in the bending magnets. When the accelerator is operated over a wide range of the magnetic field the design of the pump elements has to be optimised to take this into account.

Common designs for distributed sputter-ion pumps often use laminar anodes instead of anode tubes. The cylindrical anode cells are formed by stacked stainless-steel plates with coaxial holes [11, 12]. The anodes are stacked with a gap between the stainless-steel plates which leads to a higher conductance in the transverse direction.

## 6. OPERATIONAL ASPECTS

As mentioned above the cathode of a sputter-ion pump is sputtered away due to the impacting ions. This leads to a limited lifetime of the pumps when they are operated at high pressure. For example at a pressure of $10^{-4}$ mbar the cathode is eroded in 400 hours, but has a lifetime of over 40000 hours at a pressure of $10^{-6}$ mbar. To achieve high pump lifetimes it is recommended to start the sputter-ion pumps at pressures less than $10^{-6}$ mbar.

The lower pressure limit of sputter-ion pumps is in the range of $10^{-11}$ to $10^{-10}$ mbar. Lower pressures in the range of $10^{-12}$ mbar can only be achieved when the sputter-ion pump works in a combination with other pumping methods. Well established are the combinations of a sputter-ion pump with a titanium sublimation pump (TSP) or a non-evaporable getter module (NEG) [13].

The current measured in a sputter-ion pump is in the typical operation range, proportional to the pressure (Fig. 7). This gives additional pressure information concerning the vacuum system that results in a good overview of the pressure distribution in accelerators where usually a huge number of pumps are installed.

By connecting several sputter-ion pumps to only one power supply the number of power supplies can be reduced. Individual pump currents can be measured by using shunt resistors that are installed in a high voltage splitter between the pumps and power supply [14].

## 7. CONCLUSIONS

For most accelerators excellent vacuum conditions are required. Therefore the highest possible pumping rates are necessary. For most applications the standard diode pump meets these requirements and is optimal from a financial point of view.

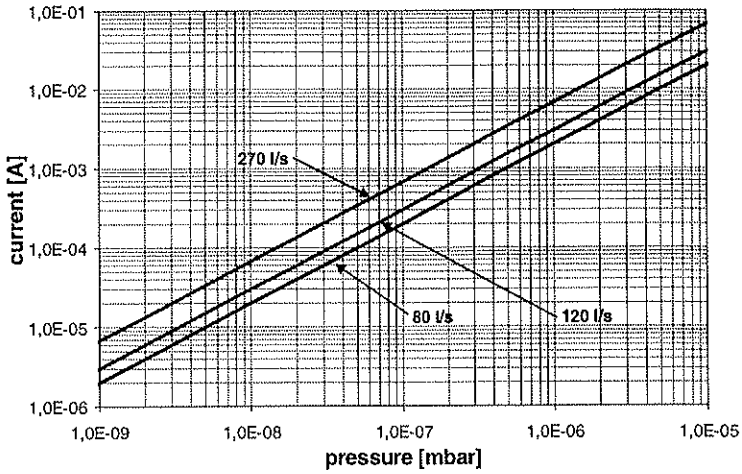



Fig. 7 Pump current vs pressure for different sputter-ion pumps.

However, to achieve the ultimate pressure regime of $10^{-12}$ and $10^{-11}$ mbar a sputter-ion pump combination with TSP or NEG-modules is needed. At high pressure ranges (>$10^{-7}$ mbar) combination pumps become less effective. These high pressure ranges are the field of the triode pump.

The diode pump with tantalum cathode or the triode pump is optimal for vacuum systems which need high pumping speed for noble gases

## REFERENCES


[1] F.M. Penning, Philips Technical Review, 2 (1937) 201.

[2] W. Schuurman, Investigation of a low pressure Penning discharge, Physica, **36** (1967) 136.

[3] L. Rutherford, Trans 10th Nat Vac Symp, AVS (1963).

[4] K.M. Welch, Capture Pumping Technology, Pergamon Press Ltd., Oxford, 1991.

[5] M.D. Malev, and E.M. Trachtenberg, Built-in getter-ion-pumps, *Vacuum*, **23** (11) 1973 pp. 403.

[6] H. Hartwig and J. Kouptsidis, A new approach for computing diode sputter-ion pump characteristics (for particle accelerators), J. Vac. Sci. Technol., **11** (6) (1974).

[7] W. Ho, R.K. Wang, and T.P. Keng, Calculation of sputter ion pump speed, J. Vac. Sci. Technol., **20** (4) (1982).

[8] M. Audi, Ion pumps, *Vacuum* **37** (8-9), (1987) pp. 629.

[9] M. Audi, Pumping speed of sputter ion pumps, *Vacuum* **38** (8-10), (1988) pp. 669.

[10] J.M. Laurent, and O. Gröbner, Distributed Sputter-Ion Pumps for use in Low Magnetic Fields, Particle Accelerator Conference, San Francisco (1979).

[11] B. Hippert and N. Marquardt, The DELTA vacuum system, European Particle Accelerator Conference, Sitges (1996).

[12] H. Pingel, and L. Schulz, Proc. 8th Intern. Vac. Congress, Cannes, 1980.

[13] M. Audi, A new ultrahigh vacuum combination pump, J. Vac. Sci. Technol., **A5** (4) (1987).

[14] F. Giacuzzo, and J. Miertusova, Total pressure measurements in the ELETTRA storage ring according to the performance of the sputter-ion pumps, European Particle Accelerator Conference, Sitges (1996).

# MOLECULAR SURFACE PUMPING: THE GETTER PUMPS

*C. Benvenuti*
CERN, Geneva, Switzerland

**Abstract**
A surface may provide a useful pumping action when able to retain adsorbed gas molecules for the duration of a given experiment. To fulfil this condition at room temperature, strong binding forces, as those resulting from chemical reactions, are required. Materials able to react with gases to form stable chemical compounds are called getters. The two main families of getters (evaporable and non-evaporable, or NEG) are presented and discussed. Special emphasis is placed on the NEG strips presently used for the vacuum systems of particle accelerators, and on the newly developed NEG thin film coatings, in view of their possible future applications.

## 1. GENERALITIES

### 1.1 Molecular mean free path λ

$$\lambda = \frac{7.3T}{10^{20}\,\pi r_0^{\,2}\,p}\ \text{cm}$$

where $T$ is the absolute temperature (K), $p$ the pressure (Torr) and $\pi r_0^{\,2}$ the collision cross section ($cm^2$). For $N_2$, $\pi r_0^{\,2} = 4.26 \times 10^{-15}$ $cm^2$. At 295 K, $\lambda = 5 \times 10^3\,p$ (i.e. 5 cm at $10^{-3}$ Torr, or $5 \times 10^6$ cm at $10^{-9}$ Torr).

Only molecular flow ($\lambda >>$ wall distance) will be considered hereafter.

### 1.2 Conductance and surface pumping

The conductance $C$ of an orifice (zero wall thickness approximation) may be expressed as

$$C = 3.64\,(T/M)^{1/2}\ \ell\ s^{-1}\ cm^{-2}$$

with $T$ = absolute temperature (K), $M$ = molecular weight.

If all molecules impinging on a surface are captured, $C$ represents the specific pumping speed $S$ of the surface. More generally

$$S = \alpha\,3.64\,(T/M)^{1/2}\ \ell\ s^{-1}\ cm^{-2}$$

where $\alpha$ is the *sticking probability* ($0 \le \alpha \le 1$).

For $\alpha = 1$, $S \cong 44\ \ell\ s^{-1}\ cm^{-2}$ for $H_2$ and $S \cong 12\ \ell\ s^{-1}\ cm^{-2}$ for $N_2$ at room temperature.

### 1.3 Surface adsorption capacity

The monolayer capacity of an atomically flat surface is of the order of $5 \times 10^{14}$ molecules $cm^{-2}$. If this monolayer of gas is released in a spherical volume of 1 ℓ (surface area ~ 500 $cm^2$) the pressure increase would be $7 \times 10^{-3}$ Torr. Therefore *surface degassing* is the main obstacle to achieving UHV conditions. Conversely under UHV conditions, *surface pumping* may be adopted.

### 1.4 Surface saturation time (monolayer formation)

Combining surface pumping speed and capacity, the surface saturation time $t_{sat}$ becomes

$$t_{sat} \cong 10^6\, p \text{ seconds } (p = \text{pressure in Torr})$$

when assuming sticking probability = 1, atomically flat surface, ambient temperature, $N_2$ gas (M = 28). Therefore, $t_{sat} \cong 1$ second at $10^{-6}$ Torr under these assumptions.

## 1.5 Sojourn time, desorption probability

For a molecule adsorbed with a binding energy $E$ on a surface at a temperature $T$ the *probability of escape* $\nu$ is

$$\nu = \nu_0 e^{\frac{-E}{RT}} \quad \text{(Frenkel, 1924)}$$

where $\nu_0$ is an attempt frequency, of the order of the vibrational energy of the adsorbed molecules ($\nu_0 \cong 10^{13}\ s^{-1}$) and $R = 1.98 \times 10^{-3}$ kcal mole$^{-1}$ K$^{-1}$.

It is often more useful to use the reciprocal of this term, $t = 1/\nu$, known as *mean sojourn time* or *mean surface lifetime*

$$\tau = \tau_0 e^{\frac{E}{RT}}$$

Obviously, a surface provides a good pumping action when $\tau$ is much larger than the duration of the experiment. Typical values of $\tau$ are given in Table 1 [1]. Large $\tau$ values imply large $E$. However, surfaces providing low $E$ values may display a useful pumping action when reducing the working temperature. The Frenkel equation is often used in its logarithmic form. By plotting the logarithm of the measured pressure versus $1/T$ a straight line is obtained the slope of which defines the energy of the considered adsorption process.

Table 1

Mean stay time for adsorbed molecules at 300 K for various values of the adsoprtion energy, assuming $\tau_0 = 10^{-13}$ s

| Energy (kcal/mol) | Typical cases | $\tau_a$ (s) |
|---|---|---|
| 0.1 | He | $1.2 \times 10^{-13}$ |
| 1.5 | $H_2$ physisorbed | $1.3 \times 10^{-12}$ |
| 3.5–4 | Ar, CO, $N_2$, $CO_2$ (physisorbed) | $1 \times 10^{-11}$ |
| 10–15 | Weak chemisorption | $3 \times 10^{-6}$ |
| | Organics physisorbed | $2 \times 10^{-2}$ |
| 20 | $H_2$ chemisorbed | 100 |
| 25 | | $6 \times 10^5$ (1 week) |
| 30 | CO chemisorbed on Ni | $4 \times 10^9$ (> 100 yr) |
| 40 | | $1 \times 10^{17}$ (≈ age of the earth) |
| 150 | O chemisorbed on W | $10^{1100}$ (≈ $10^{1090}$ centuries) |

## 1.6 Gas-surface binding energies

Two types of forces may bind a gas molecule to a surface:

- *Chemical forces* involving electrons, e.g. hydrogen bonding, covalent or metallic bonding, characterised by binding energies typically in the range of some eV per molecule or larger than 10 kcal/mole (0.4 eV/molecule ≅ 10 kcal/mole).
- *van der Waals forces* of electrostatic nature, e.g. dispersion or polar forces, characterised by smaller binding energies, lower than 0.4 eV/molecule or 10 kcal/mole.

In the first case (chemical adsorption or chemisorption) long sojourn times are possible at room temperature (*getter pumping*). In the second case (*physical adsorption or physisorption*) pumping requires surface cooling (*cryopumping*) because the mean sojourn time at room temperature is too short (see Table 1).

### 1.7 Definition of getters and getter pump types

*Getters* are materials able to fix gas molecules on their surface in the form of stable chemical compounds. To do so, their *surface* must be *clean*. There are two ways of producing a clean gettering surface

- by "in situ" deposition of a fresh getter film
- by heating an oxidised getter to a temperature high enough to diffuse oxygen from the surface into the getter bulk.

In the first case we speak about *evaporable getters*, in the second case about *non-evaporable getters (NEG)* and the required heating temperature is called *activation temperature*.

## 2. EVAPORABLE GETTERS

The two materials most widely used as evaporable getters are *barium* and *titanium* [2].

*Barium* is usually sublimated from a $BaAl_4$ alloy by heating at ~ 900°C. It is used for pumping vacuum sealed devices (electron tubes) and the sublimation is done in one single process before sealing. It is not used for UHV applications and will not be discussed here. For more details see Ref. [3].

*Titanium* is the most widely used evaporable getter for UHV applications. It is usually sublimated from filaments made of Ti alloys (with Mo or Ta) heated up to 1500°C, temperature at which the Ti vapour pressure is about $10^{-3}$ Torr. Titanium films provide *sticking probabilities* of $1-5\times 10^{-2}$ for $H_2$ and $0.4-0.6$ for CO at room temperature. Cooling to liquid $N_2$ temperature enhances these values to $0.1-0.3$ for $H_2$ and about 1 for CO [4]. Other materials (Ta, Nb, V, Zr, Mo) have also been used, but their behaviour is not as good for various reasons [5].

*Warning*: Sticking probabilities reported in the literature for a given gas-getter combination present a large spread. Besides possible experimental errors, this spread has two distinct causes, namely:

- *surface roughness* which may result in a large number of molecule-surface interactions, i.e. higher capture probability.
- *initial surface contamination* which may reduce the adsorption site density available for pumping; this effect is particularly important for experimental systems which provide a base pressure higher than about $10^{-9}$ Torr and/or when the sticking probability is measured at high pressures ($10^{-7}-10^{-6}$ Torr).

At room temperature all gases but $H_2$ adsorbed by a Ti film remain on the surface, resulting in a progressive reduction of pumping speed (surface blocking). On the contrary, $H_2$ diffuses and its pumping speed is not affected by the pumped amount. On the other hand, the high (> 30 kcal/mole) binding energies prevent the *desorption* of gases adsorbed on Ti at practically allowed temperatures. Again $H_2$ represents an exception, since its lower binding energy (~ 20 kcal/mole) allows desorption by heating (see Table 1).

The initial pumping speed of a Ti sublimation pump may be restored by a further sublimation process. The total *pumping capacity* is therefore very large and depends on the available amount of Ti in the filament. The *ultimate pressure* of a Ti sublimation pump is in principle not limited; in practice it may be spoiled by the presence of rare gases and methane if an adequate pumping for these gases is not foreseen.

In conclusion, Ti sublimation pumps provide the following advantages and disadvantages:

- *Advantages:* large pumping speed, large pumping capacity, unlimited ultimate pressure, compact, inexpensive, easily operated.
- *Disadvantages:* no pumping for inert gases, $CH_4$ production (?) and localised pumping (not very suitable for conductance-limited vacuum systems, as those of particle accelerators).

## 3. NON-EVAPORABLE GETTERS (NEG)

NEGs are usually produced by fixing a powder of the getter material to a (metal) substrate by pressing, sintering or cathaphoresis [6]. After insertion in the system to be pumped, activation is carried out by heating. Usually NEGs are alloys of the elements of the IV B column of the Periodic Table, to which some of the actinides and rare earths may be added. Also Al is often added to increase the diffusivity of the adsorbed gases when heating. For more details on this subject see [7].

The performance of a given NEG is characterised by *activation temperature, sticking probability, surface capacity, total pumping capacity* (for $H_2$ and for heavier gases) and *particulate loss*. Since NEGs are highly porous (and the small grains are in poor contact with the substrate), the danger of *pyrophoricity* imposes a lower limit to the activation temperature (350 – 400°C). If compared to Ti sublimation pumping, NEG pumping presents the risk of *powder peel-off* (excessive heating or $H_2$ embrittlement) and a *lower pumping capacity*; however, NEGs may provide *linear pumping* and *passive activation*. If the activation temperature is compatible with the baking temperature of the chamber where the NEG is inserted, the getter may be activated during bakeout. This feature is particularly attractive because it removes the need of electric feedthroughs and powering/control systems and allows increasing the NEG surface and consequently its pumping speed. All NEGs available on the market are produced by *SAES Getters*. We will consider here only the types St 101 and the St 707, which are the best suited for UHV applications.

| | St 101 [8–10] | St 707 [11, 12] |
|---|---|---|
| Composition | Zr 84%, Al 16% (at. %) | Zr 70%, V 24.6%, Fe 5.4% |
| Getter layer thickness | ~ 0.1 mm | ~ 0.1 mm |
| Activation | 750°C for about 30' | 400°C for about 1 hour |
| Porosity | ~ 10% | ~ 10% |
| Substrate | steel or constantan | steel or constantan |
| $H_2$ dissociation pressure*) | $\log(p_{H2}) = 4.28 + 2\log(q) - 7000/T$ | $\log(p_{H2}) = 5.14 + 2\log(q) - 6250/T$ |

*) pressure $p$ in mbar, $H_2$ quantity $q$ in mbar $\ell$ $g^{-1}$, $T$ in Kelvin.

*Note* that the equilibrium pressure of $H_2$ over the St 707 (same temperature and $H_2$ concentration) is about two orders of magnitude higher compared to St 101.

*Warning:* Due to its low activation temperature, the St 707 may be *ignited by spot welding*.

Both the St 101 and the St 707 have been extensively studied at CERN. The St 101 has been selected to provide the main pumping for the Large Electron Positron Collider (LEP) [13] (see Fig. 1).

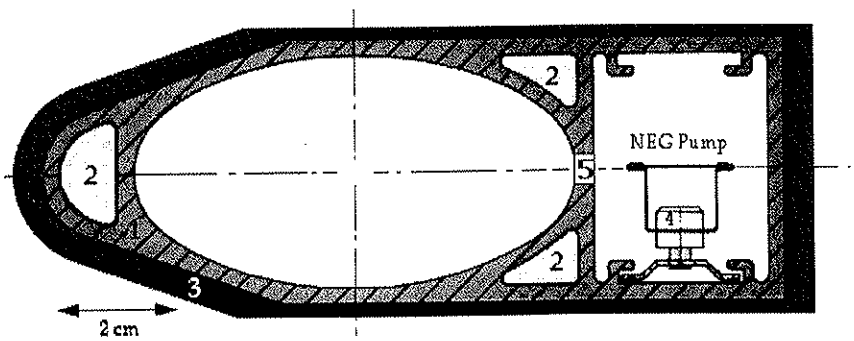



Fig. 1 Cross section of the LEP dipole vacuum chamber.

The main results are the following:

***Ultimate pressure***

A LEP chamber (12 m long, made of Al alloy, baked at 150°C pumped by St 101 and a sputter-ion pump of about 30 $\ell\ s^{-1}$ speed reaches an ultimate pressure of about 2 x $10^{-12}$ Torr, mainly due to Ar and $CH_4$, gases not pumped by the getter. By adding six additional sputter-ion pumps a pressure of about $5 \times 10^{-13}$ Torr (mainly $H_2$) is obtained [14]) see Figs. 2 and 3). Making use of the St 707 fully covering the inner walls of a 3 m long stainless steel chamber of 160 mm diameter (see Fig. 4), pressures in the low $10^{-14}$ Torr have been achieved after passive activation during a 350°C bakeout [15].

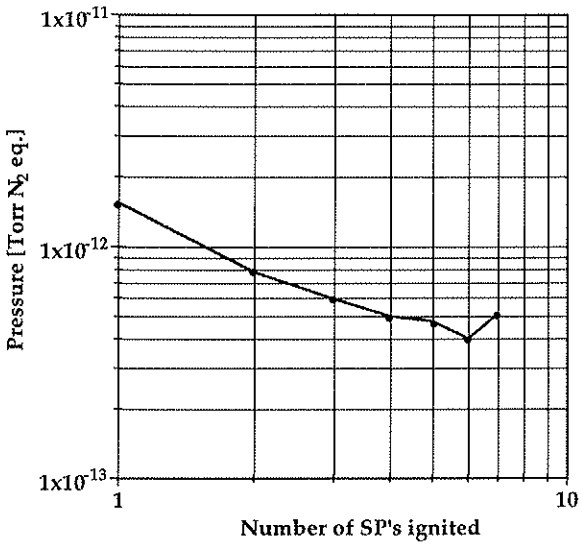



Fig. 2 Variation of the total pressure measured on a 12-m long LEP chamber equipped with a NEG pump and seven sputter-ion pumps, as a function of the number of the sputter-ion pumps ignited.

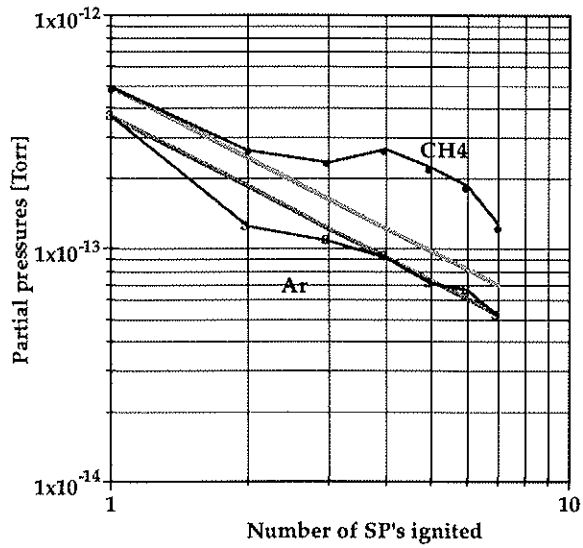



Fig. 3 Same variation as in Fig. 2, for the argon and methane partial pressures.

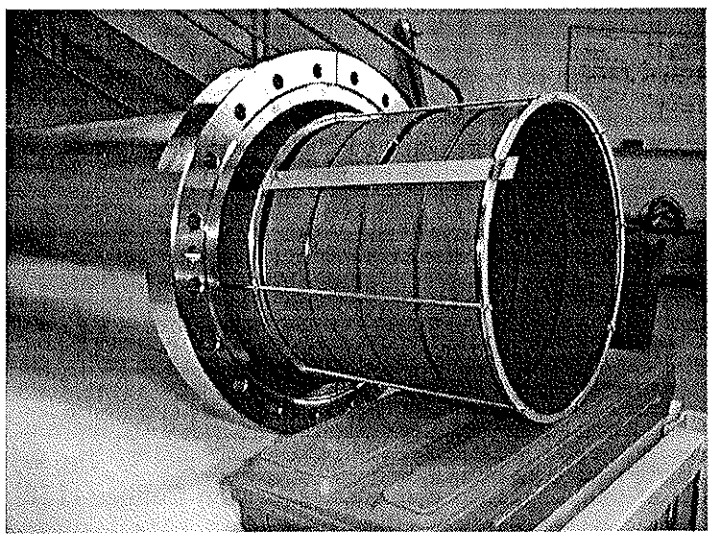


Fig. 4 Schematic view of the St 707 "total" NEG pump.

***Pumping speed variation as a function of the amount of gas pumped***

The pumping curves obtained at room temperature for the St 101 [16], represented in Fig. 5, indicate that, after saturation of the external surfaces, the pumping speed is limited by the conductance to the internal pores of the getter coating. A mathematical model of the pumping process points out that, for a given gas load in the coverage range of practical interest, the NEG pumping speed is proportional to the square of the coating porosity. This model allows estimating the porosity from the $S(Q)$ curves, in good agreement with the values obtained by direct porosity measurements. The developed model also predicts that CO is not dissociated on the St 101 (single site adsorption) while $H_2$ is dissociated (2 sites adsorption) as well as $N_2$, which however occupies six to eight adsorption sites. When *gas mixtures* are pumped [17], CO was found to inhibit the pumping of other gases, while $N_2$ has a small surface blocking effect and $H_2$ does not produce any blocking at all.

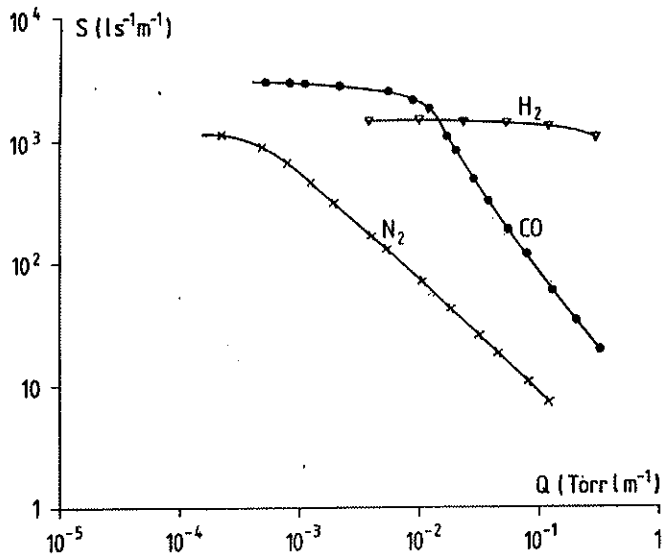



Fig. 5 Pumping speed variation of the St 101 NEG strip (30-mm wide) for $H_2$, CO, $N_2$ as a function of the pumped gas amount.

## 4. THIN-FILM GETTER COATINGS

Thin-film getter coatings, produced by sputtering, have been found to recover their chemical reactivity after exposure to ambient air followed by "in situ" baking [18, 19]. These coatings represent the last step of a process according to which the pump has been moved progressively closer to the vacuum chamber walls. The getter thin film coating actually transforms the vacuum chamber from *a gas source* into *a pump*.

Many metal coatings have been studied at CERN during the last 3 years, namely Ti, Zr, Hf, Nb, V and some of their binary and ternary alloys [18, 19]. The main goal was to reduce the activation temperature so as to allow activation also when using aluminium vacuum chambers, which cannot be baked at temperatures higher than about 200°C. This goal has been achieved with a TiZrV alloy [20]. This alloy would be highly pyrophoric when used as a powder, but does not present this risk in the form of a thin film, thanks to the thermal stabilisation provided by the much thicker substrate. The main results achieved using this coating are shown in Figs. 6 and 7. The study is still in progress in view of an application of this technique to the Large Hadron Collider (LHC) which will be built at CERN in the very near future.

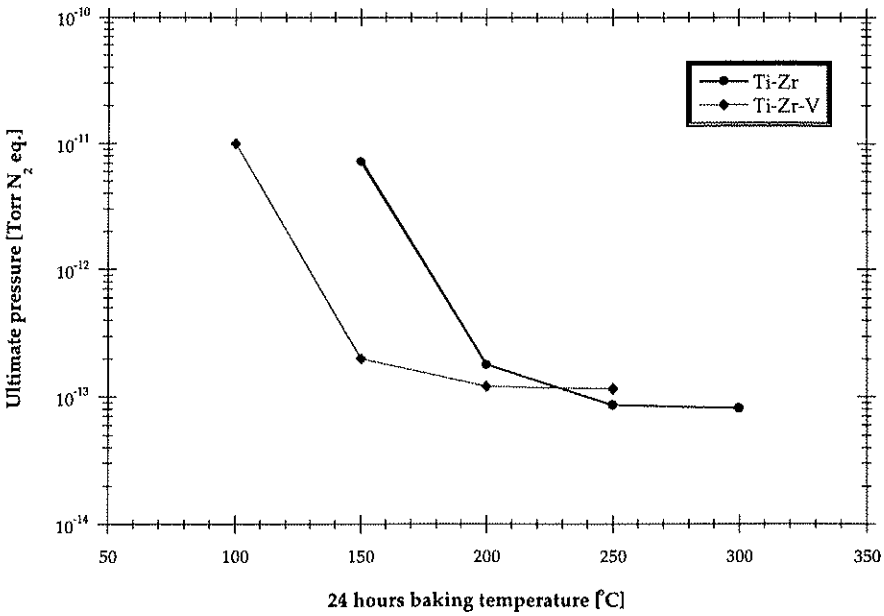



Fig. 6 comparison of the ultimate pressures achieved in stainless steel chambers coated with TiZr and TiZrV thin films. The chamber is 2-m long, its diameter is 100 mm, and it is connected at one extremity to a sputter-ion/titanium sublimation pumping station via an orifice of 25 $\ell s^{-1}$ conductance for $H_2$. The pressure is measured at room temperature at the opposite chamber extremity after a 24-hour baking at the indicated temperatures (without air exposure between bakeouts).

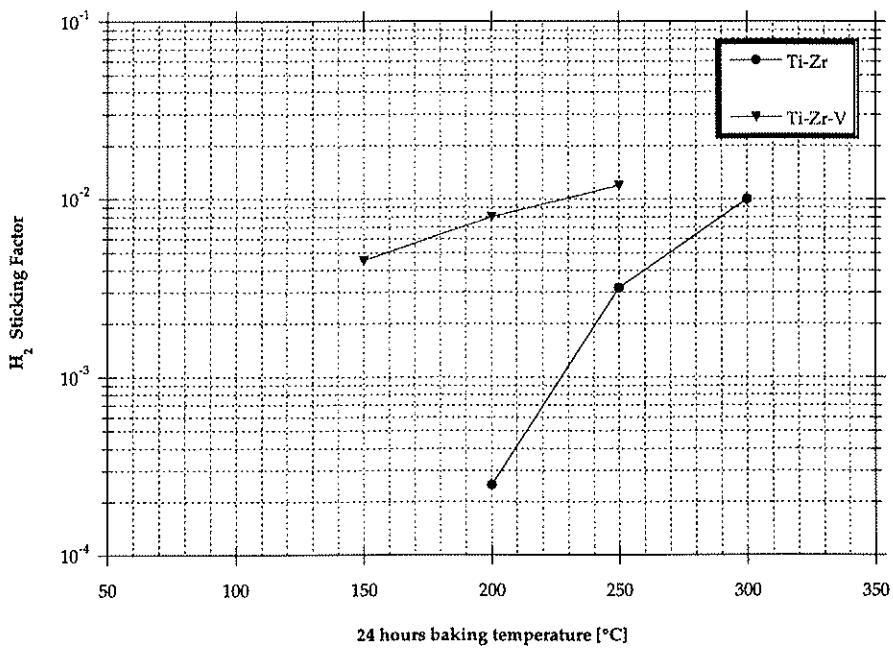



Fig. 7 Comparison of the $H_2$ sticking factors under the same conditions as described for Fig. 6.

Compared to traditional NEG strips, the thin film getter coatings provide the following *advantages:*

- reduced degassing
- lower activation temperature (200°C)
- wider choice of substrate possible

- materials may be produced which could not be produced otherwise (metastable and amorphous alloys)
- no space is required (coatings are a few μm thick)
- very low secondary electron emission.

In spite of being thinner than the powder coatings of traditional NEG strips by almost two orders of magnitude, thin film getter coatings may provide pumping speeds and total surface capacities very similar to those of the St 707 NEG. They may also withstand many activation/air venting cycles without excessive performance deterioration. A TiZr film has undergone over 20 such cycles without losing more than 50% of its original pumping speed for $H_2$.

**REFERENCES**


[1] J.M. Lafferty, Foundations of Vacuum Science and Technology, (John Wiley & Sons, Inc, New York, 1998) p. 551.

[2] Ibid., p. 275.

[3] Ibid., p. 276-290.

[4] P.A. Redhead, J.P. Hobson and E.V. Kornelsen, The Physical Basis of Ultrahigh Vacuum (American Institute of Physics, New York, 1993), p. 400.

[5] Idid., p. 402.

[6] G.L. Saksaganskii, Getter and Getter-Ion Vacuum Pumps, (Harwood Academic Publishers, Chur, 1994) p. 258.

[7] J.M. Lafferty, Foundations of Vacuum Science and Technology, (John Wiley & Sons, Inc, New York, 1998) p. 297.

[8] T.A. Giorgi, Jpn J. Appl. Phys. Suppl. 2, 1, 53 (1974).

[9] P. Della Porta, T.A. Giorgi, S. Origlio and F. Ricca, in Trans. of the 8th Nat. Vac. Symp. (Pergamon, New York, 1962) p. 229.

[10] D. Ferrario, A. Barosi, M. Borghi and T.A. Giorgi, in Proc. of the 9th Symp. on Fusion Technology (Pergamon, New York, 1976) p. 51.

[11] C. Boffito, B. Ferrario, P. Della Porta and L. Rosai, J. Vac. Sci. Technol. 18, 1117 (1981).

[12] C. Benvenuti and P. Chiggiato, J. Vac. Sci. Technol. A 14 (6), 3278 (1996).

[13] C. Benvenuti, Nucl. Instrum. Methods 205, 391 (1983).

[14] C. Benvenuti, J.P. Bojon, P. Chiggiato and G. Losch, Vacuum, 44, 507 (1993).

[15] C. Benvenuti and P. Chiggiato, Vacuum, 44, 511 (1993).

[16] C. Benvenuti and F. Francia, J. Vac. Sci. Technol. A 6, 2528 (1988).

[17] C. Benvenuti and F. Francia, J. Vac. Sci. Technol. A 8, 3864 (1990).

[18] C. Benvenuti, P. Chiggiato, F. Cicoira and Y. L'Aminot, J. Vac. Sci. Technol. A 16 (1), 148 (1998).

[19] C. Benvenuti, P. Chiggiato, F. Cicoira and V. Ruzinov, Vacuum, 50 (1-2) 57, (1998).

[20] C. Benvenuti, J.M. Cazeneuve, P. Chiggiato, F. Cicoira, A. Escudeiro Santana, V. Johanek, V. Ruzinov and J. Fraxedas, IVC14, Birmingham, (1998), proceedings to be published.

# MOLECULAR SURFACE PUMPING: CRYOPUMPING

*C. Benvenuti*
CERN, Geneva, Switzerland

**Abstract**
Weak van der Waals attractive forces may provide molecular gas pumping on surfaces cooled at sufficiently low temperatures. Cryopumps make use of a cold surface to adsorb gas molecules. They are usually classed into two families, depending whether they rely on the binding forces between gas molecules and solid surfaces (cryosorption pumps) or on the even weaker forces which bind condensed gas molecules together (condensation pumps). The design features and the possible applications of both types of cryopumps are presented. The peculiar behaviour of condensed hydrogen, i.e. its desorption induced by infrared thermal radiation, is described and discussed.

## 1. GENERALITIES

### 1.1 Gas density and pressure at T ≠ 300 K

A gas pressure is defined only if gas molecules present an *isotropic distribution of directions* and a *maxwellian distribution of speeds*. Under UHV conditions, pressure gauges do not measure pressure but *gas density*.

If two communicating vessels $V_1$ and $V_2$ are kept at different temperatures ($T_1$ and $T_2$) the gas densities ($d_1$ and $d_2$) and pressures ($p_1$ and $p_2$) are given by

$$p_1 = \sqrt{T_1/T_2}\ p_2$$

$$d_1 = \sqrt{T_2/T_1}\ d_2$$

In other words, in the colder vessel the gas density is higher but the pressure is lower. $\sqrt{T_1/T_2}$ is the *thermomolecular transpiration factor*.

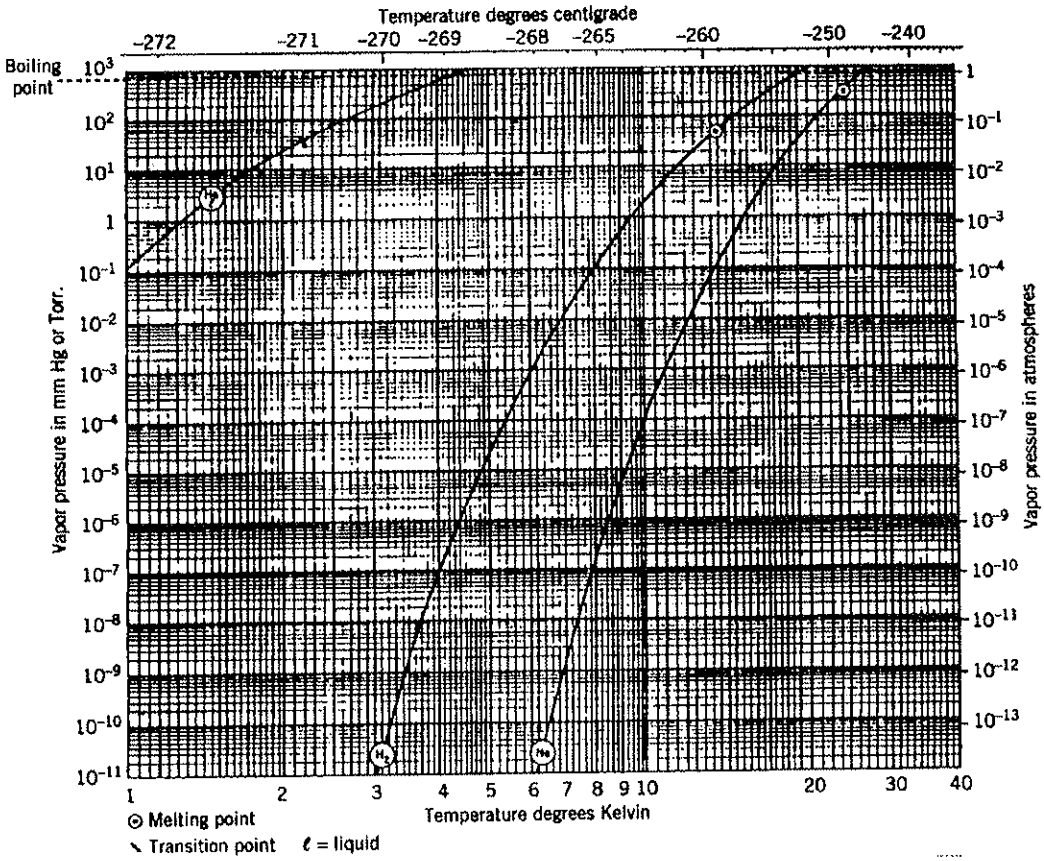
### 1.2 Gas condensation

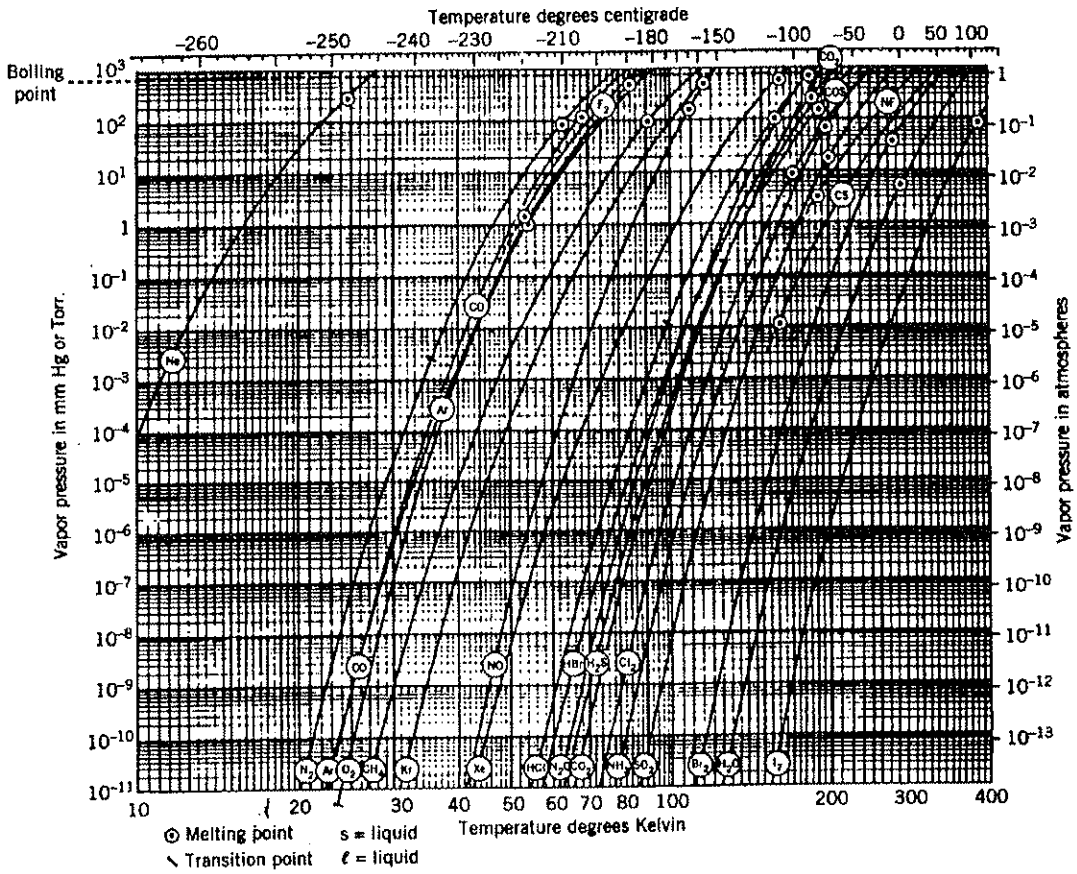
When a surface is "cold enough", the sojourn time of the molecules adsorbed on the surface, as defined by the Frenkel equation

$$\tau = \tau_0 e^{\frac{E}{RT}}$$

where $E$ is the energy of vaporisation (kcal mole$^{-1}$), $R$ the $1.98 \times 10^{-3}$ (kcal mole$^{-1}$ K$^{-1}$) and $T$ the temperature (K), may become appreciably long. The values of the *energy of vaporisation E* for the most common gases are given in Table 1 [1].

For a given gas and surface temperature, by progressively increasing the surface coverage a *saturation* equilibrium between gas adsorption and desorption is established. The corresponding gas pressure is the *saturated vapour pressure*. The saturated vapour pressure curves for the most common are given in Figs. 1 and 2 [2]. These figures indicate that most gases present a vapour pressure not higher than $10^{-11}$ Torr at 20 K, while the $H_2$ vapour pressure is in the $10^{-7}$ Torr range even at 4.2 K, the boiling temperature of He. Therefore indefinite heavy gas quantities may be effectively condensed on a surface kept at 20 K, while a temperature lower than 3 K for $H_2$ and still much lower for He would be required.




Figs. 1 and 2 Saturated vapour pressure curves for the most common gases.

Table 1
Energies of vaporisation of the common gases and energies of sublimation of common metals (kcal/mole)

| Gas | He | $H_2$ | Ne | $N_2$ | CO | Ar | $O_2$ | $CH_4$ | Kr | Xe | $CO_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Energy of vaporisation | 0·020 | 0·215 | 0·431 | 1·333 | 1·444 | 1·558 | 1·630 | 1·995 | 2·158 | 3·021 | 4·041 |
| **Metal** | **Ba** | **Ag** | **Cu** | **Ni** | **Ti** | **Pt** | **Mo** | **C** | **Ta** | **W** | |
| Energy of sublimation | 42 | 68 | 81 | 101 | 113 | 122 | 155 | 171 | 180 | 202 | |

Surface pumps relying on condensation are called *condensation cryopumps*. In a vacuum system pumped by a saturated cryopump the pressure is given by $p = Q/S + p_{sat}$ where $Q$ is the gas load, $S$ the pumping speed and $p_{sat}$ the saturated vapour pressure of the gas to be pumped.

## 1.3 Gas physisorption

For submonolayer surface coverages the gas molecules experience the attractive van der Waals forces exerted by the cold surface (*cryosorption*). These forces are larger than those existing between the gas molecules themselves (condensation). According to the Frenkel equation, cryosorption provides consequently *much lower equilibrium pressures*.

The energies of cryosorption for the most common gases on various solid substrates are given in Table 2 [3]. By comparing Tables 1 and 2, it may be seen that these energies are larger than those of vaporisation by a factor 2 to 3 for heavy gases, but *10* for $H_2$ and *30* for *He*. As a consequence, $H_2$ may be effectively cryosorbed at 20 K and He at 4.2 K. More generally, submonolayer quantities *of all gases may be effectively cryosorbed at their own boiling temperature* (1 bar). These temperatures are given in Table 3. It may be concluded that all gases except He, $H_2$, $D_2$ and Ne may be cryosorbed at liquid $N_2$ temperature. Furthermore, water vapour is adsorbed at room temperature; to remove it, heating up to above 100°C is required (*bakeout*).

Table 2
Binding energies for physical adsorption of common gases on heterogeneous surfaces at high temperatures and low coverages (kcal/mole)

| Solid | Gas | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | He | $H_2$ | Ne | $N_2$ | Ar | $O_2$ | $CH_4$ | Kr |
| Porous glass | 0·68 | 1·97 | 1·54 | 4·26 | 3·78 | 4·09 | | |
| Saran Charcoal | 0·63 | 1·87 | 1·28 | 3·70 | 3·66 | | 4·64 | |
| Carbon black | 0·60 | | 1·36 | | 4·34 | | | |
| Alumina | | | | | 2·80 | | | 3·46 |

Table 3
Boiling temperature (K) of gases at 1 bar

| | |
|---|---|
| $He_3$ | 3.2 |
| $He_4$ | 4.2 |
| $H_2$ | 20.3 |
| $D_2$ | 23.6 |
| Ne | 27.1 |
| $N_2$ | 77.3 |
| A | 87.2 |

| | |
|---|---|
| $O_2$ | 90.1 |
| $CH_4$ | 111.6 |
| Kr | 119.7 |
| Xe | 165.0 |
| $C_2H_4$ | 169.4 |
| $C_2H_6$ | 184.5 |
| | |

The curve describing the equilibrium pressure evolution obtained by progressively increasing the gas coverage on a given surface maintained at a given temperature is called the *adsorption isotherm*. The pressure described by the adsorption isotherms reaches the value of the saturated vapour pressure for gas coverages of one to three monolayers. A complete review of the different types of adsorption isotherms is given in Ref. [4].

Pumps relying on physisorption are called *cryosorption pumps*. An essential feature of these pumps is the need of a large adsorbing area. Porous materials, such as activated charcoal and zeolites are commonly used. Charcoal may provide adsorbing areas of the order of 1000 $m^2$ per gram, and may adsorb 3 x $10^{-6}$ Torr $\ell$ $cm^{-2}$ of He at 4.2 K before the pressure rises above $10^{-10}$ Torr [5].

## 2. CRYOGENIC CONSIDERATIONS

### 2.1 Surface cooling

Cryosorption pumps are usually cooled to 10–20 K by means of a closed cycle *refrigerator*. For pumps of small size (up to 5000 $\ell$ $s^{-1}$ pumping speed) the available cooling power is usually of the order of 10 W. For a given refrigerator, the temperature of the cryosurface depends on the thermal load (see Fig. 3). Since low temperatures are desired, the heat load to the cryosurface must be minimised (see next paragraph). To do so, the cold surface must be shielded at ~ 80 K, a temperature available from the first stage of the refrigerator, which provides a much larger cooling power.

For condensation cryopumping, cooling is achieved by means of *liquid helium*, usually kept at temperatures lower than 4.2 K by reducing the pressure over the He bath. The cold surface is shielded also in this case by means of liquid nitrogen. A heat load of 0.7 W vaporises 1 $\ell$ of He per hour, while ~ 45 W are required to vaporise 1 $\ell$ of $N_2$ per hour.

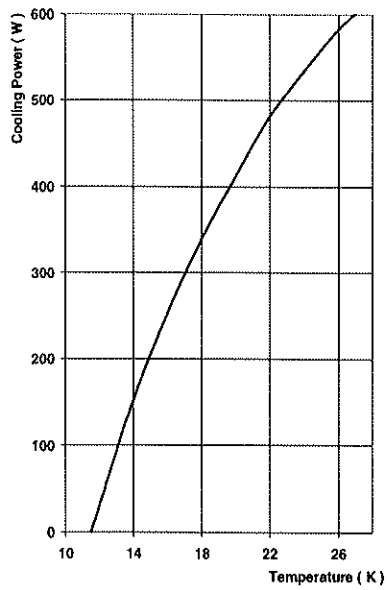



Fig. 3 Variation of the available cooling power as a function of the cold head temperature for a typical refrigerator. Since the required cooling power must be equal to the thermal load, higher values of the latter result in higher cryosurface temperatures.

### 2.2 Thermal load

A cold surface may be thermally loaded via thermal conduction, gas adsorption and thermal radiation absorption.

By properly designing the pump, the thermal conduction may be reduced to negligible values. For UHV applications also the heat of adsorption released by the pumped gas is negligible. On the contrary, the radiative losses may by very large. A surface at temperature T emits thermal (electromagnetic) radiation according to the *Stefan-Boltzmann law*

$$Q = \varepsilon \sigma A T^4$$

where $Q$ is the radiative power (W), $\varepsilon$ the surface emissivity ($0 \leq \varepsilon \leq 1$), $A$ the surface area ($cm^2$), $\sigma$ the constant = $5.67 \times 10^{-12}$ (W $cm^{-2}$ $K^{-4}$) and T the surface temperature (K). For $T$ = 300 K and $\varepsilon$ = 1 (*black body*), $Q$ = 45 mW $cm^{-2}$, a heat load which may easily reach a few watts for a surface of medium size.

Due to the fourth power dependence, the radiated power decreases quickly when reducing the temperature. At 77 K for instance the reduction factor is $(300/77)^4$ = 220. Therefore *shields at liquid $N_2$ temperature* are usually adopted to reduce the thermal load by radiation. The shield design must be optimised in order to minimise the radiation transmission without reducing too much the molecular conductance, i.e. the pumping speed. An example of this optimisation is given in Ref. [6]. By using a *"chevron type" shield*, providing a reflectivity of 0.07, the 300 K radiation load has been reduced to below $10^{-3}$ of the unshielded value, while providing 0.25 of the full molecular conductance to the pumping surface.

Another essential precaution consists in *reducing the absorptivity of the pumping surface*. By a careful choice of the surface treatment (*Ag plating*) absorptivities of about 0.01 at 4.2 K have been obtained [7]. Adsorption of water vapour on the pumping surface must be carefully avoided because it may dramatically increase the surface absorption.

## 3. CRYOSORPTION PUMPS

Cooling is usually achieved by a two stage refrigerator providing about 80 K on the first stage and 10–20 K on the second. These temperatures are transmitted to the thermal shield and pumping surface by mechanical contact. A part of the cold surface is coated with *a porous material (e.g. charcoal)* to provide a large pumping capacity for $H_2$ (a few Torr $\ell$ per gram, depending on temperature and required pressure).

Loading of the adsorbent by other gases should be avoided not to spoil the pumping of $H_2$. Since all heavier gases condense at 20 K, pump geometry should force the gas molecules to interact with the 20 K uncoated surfaces, where they are trapped before reaching the porous adsorbent material. The thermal contact with the refrigerator and the glue used to bind the adsorbent material usually do not allow baking. After use the gases are released at room temperature (*regeneration*).

Cryosorption pumps may be advantageously used when very *large pumping speeds* are needed (higher than $10^4\ \ell\ s^{-1}$) and the *vacuum* requirements are *not* very *stringent*. An interesting application of cryosorption consists in the *initial evacuation* of atmospheric air making use of a large amount of a porous material cooled at liquid $N_2$ temperature.

## 4. CONDENSATION CRYOPUMPS

A typical condensation cryopump is shown in Fig. 4. This pump, described in Ref. [7], may provide an ultimate pressure lower than $10^{-12}$ Torr and pumping speeds of 9, 3, 15 $\ell\ s^{-1}cm^{-2}$ for $H_2$, $CO(N_2)$ and $H_2O$, respectively. The thermal losses are almost completely due to thermal radiation from the $N_2$ cooled shield. The (low pressure) life-time for a He fill depends on the He vessel volume, and equals 70 days for 70 $\ell$ of helium (heat load $\cong$ 30 mW).

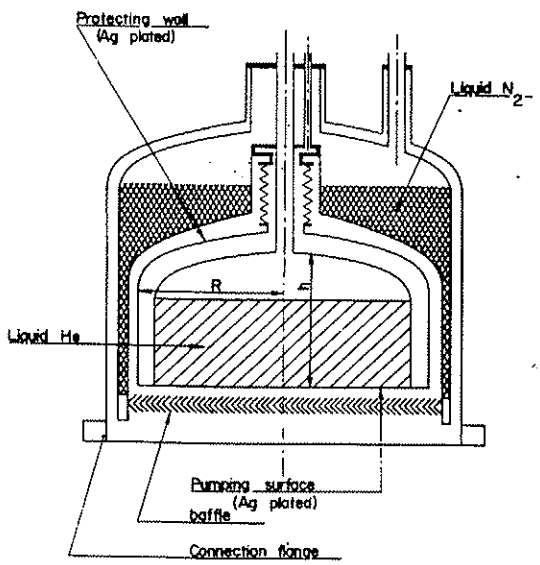



Fig. 4 First version of the CERN condensation cryopump

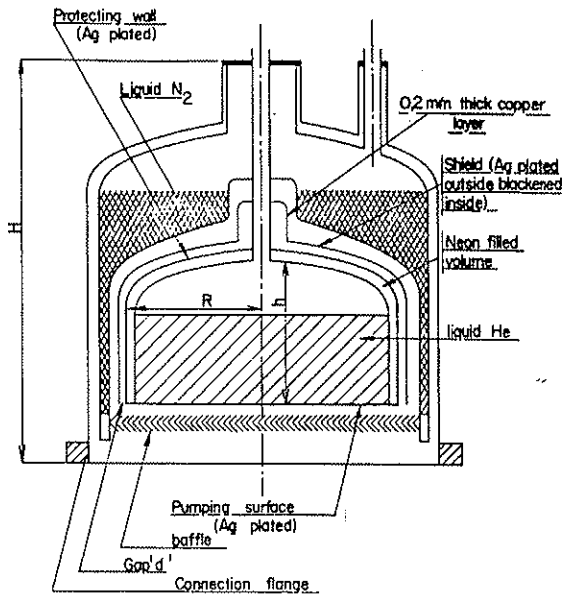



Fig. 5 Improved version of the CERN condensation cryopump

This pump model, developed at CERN for pumping the experimental regions of the ISR, has been later improved by adding a shield between the He and the $N_2$ vessels [8] (see Fig. 5). This shield absorbs the radiation from 77 K, which is then removed by the cold He gas escaping from the pump

neck, to which the shield is connected. The total heat load resulting in He boil-off is reduced in this way to a few mW, depending on pump size. For instance, a pump providing a speed of 4500 $\ell\ s^{-1}$ for $H_2$ may be operated for 190 days with 12 $\ell$ of liquid He.

The scaling laws defining how pumping speed, He boil-off rate, intermediate shield temperature and operating life-time vary when changing pump dimensions are given in Ref. [8].

## 5. THE HYDROGEN "ANOMALY"

Extrapolation from higher temperature data indicates that a saturated vapour pressure of the order of $10^{-13}$ Torr could be obtained for $H_2$ at 2.5 K. Therefore condensation cryopumping seems a reasonable choice to pump very large quantities of $H_2$ at low pressure.

However, measurements carried out in different laboratories [9] [10] revealed the existence of an unforeseen desorption mechanism resulting in a *temperature independent $H_2$ equilibrium pressure* in the $10^{-10}$ – $10^{-11}$ Torr range, depending on the experimental conditions. This "anomalous" pressure was found to be caused by thermal radiation from higher temperature surfaces (particularly those at 300 K). In a first attempt to justify this effect, it has been suggested [9] that thermal radiation absorbed by the condensed $H_2$ could produce a localised heating ("thermal spikes") resulting in enhanced $H_2$ desorption. However, more refined measurements [11] provided strong evidence against this hypothesis by showing that the $H_2$ desorption

- decreases when increasing the thickness of the condensed layer (Fig. 6)
- is proportional to the radiating power absorbed on the substrate, and not to that impinging on it (Fig. 7).

The emerging picture favours a rather indirect desorption mechanism, according to which photons absorbed on the $H_2$ condensing substrate produce phonons energetic enough to cause desorption when reaching the surface of the $H_2$ film. Further evidence in favour of this mechanism has been obtained by condensing a few monolayers of a heavier gas ($N_2$, Ne, Ar...) prior to $H_2$. Having a lower Debye temperature, these gases do not transmit the energetic phonons responsible of first order $H_2$ desorption, which is consequently strongly reduced. This model was later confirmed by other laboratories [12] [13], although a more direct desorption process seems to take place in some specific cases [14].

The main practical conclusion to be derived from these considerations is that a very careful 77 K shielding is necessary to reduce the condensed $H_2$ equilibrium pressure to below $10^{-12}$ Torr.

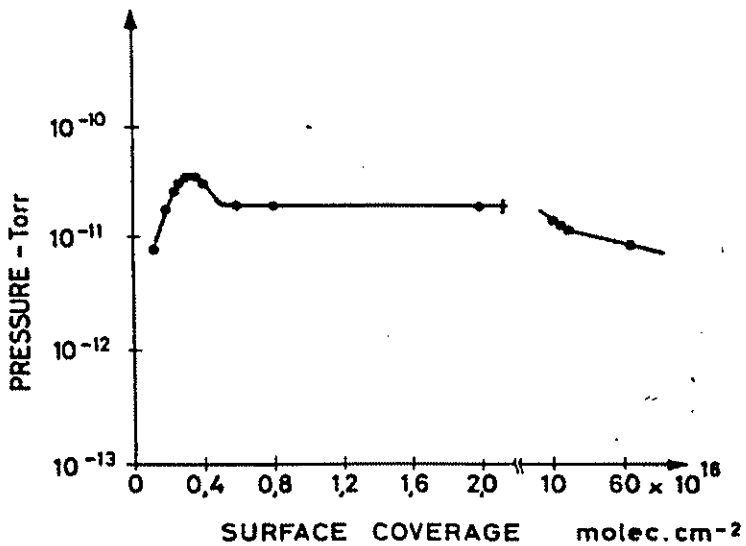



Fig. 6 Typical $H_2$ adsorption isotherm at 2.3 K showing a radiation induced desorption maximum corresponding to a coverage of one monolayer.

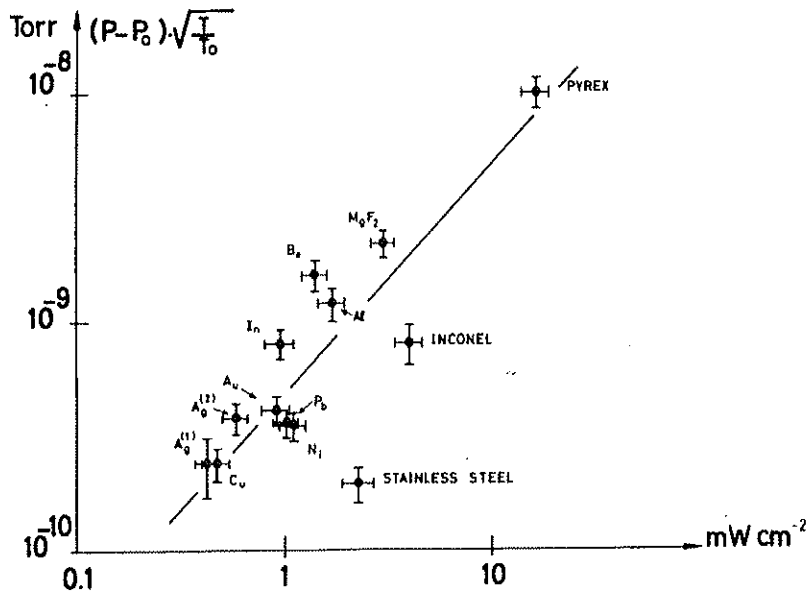



Fig. 7 Variation of the radiation induced $H_2$ desorption from various substrates at 2.3 K and fully exposed to 300 K radiation as a function of the radiating power absorbed.

## REFERENCES


[1] P.A. Redhead, J.P. Hobson and E.V. Kornelsen, The Physical Basis of Ultrahigh Vacuum (American Institute of Physics, New York, 1993), p. 25.

[2] J.M. Lafferty, Foundations of Vacuum Science and Technology, (John Wiley & Sons, Inc, New York, 1998) p. 709-710.

[3] P.A. Redhead, J.P. Hobson and E.V. Kornelsen, The Physical Basis of Ultrahigh Vacuum (American Institute of Physics, New York, 1993), p. 26.

[4] D.M. Young and A.D. Crowell, Physical adsorption of Gases, (Butterworths, London, 1962) p. 3.

[5] J.P. Hobson, J. Vac. Sci. Technol. 10, 1 (1973) 73.

[6] C. Benvenuti, D. Bleschschmidt and G. Passardi, J. Vac. Sci. Technol. 19, 1 (1981) 100.

[7] C. Benvenuti, J. Vac. Sci. Technol. 11, 3 (1974) 591.

[8] C. Benvenuti and M. Firth, Vacuum, 29, 11-12 (1979) 427.

[9] J.N. Chubb, L. Gowland and J. Pollard, Br. J. Appl. Phys. Ser 21 (1968) 361.

[10] C. Benvenuti and R.S. Calder, Phys. Lett. A 35 (1971) 291.

[11] C. Benvenuti, R.S. Calder and G. Passardi, J. Vac. Sci. Technol. 13, 6 (1976) 1172.

[12] P.M. Ferm, S.R. Kurtz, K.A. Pearlstine and G.M. Mc Clelland, Phys. Rev. Lett., 58, 24 (1987) 2602.

[13] J. Cui, S.C. Fain Jr. and W. Liu, J. Vac. Sci. Technol. A 7, 3 (1989) 1850.

[14] M. Hassel, K. Svensson, M. Persson and S. Andersson, Phys. Rev. Lett. 80, 11 (1996) 2481.

# PRESSURE MEASUREMENT – TOTAL PRESSURES

*Chr. Edelmann*
Otto von Guericke University, Magdeburg, Germany

**Abstract**
Gauges for the measurement of total pressures in the vacuum range using different physical effects are described (with the exception of ionization gauges that are discussed in a separate chapter). Advantages, disadvantages, tendencies in development, sources of error, and pressure ranges for different applications are discussed.

## 1. INTRODUCTION

The term 'vacuum' designates a pressure range which extends from pressures somewhat lower than atmospheric pressure down to pressures lower than $10^{-14}$ mbar. Thus it spans a pressure range with more than 17 orders of magnitude that can be arbitrarily subdivided into ranges of about three to five decades of pressure. According to DIN 28 400-1 (in agreement with ISO 3529-1) the overall pressure range is divided into four ranges. Since the beginning of the nineties pressures lower than $10^{-12}$ mbar have been defined as extreme high vacuum. In Table 1 this fifth range is in brackets since it is still not a standardized definition.

Table 1
Pressure ranges

| Definition | Abbreviation | Pressure range |
|---|---|---|
| Rough vacuum | | 1 mbar $\leq p <$ atm. pressure |
| Medium vacuum | | $10^{-3}$ mbar $\leq p < 1$ mbar |
| High vacuum | hv | $10^{-7}$ mbar $\leq p < 10^{-3}$ mbar |
| Ultra high vacuum | uhv | ($10^{-12}$ mbar) $\leq p < 10^{-7}$ mbar |
| (Extreme high vacuum) | (xhv) | ($p < 10^{-12}$ mbar) |

Assuming ideal thermodynamic equilibrium we can define the pressure $p$ by Eq. (1):

$$p = \frac{F}{A} \tag{1}$$

($F$ force acting perpendicularly onto an area $A$). According to the international system of units the pressure unit is:

$$1[p] = 1\,[F]/\,1\,[A] = 1\ \mathrm{N/m^2} = 1\ \mathrm{Pa} = 0.01\ \mathrm{mbar}.$$

As the ambient atmospheric air and the residual gas in a vacuum chamber both consist of $i$ different gas components each of them has a partial pressure $p_{part/i}$. The partial pressures of all gas components add linearly to give a total pressure $p_{total}$ according to Dalton's´s law:

$$p_{total} = \sum_i p_{part/i}\,. \tag{2}$$

In future I consider only the total pressure $p_{total}$ which I abbreviate by $p$. The following different physical effects, depending on the gas pressure, can be used for the measurement of total pressures:

- force effect
- heat transport by heat conductivity of the gas
- heat transport by convection of the gas
- viscosity of the gas
- radiometer effect
- scattering of electrons by collisions with gas particles
- excitation of gas particles by electron impact
- ionization of gas particles by electron impact
- deceleration of ions in an electric field by collisions with gas particles.

Gauges using these different effects for the determination of pressures will now be described.

## 2. GAUGES USING FORCE EFFECTS FOR PRESSURE MEASUREMENT

### 2.1 Piston gauges

Because of their complicated structure and operation these gauges are only used for calibration purposes. According to Eq. (1) the force $F$ acting on the area $A$ of a moveable piston placed in a cylinder can be used for the determination of the pressure. Figure 1 represents the principle. A piston P with a mass $m$ and a cross section of area $A$ limits a well defined gas volume $V$. If the pressure in $V$ is increased with the help of a gas inlet the volume can be restored by an additional weight of mass $\Delta m$. For equilibrium we find :

$$p = \frac{(m + \Delta m)g}{A} \tag{3}$$

where $g$ is the gravitational acceleration.

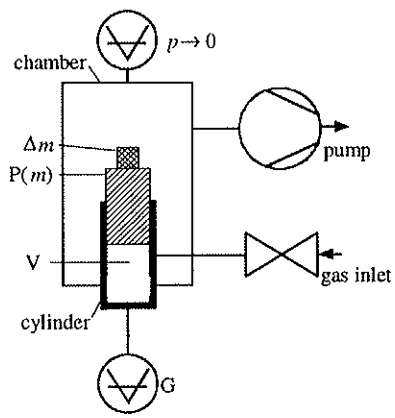



**Fig. 1** Principle of a Piston gauge

To avoid the influence of the surrounding atmosphere the cylinder and piston are placed in an additional chamber, which is evacuated. To reduce the influence of friction between the cylinder and piston the piston is made to rotate. With the help of such an apparatus as that represented in Fig. 1 it is possible to calibrate any gauge indicated in the sketch by G.

### 2.2 Bourdon gauge

The Bourdon gauge (see Fig. 2a) consists of a bent tube with an elliptic cross section closed at one end and connected at the other open end to the chamber in which the pressure is to be measured. Pressure

differences between the environment of the gauge and the interior cause forces to act on the two walls of the tube (Fig. 2b) so that it is bent by an amount that depends on the pressure difference between the environment and the interior. The bending is transformed by a lever to a pointer whose position can be calibrated. The importance of this type of gauge is that it is very robust and that it covers a range of pressure measurement from pressures higher than atmospheric pressure down to rough vacuum (about 10 mbar). The accuracy and reproducibility are relatively poor, so that it is not suitable for precision measurements, and its usefulness for vacuum measurements is limited.

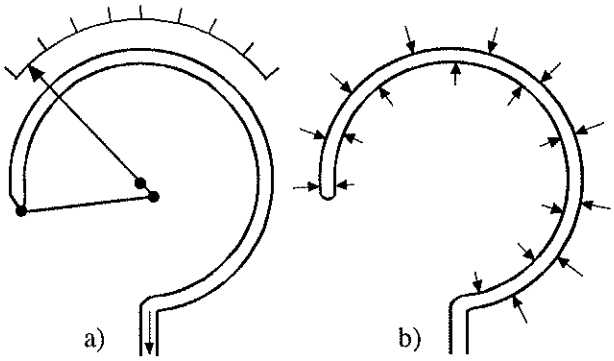



**Fig. 2** Bourdon gauge, a) principle, b) distribution of forces

## 2.3 Diaphragm gauges

If a diaphragm or a bellows separates two regions with different pressures ($p_1$, $p_2$) the difference $\Delta p$ ($\Delta p = p_1 - p_2$) of these two pressures causes a force that deforms the diaphragm or bellows. There are many possibilities for measuring this deformation, e.g. mechanically by a lever and a pointer, optically by a mirror and a light pointer, or electrically by changes of the capacity of a capacitor formed by the diaphragm and an additional electrode which is usually placed in a region of very low pressure (see Figs. 3a ... 3c). For precision measurements one side of the diaphragm is evacuated to very low pressure. This is called a reference vacuum. The other side is exposed to the pressure to be measured. The deformation of the diaphragm depends on, but is not proportional to, the pressure difference [1]. These days linearization of the pressure vs. deformation reading is mostly performed by electronic circuits. Thus it is possible to make pressure measurements in a range between some hundred mbar and $10^{-4}$ mbar with such a precision that this type of gauge can be used as a secondary standard gauge. The lower pressure limit is caused by the thermal dilatation that has the same order of magnitude as the deformation at very low pressures. Some special alloys like stainless steel or special ceramics such as $Al_2O_3$ with high density, are used as materials for the diaphragms. Generally the low pressure in the region of the reference vacuum is maintained by the use of getters. Frequently the electrodes and the circuits for the pressure reading are placed in the region of the reference vacuum. Figure 4 shows a diaphragm gauge with electrical reading. The pressure reading is independent of the gas composition. A pressure range from atmospheric pressure to $10^{-4}$ mbar is covered. The high reliability, simple handling, electrical readout that can easily be recorded, and the robust construction are reasons to use it for many practical applications.

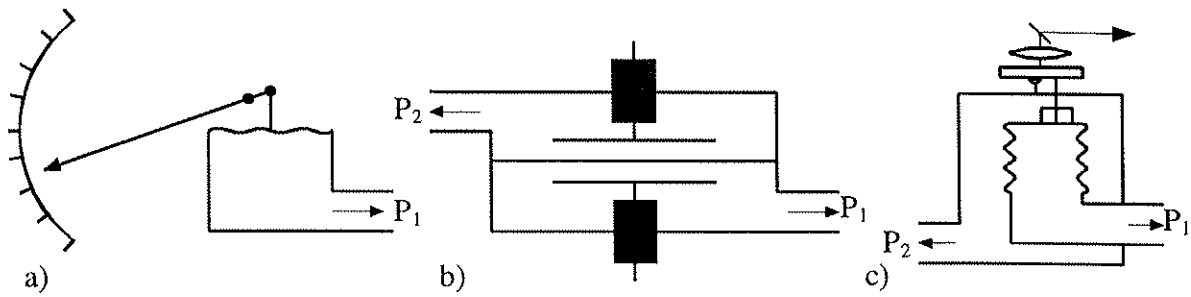



**Fig. 3** Diaphragm gauges with a) a mechanical, b) an electrical and c) an optical display

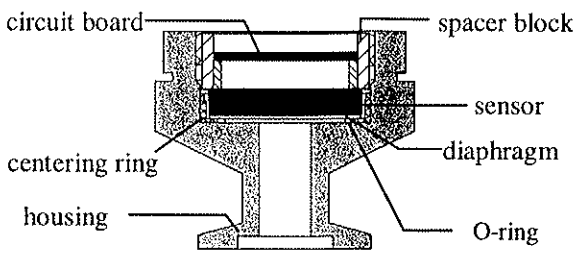



**Fig. 4** Diaphragm gauge with electrical reading

### 2.4 Piezoresistive diaphragm gauge

Similar to the diaphragm gauge the piezoresistive gauge consists of a small evacuated capsule closed with a thin silicon diaphragm at the side exposed to the space whose pressure is to be measured. On one side of this diaphragm are placed thin film piezoresists produced by evaporation and which are connected to form a bridge circuit. By deformation of the silicon diaphragm the bridge is put out of balance by an amount depending on the deformation. To avoid destruction of the silicon diaphragm by corrosive gases there exist models in which a small volume filled with special oil is placed between the silicon diaphragm and the vacuum space. This volume is closed at the side of the vacuum space whose pressure is to be measured by a thin stainless-steel diaphragm. The special oil serves as an incompressible pressure transducer.

Depending on the gauge construction the pressure reading covers a range of 0.1 to 200 mbar or 1 to 2000 mbar [2]. The pressure reading is independent of the kind of gas.

## 3. GAUGES USING THE HYDROSTATIC PRESSURE OF MERCURY OR OTHER NON-VOLATILE LIQUIDS

### 3.1 U-tube gauge

These gauges use the hydrostatic pressure of a liquid column for pressure reading. Mostly mercury is used as the liquid because of its low vapor pressure and its cohesion characteristics.

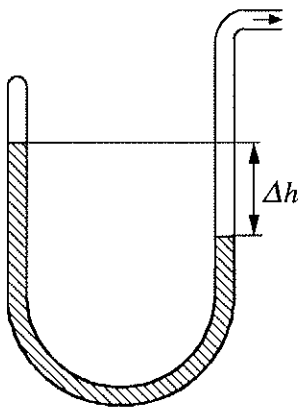



**Fig. 5** U-tube gauge with one closed glass tube

In the simplest case a gauge of this type consists of a U-shaped glass tube closed at the one end and connected at the other to the chamber whose pressure is to be measured (Fig. 5). The tube is filled with mercury so that in the volume between the closed end and the level of the mercury column there is only the vapor pressure of the mercury and is called Torricelli´s vacuum. The difference in heights of the mercury levels in the two legs of the U-shaped tube is proportional to the pressure

$$p = \rho g \Delta h \tag{4}$$

where $\rho$ is the density of the used liquid, usually mercury, g gravitational acceleration, and $\Delta h$ the difference in heights of the levels of the two columns. Because the lowest measurable difference of heights of the mercury levels is about 1 mm the lowest pressure detectable with this gauge is in the order of 100 Pa (1 mbar). Because of its uncertainty, its limited range of pressure reading, and the absence of possibilities for electronic recording, this simple type of gauge has nearly no importance for practical application today.

### 3.2 Compression gauge

In contrast to the simple U-shaped gauge, more important is the compression manometer developed by Herbert McLeod in 1874 [3]. This so called McLeod gauge shown in Figs. 6a and 6b uses mercury as a liquid piston with which the gas of a well-defined volume $V$ is compressed into a known small volume $V_M$. By this compression the pressure in the closed volume $V_M$ is increased from the unknown low pressure $p$ to the measurable higher pressure $p + \rho g \Delta h$ according to Boyle´s Law :

$$pV = const. \tag{5}$$

If the pressure $p$ can be neglected in respect to the hydrostatic pressure $\rho g \Delta h$, $p$ can be calculated by the formula:

$$p = \rho g \Delta h \frac{V_M}{V}. \tag{6}$$

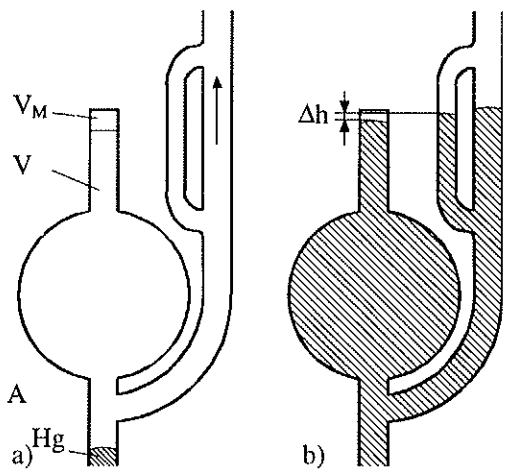



**Fig. 6** McLeod gauge

To read the pressure it is necessary to determine the ratio of the volumes $V$ and $V_M$ and the difference $\Delta h$ of the heights of the mercury levels. To avoid errors caused by different capillary depression both the capillary containing the little volume $V_M$ (in the capillary) and the comparison capillary parallel to the connection tube are made from the same piece of glass capillary. The ratio of the two volumes can be determined exactly before the first use of the gauge and represents a special constant for this gauge. Thus, the McLeod gauge substitutes the measurement of a pressure by the measurement of only one length. Therefore it was used for a long time as a primary standard for pressure measurement. To measure the pressure one has to lower the columns of mercury so that the volume $V$ is connected with the chamber in which the pressure is to be measured. After reaching the pressure equilibrium one has to compress isothermally the gas in the volume $V$ up to the volume $V_M$ and to determine the pressure in this volume. For this purpose one has to compare the heights of the mercury columns in the volume $V_M$ and in the comparison capillary. After the measurement one has to lower the mercury level so that the volume $V$ is connected again with the chamber in which the pressure is to be measured. Thus, the measurement is discontinuous and needs some dexterity. For precision measurements one has to stabilise the temperature in the surroundings of the gauge. The range of pressure measurement depends on the dimensions of the McLeod gauge. The lowest measurable pressures are about (1–0.5) $10^{-6}$ mbar. But for this purpose the volume $V$ has to be at least 1000 $cm^3$ and the volume $V_M$ about 1 $mm^3$ or smaller. In

such a gauge one has to handle a lot of mercury and any accident has terrible consequences for the laboratory where the gauge is used. Although the law of Boyle and consequently the pressure reading too do not depend on the gas composition the McLeod gauge is not able to read correctly pressures of condensable vapors because of their condensation during the compression process. Therefore the McLeod gauge is not convenient for industrial use.

## 4. MODULATION GAUGE

Differentiating Eq. (4) one finds:

$$\Delta p = -p\frac{\Delta V}{V}. \tag{7}$$

This equation explains why periodical changes in $V$ result in periodical changes of $p$ so that $\Delta p$ is proportional to $p$. The proportionality between these two quantities is used in the modulation gauge of Jurgeit and Hartung [4] shown in Fig. 7. In this gauge a small volume $V$ is periodically changed with the help of a piezoelectric oscillator (2). The resulting periodical changes in pressure can be detected with the help of a sensitive capacitor microphone (1) whose signal is proportional to the changes of the oscillating pressure. This device is able to read pressures from atmospheric pressure down to $10^{-6}$ mbar. The reading is only slightly influenced by the gas composition as a consequence of the conductance (3) between the oscillating volume of the gauge and the chamber in which the pressure is to be measured.

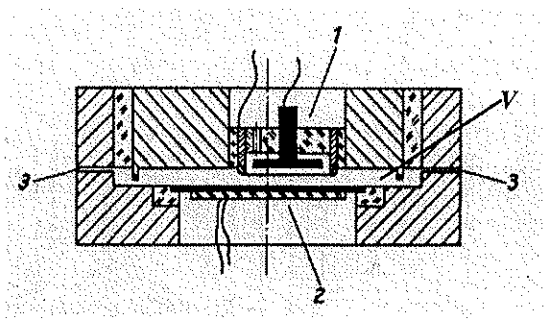



**Fig. 7** Principle of the the modulation gauge of Jurgeit and Hartung

## 5. GAUGES USING THE VISCOSITY OF A GAS FOR PRESSURE MEASUREMENT

### 5.1 The physical basis

According to the kinetic theory of gases the viscosity $\eta$ is described by Eq. (8) under the assumption that the equilibrium of the ideal gas is not disturbed:

$$\eta = \frac{1}{3}\rho\Lambda\bar{v} \tag{8}$$

with $\rho$ the density of the gas, $\Lambda$ the mean free path of a gas particle, and $\bar{v}$ the average value of the velocity of a gas particle. If the mean free path is comparable to the dimensions of the system the viscosity decreases with lowering of the density (or pressure) of the gas. Thus, the viscosity of a gas can be used to determine the pressure under the mentioned restriction. One has to bear in mind that the viscosity depends on the kind of gas.

### 5.2 Former constructions

To measure the viscosity of a gas one can use the measurement of the damping of any oscillating system. This idea is very old. Since the last century different types of gauges which used this principle were suggested. They contained an oscillating system [6], e.g. a pendulum of any shape (see Fig. 8a–c), which oscillated after a short excitation and whose damping was measured by noting the reduction of the amplitude with time. This procedure is not convenient and has the disadvantages that it is discontinuous and is disturbed by every kind of impact or vibration of the housing at which the pendulum is

fixed. Rotating oscillations (see Fig. 8d–e) were also used for the same purpose [7], but they had the same disadvantages. Therefore, in 1937 Holmes [8] suggested magnetic suspension of the moving system that was damped by the viscosity of the gas. The advantage of this equipment is that there is no mechanical connection between the housing of the gauge and the rotating or oscillating part.

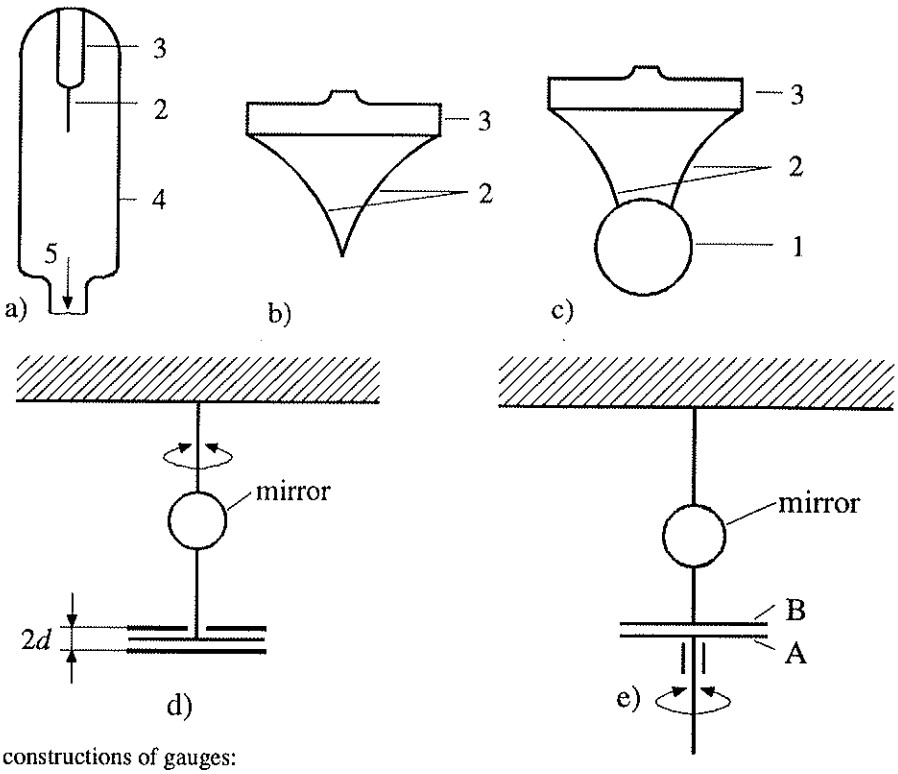



**Fig. 8** Former constructions of gauges:

a) single-string pendulum (2), suspension (3), housing (4), connection to the chamber (5);
b) double-string pendulum (2), suspension (3);
c) double-string pendulum (2) with mirror (1), suspension (3);
d) torsional vibration gauge with one moveable disc between two fixed discs;
e) viscosity gauge with one rotating disc (A) opposite to movable suspended disc (B)

## 5.3 The spinning-rotor gauge

In 1968 Fremerey [9] developed a gauge with a freely-rotating steel ball for the measurement of the viscosity of the gas. According to Fig. 9 this ball (1) was suspended by permanent magnets (2) and magnetic coils (3) whose excitation current was electronically adjusted. The ball was made to rotate by a special coil. After reaching a desired frequency the excitation was stopped and the remanent magnetism of the ball induced an electric voltage in a coil so that the frequency could be determined electronically at intervals of time.

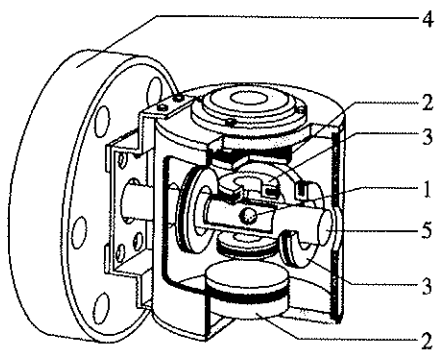



**Fig. 9** Spinning rotor gauge (after a data sheet of the Leybold company)

The viscosity and the pressure could be calculated from the decrease of the frequency $\nu$ with time from

$$p = \frac{\pi \rho_K r \bar{v}}{10 \alpha t} \ln \frac{\nu(t)}{\nu(0)} \tag{9}$$

where $r$ is the radius of the ball, $\rho_K$ its density, $\nu(t)$ frequency of the ball at the time $t$, $\nu(0)$ frequency at the beginning of the measurement, $\alpha$ the accommodation coefficient and, $\bar{v}$ is the arithmetic average velocity of the gas particles. The advantage of this gauge is the simple construction of the sensor. It consists only of a flange (4) connected with a cylindrical tube (5) closed at the end opposite to the flange and containing the ball. The arrangement of the permanent magnets and the coils is a separate structure fixed on the tube.

The lower pressure limit is $10^{-7}$ mbar or lower, the upper pressure limit is about $10^{-1}$ mbar. Due to the quantities $\alpha$ and $\bar{v}$ being specific to the gas the pressure reading depends on the composition of the gas. The precision of this gauge is so excellent that it can be used as a secondary standard for the calibration of other gauges.

### 5.4 Miniaturization of viscosity gauges

Attempts have been made to miniaturize the dimensions of viscosity gauges by using oscillator quartzes made for temperature measurement or in wrist watches. Figure 10 shows the construction principle of the so called quartz friction gauge suggested by Ono and co-workers in 1986 [10, 11]. This equipment has the form of a tuning fork, but it is much smaller. It is made of quartz by lithographic processes similar to those used for the fabrication of semiconductor devices. Two specially shaped Au/Cr electrodes are deposited by sputtering on the surface of the tuning fork. The impedance of the capacitor formed by these two electrodes changes with the frequency and amplitude of the oscillations of this tuning fork and, in turn, is influenced by the ambient gas pressure. Thus, the change in impedance $Z$ from the impedance for the resonance condition at very low pressures $Z_0$ depends on the pressure. As an example, Fig. 11 shows the difference of the impedance $(Z - Z_0)$ vs. pressure for a fork length of 2.5 mm.

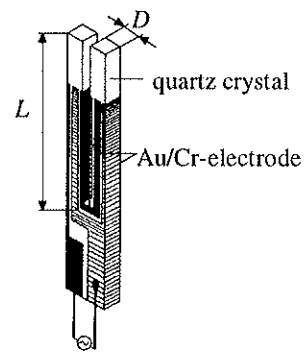



**Fig. 10** Construction principle of the quartz friction gauge

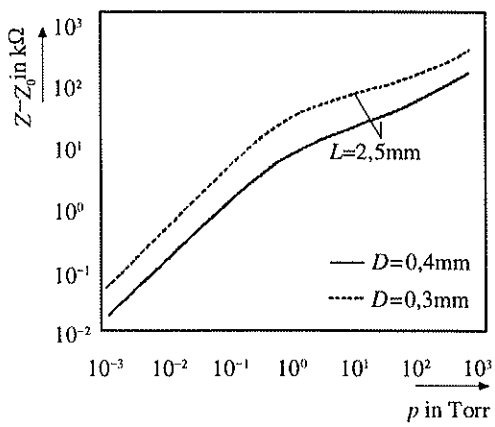



**Fig. 11** Difference of the impedance $(Z - Z_0)$ vs. pressure for a fork length of 2.5 mm

## 6. GAUGES USING THE PRESSURE DEPENDENCE OF THE SPECIFIC HEAT CONDUCTIVITY OF THE GAS

### 6.1 The physical principle

Neglecting the disturbance of the thermodynamic equilibrium the kinetic theory of gases can be used to derive Eq. (11a) for the pressure dependence of the specific heat conductivity of a gas $\lambda$ :

$$\lambda = \frac{1}{3}\Lambda\rho\bar{v}c_V \,. \tag{11 a}$$

with $\Lambda$ the mean free path length of a gas particle, $\rho$ the density of the gas, $c_v$ the specific heat capacity of the gas at a constant volume, and $v$ the average value of the velocities of the gas particles. By comparison with Eq. (8) one can write for Eq. (11 a) :

$$\lambda = \eta c_V \,. \tag{11 b}$$

Considering the disturbance of the thermodynamic equilibrium by temperature gradients Chapman found a correction factor for Eq. (11 a or 11 b)

$$\lambda = 2{,}52\eta c_V \,. \tag{11 c}$$

Independent of the special shape one can interpret Eq. (11a) in the following manner: The mean free path length of a gas is reverse proportional to the pressure, and the density is direct proportional to the pressure. Therefore, the specific heat conductivity of a gas does not depend on the pressure, because Eqs. (11a–11c) contain the product of $\rho\Lambda$. This statement is correct for higher pressures in the vacuum range, if the mean free path length $\Lambda$ is lower than the geometric dimensions of the vacuum system. But if – as a consequence of lowering the pressure – the mean free path length reaches values of the same order of magnitude as the geometric dimensions of the system, the specific heat conductivity is diminished if the pressure is reduced. This effect can be used for the measurement of the pressure.

### 6.2 The general construction of a heat-conductivity gauge

Already in 1906 Pirani [12] suggested a gauge which used the pressure dependence of the heat conductivity for the pressure measurement.

In the simplest case the heat-conductivity gauge consists (Fig. 12) of a thin wire (diameter $2r_i$) which is mounted in the axis of a cylindrical tube (diameter $2r_o$). The cylindrical tube is connected with the vacuum chamber in which the pressure is to be measured. The thin wire is heated by a constant electric power $P_{el}$. Heat transport to the walls of the tube is caused by:

- conductivity through the electric feedthroughs $I_{W1}$
- conductivity via the gas $I_{W2}$
- radiation from gas $I_{W3}$ .

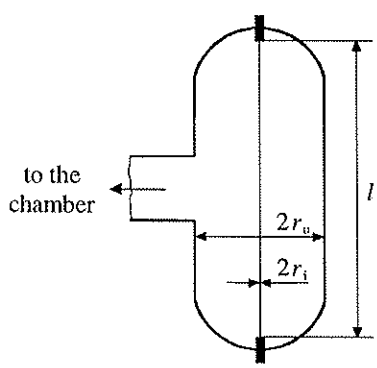



**Fig. 12** General construction of a hea- conductivity gauge

At equilibrium:

$$P_{el} = I_{W1} + I_{W2} + I_{W3} = \sum I_W \,. \tag{12}$$

Heat transfer via the electric feedthroughs depends on the temperature of the heated wire, the material, and the cross section of the feedthroughs. Heat radiation is described by the law of Stefan and Boltzmann and depends on the fourth power of the temperatures of the heated wire and its surroundings. Both the term $I_{W1}$ and $I_{W3}$ are independent of gas pressure and nearly constant.

At higher pressures, when the mean free path length $\Lambda$ is small in comparison to the diameter of the wire, the specific heat conductivity of the gas is independent of pressure. Thus a constant heat flow is transported by the gas from the heater to the wall of the gauge. It depends only on the temperatures of the heater and of the walls and on the geometric structure, but not on the pressure. But if the pressure is diminished to values at which the mean free path length $\Lambda$ is in the order of the diameter of the thin wire, the heat conductivity of the gas decreases. Simultaneously the heat transport from the wire to the wall is reduced and the temperature of the wire rises if the electric power is constant. At very low pressures the heat flows $I_{W1} + I_{W3}$ dominate and the heat flow via the gas can be neglected. Thus, the temperature of the heated wire becomes stable. This behaviour of the heat flow and the temperature of the heater operated with constant electric power is shown in Fig. 13. The pressure reading is transformed into the measurement of the heater temperature. This in turn is possible either by the determination of the electric resistance of the wire with the help of a Wheatstone bridge or by a thermocouple whose voltage is measured. Another possibility of measurement is to stabilize the temperature of the heater by variation of the heating power. At low pressure less electric power is necessary to obtain the desired heater temperature than at higher pressures.

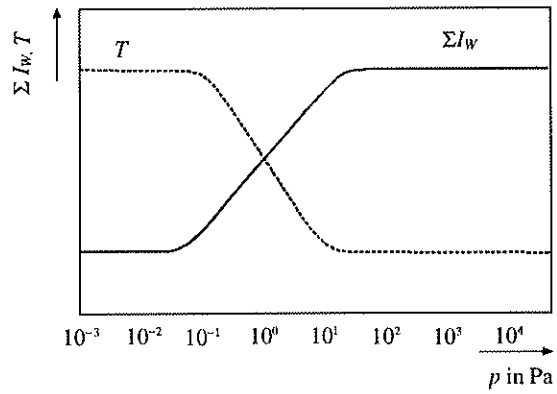



**Fig. 13** Heat flow and temperature of a heater with constant electric power

As can be seen in Fig. 13, in the case of operation with constant heating power the pressure is correctly measurable only between $10^{-3}$ mbar and 1 mbar. Above 1 mbar and below $10^{-3}$ mbar, the output vs. pressure curve is so flat that a correct pressure reading is nearly impossible. To extend the range of pressure measurement to lower pressures one can reduce the term of the heat radiation by lowering the heater temperature (i.e. lower electric power) and by cooling the tube wall. Thus, it is even possible to measure with special gauges pressures of about $10^{-4}$ mbar.

An extension of the range of higher pressures is also possible. For this purpose the heat transfer from the heater to the walls by convection can be used. This was shown by Gorski and co-workers in Poland [14].

The pressure reading depends on the kind of gas and on the accommodation coefficients of the heated wire and the inner tube wall. Thus, contamination of the tube wall or the heater surface can influence the pressure reading.

### 6.3 Miniaturization of heat-conductivity gauges

In principle it is possible to miniaturize the heat conductivity gauges by using technology used in the production of semiconductor components. For example Huang and Tong [15] have used a silicon chip with the dimension of only 1.62 x 2.02 mm. The principle of this element is shown in Fig. 14. The heater consists of a resistor spread over the whole area of the chip. It is fed via a MOS transistor which regulates the current so that the temperature of the heater is always constant. (The heating increases the temperature of the whole chip whose temperature is somewhat higher than ambient.) The temperature is

measured by a diode and stabilized by a proper integrated circuit. The voltage of the heater is used for pressure measurement. Figure 15 shows schematically the characteristics of this gauge. At pressures lower than 1 mbar the heat conductivity, and above 1 mbar the heat convection, is used to measure the pressure.

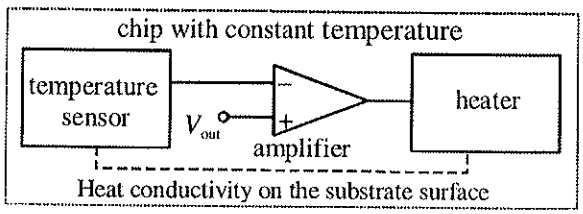



**Fig. 14** Principle of heat-conductivity gauge of Huang and Tong

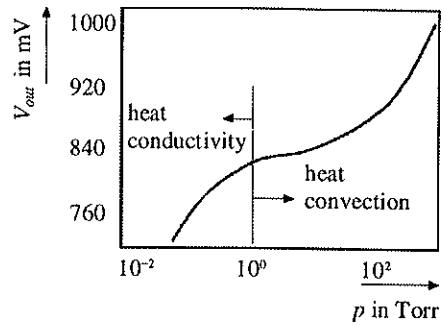



**Fig. 15** Characteristics of heat-conductivity gauge of Huang and Tong

Another possibility for miniaturization of the sensor was developed by Badinter and co-workers [13], who used a very small heater, but external circuits for the measurement. The gauge had a diameter of 3.6 mm and a height of 3 mm. The heater had a diameter of 6–7 μm and was coated with an isolating film of 5–7 μm thickness. The range of measurement was between 0.1 and 100 mbar.

## 7. RADIOMETER-EFFECT GAUGES

The thermal molecular pressure experimentally demonstrated by Crookes in 1873 was used by Knudsen [16] in 1910 for the measurement of pressures. To explain this effect we consider in Fig. 16 two plane-parallel plates 1 and 2 with the same area but different temperatures $T_1$ and $T_2$. The distance $d$ between the two plates may be small compared with the mean free path length Λ, so that $d < \Lambda/10$. Assuming $T_1 > T_2$ ($T_2$ may also be the temperature of the surroundings) gas particles travelling from plate 1 to plate 2 have a higher impact than gas particles coming from the surroundings to the other side of plate 2. Thus, different pressures seem to act onto the two sides of plate 2. Their difference can be calculated with the help of the known laws of the kinetic theory of gases by the following equation:

$$\Delta p = \frac{p}{2}\left(1+\frac{\alpha_2-\alpha_1}{\alpha_1+\alpha_2-\alpha_1\alpha_2}\right)\left(\sqrt{\frac{T_1}{T_2}}-1\right) \tag{13}$$

where $\alpha_1$ and $\alpha_2$ are the accommodation coefficients of the plate 1 and 2 respectively. According to Fig. 17 one can suspend plate 2 in the manner used for galvanometers. The pressure difference effects movement of plate 2. The elastic moment of the suspension tries to compensate the momentum caused by the pressure difference at the two sides of plate 2. Thus, in the equilibrium the torsion of plate 2 indicates the pressure. Figure 18 shows the torsion angle γ vs. pressure $p$.

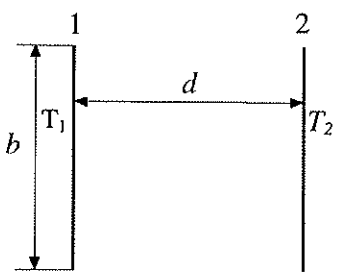



**Fig. 16** Principle of radiometer-effect gauge

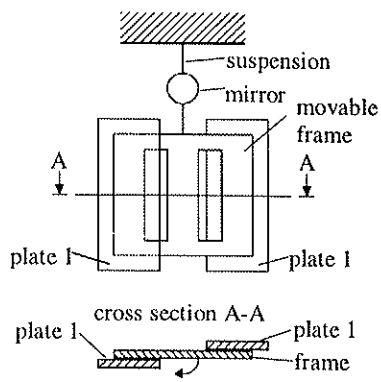



**Fig. 17** Construction of radiometer-effect gauge

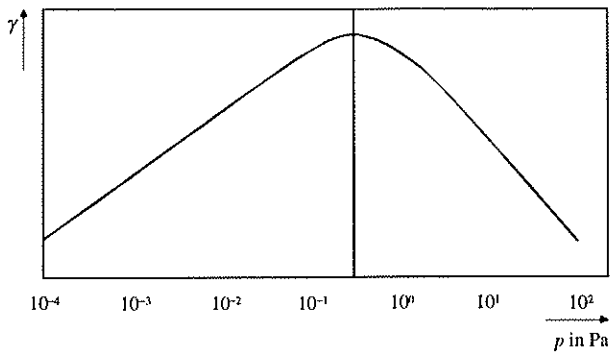



**Fig. 18** Torsion angle γ vs. pressure $p$

According to Eq. 13 this method of pressure measurement is independent of the kind of gas. Some versions of this equipment have a low-pressure limit of about $10^{-8}$ mbar [19]. Because this type of gauge is very sensitive to vibrations and impact it is not used industrially today.

## 8. GAUGES USING THE INTERACTION BETWEEN ELECTRONS AND GAS PARTICLES

Collisions between electrons and gas particles have the following effects:

- elastic scattering of the electrons
- inelastic scattering of the electrons connected with dissociation, excitation or ionisation of gas particles.

### 8.1 Gauges using the elastic electron scattering for pressure measurement

Gauges of this type can be used where the electrons have energies lower than the ionisation energy. In this case the probability for excitation is small and the probability for ionisation is zero.

If electrons are accelerated by a low-voltage, elastic collisions with gas particles cause the appearance of velocity components perpendicular to the direction of the electric field and the reduction of the accelerating effect in the electric field. In the worst case, the electrons can reach, in a homogeneous electric field, a constant drift velocity instead of an accelerated movement. Of course, these effects happen only if the gas pressure is high enough that the mean free path length of the electrons is much smaller than the distance between the electrodes producing the electric field. These effects can be used for pressure measurement.

In 1983 Lucas and Goto [18] used the appearance of velocity components of the gas particles perpendicular to a homogeneous electric field for pressure measurement. They found that electrons travelling in the direction of a uniform homogeneous electric field diffuse radially to the direction of this field. Electrons starting from a point source of the cathode arrive at the anode with a mean radial displacement $\overline{r^2}$ whose magnitude can be expressed by the formula:

$$\overline{r^2} = 4\frac{D_r}{\mu}\frac{d^2}{U} \tag{14}$$

with $D_r$ the so-called radial-diffusion coefficient, $\mu$ the mobility, $d$ the distance between cathode and anode, and $U$ the voltage between cathode and anode. The quotient $D_r / \mu$ is solely a function of $U/(pd)$ so that the mean radial displacement is a function of the pressure $p$.

The gauge of Lucas and Goto using this effect for pressure measurement has two anodes facing the cathode. One has the shape of a circular disc placed at a distance $d$ from the cathode. The other is a flat ring surrounding concentrically the disc. The distribution of the electron current to these two anodes having the same potential can be used for pressure measurement. This type of gauge is shown schematically in Fig. 19.

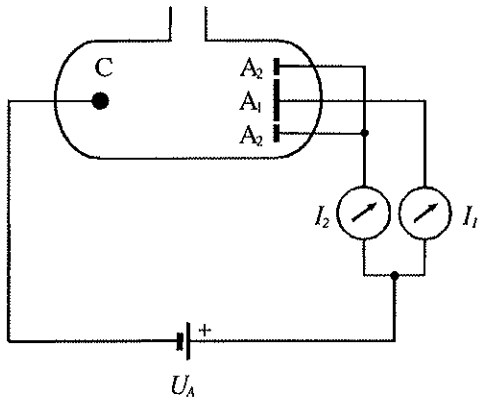



**Fig. 19** Scheme of Lucas and Goto gauge

The pressure can be measured in a range between $10^{-3}$ mbar and about 100 mbar. The curves of the current ratio vs. pressure are not linear. Their shape depends on the kind of gas. A linearization of these curves seems to be possible with the help of a microcomputer.

Another gauge construction was suggested by Edelmann in 1998 [18–21]. The original idea was to use the drift velocity $u$ of the electrons in the homogeneous electric field for pressure measurement. Kapzow [22] has shown that the drift velocity $u$ of electrons in a weak electric field depends on the pressure according to the formula:

$$u = \alpha \frac{e\Lambda_e}{mv} E, \tag{15}$$

where $e$ is the elementary charge of an electron and $\Lambda_e$ its mean free path, $E$ the electrical field strength equal to voltage $U/d$, $m$ the mass of an electron, and $v$ the (not exactly defined by Kapzov) average velocity of the electrons. The mean free path length of the electrons is inversely proportional to the pressure. Thus it should be possible to determine the gas pressure by the measurement of the drift velocity $u$. Originally, it was intended to determine the drift velocity by the measurement of the electron distribution onto two electrodes. For this purpose the gauge contained, instead of two flat anodes in one plane perpendicular to the axis between cathode and anode, a flat grid followed by a flat anode sheet as shown in Figs. 20a–c. Both electrodes are parallel and placed perpendicularly to the axis between the cathode and anode. They have the same voltage, about 10–20 V positive with respect to the cathode. The ratio of the electron current to the grid and the electron current to the anode plate can be used to determine the pressure. With this construction we have measured the pressure in the range between $10^{-1}$ and 100 mbar. The characteristics of this gauge are shown in the Fig. 21. To read the pressure it is possible to use the ratio of the grid and the anode current, while the cathode current is kept constant, or to keep either the grid or the anode current constant and to use the current of the other electrode (i.e. anode or grid respectively) for pressure measurement.

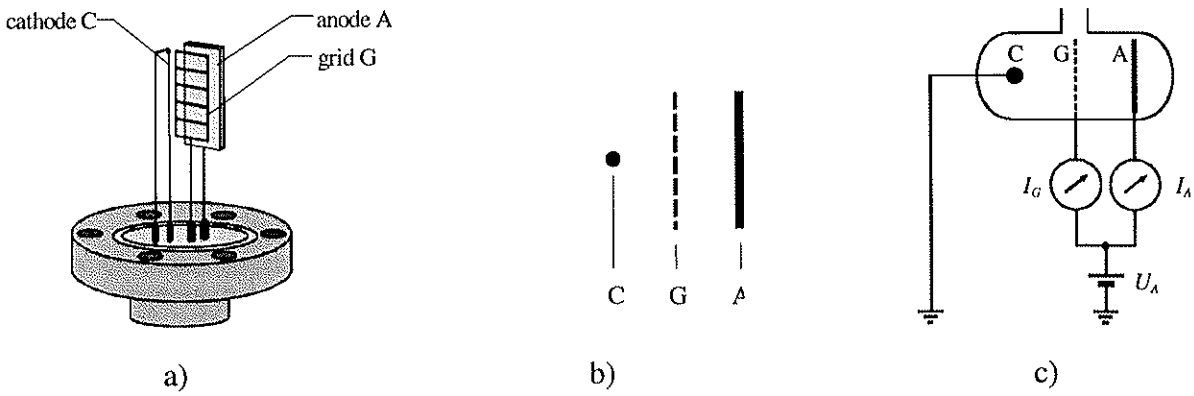



**Fig. 20** Drift-velocity gauge

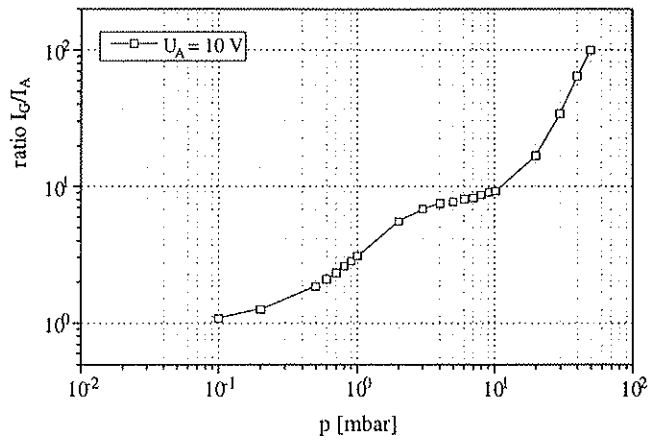



**Fig. 21** Characteristics of drift-velocity gauge

An attempt to calculate the distribution of the electrons to the two electrodes with the help of the classical methods used by Barkhausen or Rothe and Kleen for electron tubes does not give correct results. The reason may be that these methods neglect increasing velocity components perpendicular to the direction of the homogeneous electric field with increasing pressure.

Thus, for a better calculation of the current distribution only Monte Carlo calculations seem to be possible. These were performed by Kauert with the aim to optimise and to miniaturize the electrode structure.

The advantages of these types of gauges are:

- Large range of pressure measurement from $10^{-3}$ mbar (or even $10^{-4}$ mbar) to about 100 mbar .
- Fast response for pressure changes, much faster than for example diaphragm gauges.
- Possibilities to miniaturize the gauges.
- Simple electronic circuits.

Disadvantages are:

- As in the case of viscosity and thermal-conductivity gauges the pressure reading depends on the kind of gas.

However the most important problem for this type of gauge is the electron source. Lucas and Goto used a photo cathode illuminated by ultraviolet light. This is not convenient for practical applications. We used a directly-heated iridium cathode coated with yttrium oxide. In the rough vacuum range the lifetime of such a cathode is limited. It is hoped that field emitter arrays or sandwich cathodes will have a longer lifetime and will be commercially available for this purpose in the near future.

### 8.2 Gauges using the inelastic interaction between electrons and gas particles

The energy loss of electrons by inelastic collisions with gas particles can cause an *excitation of the gas particles*. The excited particles return immediately into the ground state by radiation of electromagnetic waves. Although the light emission of gas discharges can be used for a rough estimation of the pressure range, and in some cases even for information about special components of the gas mixture, this effect is not suitable for exact pressure measurement. But the energy loss of electrons can also effect the ionization of gas particles. The use of the gas ionization for pressure measurement was discovered in connection with the development of electron tubes at the beginning of this century. Today it is widely used for pressure measurement in the so called ionization gauges. These gauges are described in another chapter of these proceedings.

## 9. SURVEY OF THE OPERATIONAL RANGES OF VACUUM GAUGES

As a short summary in Table 2 are put together the operation ranges of the different gauges for the measurement of total pressures (ionization gauges are excluded).

Table 2
Short summary of the operational ranges of different gauges for the measurement of total pressures
(x, pressure reading is independent of the kind of gas; o, pressure reading depends on the kind of gas)

| **Name of the Range** | **Extreme High Vac.** | | **Ultra-High Vacuum** | | | | | **High Vacuum** | | | | **Medium Vacuum** | | | **Rough Vacuum** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| log (p/mbar) | -14 | -13 | -12 | -11 | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 10 |
| Piston gauges | | | | | | | | | | | | | xxx | xxx | xxx | xxx | xxx |
| Bourdon gauges | | | | | | | | | | | | | | | | xxx | → |
| Diaphragm gauges | | | | | | | | | | | xxx | xxx | xxx | xxx | xxx | xxx | |
| Piezoresistive diaphragm gauges | | | | | | | | | | | | | | xxx | xxx | xxx | → |
| U-Tube Gauge | | | | | | | | | | | | | | | xxx | xxx | → |
| McLeod Gauge | | | | | | | | | xxx | xxx | xxx | xxx | xxx | xxx | xxx | xxx | xxx |
| Modulation Gauge | | | | | | | | | ooo | ooo | ooo | ooo | ooo | ooo | ooo | ooo | ooo |
| Spinning Rotor Gauge | | | | | | | | ooo | ooo | ooo | ooo | ooo | ooo | | | | |
| Pirani Gauge | | | | | | | | | | | | ooo | ooo | ooo | | | |
| Radiometer Gauge | | | | | | | (xx) | xxx | xxx | xxx | xxx | xxx | xxx | xxx | | | |
| Radial Drift Gauge | | | | | | | | | | | | | | | | ooo | ooo |
| Elastic Scattering Gauge | | | | | | | | | | | (?) | (?) | (?) | ooo | ooo | ooo | ooo |

## 10. ACKNOWLEDGEMENTS

The author is grateful for the careful support in many technical activities of Mrs. U. Lehmann and Mr. Dipl.-Phys. Hauer. Both were very helpful in accurately preparing the manuscript and the slides for this lecture.

## 11. REFERENCES

### 11.1 General Reviews

S. Dushman, J.M. Lafferty : "SCIENTIFIC FOUNDATION OF VACUUM TECHNIQUE", John Wiley and Sons, Inc. New York, London, 1962

J.M. Lafferty (ed.) : "FOUNDATIONS OF VACUUM SCIENCE AND TECHNOLOGY", John Wiley & Sons, Inc. New York , Toronto 1997

Chr. Edelmann, H.-G. Schneider (editors) : "Vakuumphysik und –technik", Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1978

Chr. Edelmann: "Vakuumtechnik : Grundlagen und Anwendungen", Hüthig-Verlag, Heidelberg, 1985

Chr. Edelmann : "Vakuumphysik : Grundlagen, Vakuumerzeugung und –messung, Anwendungen", Spektrum Akademischer Verlag Heidelberg, 1998

## 11.2 Special Papers

[1] B. van Zyl : Check on the Linearity of a Capacitance Diaphragm Manometer between $10^{-4}$ and $10^{-1}$ Torr, Rev. Sci. Instrum. 47 (1976) 9, 1214

[2] Catalogue of the company LEYBOLD AG, HV 300, Part A1, date of publication: 05/92, pages 18 and 76/77

[3] H. McLeod, Apparatus for Measurement of Low Pressures of Gas, Phil. Mag. 48 (1874), 110

[4] R. Jurgeit, C. Hartung: Vorrichtung zur kontinuierlichen Messung von Gasdrücken. Wirtschaftspatent der DDR WP G 01 L/197067 Nr. 130850 vom 25 01 1977, erteilt am 3 5 1978

[5] R. Jurgeit, Dissertation, Technische Universität Dresden, 1981

[6] Chr. Edelmann: Stand und Entwicklungstendenzen der Totaldruckmessung in der Vakuumtechnik. Vakuum-Techn. 34 (1985) 6, 162 - 180

[7] R. Jaeckel: "Kleinste Drucke, ihre Messung und Erzeugung", Springer-Verlag, Berlin Heidelberg 1952

[8] F. T. Holmes: Rev. Sci. Instr. (1937) , 444

[9] J. K. Fremerey : High Vacuum Friction Manometer, J. Vacuum Sci. Technol. 9 (1972) 108

[10] T. Kobayashi, H. Hojo, M. Ono: Gas Concentration Analysis with a Quartz Friction Vacuum Gauge. Vacuum 47 (1996) 6/8, 479 – 483

[11] M. Ono, M. Hirata, K. Kokubun, H. Murakami: Quartz Friction Vacuum Gauge for Pressure Range from 0.001 to 1000 Torr. J. Vac. Sci. Technol. A 4 (1986) 3, 1728 – 1731

[12] M. Pirani. Ber. Dt. Phys. Ges. 4 (1906), 686

[13] E. Ja. Badinter, B.K. Zalewskii, I. G. Starusch: A Miniaturised Thermo-electric Pressure Gauge (in Russian), Prib. Tekh. Eksp. (1980), 248 – 250

[14] W. Gorski, A. Smiech: Convection Vacuum Gauge for Use Between $10^{-3}$ and 760 Torr, Le Vide No. 169 (1974), 221

[15] J.B. Huang, Q. Y. Tong : Constant Temperature Integrated Vacuum Sensor, Electron Lett. 24 (1988) 23, 1429 – 1430

[16] M. Knudsen: Ein absolutes Manometer. Ann. Phys. 32 (1910) 809 –842

[17] H. Klumb, H. Schwarz : Über ein absolutes Manometer zur Messung niedrigster Drücke, Z. Phys. 122 (1944), 418

[18] J. Lucas, T. Goto: A Pressure Gauge Utilising the Electron Diffusion Process and Microcomputer Control, Vacuum 34 (1984) 785 – 789

[19] Chr. Edelmann: Vakuum-Meßverfahren und Vakuum-Meßeinrichtung, Deutsches Patentamt, 02.12.1998

[20] Chr. Edelmann: Vakuum-Meßeinrichtung, Deutsches Markenamt, angemeldet 02.12.1998

[21] Chr. Edelmann, R. Kauert, St. Wilfert: Anwendung der Elektronendriftgeschwindigkeit zur Druckmessung, Frühjahrstagung der Deutschen Phys. Ges. Münster, 1999

[22] N.A. Kapzow: "Elektrische Vorgänge in Gasen und im Vakuum" (German translation from Rusian), VEB Deutscher Verlag der Wissenschaften Berlin 1955

# PRESSURE MEASUREMENT WITH IONIZATION GAUGES

*Karl Jousten*
Physikalisch-Technische Bundesanstalt, Berlin, Germany

**Abstract**
A brief history, the design, the use, and the calibration of ionisation gauges are described in this article.

## 1. INTRODUCTION

The pressure $p$ in an enclosed gaseous system is defined as the force $dF$ per area $dA$ exerted by the gas in the chamber. In a fundamental manner, forces can be measured for practical areas of a few square centimetres down to about 1 Pa, for example with elaborated U-tube manometers, filled with mercury or oil. In capacitance diaphragm gauges or membrane gauges the force is used to bend a membrane due to a differential pressure, but the force cannot be determined in a fundamental way and the gauge has to be calibrated. In the high and ultrahigh vacuum regime, however, it is no more possible to use the force on a certain area as indicator for pressure and other physical properties of the gas like gas friction, viscosity, thermal conductivity, or particle density are used to indicate pressure.

In ionisation gauges (IG) the particle density $n$ in their gauge volume is measured. Therefore it is important to remember the ideal gas law for an enclosed system in equilibrium

$$p = nkT\,. \tag{1}$$

It is not sufficient to measure $n$ with an ion gauge, but also the temperature $T$ of the gas has to be known to indicate pressure with an IG.

How is $n$ measured with an IG? As the name implicates, neutral gas molecules are ionised and then counted, usually by measuring a current. The ionisation normally takes place by electrons, but also photons (high intensity lasers) or ions can be used (Fig. 1).

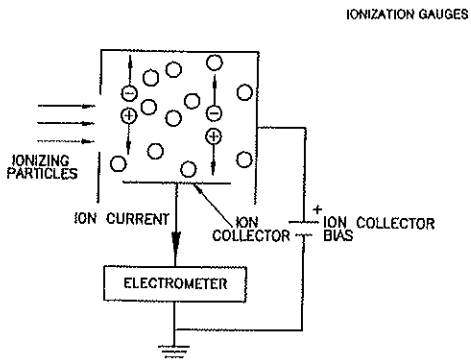



Fig. 1 The basic measuring principle of ionisation gauges. From James M. Lafferty, *Foundations of Vacuum Science and Technology*, John Wiley & Sons, New York, 1998.

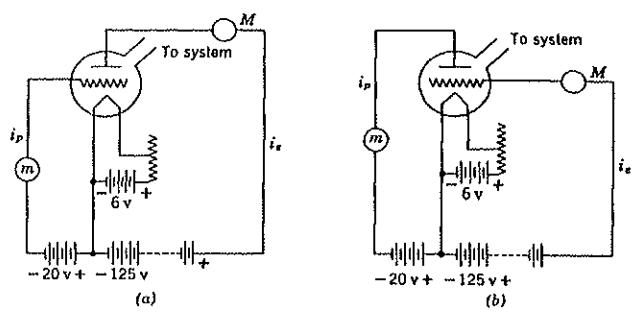



Fig. 2 Electrical circuits for triode ionisation gauges: (a) Internal control type. (b) External control type. From Saul Dushman, James M. Lafferty, *Scientific Foundations of Vacuum Technique*, 2nd edition, John Wiley & Sons, 1962.

## 2. BRIEF HISTORICAL REVIEW

The history of the IG dates back to 1909, when Baeyer showed that a triode vacuum tube could be used as a vacuum gauge. As inventor of the triode gauge, however, is usually named Buckley in 1916, who later improved the gauge to a lowest pressure measurement limit of about $10^{-6}$ Pa.

Three electrodes, sealed in a glass bulb, were needed for an IG: The cathode, as the source of electrons, the anode, and the collector of positive ions (Fig. 2). It was possible to use the grid as ion collector as shown in Fig. 2 (a), but to use the anode plate as collector (b) was customary because it was more sensitive. More ions were collected.

A few basic ideas shown in Fig. 2 are identical in today's gauges. That is, the ion collector has to be negative with respect to the cathode, so as to pick only ions and no electrons, and the acceleration voltage for the electrons has to be roughly 100 Volt. The reason is that the ionisation probability of a neutral gas molecule by an electron is energy dependent, and close to 100 eV there is a maximum for most gases as can be seen in Fig. 3.

Because the electron energy should be high on a their total path length, the acceleration voltage is usually tuned somewhat higher than 100 V. This has also the advantage that the ionisation cross-section differences between different gases are less emphasised.

The basic design of the triode gauge remained unchanged for more than 30 years, although the physicists wondered why all vacua stopped at about $10^{-6}$ Pa. The pumps improved continuously and in the 1930s and 40s there was considerable evidence from measurements of the rate of change of surface properties like the work function and thermionic emission that much lower pressures were actually obtained than were indicated by the IG.

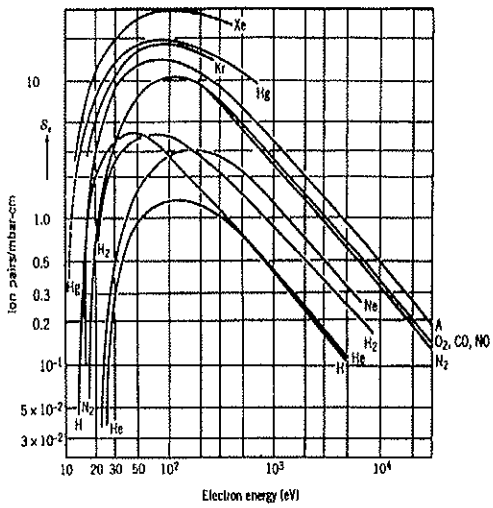



Fig. 3 Generated ions per cm electron path length per mbar at 20°C versus kinetic energy of incident electrons for various gases. From A. von Engel, *Ionized Gases*, AVS Classics Series.

At the 1st International Vacuum Congress in 1947, Nottingham suggested that the limit to the lowest measurable pressure was not caused by the pumps, but by an X-ray effect in the IG: He proposed that soft X-rays, produced by electrons impinging on the anode, released photoelectrons from the ion collector; this photocurrent was indistinguishable in the measuring circuit from the current due to positive ions arriving at the ion collector. This hypothesis was soon confirmed by Bayard and Alpert [1] who reduced the size of the ion collector from a large cylinder surrounding the other electrodes to a fine wire on the axis of a grid anode. This elegant solution reduced the lowest measurable pressure by a factor of 100 and is still the most common design in today's commercial IG: The Bayard-Alpert gauge or just BA gauge.

The history of the so-called cold cathode ionisation gauges is briefly outlined in section 3.2.

## 3. THE DESIGN OF TODAY'S MOST COMMON IONIZATION GAUGES

### 3.1 Introduction

The design of an IG must reply to four questions:

1. *How are the electrons produced*? Mainly, the electrons can be emitted from a heated wire or are being produced by discharge in a crossed high electric and magnetic field. These two types of electron production lead to two different types of ion gauges. It has become common notation to call the first one hot-cathode ionisation gauges (HCIG) and the latter cold-cathode ionisation gauges (CCIG). However, this distinction is no longer clear, because there exist nowadays so called hot-cathode ionisation gauges which have a field emitter source that operate at room temperature. Usually these field emitters are arrays manufactured from a silicon chip. The design of these gauges, however, is the same as in the true HCIG. It seems therefore to be better to distinguish between 'emitting cathode gauges' and 'crossed-field gauges' or 'discharge gauges'.

2. *How is the path of the electrons shaped?* There are long, short and oscillating paths.
3. *How is the path of the ions is shaped?* There can be a direct path from the ionisation place to the collector, but also longer and indirect paths, perhaps with an energy or impulse selection inbetween.
4. *What is the position and shape of the ion collector?*

### 3.2 Cold cathode gauges or, better, crossed-field ion gauges

The inventor of this type of gauge was Penning in 1937. He used a high voltage of up to 2kV to generate a discharge between cathode and anode. At low pressures (< 1 Pa or so) this discharge could only be maintained, if a magnetic field crossed the electrical field. The magnetic field greatly increased the path length of an electron from cathode to anode, so that it could generate another electron by impacting on a gas molecule to maintain the discharge.

It turned out that the discharge current was almost linearly proportional to the pressure in the gauge from 1 mPa to 0.1 Pa. Due to the magnetic field the electron is prevented from going directly to the anode and moves instead in helical paths through the gauge. The ions, because of their large mass, are virtually unaffected by the magnetic field and travel directly to the cathode. Secondary electrons released from the cathode by ion bombardment serve to build up and maintain the discharge (Fig. 4).

The discharge is generally not stable in crossed-field gauges. In the early designs the discharge became erratic below $10^{-3}$ Pa, and was often extinguished completely at $10^{-4}$ Pa. Therefore better designs were invented with the aim to increase the active volume of the discharge and so reduce discontinuities.

In Fig. 4 can be seen Penning's version of 1949, where the anode was changed from a ring in his original version to an open cylinder. This geometry is now widely used in ion pumps, but only for rugged and simple vacuum gauges.

A kind of breakthrough was accomplished by Redhead and Hobson, who invented the so-called magnetron and inverted magnetron gauge, the latter earlier designed by Haefer in 1955.

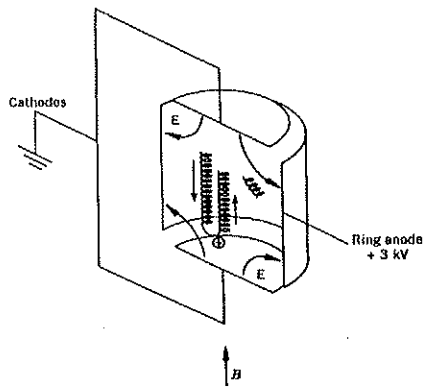



Fig. 4 Electrode arrangement, fields, and trajectories in the Penning gauge. From James M. Lafferty, *Foundations of Vacuum Science and Technology*, John Wiley & Sons, New York, 1998.

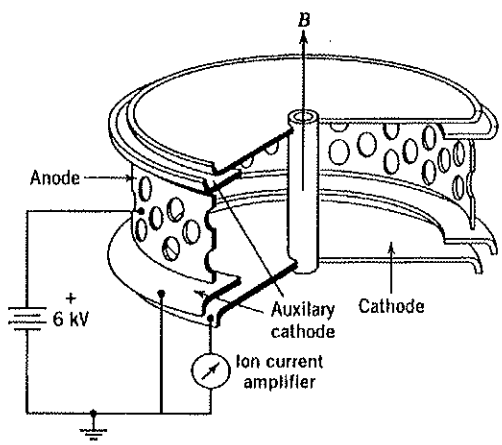



Fig. 5 Schematic diagram of the magnetron gauge. From ref. [2].

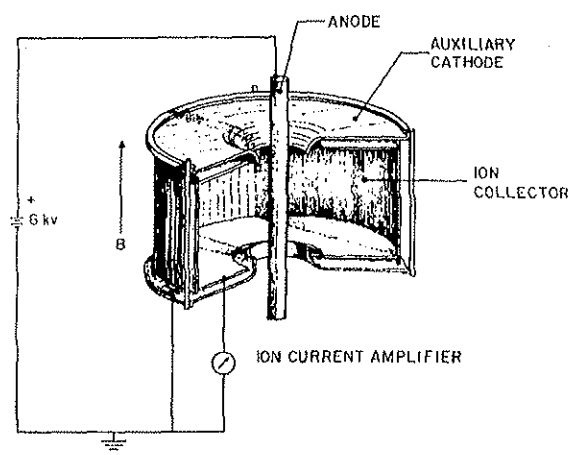



Fig. 6 Schematic diagram of the inverted magnetron gauge. From ref. [3].

In the magnetron gauge [2] (Fig. 5) the anode is an open cylinder with the cathode on axis and as endplates, in the inverted magnetron gauge [3] (Fig. 6) the anode is a rod on the axis of an almost closed cylinder as cathode. In the magnetron gauge, the end discs of the cathode are shielded from high electric fields by two annular rings held at cathode potential. Any field emission that can occur from the shield electrodes is not measured by the ion current amplifier.

One of the important features in the inverted magnetron gauge (IMG) is the use of guard rings held at cathode potential to prevent field emission currents from the cathode to the anode. The magnetic field is parallel to the anode axis. This gauge can be operated at up to 6 kV with 0.2 Tesla. In these gauges the electrons are trapped more efficiently than in the original Penning design. Due to this, the starting conditions are improved, the relations between $p$, $B$, $V$ follow reasonably the theoretical predictions, and the discharge is stable to much lower pressures. Redhead and Hobson claimed that their gauges could be used from $10^{-11}$ Pa up to $10^{-2}$ Pa.

Almost all available commercial crossed-field gauges are of the Penning design or of the Redhead and Hobson design as Magnetron or Inverted Magnetron. Normally, at low pressures, the gauges are operated with constant voltage, measuring the discharge current, while at higher pressures (> 10 mPa) they are operated at constant discharge current with accordingly reduced voltages. Otherwise, at constant voltage, the discharge current would be so high at higher pressures that heating and sputtering of material on the electrodes would become a problem.

In crossed-field gauges, the ion current vs. pressure relationship follows the equation

$$I^{+} = K \cdot p^{m}, \tag{2}$$

where $m$ depends on the type of gauge and varies between $m = 1...1.4$. However, $m$ also depends on pressure (Fig. 7) and this makes the situation rather complicated for reliable measurements. Generally, $m$ is higher for lower pressures than for higher and may reach values up to 2. If therefore in gauge controllers the relation for higher $p$ is extrapolated to very low pressures (< $10^{-7}$ Pa), the gauge will indicate at these small pressures lower pressures than actually present. At pressures as high as $10^{-10}$ Pa this error may be an order of magnitude.

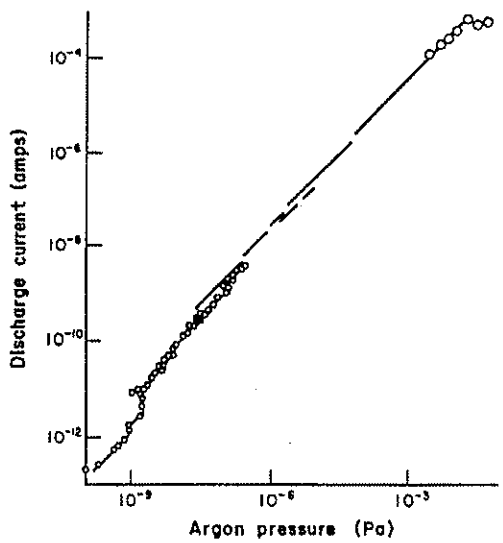



Fig. 7 Discharge current on a Penning gauge vs. argon pressure. From W.J. Lange, J.H. Singleton and D.P. Eriksen, JVST 3 (1966), 338.

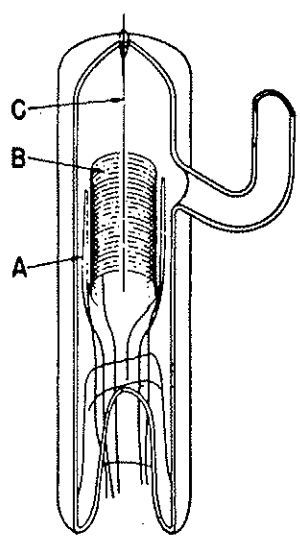



Fig.8 The original design of the Bayard-Alpert gauge. From R.T. Bayard and D. Alpert, Rev. Sci. Instr. 21 (1950), 571

Theoretical descriptions of characteristics of crossed-field Townsend discharges including electron space charge, which controls the discharge, have been given. However, Redhead [4] has pointed out that these theories have ignored the dynamics of dense electron space charge. The long trapping times of electrons allow large amplitude rf oscillations to build up. These oscillations modify

the static characteristics of the discharge and low frequency instabilities that are associated with mode jumping of the rf oscillations. Due to interaction of the electrons with the produced AC fields, excess energy electrons are generated which easily cross the magnetic field and hit the cathodes (Penning gauge) or the cathode end plates (magnetron or inverted magnetron). They falsify the discharge current or ion current. Since this effect is pressure independent it causes non-linearities in the current pressure curve. The rf oscillations may also cause serious measurement errors if unintentionally rectified in the ion-current amplifier.

In a summary comparison between crossed-field gauges and emitting-cathode gauges we will came back to this point.

### 3.3 Hot-cathode ionisation gauges or, better, emitting-cathode ionisation gauges

In our brief historical review we have already come to the early design of BA gauges and we will continue from there.

One of the main problems in the beginning of the BA gauge (Fig. 8) was instabilities in the gauge due to the floating potential of the glass envelope. Therefore the glass was furnished with a conductive layer which could be grounded or set to a defined positive potential. Also, it was soon noticed that ions could be lost through the open ends of the cylinder so the grid was closed at its ends to reduce this effect. The disadvantage of closing the grid seems to be that the pressure versus ion current ratio becomes non-linear for higher pressures at about 1 mPa, while this is only the case for the open cylindrical grid at pressures of 10 mPa or more [5].

In order to further reduce the X-ray limit it was tried to reduce the thickness of the collector wire. For example Van Oostrom [6] reduced its diameter to about 4 μm. Although with this method the X-ray limit is reduced, it has the disadvantage that the sensitivity is also reduced: ions formed inside the grid experience a radially inward force. Since angular momentum must be conserved, an ion with initial kinetic energy may not strike the collector wire, but rather go into orbit around it and tend to drift out axially from the electrode structure.

When the X-ray limit was reduced in this manner, another component to the background current became evident. Electrons hitting the anode may ionise molecules adsorbed on the surface causing their release. Ions generated in this manner cannot easily be distinguished from those generated in the gas phase. Since a grid structure of a BA gauge has a surface area of about 10 $cm^2$ the amount of adsorbed molecules can be rather high ($10^{16}$). Therefore it is important that the grid structure is very clean. Two measures are usually taken to cure this problem: The grid is cleaned by electron bombardment after the gauge has been exposed to high pressures or to atmosphere, and the electron current to the anode should not be too small during operation so that the gauge is continuously 'self-cleaning'.

By using the smallest practical diameter for the collector wire and increasing the sensitivity of the BA gauge by maximising the volume enclosed by the anode and using end caps the residual X-ray limit can be reduced to the low $10^{-9}$ Pa regime.

To measure pressures lower than $10^{-9}$ Pa, different approaches have been made:

- The X-ray current is measured so that it can be subtracted from the signal.
- Changes in the geometry of the gauge.
- Increasing the sensitivity by several orders of magnitude without reducing the background level.

The first two techniques have been found to be reliable and relatively easy to use in laboratory applications. The third method, however, has been disappointing, because reliable operation could not be demonstrated. Thus, there has been no widespread commercial development.

The first technique evolved when Redhead [7] suggested ion-current modulation by introducing an extra electrode into the grid space. This could be a wire close to the grid and parallel to the collector. When the wire is at grid potential, there is little or no effect on the gauge operation, but

when its potential is lowered by 100 V it seriously distorts the ion trajectories. The total ion current with the modulator at grid potential is

$$I_1 = I_r + I_g \tag{3}$$

where $I_g$ is the current from the ions and $I_r$ the residual current due to the X-ray effect. With the modulator at lower potential the true ion current $I_g$ falls to $m\,I_g$:

$$I_2 = I_r + mI_g \tag{4}$$

so that

$$I_g = \frac{I_1 - I_2}{1-m} \tag{5}$$

and

$$I_r = \frac{I_2 - mI_1}{1-m} \tag{6}$$

where $m$ may be determined at pressures where $I_r$ can be neglected. Usually $m$ was in the range 0.4 to 0.6, but Watanabe [8] succeeded in building a modulator gauge with $m = 0.05$.

Unfortunately it turned out that the residual current was modulated to a significant extent, because the electron trajectories were also modulated. Hobson [9] estimated that due to this, an error of $3 \cdot 10^{-10}$ Pa would be introduced in the measured pressure.

The second method of realising lower pressure measurement led, about 30 years ago, to the development of the so-called extractor gauge (Fig. 9). In this approach the ion collector is removed from sight of the grid. A simple lens is introduced between the grid and the collector to remove the ions to the collector. An ion reflector is used to reflect the ions onto the collector tip to increase the sensitivity in a way similar to that of a conventional BA gauge. In this way pressures from about $10^{-10}$ Pa can be measured.

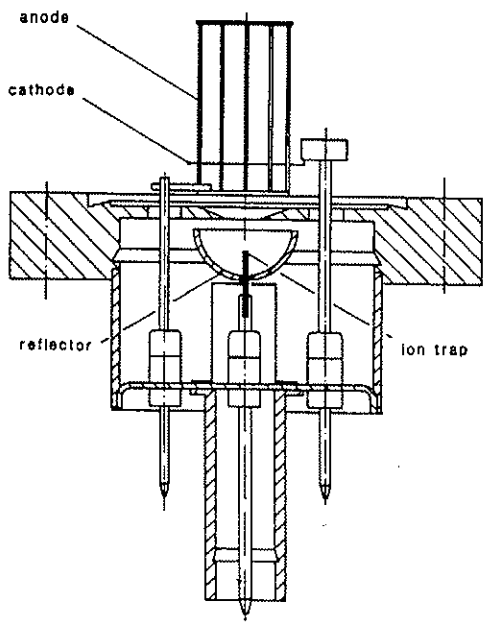



Fig. 9 Design of the extractor gauge manufactured by Leybold.

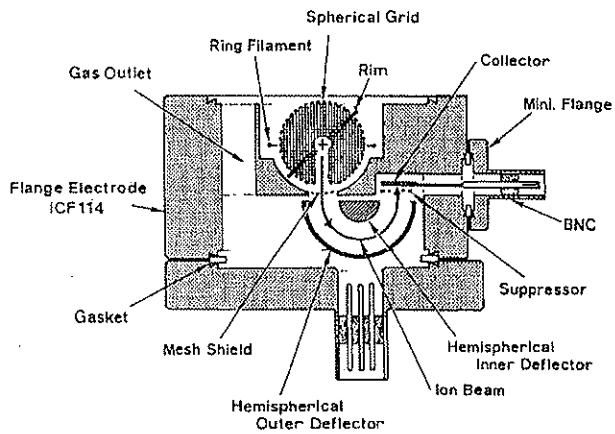



Fig. 10 Cutaway drawing of the ion spectroscopy gauge by F. Watanabe. From ref. [10].

A very sophisticated ion gauge was invented by Watanabe in 1992 [10], which he called the ion spectroscopy gauge (Fig. 10). This gauge has so many features that only the most important ones can

be mentioned. The gauge uses the extractor scheme, but with a hemispherical deflector so that the ion collector plate is completely out of sight of the grid. The collector is equipped with a suppresser electrode to inhibit electrons that are generated by reflected X-rays leaving the collector. With the hemispherical deflector where the inner electrode is at ground potential, and the outer at a variable positive potential, it is possible to separate the ions generated according to their energy. Ions which are generated at the anode grid (electron stimulated desorption effect) have higher energies than ions created in the gas phase due to a potential gradient from the grid to the extractor and due to space charge effects. The gauge uses a spherical grid in order to increase the space charge of the electrons in its centre. By this means ESD ions and gas ions can be separated (Fig. 11). Those parts of the gauge close to the hot filament could be outgassed by resistive heating or electron bombardment. Additionally, the housing of the gauge was made of high-thermal-conductance materials such as copper or aluminium in order to reduce warming of the gauge, which would stimulate hydrogen outgassing. Watanabe claimed a residual measurement limit of $2 \cdot 10^{-12}$ Pa for this gauge.

Probably the most well known example for the third method of approaching lower pressure limits is the so-called Lafferty gauge [11]. Lafferty adopted the diode magnetron principle by placing the filament along the axis of a cylindrical anode (see Fig. 12). An axial magnetic field provided by a magnet outside the enclosure forces the electrons to follow circular paths and increases their path length by several orders of magnitude. The electron emission current had to be very low (10 μA) to ensure stable operation. An X-ray limit of about $3 \cdot 10^{-12}$ Pa was calculated for this gauge.

It is not possible to mention all the approaches to reach lower pressure limits, or all designs of ionisation gauges, instead the reader is referred to text books or review articles.

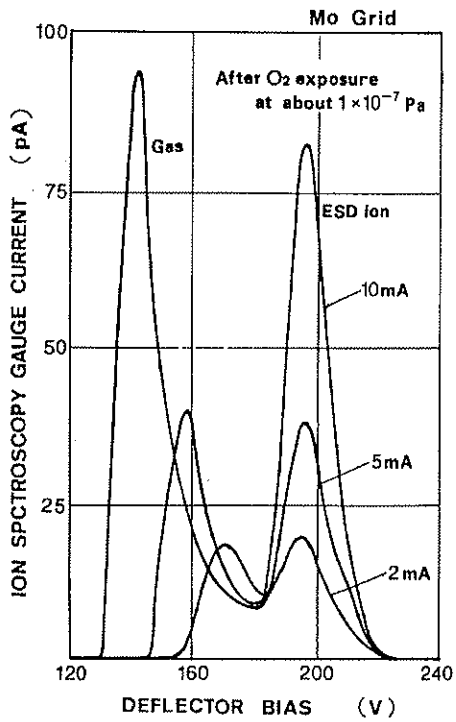



Fig. 11 Ion current vs. deflector bias voltage in the ion spectroscopy gauge after oxygen exposure at $10^{-7}$ Pa. From ref. [10].

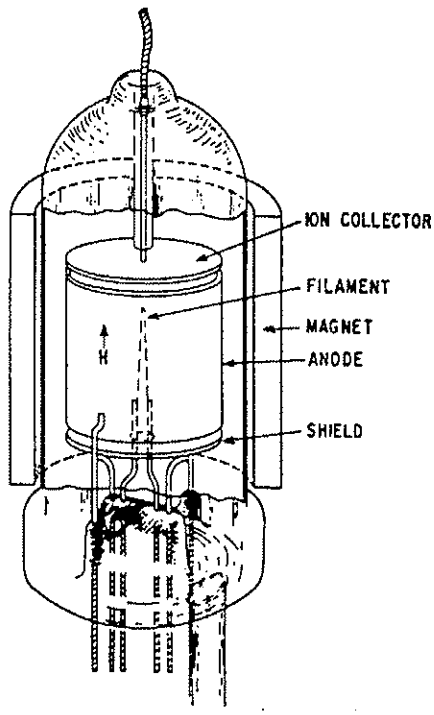



Fig. 12 Ionisation gauge designed by Lafferty to increase the electron path length. From Ref [11].

### 3.4 Summary comparison of crossed-field (CFG) and emitting-cathode (ECG) ionisation gauges

CFGs have the general advantage that they have no X-ray limit and electron stimulated desorption effects are small and cause little error. Also, because they already have a strong magnetic field, they are less affected by an outside magnetic field than an ECG. Where there is a suitable magnetic field,

for example in bending magnets in accelerators, this field can be used for the gauge. On the other hand, for example in electron microscopes, the magnetic field of a CFG may disturb the electron optics and must be carefully shielded.

CFGs have three generic disadvantages:

- Generally their output varies non-linearly with pressure.
- The very dense electron space charge trapped in these gauges leads to instabilities associated with mode jumping of the high-frequency oscillations.
- Their pumping speed is usually one or two orders of magnitudes higher than in ECGs and cannot be controlled.

That an ion gauge has a pumping action is a classical disturbing effect of many measuring devices, namely that it changes the value of the quantity it is designed to measure. The problems in CFGs with starting discharges at low pressures or extinction at low pressures are mostly solved in today's magnetrons or inverted magnetrons. In ECGs the electron emission current can be controlled, stabilised and varied. Mainly for this reason, ECGs are more stable and accurate, when they are conditioned before measurement.

When measuring pressures in the HV and UHV regions, a decision has to be made whether a CFG or a ECG should be bought. For this decision, the following points should be considered:

- Pressure range
- Gauge pumping speed
- Gas species to be measured
- Accuracy and stability
- Size and mechanical stability
- Interferences with magnetic fields
- Price

The available pressure ranges are very much the same for both types of gauges in the sense that there are gauges of either type for very low pressures ($<10^{-8}$ Pa) and relatively high pressures ($> 10^{-2}$ Pa). However, the accuracy of the ECG is significantly better at very low pressures. But, even at higher pressures, in all CFGs discontinuities in the pressure-vs-ion-current characteristic occur and these discontinuities are not reproducible. When a calibration is repeated, they may not occur.

Hot cathodes are extremely disturbing when other than rare gases or nitrogen, hydrogen etc. have to be measured. For chemically active gases, CFGs should be used since they can be cleaned more easily than ECGs. Very important, the price of a CFG is usually lower than for an ECG.

## 4. PROBLEMS IN USING IONIZATION GAUGES

Suppose there is a sealed-off chamber containing pure hydrogen at room temperature and which is not pumped. An ionisation gauge is installed to measure the pressure $p_1$ inside it. Now let us immerse the whole chamber in liquid nitrogen. The pressure will drop by

$$\frac{p_2}{p_1} = \frac{T_2}{T_1} = \frac{77}{300} = 0.257 \tag{7}$$

but the reading of the IG will be unchanged since the gas density is the same as before. This example shows how important it is to determine also the temperature during a measurement. Even gas temperature variations caused by room temperature variations have to be considered when gauges are accurately calibrated [12].

In other cases, when a chamber is continuously pumped, the molecular flow will adjust itself such that the law of continuity holds. For example, installing a gauge with a hot cathode in a tube (Fig. 13) results in the so-called thermal transpiration effect, where

$$\frac{p_1}{p_2} = \sqrt{\frac{T_1}{T_2}} \tag{8}$$

Since the hot cathode heats up its enclosure, the temperature $T_2$ will be larger than in the chamber $(p_1, T_1)$ and the pressure $p_2$ will be accordingly higher, but the reading of the ion gauge will be lower, because $n_2 = n_1 \cdot (T_1/T_2)^{1/2}$.

Another disadvantage of installing IG in tubes are problems associated with their pumping speed (Fig. 13). All IGs do pump, at least the ionised gas molecules, but pumping effects due to adsorption and dissociation can be much higher. If the conductance of the tube $C$ to the IG is comparable to the pumping speed $S$ of the gauge, the pressure in the IG is lower than that at the entrance of the tube. The advantage of installing a gauge in a tubulation is that the electrical field inside the gauge is not altered by different enclosures. Considerable sensitivity changes can be observed, when gauges are calibrated in the so-called nude configuration (Fig. 13) (no tubulation, but immersed in a large chamber) or in tubes of various inner diameters. Another advantage of tubulated gauges is that they are less sensitive to stray ions from a plasma process or other gauges.

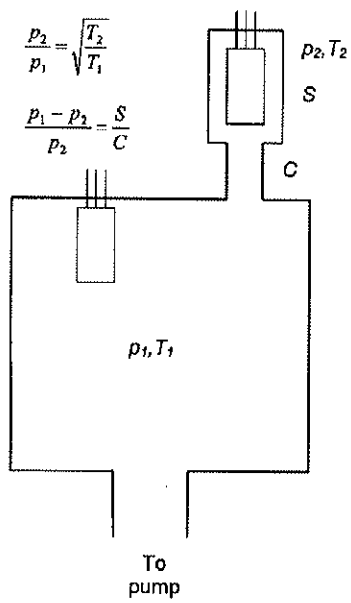



Fig. 13 Effects of tubulation of a gauge.

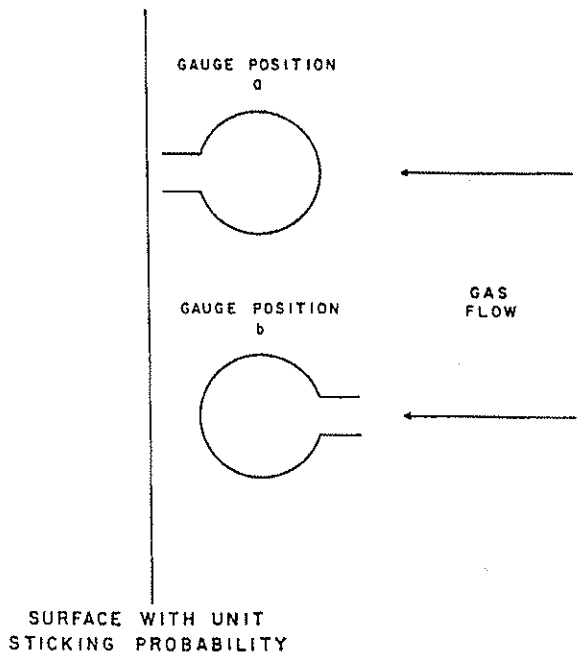



Fig. 14 Example of orientation effects when measuring gas pressures with vacuum gauges. From P.A. Redhead, J.P. Hobson, E.V. Kornelsen, *The Physical Basis of Ultrahigh Vacuum*, Chapman and Hall Ltd, London, 1968.

Other problems when measuring pressure are due to non-uniform pressure distributions inside chambers or net fluxes of molecular flow. Consider the example of Fig. 14, where gas flows from the right to the left and suppose the left wall is a cryo surface with a sticking probability of 1. The upper (a) ideal gauge (no internal gas source) will read zero, while in orientation (b) it will read an equilibrium pressure, which is determined by the differences is the rate of influx and the rate of return flow through the tubulation. Neither of these gauges gives the true pressure.

Problems with HIG arise with dissociation and enhanced chemical reactions on the hot cathode surface. For example with tungsten filaments, there is always carbon present on the surface that has diffused out of the bulk. Also, oxygen is present on its surface. Some reactions that can take place after dissociation of hydrogen are shown in Fig. 15 [13].

Outgassing and re-emission of molecules previously pumped by the gauge are significant problems in IG. A gauge operated at higher pressure will have a long relaxation time of hours or days, until a stable pressure at very low pressures is achieved. Outgassing rates of HIGs vary typically from $10^{-9}$ Pa l/s to $10^{-7}$ Pa l/s and are often the main source of gas when very low pressures must be achieved.

To obtain a reliable and long-term stable gauge reading, the gauge electrode surfaces must have a stable surface structure and composition. Not only the secondary electron yield on the collector changes with the surface composition, but also the numbers of secondary electrons generated by electrons hitting the anode grid are dependent on the anodes surface composition. Higher energy electrons (> 20 eV) also contribute to the number of ions generated in the gauge, hence the gauge sensitivity.

The gauge sensitivity depends on the gas species. Attempts to correlate this gas specific sensitivity accurately with ionisation cross sections fail due to other gas specific effects like ion capture probability, dissociation effects and secondary electron generation. Values for relative ionisation sensitivities (normalised for nitrogen = 1) presented in tables [14, 15] can be applied with some confidence while going from one gas to another, but the level of accuracy is only 10 to 20%. Where greater precision is required, gauges must be calibrated individually and for the gas in use.

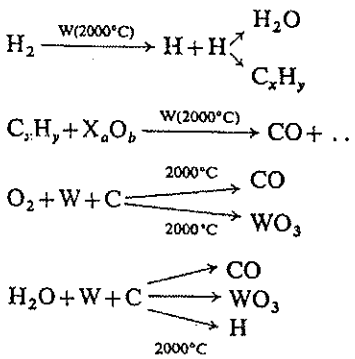



Fig. 15 Some chemical reactions which can occur at the hot tungsten filaments in ionization gauges [13].

As a final example of what can happen in a hot-cathode ionisation gauge (HIG), Fig. 16 shows calibrations of various HIGs, including BAG, EG, the Watanabe gauge, and BAG with gold electrodes for $H_2$ and $D_2$ [16]. Since the electronic structure of $H_2$ and $D_2$ is identical for the purpose of an IG, it could be expected that the relative sensitivity of $H_2$ to $D_2$ is exactly 1. As can be seen on Fig. 16, this is only true in some cases. Moreover, as can be seen on Fig. 17 the ratio is not even a constant for a single ion gauge. It varies with the treatment and history of the gauge. This is very surprising, since neither the potentials nor the geometry in the gauge were changed.

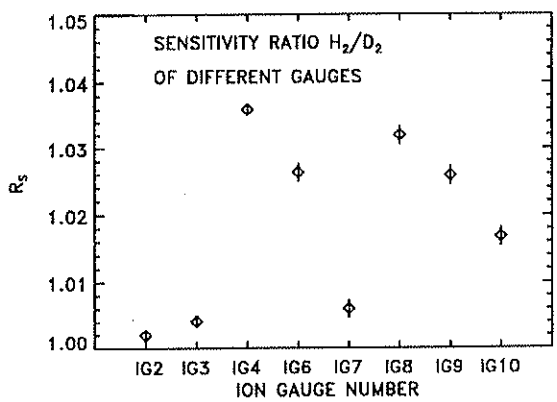



Fig. 16 Sensitivity ratio hydrogen/deuterium for various gauges. From ref. [16].

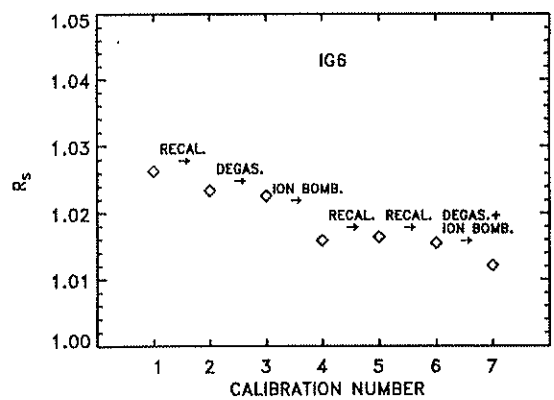



Fig. 17 Sensitivity ratio hydrogen/deuterium for a single gauge after various treatments. From ref. [16].

The reasons for the difference in the sensitivity for $H_2$ and $D_2$ is that $H_2$ due to its smaller mass and higher velocity in the same electric field gives a larger secondary electron yield at the collector than $D_2$. This higher secondary electron yield results in a higher current on the collector and therefore a higher sensitivity. The secondary electron yield on the collector depends strongly on the surface condition, so explaining why the ratio changes with treatments and history of the gauge.

If the secondary electrons are completely repelled to the collector by applying a negative potential on a suppressor grid in front of the collector, as in the ion spectroscopy gauge of Watanabe, the sensitivity ratio for $H_2/D_2$ equals indeed 1 (Fig. 18).

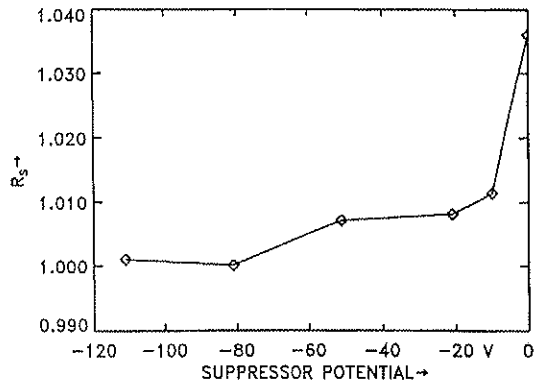



Fig. 18 Sensitivity ratio hydrogen/deuterium for the ion spectroscopy gauge in dependence of suppressor potential in front of ion collector plate. From Ref. [16].

## 5. IONIZATION GAUGE CALIBRATION

Most manufactures calibrate ion gauges in a rough manner before they leave the factory. This calibration gives typically an accuracy of within 20 to 50% for the calibration gas. If better accuracy is required, the ion gauge has to be calibrated with a primary standard or a secondary standard for vacuum pressures.

The calibration constant of an emitting cathode ionisation gauge is the so-called ionisation gauge constant or, less accurately, but now widely used, the sensitivity of an ion gauge. This is defined by

$$S = \frac{I^+ - I^+_{res}}{I^-(p - p_{res})} \tag{9}$$

where $I^+$ is the collector current at pressure $p$ and $I^+_{res}$ the collector current at the residual pressure $p_{res}$ and $I^-$ is the electron current. Simplified equations such as

$$S = \frac{I^+}{I^- p} \tag{10}$$

should not be used, because when $p$ is so low that $I^+$ is approaching its lower limit $I^+_{res}$ (X-ray limit, electron stimulated desorption and outgassing of the gauge) the sensitivity goes to infinity, which makes no sense (A high sensitivity is usually considered as something desirable).

In CFGs the ionising electron current cannot be measured and in this case the sensitivity is usually defined as [17]

$$S = \frac{I^+}{p^m} \quad S = \frac{I^+}{p^m} \tag{11}$$

where $m$ is a numerical exponent. This equation for a CFG is more simple than the one for ECG (Eq. (9)), because it is assumed that there is no residual collector current (field emission, however, may occur or voltage insulation problems may be present).

It is widely assumed that the collector current of the ECG is strictly linear with pressure, hence that $S$ as defined in Eq. (9) is pressure independent. This is generally not true. In cases, where high precision current meters are being used to determine $S$, typical relative variations of $S$ of a few % are found. In cases where lower-quality current meters are used typical for built-in devices for ion gauge control units, differences of $S$ between different pressure decades of 10% or more can be found. These differences are mainly due to imprecise resistors and rarely due to effects in the gauge itself.

The reason for the gauge inherent pressure dependence lower than about 1 mPa is unknown, but several effects may be responsible: space-charge effect may vary with pressure, secondary electron yield on the collector can be pressure dependent and also the electron emission distribution from the cathode may be pressure dependent. Above about 10 mPa it can be expected that the sensitivity will be pressure dependent due to intermolecular collisions and ion-neutral collisions, but also due to changes in space charge.

The accuracy of pressure measurement with calibrated ionisation gauges is mainly determined by long term instabilities of their sensitivity. Typically, high quality BA gauges have long-term instabilities of 2 to 5%.

## 6. IONIZATION GAUGE CALIBRATION METHODS

Two basic calibration methods exist for the calibration of ionisation gauges: calibration by comparison with a reference gauge or calibration on a primary standard for high and ultrahigh vacuum pressures.

The calibration by comparison is the less accurate method, mainly because the measurement uncertainty and the long-term instability of the calibrated reference gauge has to be taken into account. The calibration by comparison has to be carried out in an apparatus that ensures that the pressure or gas density is the same at the position of the test gauge and the reference gauge. In a recent review [18] systems for calibration by comparisons have been described. If available, it is recommended that a spinning-rotor gauge is used for the calibration of an ionisation gauge between $3 \cdot 10^{-4}$ Pa and $10^{-2}$ Pa. It is much more accurate than the calibration with an ionisation gauge due to the better stability of the spinning-rotor gauge compared to the ionisation gauge.

The calibration of an ionisation gauge on a primary standard is the most accurate calibration method, because a primary standard has the highest possible metrological quality and deduces the pressure unit to the corresponding SI-units. Primary standards for high and ultrahigh vacuum pressures are normally pressure generators, i.e. well-known pressures with a correlated uncertainty are generated in there. The methods of pressure generation have recently been reviewed [18] together with the procedures for calibrating ionisation gauges. Primary standards for vacuum pressures are available in the major National Metrological Institutes of the world, among them the Physikalisch-Technische Bundesanstalt (PTB, Germany), the National Institute of Standards and Technology (NIST, USA), the National Physical Laboratory (NPL, England) and the Istituto di Metrologia G. Colonetti (IMGC, Italy).

## REFERENCES


[1] R.T. Bayard and D. Alpert, Rev. Sci. Instr. 21 (1950), 571.

[2] P.A. Redhead, Can. J. Phys. 37 (1959), 1260.

[3] J.P. Hobson and P.A. Redhead, Can. J. Physics 36 (1958), 271.

[4] P.A. Redhead, J.P. Hobson and E.V. Kornelsen, *The Physical Basis of Ultrahigh Vacuum*, Chapman and Hall Ltd, London, 1968, p.335.

[5] R.N. Peacock and N.T. Peacock, J. Vac. Sci. Technol. A 8 (1990), 3341.

[6] A. Van Oostrom,. *Vac. Symp. Trans. Comm. Vac. Tech., Pergamon* 1 (1961), New York, 443.

[7] P.A. Redhead, Rev. Sci. Instr. 31 (1960), 343.

[8] F. Watanabe, J. Vac. Sci. Technol. A 5 (1987), 242.

[9] J.P. Hobson, J. Vac. Sci. Technol. 1 (1964), 1.

[10] F. Watanabe, J. Vac. Sci. Technol. A 10 (1992), 3333.

[11] J.M. Lafferty, J. Appl. Phys. 32 (1961), 424.

[12] K. Jousten, Vacuum 49 (1998), 81.

[13] D. Alpert, Le Vide 17 (1962), 19.

[14] J.H. Leck, *Total and Partial Pressure Measurement in Vacuum Systems*, Blackie, 1989, Glasgow and London, p.73.

[15] R.L. Summers, *NASA Tech. Note* NASA TN D-5285 (1969).

[16] K. Jousten and P. Röhl, Vacuum 46 (1995), 9.

[17] A. Berman, *Total Pressure Measurements in Vacuum Technology*, Academic Press, Orlando, 1985, p. 45.

[18] James M. Lafferty, *Foundations of Vacuum Science and Technology*, John Wiley and Sons, New York, 1998, Chapter 12.

**BIBLIOGRAPHY**

A. Berman, *Total Pressure Measurements in Vacuum Technology*, Academic Press, Orlando, 1985.

Saul Dushman, James M. Lafferty, *Scientific Foundations of Vacuum Technique*, 2nd edition, John Wiley & Sons, 1962.

James M. Lafferty, *Foundations of Vacuum Science and Technology*, John Wiley & Sons, New York, 1998.

J.H. Leck, *Total and Partial Pressure Measurement in Vacuum Systems*, Blackie, Glasgow and London, 1989.

P.A. Redhead, J.P. Hobson, E.V. Kornelsen, *The Physical Basis of Ultrahigh Vacuum*, Chapman and Hall Ltd, London, 1968. This book has recently been re-edited by the American Vacuum Society.

# PARTIAL PRESSURE MEASUREMENT

*J.H. Leck*
University of Liverpool, Liverpool, UK

**Abstract**
The quadrupole mass spectrometer is now well established as the instrument used almost universally for partial pressure measurement. This spectrometer can be considered as a hot-cathode extractor ionisation gauge with the addition of an electrostatic mass filter between the ion source and the ion collector. This filter consists of a square array of circular rods approximately 6 mm diameter and between 100 to 150-mm long. Positive ions are injected parallel to the central axis towards the ion collector. A filtering action is achieved by applying combined radio frequency and dc potentials between opposite pairs of rods. The characteristics of the filter can be controlled by small changes in the relative values of the radio frequency and dc voltages and the mass range by changes in the frequency of the ac signal. Care must always be taken when operating this mass spectrometer in an analytical mode because of non-linearities caused by space charge and other effects which are important, particularly in the injection region between the ion source and filter. Also the filter is subject to loss of sensitivity over the long term, generally caused by contamination from hydrocarbon impurities in the vacuum system. This is not normally a problem in ultra-high vacuum operation. Although almost all filters have been designed to operate with relatively low voltages applied to the rods, alternative modes are being considered which require distinctly high voltages or lower frequencies. There are hopes that these will have advantages for general operation.

## 1. INTRODUCTION

There are a number of different types of mass filter — mass spectrometer — that have been considered for the analysis of the gases found in vacuum systems. All are based upon the hot cathode ionisation gauge, with the addition of some form of mass filter placed between the ionisation chamber and the ion collector. The pressure range to be considered is from approximately $10^{-4}$ mb down to ultra-high-vacuum. The four instruments that have been responsible for the major share of the development are:

- The magnetic deflection mass spectrometer
- The Omegatron
- The time-of-flight mass spectrometer
- The quadrupole mass filter

All have the advantages and disadvantages of hot-cathode devices (e.g. problems due to filament outgassing) and all operate over roughly the same pressure range.

Not surprisingly the magnetic deflection mass spectrometer, because of the general interest in its performance, was the first instrument to be used. Developments were particularly significant during the period 1940 to 1960. Figure 1 illustrates the principles of operation of the $180^0$ deflection instrument. The positive ions generated in the electron stream are formed into a "ribbon-like" beam and ejected into the magnetic field through a narrow slit in the base of the ionisation chamber. In this field, usually generated by a permanent magnet, they are bent in circular paths through $180^0$ towards the defining slit in the collector plate. The radius of curvature of paths ($R$) depends upon (i) the

strength of the magnetic field, (ii) the charge to mass ratio of the ions and (iii) the energy of the ions. For ions of a given charge-to-mass ratio the radius of curvature is directly proportional to the square root of their kinetic energy and inversely proportional to the magnetic field strength. As illustrated in Fig. 1, some ions will have a radius exactly equal to $R_0$ and reach the collector plate through the collector defining slit. The "heavier" and "lighter" ions have respectively radii greater or less than $R_0$ and, hence, fail to reach the collector. The various components of the ion beam can be scanned across the collector defining slit by varying the kinetic energy of the ions (i.e. by controlling the bias voltage $V$ of the ion source).

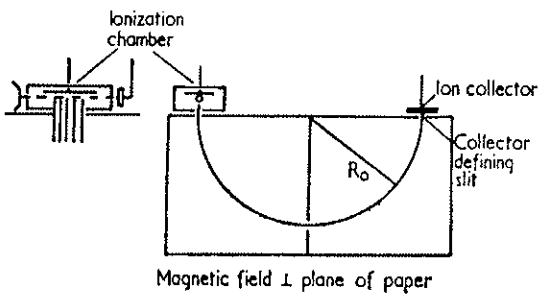



Fig. 1 180° deflection magnetic mass spectrometer.

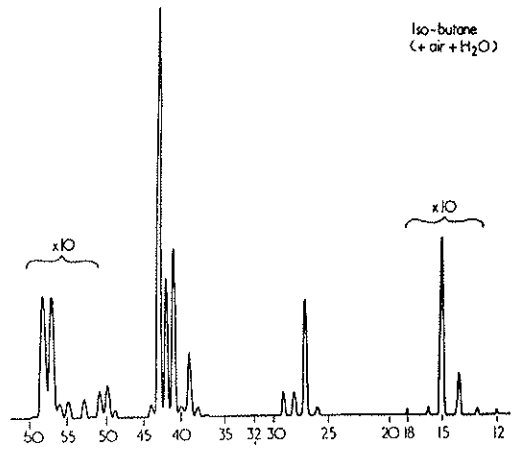



Fig. 2 Typical spectrum for magnetic deflection mass spectrometer showing iso-butane with a small impurity of water vapour and air [1].

Figure 2 depicts a typical spectrum for this instrument. In this diagram the ion current to the collector is plotted as a function of the source bias voltage (i.e. the kinetic energy of the ions). It will be noticed that this method of scanning results in a non-linear display; the value of $V$ being very low for ions of high charge-to-mass ratio. In fact a useful output is obtained with $V$ kept constant and scanning effected by varying the magnetic field strength. Unfortunately the electro-magnets required to generate variable magnetic fields tend to be both bulky and unrealistically expensive for general vacuum use.

An alternative to the magnetic deflection spectrometer is the Omegatron which, unfortunately after extensive development showing promise, has not found significant practical applications. The electrodes of this analyser form a small cube with insulated plates at top and bottom across which a radio frequency potential difference can be applied. This is illustrated in Fig. 3. Positive ions formed in the central electron beam are forced, by the strong magnetic field parallel to this axis, to travel in circular paths, as indicated in Fig. 3. The time of revolution $T$ in the magnetic field is given by

$$T = \frac{2\pi M}{eB}$$

with the usual notation. It is significant that this time is dependent upon the magnetic field strength and the mass of the ions, but not their energy. Thus ions with one particular mass will resonate with the field applied across the RF plates and, therefore, gain energy continuously. The radius of curvature of their paths will increase until they strike the collector. A complete e mass spectrum can be obtained by varying either the magnetic field strength or the frequency of the RF signal. Unfortunately, although this instrument is small and easily fitted to a vacuum system, it requires a very strong magnetic field for effective operation. Resolution is inversely proportional to mass in the omegatron; making performance poor at the high end of the spectrum. This does however mean that, although not practical for general use, it could have applications where interest is confined to "light" ions; resolution is certainly excellent in the range 1–4 amu.

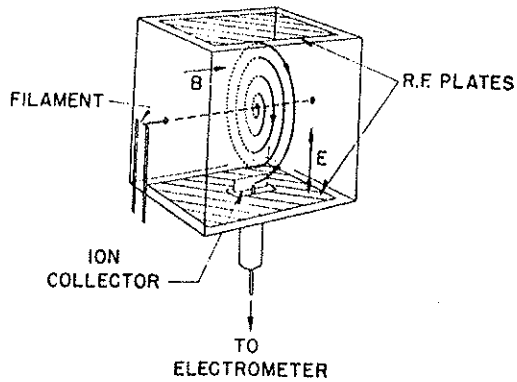



Fig. 3 The Omegatron [2].

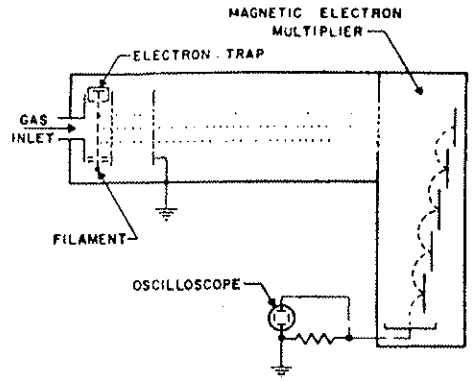



Fig. 4 The time-of-flight mass spectrometer [3].

The time-of-flight mass spectrometer has the advantages of a relatively small size and, more important, does not require any magnetic field for its operation. The principle of operation is simple; ions are formed as a bunch in a pulsed electron beam (say of 1μsec. duration). They are all accelerated through the same potential difference (therefore receiving the same kinetic energy) and are constrained to drift in a field-free space to an ion collector, as shown in Fig. 4. Obviously mass separation will occur because light ions, travelling faster, will reach the collector before the heavy ions. With a sensitive multiplier and an amplifier with a small time constant the arrival time of the various ions can be measured and displayed. This instrument gives an intermittent output signal that limits its value. It also has the disadvantage of low sensitivity and should high performance be required a long drift tube must be used, thus negating the advantage of small size. This has inhibited its development for general vacuum use.

For practical operation it has been established that none of the above instruments, with the possible exception of the magnetic deflection mass spectrometers, can mount a serious challenge to the quadrupole mass filter which is now used almost universally in general vacuum application.

## 2. THE QUADRUPOLE MASS FILTER

In its simplest form this filter (or analyser) can be considered as a conventional extractor ionisation gauge with an electrostatic mass filter interposed between the ion source and the ion collector. The filter takes the form of four parallel rods of circular, or near circular, cross-section mounted symmetrically about a central axis. In a typical instrument for general vacuum use these rods are between 5 and 10-mm diameter and 50 to 150-mm long. Precision in mounting the rods is most important. Opposite rods in the assembly are connected together electrically, the structure being shown schematically in Fig. 5. Ideally the rods must have the hyperbolic profiles illustrated in Figs. 6a and 6b. Practical experience has shown that the actual field is sufficiently close to hyperbolic when circular rods are substituted for those shown in Fig. 6a, provided the radius r of the rods is chosen to be such that $\boldsymbol{r = 1.147 r_0}$, $r_0$ being the inscribed radius shown in Fig. 6b. With a suitable electrical potential applied between the two pairs of rods a filtering action takes place for ions injected from the source along the central axis of this symmetrical structure. The electric field (in the plane at right angles to the z axis of Fig. 6a) must be hyperbolic in order to obtain the best filtering action.

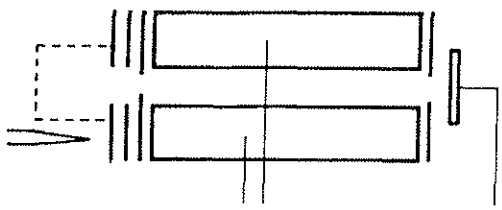


Fig. 5 The quadrupole mass spectrometer.

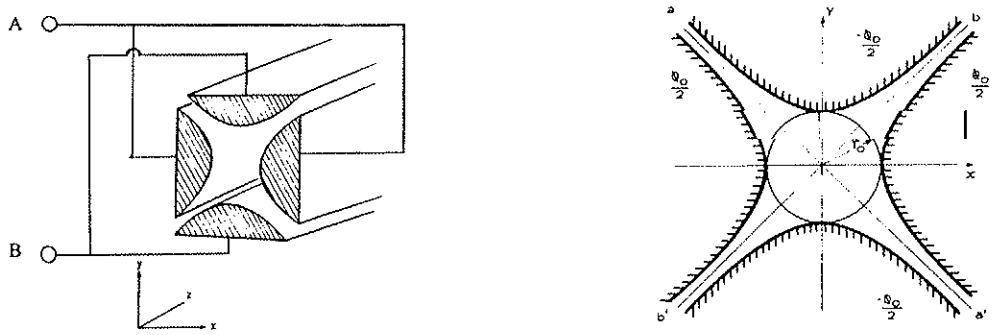



Fig. 6 Rod assembly in quadruple filter. (a) Showing electrical connections. (b) Defining inscribed radius.

In order to obtain the filtering action an RF voltage superimposed upon a dc voltage must be applied between the two pairs of rods, i.e. a voltage $U + V \sin \omega t$ must be applied between terminals A and B in Fig. 6a. Theory indicates that $U$, $V$ and $\omega$ can be chosen such that, whilst the majority of ions will be unstable in the $x$ or $y$ directions (or both), a small fraction may be stable in both $x$ and $y$ directions and therefore travel the full length of the filter ($z$ direction in Fig. 6a) to reach the collector. The criteria for determining stability can be obtained from the diagram presented in Fig. 7. This maps out the boundaries of stability/instability for what is in effect the (U V) plane. Because the stability criteria depends upon $M$, $\omega$ and $r_0$ as well as $U$ $V$, Fig. 7 is presented in terms of the more general co-ordinates $a$ and $q$ which are defined as follows:

$$a = \frac{4eU}{m\omega^2 r_0^2} \qquad q = \frac{2eV}{m\omega^2 r_0^2}$$

If a point plotted in Fig. 7 for particular values of $a$ and $q$ lies inside the triangle then the ion paths are stable; if a point lies outside the triangle then the paths are unstable. It has to be pointed out that the boundaries of the triangle in Fig. 7 are strictly speaking only correct for very long filters; as basic theory only concerns itself with ions that have exposure to virtually an infinite number of cycles of the RF field. However, Fig. 7 is a sufficiently good approximation provided the ions are subject to more than about 100 cycles of the RF voltage in the filter.

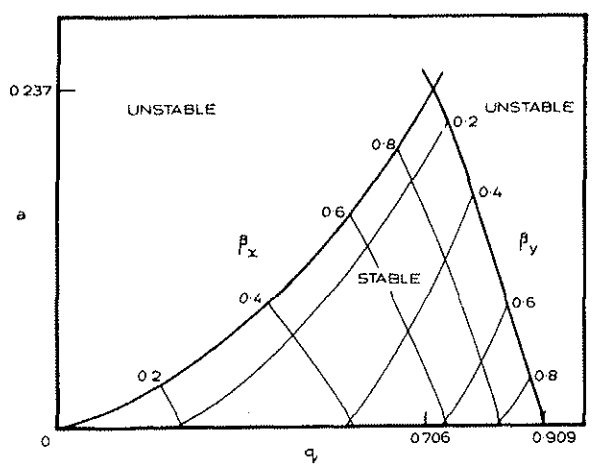



Fig. 7 The first stability zone.

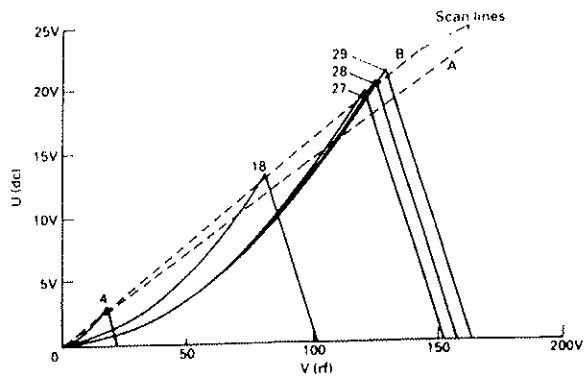



Fig. 8 Stability diagram plotted for singly charged ions of mass 4, 18, 27, 28, & 29 amu [4].

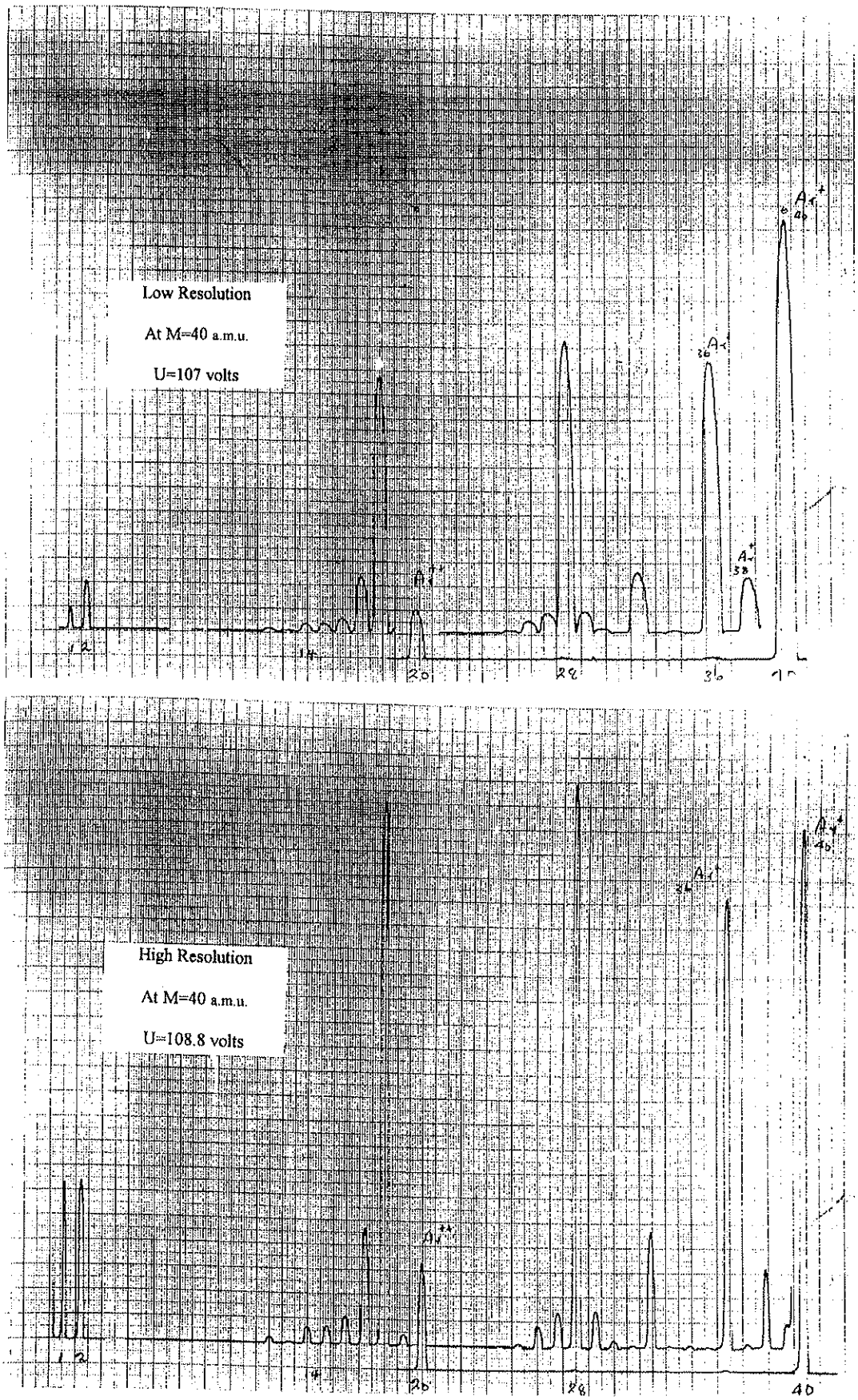



Fig. 9 Mass spectra obtained at both low and high resolution.

The significance of the "stability triangle" is illustrated clearly in Fig. 8. In this figure triangles are drawn in terms of $U$ and $V$ (for fixed values of $\omega$ and $r_0$) for five different values of ion mass $M$ (singly charged ions with mass 4, 18, 27, 28 & 29 a.m.u). These five triangles are computed for an ac frequency of 2 MHz and a rod radius of 3.25 mm. A family of similar triangles is obtained with the tips of the triangles all lying in a straight line that passes through the origin. Obviously mass spectra can be obtained if the applied voltage is increased from zero to a high value, keeping the ratio of $U/V$ constant. Two such "scan lines" are shown in Fig. 8. Obviously by increasing the slope of this line an increase in resolution will be obtained. Corresponding experimental results are given in Figs. 9a and 9b. These two diagrams show clearly how the resolution increases significantly for only a 2% change in the slope of the scan line (i.e. an increase in $U$ at mass 40 amu from 107 to 108.8). The ability to control resolution by a simple adjustment of the electrical parameters is an important advantage of the quadrupole over the magnetic deflection instruments. (In the magnetic instruments resolution can only be changed by physically altering the defining slit widths at source and collector). A further advantage — not immediately obvious — is the requirement of the quadrupole for increased applied voltage levels at the highest masses. This means that, unlike the magnetic deflection spectrometers and the omegatron, the quadrupole is least sensitive to electrode surface contamination at the high mass end of the scale. This is a positive advantage because usually it is at the highest masses where the highest precision is required.

There is good agreement between the experimentally measured performance, as typified by the data presented in Fig. 9, and the theoretical predictions of Figs. 7 and 8. This is true in spite of the small, but numerous, approximations in the mathematical model. For example, as already noted, the ions are subject to only a finite number of cycles of RF field. Also the fact that the ions have to travel through fringing fields to enter and exit the filter is completely ignored in the basic theory. It is also true that there must be some deviations from a true hyperbolic field because (i) the electrode surfaces are almost always circular in section rather than hyperbolic, (ii) there must be some errors in mounting the rod assembly and (iii) there will occasionally be some spurious electrical charge due to insulating layers building up on the electrode surfaces.

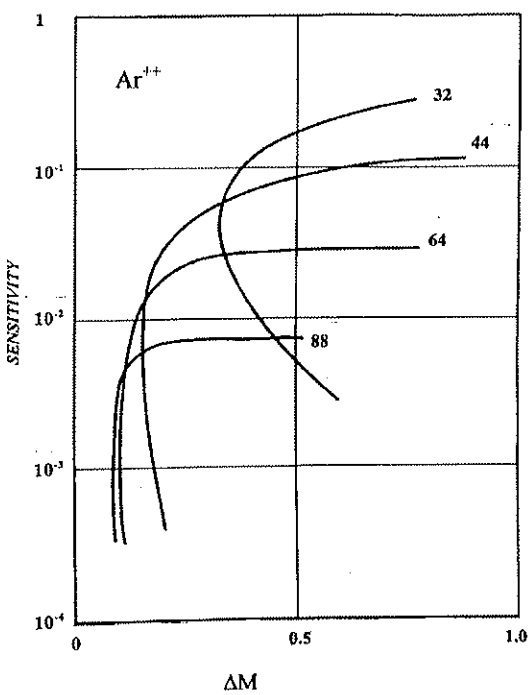



Fig. 10 Experimental results showing relation between sensitivity and resolution for different values of $N$ the number of cycles the ions spend in the filter (the values of $N$ are marked on each curve).

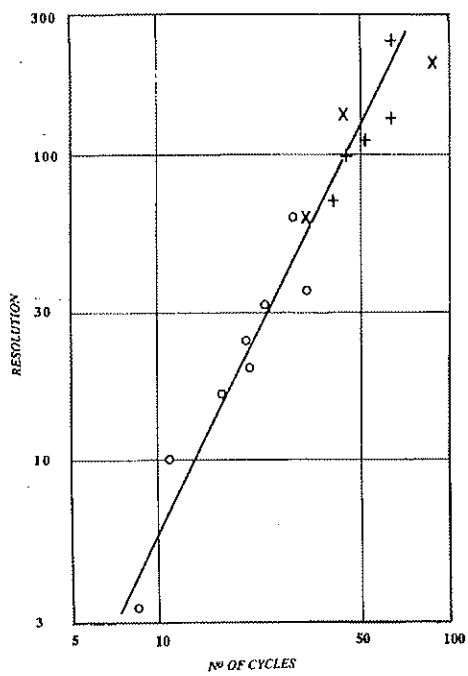



Fig.11 Relation between maximum resolution and the number of cycles the ions spend in the lens.

From data obtained in experiments of the type depicted in Figs. 9a and 9b families of curves, such as that shown in Fig. 10, can be built up. These demonstrate how sensitivity and resolution depend very much upon $N$, the number of cycles of RF which the ions "experience" in the filter. In Fig. 10 the number of cycles $N$ is indicated on each curve; thus in this particular filter maximum resolution attainable increases from about 50 to 200 when $N$ is increased from 32 to 88 cycles. Resolution is defined here, and throughout the paper, by the ratio $M/\delta M$ where $\delta M$ is the peak width at 10% of the peak height. This is conventional.

Figure 11 obtained from many families of curves similar to those in Fig. 10 shows the simple relationship between maximum resolution and number of cycles $N$ over a range of values of ion charge to mass ratio. Although there is considerable scatter in the experimental points (bearing in mind that this is a log-log plot), the relationship between $R$ and $N$ is given by $R = N^n$ where $n$ is close to 2.

## 3. THE DESIGN OF PRACTICAL QUADRUPOLE FILTERS

The requirements for optimisation of the four important parameters are as follows:

**Increase Resolution:**

1. Increase length of filter
2. Increase frequency of RF
3. Decrease ion injection energy

**Decrease cost**

1. Decrease frequency of RF
2. Decrease length of filter

**Increase sensitivity**

1. Increase ion injection energy
2. Introduce a multiplier at the collector
3. Increase electron current

**Increase mass range**

1. Reduce frequency
2. Increase the ac and dc voltages

Unfortunately there are serious clashes in the requirements, which means that compromises are necessary in virtually all designs. For example a decision to increase the resolution of a given filter by increasing either its length or the frequency of the RF supply incurs a cost penalty. The manufacture of a long filter is obviously more expensive than a short filter (also it can be less convenient to attach to a vacuum system). An increase in RF frequency leads to an increased requirement from the ac and dc voltages ($U$ and $V$). This means more expensive and larger power supplies. Increasing resolution by reducing the speed with which the ions travel through the filter in the $z$ direction (i.e. by reducing their injection energy) appears attractive but, unfortunately, a lower sensitivity must be accepted. This might be compensated by introducing an electron multiplier into the collector assembly as shown in Fig. 12. Again there is a cost penalty.

Stability of performance over a long period, a factor of considerable practical importance, is difficult to quantify. It is well established that stability depends significantly upon the energy with which the ions are injected into the filter. Obviously changes in the performance of the instrument with time depend critically upon the atmosphere in the vacuum system. Whilst in ultra-high vacuum systems instability will probably not be a problem it can be significant even in relatively clean, but unbaked, systems.

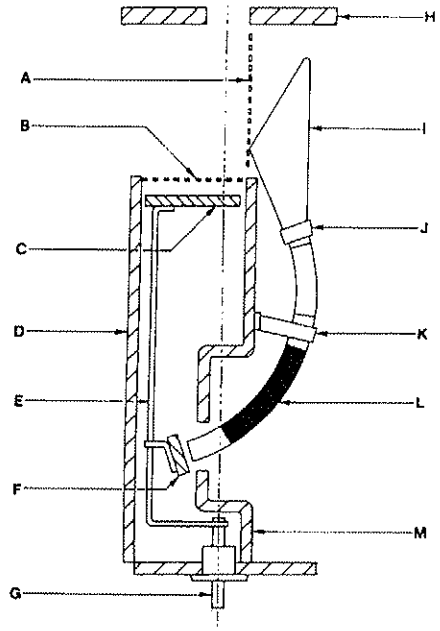



Fig. 12 Channel electron multiplier connected to quadrupole exit plate H. A, high-voltage; B, grounded mesh; C, ion collector; D, grounded shield; E, signal lead; F, electron collector; G, output connector; I, CEM cone; J, high-voltage connection; K, ground connection; L, bias resistor; and M, grounded shield, [5].

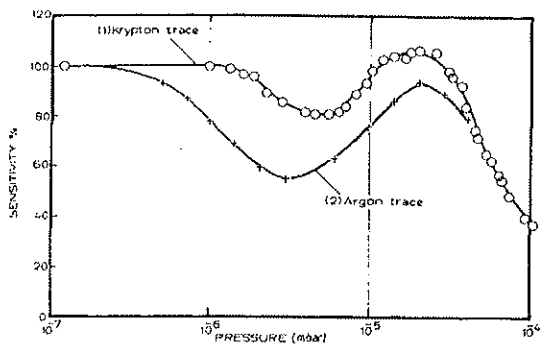



Fig. 13 Observed variation in sensitivity over the pressure range $10^{-7}$ to $10^{-4}$ mbar for a residual gas analyser [6].

Since space charge fields, both negative and positive, can be more important in the mass filter than in the ionisation gauge it is important to recognise that (i) because of the negative space charge the sensitivity, i.e. (output current)/(partial pressure) is not necessarily proportional to electron current and (ii) because of positive space charge, sensitivity is not necessarily independent of pressure even when mean-free-path effects can be neglected. The magnitude of the non-linearity depends very much upon individual instrument design.

A simple technique for measuring non-linearity in the output vs pressure characteristics is illustrated in Fig.13. This presents the data obtained from two sets of measurements obtained with a dynamic calibration plant where two gases were introduced under controlled conditions through separate inlet valves Ref. [6]. In one experiment a flow of argon was first introduced so as to maintain a constant partial pressure of the order of $10^{-8}$ mb then krypton was introduced at a steadily increasing rate, taking the total pressure in the system from below $10^{-7}$ mb to about $10^{-4}$ mb. The variation in the indicated partial pressure for argon demonstrates the change in sensitivity for argon over this pressure range $10^{-7}$ mb to $10^{-4}$ mb. For completeness the experiment has been repeated with the two gases reversed, i.e. krypton now becoming the trace gas. Both sets of results indicate a significant dip in sensitivity at comparatively low pressures, where effects of mean free path can be neglected. This experiment indicates the need for caution in interpreting data from these instruments, because the instabilities in the signals from the trace gas could erroneously be interpreted as a fall in pressure for these gases.

## 4. THE FOUR OPERATING ZONES FOR THE QUADRUPOLE FILTER

All references to theoretical analysis and experimental data in the above sections refer to the filter in the "first operating zone". Extending the analysis from the very brief presentation above shows clearly the existence of a number of additional stability zones; which may well have practical significance.

Meaningful theoretical analysis is based upon a solution of the Mathieu equation which defines stability in both the $x$ and $y$ directions. Analysis and its interpretation is simplified because motion in the $x$, $y$ and $z$ directions ($x$, $y$ and $z$ as defined in Fig. 6) is uncoupled and, therefore, the ion motion in

these three directions including the conditions for stability can be considered independently. Further the $z$ direction can be discounted as the electric field in this direction is, in theory, always zero. Thus from the analysis point of view this filter presents a two dimensional problem, with displacements as a function of time in the $xz$ and $yz$ planes being the defining factors. Results of the analysis of the Mathieu equation show clearly defined stability zones in the $aq$ diagram for motion in the $xz$ and $yz$ planes. These zones are shown for a relatively large range of $a$ and $q$ for motion in the $xz$ plane (Fig. 14).

Stability in the $yz$ plane is obtained by reversing the $a$ axis, i.e. $+a$ becoming $-a$ and vice-versa. Figure 15 presents the stability characteristics for both the $xz$ and $yz$ planes superimposed in one diagram. It shows four regions where motion is stable in both planes. These are regions where ions will be able to pass through the filter. The first stability region (the region depicted in Fig. 7) is the small triangle near the central origin, difficult to distinguish in this larger scale diagram.

Although operation in the second and higher stability zones have been neglected in the past, there is now indication of an interest in developing filters for operation in zone III of Fig. 15.

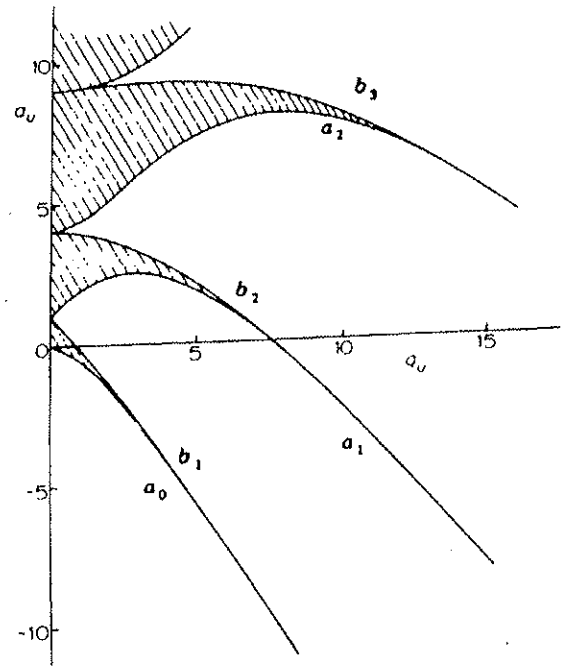



Fig. 14 Stability diagram for the Mathieu equation for either the $xz$ or $yz$ directions.

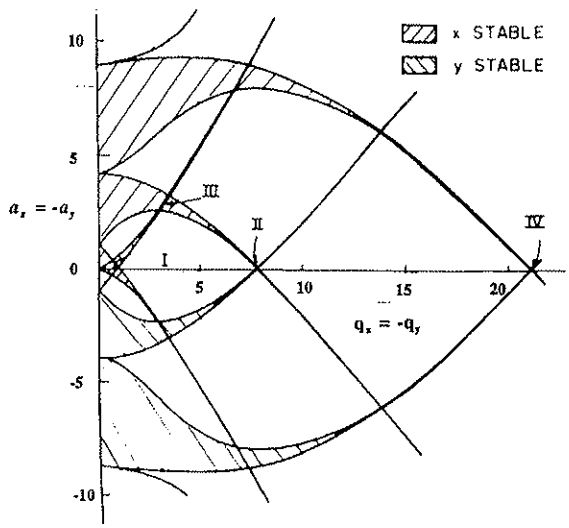



Fig. 15 Stability shown for both $xz$ and $yz$ planes indicating four regions where ions may pass through the filter.

## 5. REFERENCES


[1] J. Yarwood, (1959) Br. J. Appl. Phys. **10** 383.

[2] D. Alpert and R.S. Buritz, (1954) J. Appl. Phys. **25**, 202.

[3] W.C. Wiley, and I.H. McLaren, (1955) Rev. Sci. Instrum. **26** 1150.

[4] J.H. Batey, (1987) Vacuum. **37**, 659.

[5] N.R. Reagan, L.C. Frees, and J.W. Gray, (1987) J. Vac. Sci. Technol. **A5**, 2389.

[6] F.M. Mao, J.M. Yang, W.E. Austin, and J.H. Leck, (1987) Vacuum. **37**, 335.

# PHYSICS OF OUTGASSING

*J.L. de Segovia*
Instituto de Física Aplicada, CETEF "L. Torres Quevedo", CSIC, Madrid, Spain

**Abstract**
This paper relates the physical bases of the gassing phenomena in ultra-high vacuum. The main phenomena discussed are: diffusion of gases through system walls and solids in vacuum, solubility of gases on solids and permeation as well as the phenomena of adsorption, desorption and absorption.

## 1. INTRODUCTION

It has been recognised for many years that gassing and outgassing are processes controlling the ultimate pressure and gas composition in high- and ultra-high vacuum systems [1, 2]. The mechanisms to stimulate outgassing are diverse:

(i) Thermal desorption.

(ii) Desorption induced by electronic transitions.

(ii) Vaporisation of materials.

(iv) Gas diffusion from the bulk and subsequent desorption.

(v) Gas permeation through the walls.

In Fig. 1 the several mechanisms producing outgassing have been depicted.

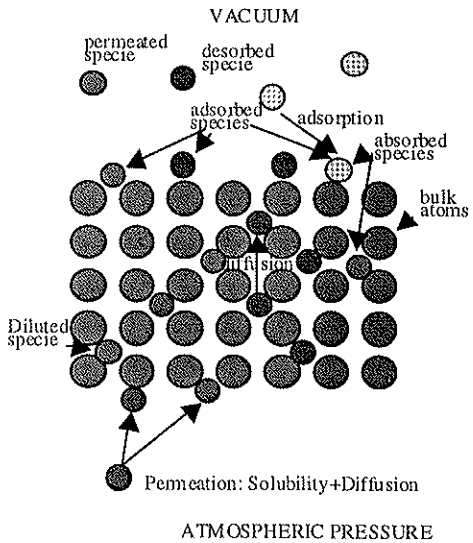



Fig. 1 Surface and bulk phenomena in vacuum.

## 2. PERMEATION OF GASES THROUGH MATTER

Gases permeate through most materials, so gases present in the ambient air can permeate to vacuum chambers through the system walls. Permeation of gases is a combination of two physical processes: dissolution and diffusion. First, the gas **dissolves** in the solid and then it **diffuses** to the inner wall of the system. When the gas arrives at the inner surface, it **desorbs** to the vacuum volume.

The dissolution phenomenon obeys Henry's law:

$$c = sP^n$$

where $c$ is the concentration, $P$ the gas pressure, $s$ the solubility, and $n$ depends on the material and is 1 for non-metals. Diatomic molecules dissociate upon dissolving.

The concentration is measured in mbar.l or Pa.m$^3$ and it is the amount of gas expressed in mbar.l or Pa.m$^3$ measured at 293 K that dissolves in 1l or 1 m$^3$ of substance. $s$ is the amount of gas measured in STP that dissolves in 1 l or 1 m$^3$ of substance at a pressure of 1033 mbar or $1.033 \times 10^5$ Pa, for $n = 1$ it is dimensionless and for $n = 1/2$ it has the dimensions of mbar$^{0.5}$ or Pa$^{0.5}$. After dissolving, the gas diffuses to the inner wall of the system according to Fick's first law of diffusion, for the stationary case. The gas flow, $Q$, through an area of unit cross section and time unit, is given by:

$$Q = -D\frac{dc}{dx}$$

where $D$ is the diffusion coefficient, which depends exponentially on temperature.

$$D = D_0 e^{\left(-\frac{E}{kT}\right)}$$

where $k$ is the Boltzmann constant; $E$ is the activation energy for diffusion, $T$ is the wall temperature and $D_0$ is a coefficient of proportionality.

We consider a wall of thickness $d$ and very large area and at time $t = 0$, the two lateral surfaces are exposed to pressures $P_1$ and $P_2$. The concentrations on both faces due to solubility will be:

$$c_1 = sP^n \quad c_2 = sP^n$$

In addition, the gas flow will be:

$$Q = \frac{Ds\left(P_1^n - P^n\right)}{d}$$

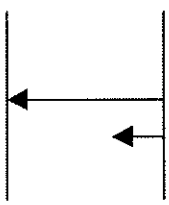
where $n = 1$ for non-metals, and $n = 1/2$ for diatomic molecules in metals. $Ds = K$ is the permeation constant.

As the equilibrium is reached in many cases after a long period of time, so the transition regime is determined by applying Fick's second law:

$$D\frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t}.$$

This equation can be solved for specific cases useful in vacuum technology.

Let's consider a wall as before with the following boundary conditions:



| | | | | |
|---|---|---|---|---|
| $c = 0$ | $0 < x < d$ | $t = 0$ | $P = P_1$ | $P_2 = 0$ |
| $c = c_1$ | $x = d$ | $t > 0$ | | |
| $c = 0$ | $x = 0$ | $t > 0$ | | |

The total amount of gas permeated into the vacuum is given by:

$$Q = \frac{c_1 D}{d}\left(t - \frac{d^2}{6D}\right)$$

and the time constant:

$$t_a = \frac{d^2}{6D}$$

is the required time where $Q$ is valid. This time constant is important to determine $D$. By representing $Q$ as a function of $t$, the value of $t_a$ can be determined by extrapolating the straight line and then $D$ can be found. Table 1 summarises the permeation properties of some materials and it can be seen that elastomers should be avoided at pressures lower than $10^{-6}$ due to the high permeation rates for most

gases, helium permeation is important only in glass systems while gas permeation is not of importance in metal systems at room temperature except for $H_2$ but this gas is not in high concentrations in air.

Table 1
Permeation of gases through materials [3]

| Metals | Semiconductors | Polymers | Glasses |
|---|---|---|---|
| Rare gases do not permeate.<br>$H_2$: high rate through palladium, through Fe by corrosion and electrolysis.<br>$O_2$: permeates Ag.<br>Permeation rates vary as $(\text{pressure})^{0.5}$. | He and $H_2$ through Ge and Si.<br>Ne and A not measurable.<br>$H_2$ varies as $(\text{pressure})^{0.5}$. | All gases permeate all polymers.<br>Water rate high.<br>Many particularities with the material.<br>All rates vary as $(\text{pressure})^{0.5}$. | $H_2$, He, $D_2$, Ne, A, $O_2$ are measurable through $SiO_2$.<br>Vitreous silica.<br>All rates vary as (pressure). |

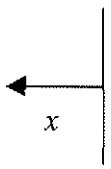
Table 2 shows an example of the permeation of an O-ring of standard dimensions: 0.5-m length and 4-mm diameter for the indicated materials. The O-ring when compressed takes a squared form 7-mm thick by 2-mm high so that the area of permeation is $0.5\ \text{m} \times 0.002\ \text{m}^2$.

Table 2
Permeation of air through an O-ring

| MATERIAL | Permeation constant at 20°C ($\text{mbar.l.s}^{-1}.\text{mm.m}^{-2}.\text{bar}^{-1}$) | Air rate ($\text{mbar.l.s}^{-1}$) | Ultimate pressure with a 1000 l/s pump (mbar) |
|---|---|---|---|
| Perbunam | $2 \times 10^{-2}$ | $2 \times 10^{-2} \times 0.002 \times 1/8 = 5.0 \times 10^{-6}$ | $5.0 \times 10^{-9}$ |
| Viton | $3 \times 10^{-3}$ | $3 \times 10^{-3} \times 0.002 \times 1/8 = 7.7 \times 10^{-7}$ | $7.5 \times 10^{-10}$ |
| Silicone | $2 \times 10^{-1}$ | $2 \times 10^{-1} \times 0.002 \times 1/8 = 7.5 \times 10^{-5}$ | $7.5 \times 10^{-10}$ |

## 3. DIFFUSION FROM A SEMI-INFINITE WALL

An important case for vacuum technology is the degassing of a semi-infinite wall with an initial concentration $c_1$ that is exposed at $t = 0$. The boundary conditions are now:




$P = 0 \qquad c = c_1 \qquad x > 0 \qquad t > 0$

$c = 0 \qquad x = 0 \qquad t > 0$

The gassing rate at time $t$ is given by:

$$Q = c_1 < d^{-\frac{1}{2}} \left( (\pi t)^{-1/2} \right).$$

When the vacuum is connected to a pump of speed $S$, the pressure is given by:

$$P = \frac{Q}{S} = c_1 D^{\frac{1}{2}} S^{-1} (\pi t)^{-\frac{1}{2}} .$$

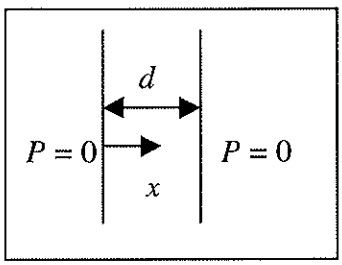
The total amount of gas removed from the slab is:

$$Q_T = \int_0^t D\left(\frac{\partial c}{\partial t}\right)_{x=0} dt = 2\pi^{-\frac{1}{2}} c_1 (Dt)^{\frac{1}{2}} .$$

Note that in a log-log scale the slope is 0.5. The same amount of gas will diffuse from the atmosphere in an initially clean wall with a final concentration of $c_1$.

## 4. DIFFUSION FROM A FINITE WALL

Another important case in vacuum engineering is the degassing from a finite wall, with uniform initial concentration $c_1$ and both faces are exposed to vacuum at time zero.




In this case the boundary conditions to solve Fick's second law are:

| | | |
|---|---|---|
| $c = c_1$ | $0 < x < d$ | $t = 0$ |
| $c = 0$ | $x = 0, x = d$ | $t > 0$ |

The instantaneous gassing rate is given by:

$$Q_{x=d} = \frac{2c_1 D}{d} \sum_{i=0}^{\infty} e^{\left(-\frac{(2i+1)\pi^2 Dt}{4d^2}\right)}$$

with

$$t_a = \frac{4d^2}{\pi^2 D}$$

which is the characteristic parameter.

According to the value of time with respect to this parameter, the instantaneous gassing rate is given by:

(1) $t > 0.5t_a$

$$Q_{x=0,d} = \frac{2Dc_1}{d} e^{\left(-\frac{\pi^2 D}{4d^2}.t\right)}.$$

(2) $t < 0.5t_a$

$$Q_{x=0,d} = \frac{2Dc_1}{d}\sqrt{\frac{\pi t_a}{16t}} .$$

The degree of outgassing e.g. the remaining gas in the wall with respect to the initial number of molecules, is given by the expression:

$$f = \frac{N(t)}{N_0} = \frac{8}{\pi^2}\left( e^{\frac{t}{t_a}} + \frac{e^{-9\frac{t}{t_a}}}{9} + \frac{e^{-25\frac{t}{t_a}}}{25} + \ldots\ldots \right)$$

Figure 2 shows the required time to reduce the initial outgassing rate by $e$ for the case of $t > 0.5t_a$.

***Example***

Let's determine the gas diffusing from a finite nickel wall inside a vacuum system with the following parameters [4]:

| | |
|---|---|
| Thickness: | $2.10^{-3}$ m |
| $\Delta$: | $8.10^{3}$ kg.m$^{-3}$ |
| $H_2$ concentration: | $10^{-5}$ (10 ppm) |

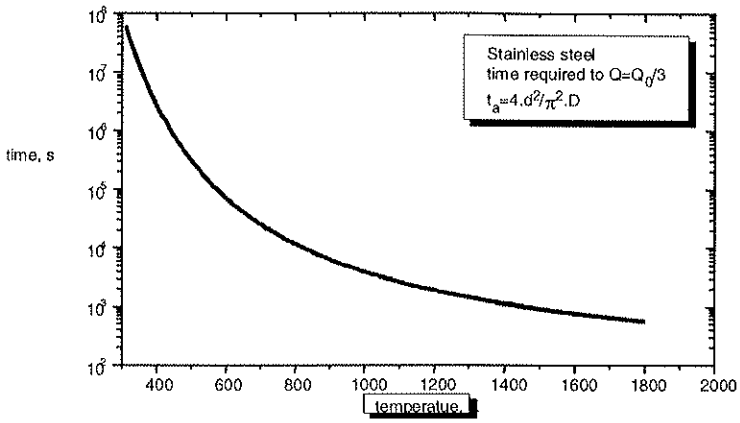



Fig. 2 Time required to reduce $Q_0$ by a factor of $e$

***Determination of the outgassing rates***

| | |
|---|---|
| Mass of dissolved $H_2$ | $10^{-5} \times 8.10^{-3} = 8.10^{-2}$ kg.m$^{-3}$ |
| Molar concentration: | $8.10^{-2}$ kg.m$^{-3}$/2 kg.kmol$^{-1}$ = $4.10^{-2}$ kmol.m$^{-3}$ |
| Particle density number: | $4.10^{-2}$ kmol.m$^{-1}$.$6.10^{26}$ kmol$^{-1}$ = $2.4.10^{25}$ m$^{-3}$ |

Table 3
Calculation of the diffusion rate from a nickel wall

| **Temp. (K)** | **Diffusion coefficient** $D = 2.10^{-7}.e^{(-4350/T)}$ (m$^2$.s$^{-1}$) | **Time constant** $t_a = 4d^2/BD^2$ (s) | **Time to reduce concentration,** $f = N/N_0$, **by** (s) | | | **Gas flow at** $t = 0$ $Q_0 = (2D/d)n_0$, (kg.m$^{-2}$.s$^{-1}$) | **Gas flow at** $t$ $Q(t) = Q_0.e^{-t/t_a}$ (kg.m$^{-2}$.s$^{-1}$) |
|---|---|---|---|---|---|---|---|
| | | | **0.1** | **$10^{-3}$** | **$10^{-6}$** | | |
| 300 | $1.01.10^{-13}$ | $1.6.10^{7}$ | $3.3.10^{7}$ | $1.01.10^{8}$ | $2.17.10^{8}$ | $2.42.10^{15}$ | $Q(t) = 2.42.10^{15}.e^{-t/t_a}$ |
| 500 | $3.33.10^{-11}$ | $4.87.10^{4}$ | $1.05.10^{5}$ | $3.26.10^{5}$ | $6.62.10^{5}$ | $7.99.10^{17}$ | $Q(t) = 7.99.10^{17}.e^{-t/t_a}$ |
| 700 | $4.10^{-10}$ | $4.05.10^{3}$ | $8.47.10^{3}$ | $2.71.10^{4}$ | $5.5.10^{4}$ | $9.6.10^{17}$ | $Q(t) = 9.6.10^{17}.e^{-t/t_a}$ |
| 1000 | $2.6.10^{-9}$ | $6.23.10^{2}$ | $1.3.10^{3}$ | $4.17.10^{3}$ | $8.47.10^{3}$ | $6.24.10^{19}$ | $Q(t) = 6.24.10^{19}.e^{-t/t_a}$ |
| 1300 | $7.10^{-9}$ | $2.3.10^{2}$ | $4.81.10^{2}$ | $1.54.10^{3}$ | $3.13.10^{3}$ | $1.68.10^{20}$ | $Q(t) = 1.68.10^{20}.e^{-t/t_a}$ |

***Expression of outgassing as P.V.***

In practical cases, as well as in vacuum engineering, it is convenient to express the outgassing rate as $P.V.$

$$Q_{P.V} = P.\frac{dV}{dt} = \frac{RT}{M_{mol}}.\frac{dm}{dt} = \frac{RT}{M_{mol}}Q_m$$

***Exercise***: Let's apply this expression to the outgassing of a cylindrical chamber:

Concentration of $H_2$: $c_{H2} = 8.\ 10^{-2}\ \text{kg.m}^{-3}$

Surface coverage: $2_{H2} = 10^{19}\ \text{m}^{-2}$

Areas of the inner wall: $A = 5.10^{-2}\ \text{m}^2$

Temperature: $T = 300\ \text{K}$

Binding energy: $E = 83\ \text{kJ.mol}^{-1}.\text{K}^{-1}$

$$\frac{RT}{M_{mol}} = \frac{8.3\frac{\text{J}}{\text{mol.K}}\cdot 300\text{K}}{2\frac{\text{kg}}{\text{kmol}}} = 1.25.10^6\ \frac{J}{\text{kg}}$$

$$1.125.10^6.5.10^{-2}\ \frac{J.m^2}{\text{kg}} = 6.15.10^4\ \text{Pa.m}^3\ \frac{\text{m}^2}{\text{kg}}$$

$$Q_m = m_{H_2} v e^{-\frac{E}{RT}} = 10^{19}.\frac{2}{6.10^{26}} 10^{13}\, e^{-\frac{83000\frac{\text{J}}{\text{mol.K}}}{8{,}3\frac{J}{\text{mol.K}}300\text{K}}} = 1.1.10^{-9}\ \frac{\text{kg}}{\text{m}^2.\text{s}}$$

$$Q_{PV} = 6.15.10^4\ \text{Pa.m}^3\ \frac{\text{m}^2}{\text{kg}} 1.1.10^{-9}\ \frac{\text{kg}}{\text{m}^2.\text{s}} = 6.8.10^{-4}\ \frac{\text{mbar.l}}{\text{s}}$$

Despite this high value it is possible to obtain ultra-high vacuum conditions in a short time because $Q$ decreases with a time constant of 50 s. However the outgassing from diffusion has a value of:

$$Q_{PV} = 2.10^{-6}\ \frac{\text{mbar.l}}{\text{s}}$$

After only two months it will be possible to obtain ultra-high vacuum conditions because $Q$ decreases with a very small constant time.

## 5. VAPOUR PRESSURE

All materials evaporate and have a saturation pressure given by:

$$\text{Ln}P_s = A.\frac{B}{T}$$

with $A$ and $B$ being constants not depending on $T$.

The evaporation rate is given by:

$$Q_{\max} = 0.438\sigma\left(\frac{M}{T}\right)^{\frac{1}{2}} P_s \qquad \frac{\text{kg}}{\text{m}^2.\text{s}}$$

where $\sigma$ is the condensation coefficient, $T$ the surface temperature and $P_s$ the vapour pressure at $T$. Table 4 shows values for the evaporation rate of some metals.

Table 4
Outgassing rates of some metals

| Metal | Mol. weight | Temperature (K) | Evaporation rate, ($kg.m^{-2}.s^{-1}$) |
|---|---|---|---|
| Cd | 112.4 | 340 | $2.50.10^{-11}$ |
| Zn | 65.37 | 390 | $1.70.10^{-11}$ |
| Ca | 40.08 | 550 | $1.18.10^{-11}$ |
| In | 114.82 | 760 | $1.70.10^{-11}$ |
| Ag | 107.87 | 830 | $5.70.10^{-12}$ |
| Cu | 63.54 | 990 | $1.10.10^{-11}$ |
| Pt | 195.09 | 1550 | $1.55.10^{-11}$ |
| Ti | 47.96 | 1400 | $8.10.10^{-11}$ |
| Nb | 92.906 | 2050 | $9.30.10^{-12}$ |
| Ta | 180.948 | 2300 | $1.23.10^{-11}$ |
| W | 183.85 | 2400 | $1.20.10^{-11}$ |

The materials used in ultra-high vacuum devices have vapour pressures that are negligible. However, they should be remembered when using filaments at high temperatures, such as in hot-cathode gauges and mass spectrometers.

## 6. ADSORPTION

All solid, as well as liquid surfaces present attraction forces normal to the surface, so molecules landing on the surface can be **adsorbed**. Adsorbed gases under certain conditions of temperature and pressure can be **desorbed**, and are the main source of gas in vacuum systems. Note that the number of molecules adsorbed is much higher than the number of molecules in the volume. Gases can be removed from the volume by the adsorption process as in sorption pumps.

The **adsorption-desorption** process is controlled by the interaction energy between the adsorbed molecule or atom with the surface atoms. According to this "**binding energy**", the adsorption can be:

Weak: physical adsorption. $E < 30$ kJ/mol

Strong: chemical adsorption. $E > 100$ kJ/mol

Figure 3 shows the potential energy curve between the surface and the adsorbate for the different kinds of adsorption [5].

### 6.1 Adsorption physics

*6.1.1 Physisorption*

When the energy of adsorption is less than 30 kJ/mol, the energy of interaction is similar to the energy between the free molecule and the molecule in the liquid phase, i.e. van der Waals forces. In this case the potential energy between the molecule and the surface is given by the Lenard-Jones potential:

$$E = E_m\left[\left(\frac{d_0}{d}\right)^{12} - \left(\frac{d_0}{d}\right)^{6}\right]$$

where $E_m$ is the energy minima at the equilibrium distance, $d_0$ the equilibrium distance from the surface and $D$ the distance from the surface. Table 5 gives some values of the physisorption energy for glass and carbon while Fig. 4 shows a typical curve of potential energy.

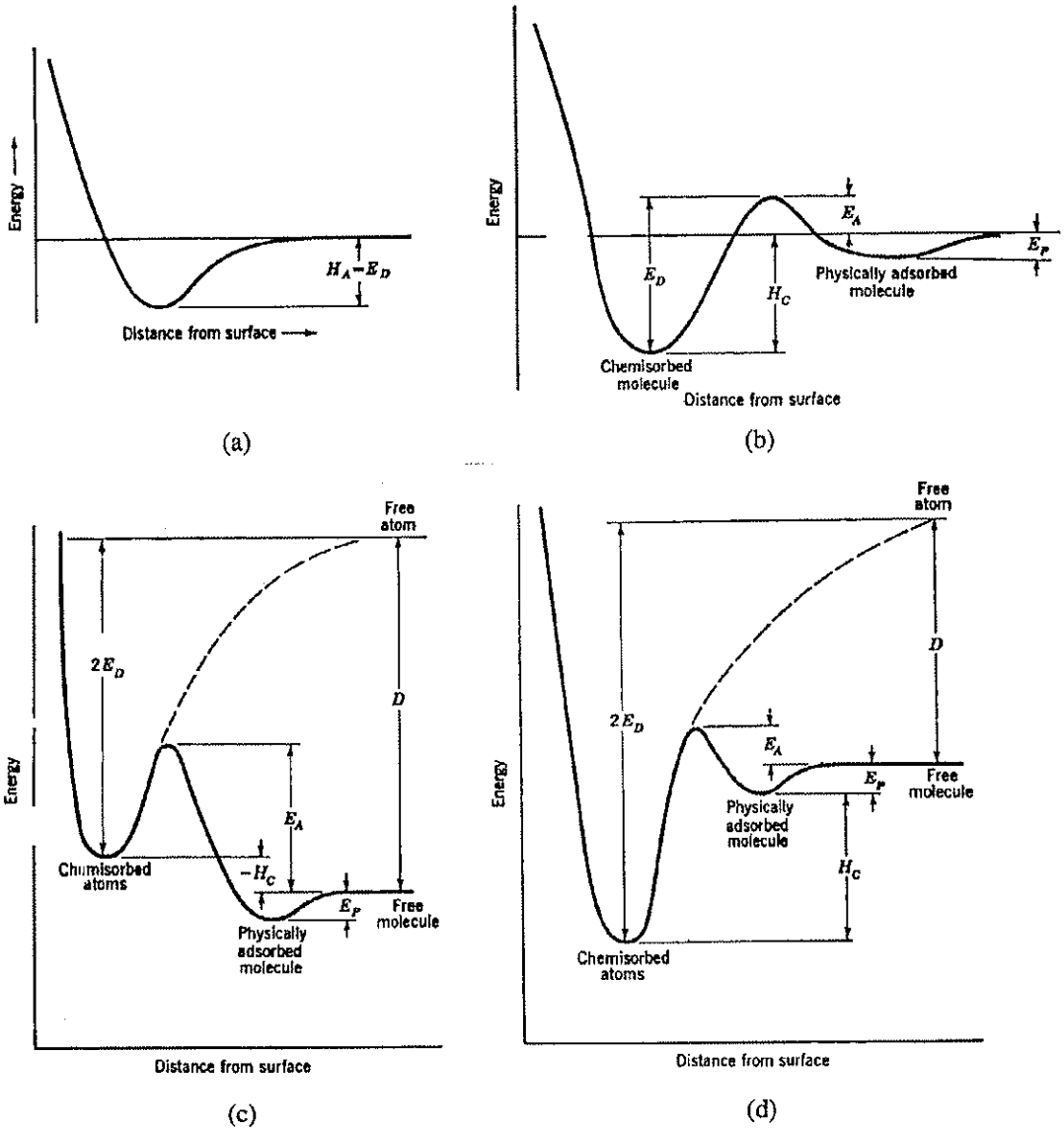



Fig. 3 Curves of potential energy surface-adsorbate. (a) is for physical adsorption. (b) Molecular chemical adsorption. (c) Endothermic dissociative chemical adsorption. (d) Exothermal dissociative chemical adsorption. $H_a$ is the adsorption energy, $E_D$ the desorption energy, $E_p$ the energy for physical adsorption, $E_A$ the activation energy for chemisorption, $H_c$ the chemisorption energy and $D$ the dissociation energy.

Table 5
Values of some physisorption energies

| Material \ Gas | Physisorption energy (kJ/mol) | | |
|---|---|---|---|
| | $O_2$ | $H_2$ | $N_2$ |
| Glass | 17.10 | 8.23 | 17.81 |
| Carbon | | 78.17 | 15.47 |

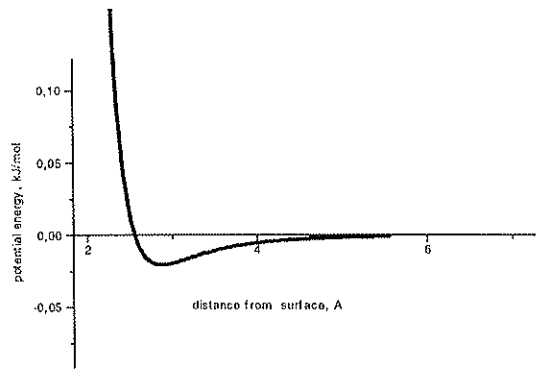



Fig. 4 Curve of potential energy for physical adsorption

*6.1.2 Strong chemisorption*

Strong chemisorption occurs when an alkaline atom is adsorbed on a metal surface. In this case there is transference of electrons from the adsorbate to the substrate. The interaction force will be given by:

(1) The imaginary potential when the adsorbate is at equilibrium distance $d_0$ from the surface.

(2) The energy derived from the electron transference

$$E_{ads} = e(\phi - I) + \frac{e^2}{4d_0}$$

where $e$ is the electron charge, $N$ the work function of the metal surface, $I$ the ionisation potential of the adsorbate and $d_0$ the equilibrium distance.

The equilibrium distance is given by:

$$\Delta V = 4e\pi N_s d_0$$

where $N_s$ is the number of available sites for adsorption, and ΔV the contact potential.

*6.1.3 Weak chemisorption*

This case occurs when the interaction between the adsorbate and the surface is of the covalent type. The interaction energy is given by:

$$E_{ads}(A-M) = \frac{1}{2}[E_{dis}(A-A) + E(M-M) + \Delta s]$$

where $E_{ads}(A - M)$ is the energy metal-adsorbate, $E_{dis}(A - A)$ the dissociation energy of the asdsorbate $A_2$, $E_{dis}(M - M)$ the dissociation energy of the atom of the metal substrate (approximately the sublimation heat) and Δs the electronegativity.

The electronegativity is given by:

$$\Delta s = (\chi_M - \chi_A) \approx \mu \text{ (dipole momentum)}$$

*6.1.4 Kinetics of adsorption*

The number of molecules incident on a surface at a pressure $P$ and gas temperature $T$ is given by:

$$\frac{dn_a}{dt} = sP(2\pi mkT)^{-\frac{1}{2}} = 8.32.10^{27} sP(MT)^{-\frac{1}{2}} = k_1 sP \qquad \text{s}^{-1}\text{m}^{-2}$$

where $s$ is the sticking coefficient, $P$ is the pressure in mbar, $M$ the molar mass in kg, and $T$ the gas temperature, K.

or

$$\left(\frac{dn_a}{dt}\right) = 27.4sP\left(\frac{T}{M}\right)^{\frac{1}{2}} \text{ mbar.m}^3\text{.s}^{-1}\text{.m}^{-3}$$

The sticking coefficient is the probability for one molecule or atom to be adsorbed. Once the molecule is adsorbed it can undergo migration if the system supplies the activation energy. It is a function of the surface temperature and of surface coverage:

$$s = s_0(T)f(\theta)$$

Surface coverage, $\theta$, is the fraction between the number of adsorbed species and the available site for adsorption. In a first approach it can be assumed that $s_0$ is independent of the temperature. Langmuir proposed for the dependence of $s$ with coverage that a molecule is adsorbed when it arrives at an unoccupied site.

$$f(\theta) = 1 - \theta$$

$$s = s_0(1-\theta)$$

and $s_0$ is the initial sticking coefficient at $\theta = 0$. The number of adsorbed species on a surface at a pressure $P$ and temperature $T$ will be the equilibrium between the rate of adsorbed molecules and the rate of desorbed molecules [6]

$$k_1 s_0 (1-\theta) P = k_2 \theta$$

The constant of desorption rate (desorption phenomena are detailed in another chapter of these proceedings) is given by:

$$k_2 = \nu\, e^{-\frac{E}{RT}}$$

where $\nu$ is the frequency factor, $10^{13}$ $s^{-1}$, $E$ the desorption energy, J/mol, $R$ the gas constant, J/mol.K, and $T$ the surface temperature, K.

The amount of adsorbed gas can be given in the form of:

$$\theta(k_2 + k_1 s_0 P) = k_1 s_0 P$$

$$\theta = \frac{k_1 s_0}{k_2} \cdot \frac{P}{1 + \frac{k_1 s_0}{k_2} P} = c\frac{P}{1+cP}$$

which is the well-known Langmuir adsorption isotherm or the hyperbolic adsorption isotherm. Figure 5 shows a typical representation of the Langmuir isotherm and Table 6 gives some of the parameters of the adsorption kinetics.

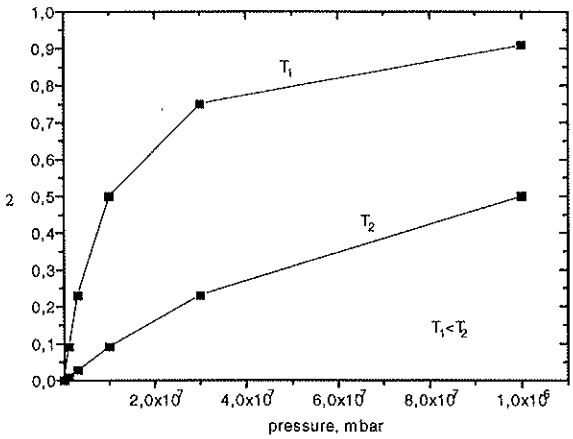



Fig. 5 Representation of the Langmuir isotherms, $T_1 < T_2$.

Table 6

Relevant parameters related to adsorption of some gas surface combinations

| Gas | Surface | $T_{surface}$K | $T_{gas}$ (K) | $s_0$ | $N_{maxi} \times 10^{14}$ ($cm^{-2}$) | θ (where $s_0$ starts to decrease) | Activation energy for desorption (eV) |
|---|---|---|---|---|---|---|---|
| $H_2$ | W(poly) [7] | 300 | 300 | 0.,064 | 2.24 | 0.40 | 0.60 |
| | Nb (100) [8] | 90 | | 0.23 | 3.20 | | 0.03 |
| | Ni(111) [9] | 300 | | 0.10 | 0.10 | | |
| CO | W(120) [10] | 300 | | 0.53 | 4.47 | 0.60 | 0.98 |
| $CO_2$ | W(100) [11] | 300 | | 0.85 | 0.42 | 0.20 | |

## REFERENCES


[1] G. Lewin, Fundamentals of Vacuum Science and Technology, McGraw Hill Inc. (1965).

[2] P.A. Redhead, J.P. Hobson and E.V. Kornelsen, The physical Bases of Ultra High Vacuum, Chapman and Hall Ltd. (1968).

[3] F.J. Norton, 1961 Vacuu Symmp. Trans., p. 8, 1962.

[4] M. Wutz, H. Adam, W. Walcher, Theory and Practice of Vacuum Technology, Friedr. Vieweg and Son, (1989).

[5] G. Ehrlich, J. Chem. Phys. 31 (1959) 1111.

[6] S. Dushman, Scientific Foundations of Vacuum Technique, Ed. J.M. Lafferty, John Wiley and Sons, (1962).

[7] P.W. Tamm, L.D. Schmidt, J. Chem. Phys. 55 (1971) 4253.

[8] D.I. Hagen and E.E. Donaldson, Surface Sci. 45 (1974).

[9] J. Lapoujalade and K.S. Neil, J. Chem Phys. 57 (1972) 3535.

[10] D.L. Adams and L.H. Germer, Surface Sci. 32 (1972) 205.

[11] B.J. Hopkins, A.R. Jones and R.I. Winton, Surface Sci. 57 (1976) 266.

# THERMAL OUTGASSING

*Karl Jousten*
Physikalisch-Technische Bundesanstalt, Berlin, Germany

**Abstract**
The physics and practical consequences of thermal outgassing in vacuum chambers are described in this article.

## 1. INTRODUCTION

Outgassing means usually two things (Fig. 1):

1. Molecules diffusing through the bulk material of a vacuum chamber, entering the surface and desorbing from it.
2. Molecules which have been adsorbed previously, usually during venting of the vacuum chamber, that desorb again, when the chamber is pumped to vacuum.

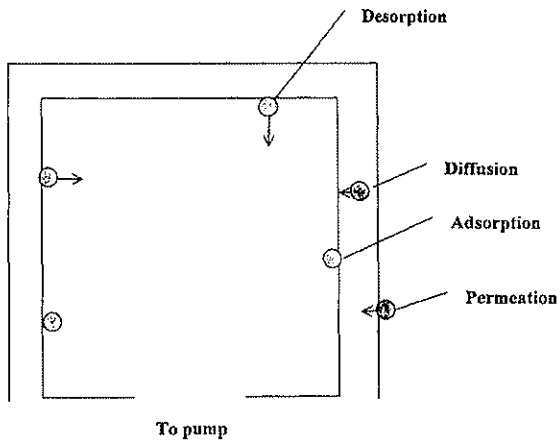



Fig. 1 The main physical effects which occur for thermal outgassing in a vacuum chamber.

Both effects have the same consequences: they limit the lowest achievable pressure in a vacuum chamber, they considerably extend the time for high and ultrahigh vacuum to be reached, and the outgassing molecules are a source of impurities in a vacuum chamber which is, for example, extremely inconvenient in the semiconductor industry. Especially in metallization processes, oxygen contaminants are disastrous because they reduce conductivity, and the smaller the structures on the integrated circuit (IC), the fewer particles can make it fail. In accelerators the residual gas particles limit the performance and its life time and can finally make high energy physics experiments impossible. In gravitational wave observatories the residual gas may limit the resolution of Michelson interferometers.

A further source of outgassing is the vaporisation or sublimation of atoms or molecules from a material with a vapour pressure higher or comparable to the vacuum pressure to be applied to a chamber. Although this fact is trivial, it greatly limits the number of materials that can be used in high and ultrahigh vacuum chambers and can sometimes be a difficult problem to overcome. But the careful selection of materials will not be the subject of this paper and we restrict ourselves to the effects of desorption, out-diffusion and permeation of materials suitable for UHV applications.

## 2. PUMP-DOWN BEHAVIOUR AND DESORPTION

Consider a perfect vacuum vessel, meaning a vessel with zero leakage, gas permeability and outgassing. Suppose that it is of volume $V$ and pumped by a pump with zero ultimate pressure and a volumetric pumping speed $S$. If at time $t$ the pressure is $p$, the pressure-time relation is

$$V\frac{dp}{dt} = -Sp(t) \tag{1}$$

or

$$p = p_0 \exp(-\frac{S}{V}t) \tag{2}$$

where $p_0$ is the pressure at the beginning of pumping. On a ln$p$ versus time plot the pressure decrease will be linear. *V/S* can be considered as the time constant of the system.

In a real chamber, in the initial phase of pumping there will be this linear behaviour in a ln$p$ versus time plot [1], but then there will be a much slower decrease with much longer time constants than *V/S* with finally an asymptotic behaviour. If the chamber had been exposed to atmospheric pressure before, this is due to gas adsorbed on the surface that desorbs only slowly under vacuum. The rate of desorption will depend on the binding energies of the various gases to the surface, the surface temperature and surface coverage. Desorption outnumbers the effect of out-diffusion greatly in this initial phase.

Adsorption takes place by physical adsorption or chemical adsorption. In physical adsorption gas molecules are attracted weakly by van der Waals forces with binding energies of less than 40 kJ/mol (10 kcal/mol or 0.4 eV), in chemisorption actual chemical bonding occurs between the gas molecules and the molecules and atoms on the surface of the vacuum material. These binding energies vary between 80 kJ/mol and 800 kJ/mol (20 to 200 kcal/mol or 0.8 to 8eV).

The quantity of gas adsorbed at constant temperature on a given surface increases with the pressure of gas in the gas phase. The relationship between the quantity adsorbed and the pressure is called the adsorption isotherm. There exist several models to describe these isotherms, the three most important can be attributed to Langmuir, Freundlich, and Temkin (Fig. 2).

*Langmuir-Isotherm*

The Langmuir Isotherm assumes that the adsorption energy is independent of surface coverage. It is therefore applicable, where surface coverage $\theta$ is low ($\theta < 0.3$).

*Freundlich-Isotherm*

This assumes that the adsorption energy decreases exponentially with surface coverage. It is applicable to higher surface coverages.

*Temkin-Isotherm*

This assumes that the adsorption energy decreases linearly with surface coverage.

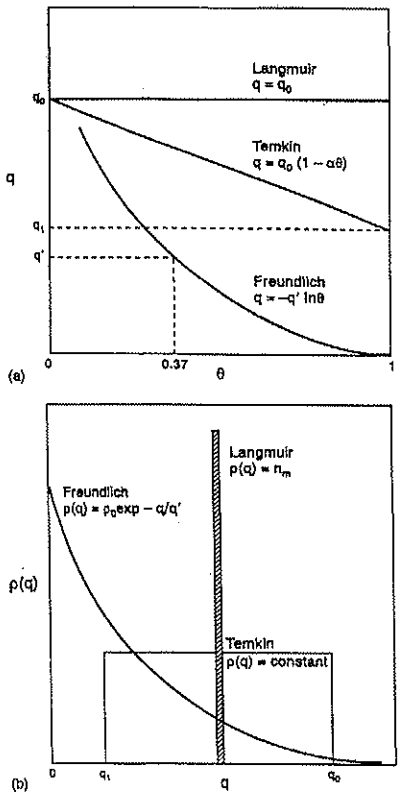



Fig. 2 Heat of adsorption $q$ versus relative surface coverage ($\theta = 1$ monolayer coverage) for three major isotherms (top) and the corresponding density of adsorption sites ρ as a function of $q$ (bottom). From Ref [5]. Copyright American Vacuum

The desorption rate can be described by

$$\frac{dn_S}{dt} = -\nu_0 n_{mon} f(\theta) \exp(E_{de} / RT) \tag{3}$$

With $n_S$ the number of molecules on the surface, $n_{mon}$ number of molecules in a monolayer, $\nu_0$ frequency of oscillation of a molecule on the surface, $f(\theta)$ function of surface coverage (isotherm function), and $E_{des}$ the activation energy for desorption.

The residence time τ of a particle on the surface can be described as

$$\tau = \frac{1}{\nu_0} \exp(E_{des} / RT) \tag{4}$$

For $\nu_0 = 10^{-13}$ s$^{-1}$, E = 42 kJ/mol and T = 300 K this gives τ = 2 μs. Gases that are physisorbed desorb very quickly at room temperature.

It is possible to derive an equation for the pump down:

- Gas removed by the pump: $n_V\ c/4\ A_p$ ($n_V$ molecule density in the volume, $c$ mean thermal velocity, $A_p$ effective pump area.
- Gas desorbed from the surface: $n_S/\tau\ A_S$ ($n_S$ number density of molecules on surface $A_S$ surface area).
- Gas being re-adsorbed on the surface: $n_V\ c/4\ sA_S$ ($s$ sticking probability).

So, finally

$$V\frac{dn_V}{dt} = -\frac{c}{4}n_V A_p + \frac{n_S A_S}{\tau} - \frac{c}{4}n_V sA_S \tag{5}$$

where at the same time the equation

$$A_S\frac{dn_S}{dt} = -\frac{n_S A_S}{\tau} + n_V\frac{c}{4}sA_S \tag{6}$$

has to be fulfilled for the surface. From Eq. (6) it can be seen that the net outgassing rate for desorbing gas particles depends generally on pressure.

Combining the two equations gives a differential equation for $n_V$

$$\frac{d^2 n_V}{dt^2} + \left(\frac{c}{4V}\left(sA_S + A_p\right) + \frac{1}{\tau}\right)\frac{dn_V}{dt} + \frac{c}{4V}\frac{A_p}{\tau}n_V = 0 \tag{7}$$

With some simplifications, Hobson [2] has plotted a solution of this equation for various values of $E_{des}$ that you can often see in the literature. The system shown in Fig. 3 has been calculated for $V = 1$ L, $A_S = 100\ \text{cm}^2$ and $S = 1$ L/s, first-order desorption and T = 295 K. Hobson also assumed a pressure independent sticking probability, which is a simplification that limits the usefulness of this approach. As in a pump down of a system without desorption (Eq. (2)) the logarithmic pressure falls with the inverse of time, but with time constants generally much greater than $V/S$.

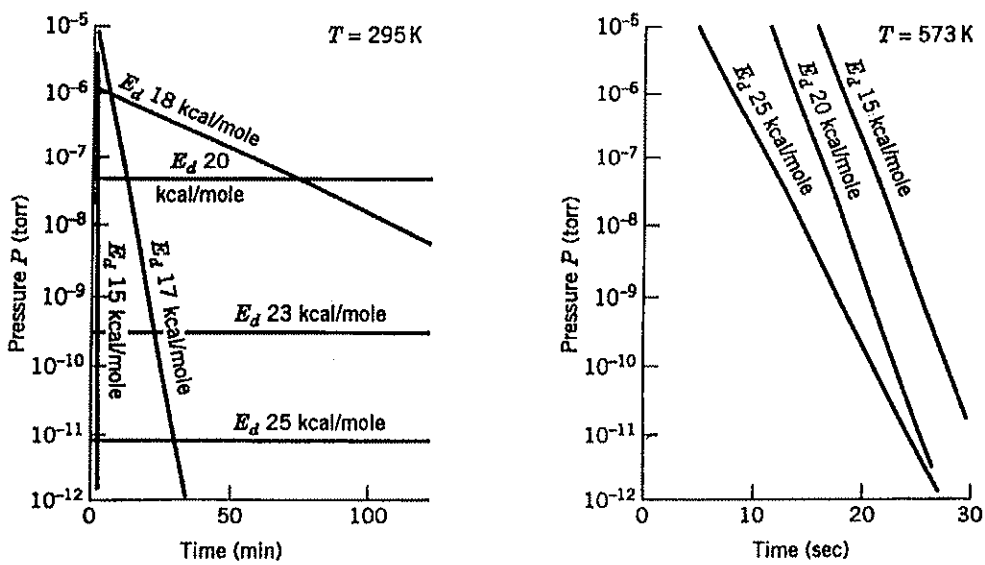
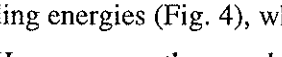



Fig. 3 Calculated pressure vs. time curves for the pumpdown of vacuum chambers having adsorbed layers with the adsorption energies indicated. From Ref. [3]. Copyright 1961 American Vacuum Society. Reprinted by permission of John Wiley & Sons, Inc

Two points can be concluded from Fig. 3. For small values of $E_{des}$ the desorption rates are so high that the corresponding molecules quickly disappear out of the volume and cause no further problem. For very high values of $E_{des}$ the desorption rates are too small to be of any consequence. It is

only the intermediate range that is really troublesome since gas cannot be eliminated from the vacuum system in a reasonable time.

And what nature did was to put the desorption energy of water on stainless steel and aluminium exactly in this intermediate range. The desorption energy of water on these surfaces lies in the range of 80 to 92 kJ/mol (19 to 23kcal/mol). So, at the start of pumping, $H_2O$ will be the major component (about 85% [3]) in the system and this will continue to be for days and weeks, if nothing is changed. Other minor outgassing molecules are $H_2$, CO, $CO_2$ and $CH_4$ [2]. We note at this point that $H_2O$-outgassing is not only generated by molecules previously adsorbed, but also by the reaction of iron oxides [4] in the case of stainless steel:



$$Fe_xO_y + 2H \rightarrow Fe + H_2O$$

In real systems the net desorption rate will follow a power law

$$q_d = Kt^{-n} \tag{8}$$

and when the pumping speed is constant, the pressure in the system will follow

$$p = kt^{-n} \tag{9}$$

Several authors have measured $n$ and it typically varies between 0.9 and 1.3. This indicates that there is more than one type of adsorption site for water as is also clear from the fact that a technical surface is not uniform but very complex and it is rather improbable that there would be only one desorption energy possible. Recently, Redhead [5] has explained the approximate $t^{-1}$ behaviour by a distribution of binding energies (Fig. 4), whereby several isotherm gave the same general behaviour.

However, another model has been proposed to explain the $t^{-1}$ behaviour of the pump-down curve. Dayton [6] suggested a mechanism based on the diffusion of water vapour molecules out of capillaries in a thin oxide layer covering the metal surface. We will see later that for the case of diffusion the outgassing rate or pressure varies with the square root of the pumping time, whereas a $t^{-1}$ behaviour has to be explained. The Dayton theory assumes that diameter and length of pores and capillaries in the surface oxide layer vary, giving rise to a range of diffusion coefficients. A time constant $\tau$ is introduced and when the pumping time is greater than $\tau$, the outgassing should start to decay exponentially. With a distribution of $\tau$, the final sum curve will also follow a $t^{-1}$ behaviour (Fig. 5).

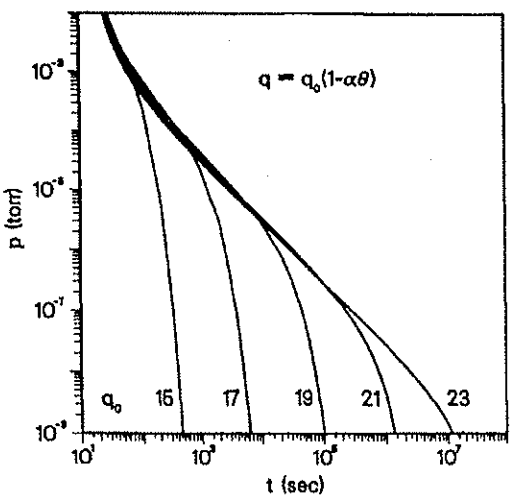



Fig. 4 Calculated pressure vs. time for a chamber with an adsorbed layer obeying the Temkin isotherm. From Ref. [5]. The sum of the curves follows approximately the experimentally observed $t^{-1}$ behaviour. Copyright American Vacuum Society.

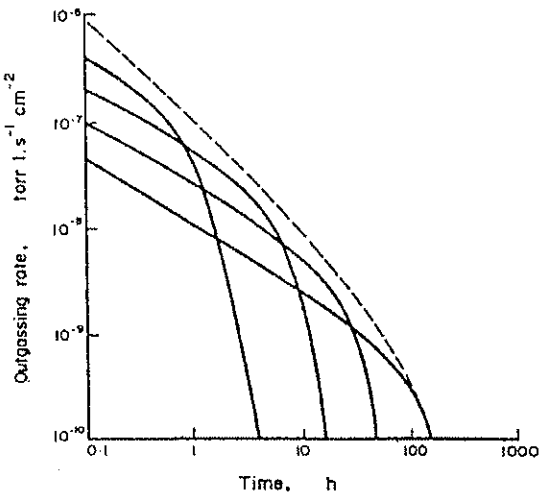



Fig. 5 Temperature dependence of CO outgassing of stainless steel showing different activation energies above and below about 250°C. From Ref. [7]. Copyright American Vacuum Society.

In conclusion, from the very easily measurable pumping curve useful information can be obtained, especially in practice where after a while one has specific experience how the system pumps down. By comparison of a pumping curve to a normal one, one can easily detect problems such as a virtual or real leak, because in this case the logarithmic pump down curve versus time will approach asymptotically a constant pressure value, determined by the leak rate divided by the pumping speed, instead of falling with $t^{-1}$. But also theory can draw useful information from a pumping curve, because different isotherms give different pumping curves. Redhead [5] found that the Temkin isotherm without any approximations gave the best results compared with experimental data.

## 3. VACUUM SYSTEMS AFTER BAKE-OUT AND OUT-DIFFUSION

The clear conclusion from Fig. 3 is that a vacuum chamber has to be heated up to remove the adsorbed gas molecules with binding energies in the range of 17 to 25 kcal/mol or somewhat higher. To what temperature and for which time length a system has to be raised to remove adsorbed molecules, namely $H_2O$, is dependent on the ultimate pressure that is desired after bake-out and cool-down to room temperature, on the binding energies of the adsorbed molecules, the material of the vessel and the sealings and on the surface of the vessel and its pre-treatment.

The latter two points are very important in the author's opinion, because the surface chemistry greatly depends on the material, its pre-treatment and temperature since the surface before, during and after a bake-out will not be the same. As an example, we report on measurements of Ishikawa and Yoshimura [7], who have analysed the temperature dependence of $H_2$, $H_2O$, CO, $CO_2$, and $CH_4$ outgassing of 316 L stainless steel (Fig. 6). They found a clear change of activation energy at 250°C for CO that also occurred for $H_2O$, $CO_2$, and $CH_4$. At temperatures > 250°C apparently these molecules are chemisorbed whereas they are only physisorbed below 250°C.

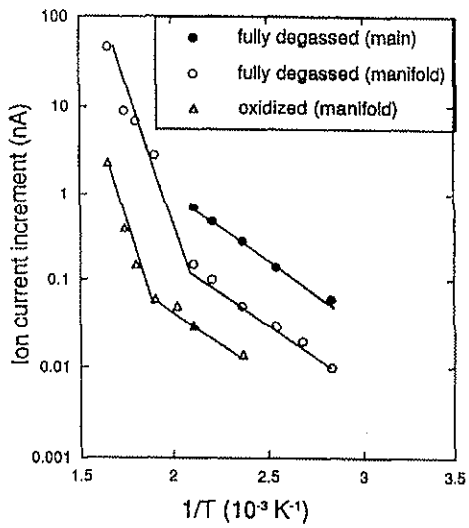



Fig. 6 Temperature dependence of CO outgassing of stainless steel showing different activation energies above and below about 250°C. From Ref. [7]. Copyright American Vacuum Society.

It is not as simple as was thought with early UHV metal systems that a system has to be baked to as high a temperature as possible, usually to more than 350°C, to remove all water from the system to obtain the lowest outgassing rates. Perhaps this view came from experience with the previously used glass systems, which contained a large amount of dissolved water which had to be removed at temperatures up to 500°C. We will come back to this point when we have further discussed the problems of out-diffusion, permeation and pre-treatment. Because today's vacuum systems are almost exclusively manufactured from stainless steel or aluminium, only these materials will be discussed in the following.

It should be noted here that for bake-out ion pumps should not be used since they soon saturate and give a poor ultimate pressure after bake-out, and re-emission of pumped gases may also occur.

Let us suppose that, after a successful bake-out of a metal vessel, all physisorbed and chemisorbed molecules have been desorbed and the surface is clean in this sense. Now, a different mechanism of outgassing comes into action. Hydrogen atoms dissolved in the bulk diffuse to regions where the hydrogen concentration is lower. The region under the vacuum surface is such a region, because eventually hydrogen atoms diffuse to the surface and recombine with a second hydrogen atom by surface diffusion which enables the molecule to desorb from the surface since the desorption activation energy of $H_2$ is rather low. In a quasi-static equilibrium there will be a constant net flow of hydrogen atoms to the vacuum surface which then desorb.

Two questions have to be answered at this point:

1. *Why is hydrogen present in the metal and can one avoid it being dissolved there?*

Hydrogen has a very small atom with the lowest possible mass, has a high solubility in most metals and diffuses easily through them. During the production of stainless steel, when it is molten, hydrogen, which is contained in air with about $4 \cdot 10^{-2}$ Pa diffuses quickly into the metal. An obvious cure would be to carry out this process under vacuum. This is actually done for some alloys, but one has to remember that this vacua, in a rough industrial environment at very high temperatures is only of the order of 10 Pa. Additionally, the material after cooldown is processed with acids to remove thick oxide layers in order to be able to roll the steel and, probably, allows hydrogen to diffuse into the material.

2. *What is the rate of diffusion to the surface?*

The theory to solve this problem is discussed in another article of this volume. The result for an infinitely large sheet of metal (a sufficient approximation, because the thickness of all metal sheets is much smaller than the lateral dimensions). With thickness $2d$ (defined in this manner for symmetry reasons) and initial concentration $c_0$ the flow of atoms to the surface is

$$j_{diff} = \frac{2D}{d} c_0 \sum_{i=0}^{\infty} \exp\left( -\frac{(2i-1)^2 \pi^2}{4d^2} Dt \right) \tag{10}$$

with the temperature dependent diffusion coefficient

$$D = D_0 \exp\left(-E_{diff} / kT\right) . \tag{11}$$

The characteristic time constant

$$t_c = \frac{4d^2}{\pi^2 D} \tag{12}$$

is strongly temperature dependent. For times $t > 0.5\ t_c$ the higher orders of $i > 0$ contribute with less than 2% to the total sum [8] and Eq. (10) can be simplified

$$j_{diff} \approx \frac{2D}{d} c_0 \exp\left(-t / t_c\right) . \tag{13}$$

For much smaller times

$$j_{diff} = \frac{2D}{d} c_0 \sqrt{\frac{\pi t_c}{16t}} . \tag{14}$$

We firstly estimate $t_c$ for stainless steel with $D_0 = 0.012$ cm²/s, $E_{diff} = 0.56$ eV, $T = 300$ K, $d = 1$ mm to $t_c(300\text{ K}) = 1.8 \cdot 10^9$ s (57 years), while $t_c(1200\text{ K}) = 92$ s.

Therefore, for the room temperature case and the stainless steel never heated since production ($c_0 = 0.4$ mbar l/cm³)

$$j_{diff}(10^7 s) \approx 10^{-10} \text{mbarl/scm}^2$$

We have chosen $10^7$ s (116 days) as an average time between manufacturing the stainless steel and first pumping a vessel made from it. Although this outgassing rate is sufficient to reach HV, it is not for UHV (consider $A_s = 10000$ cm², $S = 100$ L/s, it follows that $p = 10^{-8}$ mbar or $10^{-6}$ Pa). It would take a factor of 100 longer time, that means 30 years, to decrease this outgassing value by a factor of 10!

However, because we want to desorb adsorbed molecules from the surface, let us suppose, we made a 24-h in-situ bake-out at 250°C of our vacuum chamber (in-situ bake-outs are carried when the vacuum system is completely installed and in operation). Due to the much higher temperature the concentration profile of the hydrogen will be determined during the bake-out time which is then 'frozen-in' at room temperature. Again, for $t > 0.5\ t_c$ one obtains

$$j_{diff} = \frac{2D_r}{d} c_0 \exp\left(-\frac{\pi^2 D_1 t_1}{4d^2}\right) \tag{15}$$

where $D_r$ is the diffusion coefficient at room temperature, $D_1$ at bake-out temperature and $t_1$ the bake out time. $t_e$(523 K) = 25 h, so that Eq. (15) can be applied. We obtain under the described bakeout conditions

$$j_{diff} \approx 6 \cdot 10^{-12} \text{mbarl/scm}^2 \tag{16}$$

which is constant for practical times. This value is already acceptable for modest UHV applications, and an ultimate pressure of $6 \cdot 10^{-10}$ mbar ($6 \cdot 10^{-8}$ Pa) can be expected ($A_s$= 10000 cm$^2$, $S$= 100 L/s as above). A second bake-out would reduce the outgassing rate by a factor of 2.6 according to this calculation.

## 4. PRE-BAKE-OUTS TO REDUCE OUTGASSING

To further reduce outgassing rates below a level of $10^{-12}$ mbar l/s cm$^2$, Calder and Lewin [9] investigated in the 1960s, how high-temperature pre-bake-out in a vacuum furnace could be used to degas hydrogen from stainless steel to reduce its bulk concentration (pre-bake-outs are carried out before the vacuum system is installed). Since $Dt$ appears in the exponent and $D$ is exponentially dependent on temperature it seems much more effective to increase the temperature of the bake than to increase the bake-out time.

With their available data, Calder and Lewin calculated that a 1-h bake at 1000°C would decrease the outgassing rate at room temperature to the same level as a 300°C bake for 2500 h. From the permeation of hydrogen from the atmosphere into the material during in-situ bake-outs and the resulting outgassing rate at room temperature, they concluded that outgassing rates as low as $10^{-16}$ mbar l/s cm$^2$ could be achieved. Their experimental values were two orders of magnitudes larger, but were attributed to limitations in their apparatus. However, many other investigators after Calder and Lewin failed to reach lower outgassing rates than $10^{-14}$ mbar l/s cm$^2$.

From our present knowledge, Calder and Lewin were in error for several reasons:

1. Their diffusion coefficient values were based on an investigation by Eschbach [10] in 1963. These values were determined by measurements of the permeation of hydrogen through stainless steel. In later investigations it became clear that the surface oxide plays a very significant role when permeation rates are measured. Louthan and Derrick [11] have carried out a very careful experiment to measure the permeation rate of deuterium through austenitic stainless steel. The diffusion coefficient of hydrogen and deuterium is very much the same and should scale with the square root of the mass, which is probably within the uncertainty of the experimental values. Hence the hydrogen diffusion coefficient should be larger by 1.4. Louthan analysed the time-temperature dependence of deuterium flux through austenitic stainless steel foils which were in contact with lithium deuteride (LD) powder to prevent oxidation of the steel surface by $Cr_2O_3 + 6LiD \rightarrow 3Li_2O + 2Cr + 3D_2$. There are also similar reactions to reduce iron and nickel oxides. Louthan found large differences of permeation rates depending on surface preparation which, at a given temperature, were orders of magnitude different. Their final result for the diffusion in 304 type steel (304, 304N, 304L) was:

$$D = 6.6 \cdot 10^{-3} \exp\left(-\frac{12.9 \text{ kcal/mol}}{RT}\right) \text{cm}^2/\text{s} \tag{17}$$

where 12.9 kcal/mol correspond to 0.56 eV. At room temperature this gives a value of $2.4 \cdot 10^{-12}$ cm$^2$/s to be compared with Eschbach's value of $5 \cdot 10^{-14}$. This is consistent with the observation of Louthan that oxide layers as probably present in Eschbach's experiment reduce the permeation rate significantly.

The consequence of higher diffusion of hydrogen in stainless steel is that Calder and Lewin underestimated the total outgassing rates at room temperature according to their pure diffusion

limited theory, but their general conclusions are true in both cases (for example that a 1-h bake out at 1000°C is completely sufficient).

2. Their second error is related to the fact that surface layer effects were neglected. For the example of the permeation rate measurements, we have already seen that oxide layers play a significant role. If we believe that Eschbach's measurements were correct, but taken from a technical surface with a normal oxide, and consider the results of Louthans from an oxygen-free stainless steel surface, than the effect of the surface layer is about a factor of 40 at room temperature. Other investigators have collected data that surface layers significantly affect outgassing rates. Odaka and Ueda (Fig. 7) have determined outgassing rates depending on the Cr-oxide layer thickness on stainless steel [12]. The outgassing rate at 3.6 nm thickness was about a factor of 5 lower than with 1.6 nm thickness.

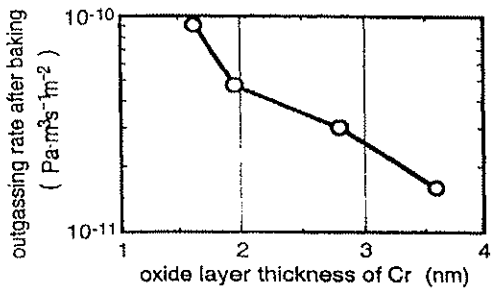



Fig. 7 Outgassing rate after in-situ baking as a function of the oxide layer thickness on stainless steel. Reprinted from Vacuum, 47, K. Odaka and S. Ueda, p689, Copyright (1996), with permission from Elsevier Science

3. An error, closely related to this, is that the permeation measurements were carried out with pure hydrogen. Eschbach, however, noted that adding a few Torr of oxygen to the hydrogen caused the permeation to decrease by a factor of 10 or more. Considering that in air 159 Torr (21 kPa) are present the reduction effect is probably much greater than a factor of 10, perhaps 100.

4. When a hydrogen atom has moved to the vacuum surface, it needs another hydrogen atom on the surface to recombine. The smaller the number of freely-movable hydrogen surface atoms, the less probable is recombination. This means that when the out-diffusion of hydrogen atoms to the surface becomes small, the recombination of the hydrogen atoms might be the limiting step. The outgassing from the surface is then no longer given by the out-diffusion rate, but by the rate at which the hydrogen atoms can recombine. The assumption of Calder and Lewin in solving the diffusion equations, that the surface density of hydrogen is always zero, is therefore very questionable. In contrast, if the surface density was zero, the recombination probability would also go to zero and the number of hydrogen atoms would accumulate on the surface, until the rate of recombination equalled the rate of diffusion to the surface. Therefore, the surface density can never be zero and the diffusion equations have to be solved by a boundary condition, which takes the recombination into account. Moore [13] calculated this numerically and Fig. 8 shows the H concentration profile in the cross section of an infinite plate 1.9 mm thick during a vacuum furnace bake-out at 950°C. The label numbers are the bake-out time in seconds. Due to the finite concentration on the surface there is a major departure from the diffusion limited profile. As time increases, the distribution approaches a sine function plus a constant, instead of the pure sine function with pure diffusion limited outgassing. This sine component decreases as time goes by and after 2000s the concentration is practically uniform. This is the time where the outgassing is limited

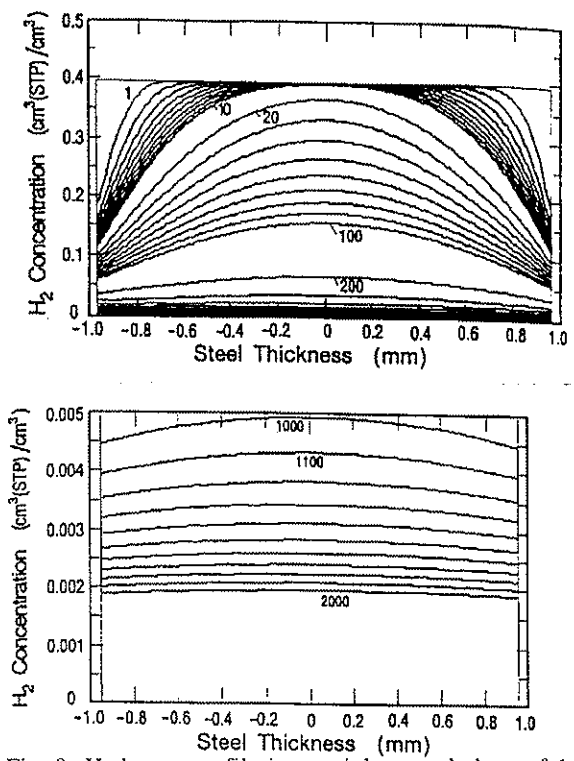



Fig. 8 Hydrogen profile in a stainless steel sheet of 1.9 mm thickness at 950°C with time as parameter in seconds. Calculations of Moore for recombination limited outgassing. The lower figure is an enlargement of the upper one for larger times. From Ref. [13]. Copyright American Vacuum Society.

by the recombination while the diffusion is still fast enough that every two surface atoms which recombine are immediately replaced.

When Moore analysed the outgassing rate versus time, he noted, that after about 1000 s the recombination limited outgassing will be orders of magnitudes higher than the results with the pure diffusive model. It is usually assumed that the recombination rate is proportional to the square of the number density of surface atoms $n_s$ times a recombination coefficient $K_{rec}$

$$j_{rec} = K_{rec} n_S^2 \tag{18}$$

By fitting this model to experimental values [14], Moore was able to determine $K_{rec}$ to $3 \cdot 10^{-22}$ cm²/s at 950°C.

If the desorption of $H_2$ is governed by the recombination rate, the surface roughness, which affects surface diffusion, should also change the effective $H_2$ desorption energy. This was indeed indicated by experimental results from Chun et al. [15]

From Calder and Lewins work and their errors that have been described above, the following conclusions can be drawn:

1. Pre-bakes at high temperature are a very efficient way to reduce outgassing of hydrogen from stainless steel. For aluminium as vacuum material for XHV it is probably a great disadvantage that high temperature bakes (at least not with currently available alloys) are not possible.
2. The diffusion model which describes the outgassing only by the diffusion rate without considering the recombination process at the surface and the role of the oxide layer is not sufficiently accurate to predict temperatures and times needed for very low outgassing rates. Much longer bake-out times than calculated by Calder and Lewin are necessary to obtain outgassing rates < $10^{-14}$ mbar l/s cm².
3. Permeation rates as calculated by Calder and Lewin are probably too high by a factor of 100, maybe 10000.

That these conclusions are actually true, has recently been confirmed by Nemanic, Setina and Bogataj [16, 17]. They wanted to produce vacuum insulating panels without getters. For this reason, a material had to be found with outgassing rates so small that high vacuum could be kept for 10 years without pumping. This required a search for materials with outgassing rates below $10^{-15}$ mbar l/s cm². Their solution was to use a 0.15-mm austenitic stainless steel foil which was baked for 68 h at 265°C. They obtained outgassing rates of $10^{-17}$ mbar l/s cm². A thin material has the great advantage that much shorter times or lower temperatures can be applied, which can be seen from Eq. (15). If $d$ is reduced by a factor of 10, $D_l t_l$ can be reduced by a factor of 100 to get the same outgassing rate with the diffusion model. As we have learned, this model is insufficient, but predictions with the recombination model are rather difficult, because the extrapolated or measured recombination coefficients $K_{rec}$ differ by 4 orders of magnitude [17]. Nemanic has used the dimensionless Fourier number $F_0$

$$F_0 = \frac{Dt}{d^2} \tag{19}$$

to characterise bake-outs. He found that $F_0$ of about 50 is required for outgassing rates as low as $10^{-15}$ mbar l/s cm² and with $F_0$ = 160 he was able to obtain $2 \cdot 10^{-17}$ mbar l/s cm². For comparison, $F_0$ = 13 for a 2-h bake at 1000°C of a 3-mm thick stainless steel plate. Fremerey [18] used a 25-h vacuum firing at 960°C instead of only 2 h, as was normal in the past, and obtained an outgassing rate of $10^{-15}$ mbar l/s cm².

In conclusion, if very low outgassing rates must be achieved, much longer bake-out times than in the past should be used for 2 mm to 3 mm thick stainless steel, or alternatively much thinner material should be used. Perhaps in future XHV chambers or CERN accelerator tubes will be built of

0.15 mm stainless steel foil, just enough to withstand atmospheric pressure, welded to stronger supports to support the structure.

We mentioned in the last section that the optimum value of bake-out temperature of an in-situ bake-out may strongly depend on the pre-treatment. Jousten [19] has measured outgassing rates depending on the bake-out temperature of a previously vacuum-fired stainless steel chamber and found the surprising result that lower bake-out temperatures gave lower outgassing rates (Fig. 9). The measurements were carried out with the pressure-rise method (see next section). The pressure rise was linear indicating that no significant amount of water desorbed from the surface but only hydrogen (see next section).

There are other effective procedures to reduce outgassing, for example oxidation of stainless steel in air at 200°C to 250°C [16, 17, 20, 21]. These generally give very low outgassing rates for hydrogen, but for thicker samples (> 2 mm) not as low as the described high-temperature bake-outs. An advantage of oxidation is that the number of carbon impurities present on the surface is greatly reduced [7] with the consequence that the production and outgassing of CO, $CO_2$, and $CH_4$ may be also reduced. Additionally, after a 250°C bake-out in air, iron oxide dominates $Cr_2O_3$ at the surface [7]. Iron oxide has an even lower diffusion coefficient than Cr oxide. However, if also the structure of iron oxide is of advantage (pores?, density?) is yet unclear and one has to be concerned about the loss of anti-corrosion effect. Another beneficial effect of the oxide layer is that the number of adsorption sites which have a high desorption energy are considerably reduced.

Another method to reduce outgassing is glow-discharge cleaning which has mainly been developed for cases where high-temperature baking was not possible, for example in fusion devices. This method seems to be beneficial when a chamber is made for plasma processing, because a plasma may liberate much more highly-bonded carbon oxides and hydrocarbons than can be removed by thermal methods. More information can be found in the 'Outgassing III' chapter by O. Gröbner in these proceedings. Generally, the outgassing rates at room temperature after glow-discharge cleaning are considerably higher than with the above mentioned thermal pre-bake-outs. Glow discharge cleaning can also be combined with oxidations.

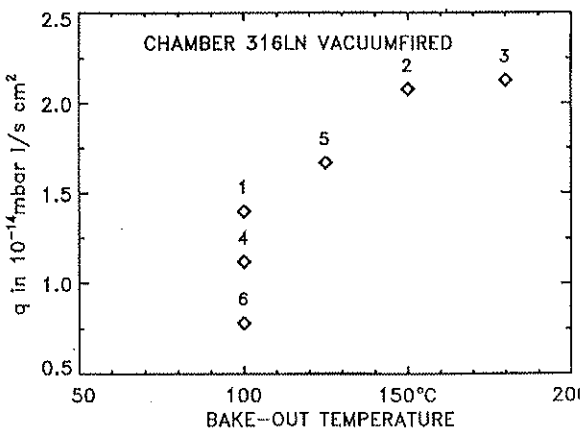



Fig. 9 Measured outgassing rate vs. bake-out temperature for a vacuum fired stainless steel chamber. The numbers follow the chronological order. Between each measurement the chamber was exposed to air and subsequently baked for 60h. From Ref. [19].

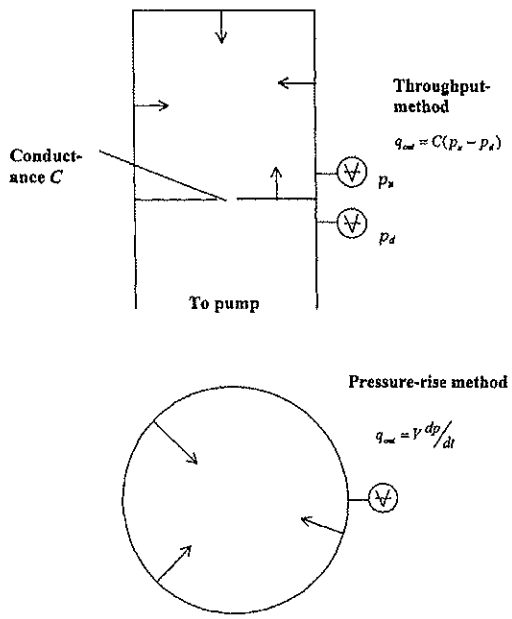



Fig. 10 Scheme showing the principal experimental set-up for the two main methods of outgassing rate measurement. In the throughput method the pressure drop across an orifice of known conductance is measured, in the pressure rise-method the accumulated gas pressure is measured in time.

## 5. OUTGASSING-RATE MEASUREMENT METHODS AND THE PROBLEM OF RE-ADSORPTION

Outgassing rates are measured by two general methods (Fig. 10): in the first, the chamber under test is pumped through an orifice of known conductance and the pressure drop across the orifice is measured as a function of time (outgassing rate may depend on time, as we have learned in sections 3 and 4). This is called the throughput method. In the second method, the pressure rise with time is measured in a closed system after evacuation to a low pressure. This is called the pressure-rise or gas-accumulation method. Both methods have a common problem: the measured outgassing rate is a net rate being the difference between the intrinsic outgassing rate of the surface and the readsorption rate. Papers containing published measurements rarely mention the difference between the two rates. It is widely assumed that outgassing rates measured by the throughput method can be applied to the design and analysis of vacuum systems with pumping parameters *S/A* (pumping speed divided by total area) and sticking probabilities that differ from the parameters of the system in which the outgassing rate was measured. So, be careful with the interpretation of results!

There is a third method using a direct molecular beam from a sample [22] and has the advantage that it directly measures net outgassing rates. Due to the need for cryogenic apparatus, it is however rarely applied.

Before discussing the problem of re-adsorption, however, let us first study the experimental methods in more detail. Figure 10 shows the basic scheme of an apparatus applying the throughput method. Two gauges can be used to determine the pressure drop across the conductance *C*, but very often it is assumed that the downstream pressure is negligible. This is a rather crude assumption when low specific outgassing rates ($<10^{-12}$ mbar l/s cm$^2$) are measured and should not be accepted in these cases. The gauges should be calibrated. The test sample can be either the chamber itself or a sample that is put into it. If a sample is used, a background measurement must be taken first without the sample. The total outgassing rate must be higher than the outgassing rate from the chamber to get a reasonably high signal. If this is not the case, the sample can be heated.

Since the pressures to be measured are usually very low in the throughput method ($<10^{-5}$ Pa), only ionisation gauges can be used for this purpose. This involves a big problem. Crossed-field ion gauges have the problem of a relatively high pumping speed that may be comparable to the conductance value of the orifice and would therefore significantly falsify measurements (indicating lower outgassing rates). On the other hand, you will hardly find a hot-cathode ionisation gauge with an outgassing rate $<10^{-11}$ mbar l/s. To get this low value, the gauge has to be specially prepared, degassed and evacuated over a long time. An outgassing rate of $10^{-11}$ mbar l/s can also be produced by 1000 cm$^2$ surface area with a specific outgassing rate of $10^{-14}$ mbar l/s cm$^2$. In this case an error of 100% would occur (indicating higher outgassing rates). Considering that normal hot ionisation gauges have outgassing rates of about $10^{-10}$ mbar l/s, it becomes clear that one of the first things that should be checked, when measuring total outgassing rates of this order, is the outgassing of the gauge.

Yang, Saitoh and Tsukahara [23] have proposed the following improvement to the throughput method (Fig. 11). They used two symmetrically arranged gauges, each separated by two valves from the upper and lower chamber. A single gauge can be used to measure both the upstream and downstream pressure. With the second gauge it is possible to measure the outgassing of the first gauge by taking the difference between the pressures measured with the valve $V_{u1}$ open and closed.

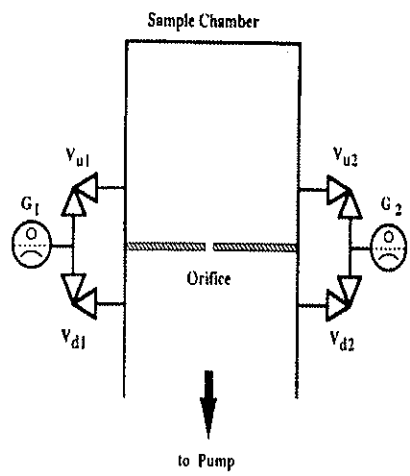



Fig. 11 Experimental set-up for the improved throughput method according to Yang et al. [23]. Two ionization gauges are used to eliminate effects due to outgassing of the gauges. Reprinted from Vacuum, 46, Y. Yang, K. Saitoh and S. Tsukahara, p1371, Copyright (1995), with permission from Elsevier Science.

Another variation of the throughput method is the so-called variable conductance or Oatley method [24]. The fixed conductance is replaced by a variable one or it is possible to replace the orifice in situ by a number of different conductances. The upstream pressure is measured depending on the conductance. If the outgassing rate is independent of the pumping speed, $p$ versus $1/C$ will give a straight line. Deviations from it may indicate outgassing of the gauge or dependence of the outgassing rate on pumping speed or pressure (readsorption).

The pressure-rise method has similar problems to the throughput method if ionisation gauges are used. Outgassing of the gauge may indicate higher outgassing rates for the surface under investigation, while pumping by the ion gauge indicates lower outgassing rates. Therefore, usually, ionisation gauges are only switched on intermittently. But even in this case disturbances cannot be excluded. Another possibility is to accumulate the pressure in a chamber without any measuring device. Instead, a valve is opened after a certain time, and the pressure burst is measured by a total pressure gauge or QMS, that has to be calibrated for similar pressure bursts. The ideal choice of a gauge for applying this method is an inert gauge, which is now available in the form of the spinning rotor gauge. This gauge neither pumps nor outgasses. Although the residual pressure reading is only $10^{-6}$ Pa, in a pressure rise system this pressure is soon reached after only a few minutes, even when the outgassing rates are very low. When the pressure rise method is continued for days or even weeks, accurate measurements can be made. The pressure rise method has recently been successfully applied to determine very low hydrogen outgassing rates of UHV and XHV systems. The main problem in this method, however, is the effect of readsorption. If readsorption occurs, interpretation of results is even more difficult than in the case of throughput method, because the pressure and so the surface coverage is changing with time.

Redhead [25] has recently analysed the problem of readsorption on outgassing measurements, both for the throughput method and the accumulation method. For the throughput method he started with the conservation of mass, which requires that the flux of molecules leaving the surface is equal to the sum of the flux through the orifice, the flux of readsorbed molecules and the change in the number of molecules in the volume.

$$Aq_i = Aq_m + A\frac{p}{\sqrt{2\pi mkT}}s + KV\frac{dp}{dt} \tag{20}$$

where $q_i$ is the intrinsic outgassing rate in molecules/s cm$^2$, $q_m$ the measured outgassing rate (or the flux through the conductance), $A$ total surface, $s$ sticking probability, and $K$ the number of molecules in a mbar l at 295 K.

The measured specific outgassing rate is given by

$$q_m = \frac{p}{\sqrt{2\pi mkT}}\frac{a}{A} \tag{21}$$

where $a$ is the effective pump area.

In stationary state $dp/dt = 0$, so that

$$q_i = q_m + \frac{p}{\sqrt{2\pi mkT}}s \tag{22}$$

$$q_i = \frac{p}{\sqrt{2\pi mkT}}s\left(1 + \frac{a}{As}\right) \tag{23}$$

$$\frac{q_m}{q_i} = \frac{\frac{a}{As}}{1 + \frac{a}{As}} \tag{24}$$

This result is plotted in Fig. 12 (top) with the sticking probability as parameter. Typical experimental values of $a/A = 10^{-4}$, while sticking probabilities for water are between $10^{-1}$ and $10^{-3}$. In this case,

ratios of $q_m/q_i = 10^{-3}$ to $10^{-1}$ can be expected! For hydrogen the sticking probability is smaller, about $10^{-4}$, so that the error in this case is less.

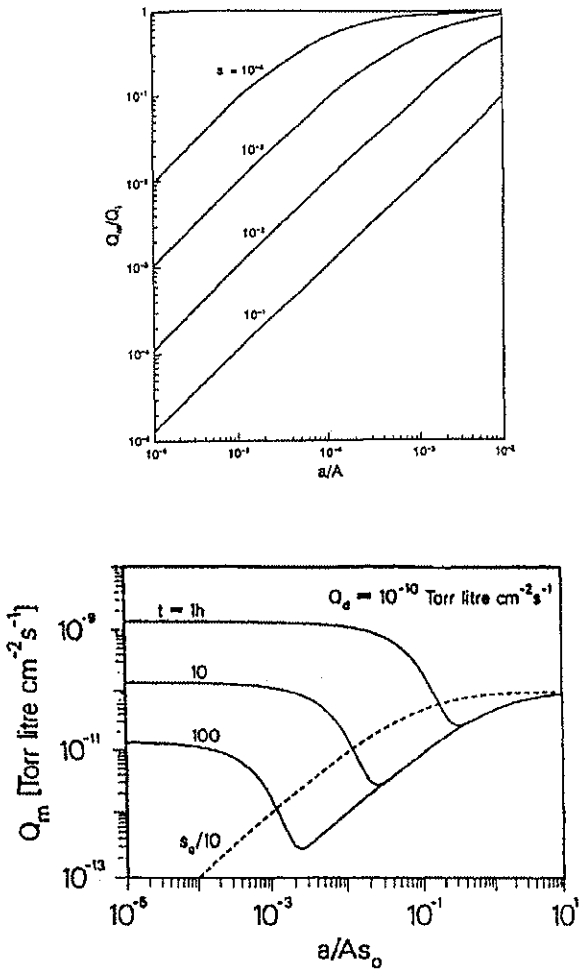



Fig. 12 Calculated ratio of the measured net outgassing rate to the intrinsic outgassing rate ($Q_m/Q_i$) in the throughput method as a function of the area ratio (open area to an ideal pump to total area) $a/A$ with the sticking probability $s$ as parameter (top) and the measured outgassing rate as function of $a/As$ with pumping time as parameter (bottom). Outgassing by diffusion $Q_d$ has been included in the lower figure. From Ref. [25]. Copyright American Vacuum Society.

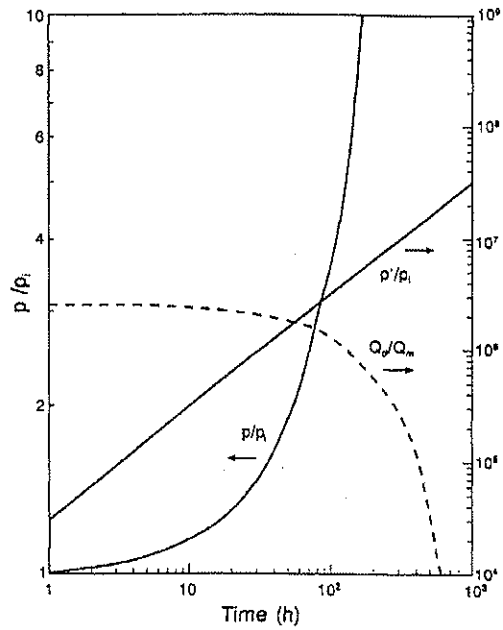



Fig. 13 Pressure rise-method: Calculated relative pressure rise (in units of initial pressure after closing the valve) versus time. From Ref. [25]. Copyright American Vacuum Society.

Another problem is how results from an outgassing measurement can be used under different pumping conditions. Figure 12 (bottom) shows the outgassing rates of $H_2O$, as they would be measured under different pumping conditions ($a/A$) and various pumping times. Also diffuse outgassing of $10^{-10}$ Torr l/s cm$^2$ was considered here. It is clear that the measured outgassing rate is a function of the pumping parameter and sticking probability for $a/As_0$ greater than some critical value. And the changes can be very significant (> factor 10). In the case of the accumulation method the pumping speed is zero and equilibrium adsorption isotherms can be used.

Figure 13 shows the increase of hydrogen pressure with time. $P'$ is the pressure, if no readsorption occurs, $p$ with adsorption, $p_i$ is the pressure in the beginning of measurement when the pump valve is closed. It shows that the increase of $p$ with readsorption is smaller by orders of magnitudes than without adsorption and far from being linear in time. $Q_d/Q_m$ is the ratio of true outgassing rate to the measured outgassing rate.

Experimental curves of hydrogen outgassing, however, show a completely linear pressure rise over long periods of time (Fig. 14). What is the solution? It is well known that $H_2$ adsorbs significantly on stainless steel walls, when injected into a evacuated vessel, so an adsorption rate = 0 can be excluded, if the surface is empty. So, is the conclusion that all sites available for dissociative adsorption of hydrogen molecules are occupied? Redhead [25] could explain the experimental results with the following model:

1. The rate of hydrogen desorption is limited by the rate of recombination and by thermal desorption.
2. There is a number of type A sites accessible to H atoms from the interior. The diffusion rate is so high that, as soon as sites are emptied by desorption, they are immediately refilled from the interior.
3. There a number of type B sites that are not accessible to H atoms from the interior, for example due to an effective oxide layer, but these sites are involved in the ad/desorption cycles with $H_2$ in the gas phase.

After starting a pressure rise these type B sites will lead to readsorption, but will, because of their small number soon saturate and readsorption will cease. Redhead has calculated (for the experimental conditions of Fig. 14) what the ratio between the sites A and B may be to see a linear pressure rise. If $\sigma_B/\sigma_A \leq 30\%$, the pressure increases linearly after 1 h (Fig. 15).

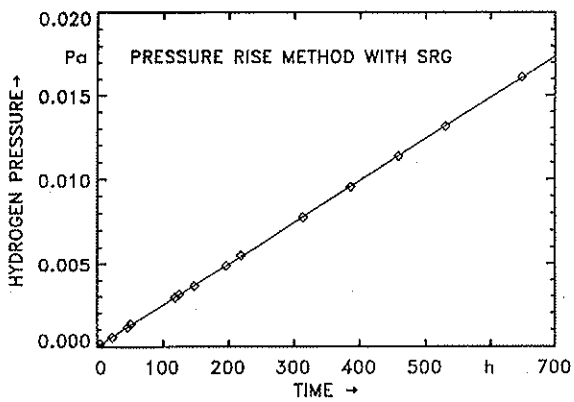



Fig. 14 Pressure rise method: Measured pressure rise in a vacuum fired stainless steel chamber with a spinning rotor gauge. Experiment carried out by the author.

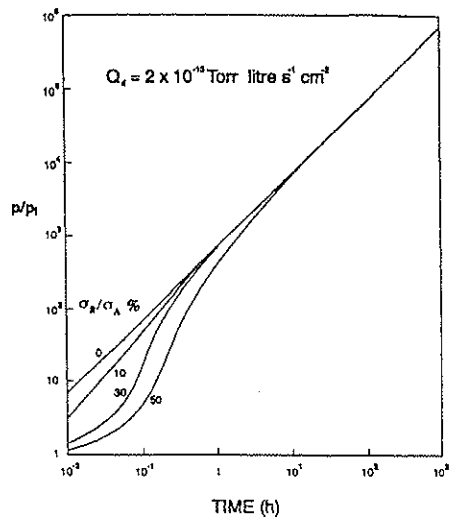



Fig. 15 Calculated relative pressure rise (in units of initial pressure after closing the valve) with the model introduced by Redhead. From Ref. [25]. Copyright American Vacuum Society.

From this model, which explains experimental results, we have another indication that at low outgassing rates the adsorption of hydrogen from a stainless-steel surface is not diffusion rate limited, but recombination limited. Otherwise it could not be explained why all available sites are always full. From the experimental data, and assuming desorption energies of $H_2$ between 18 and 24 kcal/mol, Redhead calculated the number of the available type A sites. This ranged from $10^{-4}$ to$10^{-8}$ monolayers. This low value means that the distance between neighbouring sites is rather large and makes obvious the recombination as being the rate limiting process.

## 6. CONCLUSION

At the end of this review, it should be noted that the outgassing of materials both in theory and experiment is a very vivid subject, heavily discussed, and very complex, and there is still much to learn about it. The important problems will only be solved, if metal/material science, surface science and vacuum science work closely together.

## REFERENCES


[1] P.A. Redhead, J. Vac. Sci. Technol. A 13 (1995), 2791.

[2] J.P. Hobson, Trans. 8th AVS. Vac. Symp. (1961), 26.

[3] H.F. Dylla, D.M. Manos and P.H. LaMarche, J. Vac. Sci. Technol. A 11 (1993), 2623.

[4] H.F. Dylla, J. Nucl. Mater. 93/94 (1980), 61.

[5] P.A. Redhead, J. Vac. Sci. Technol. A 13 (1995), 467.

[6] B.B. Dayton, Trans 8th AVS Vac. Symp. (1961), 42.

[7] Y. Ishikawa and T. Yoshimura, J. Vac. Sci. Technol. A 13 (1995), 1847.

[8] M. Wutz, H. Adam, und W. Walcher, *Theorie und Praxis der Vakuumtechnik,* 1988, 4. Aufl., Friedrich Vieweg & Sohn, Braunschweig/Wiesbaden p.68.

[9] R. Calder and G. Lewin, Brit. J. Appl. Phys. 18 (1967), 1459.

[10] H. L. Eschbach, F. Gross, and S. Schulien, Vacuum 13 (1963), 543.

[11] M.R. Louthan and R.G. Derrick, Corrosion Science 15 (1975), 565.

[12] K. Odaka and S. Ueda, Vacuum 47 (1996), 689.

[13] B.C. Moore, J. Vac. Sci. Technol. A 13 (1995), 545.

[14] H.C. Hseuh and Xiahua Cui, J. Vac. Sci. Technol. A 7 (1989), 2418.

[15] I. Chun, B. Cho, and S. Chung, J. Vac. Sci. Technol. A 14 (1996), 2636.

[16] V. Nemanic and J. Setina, Vacuum 49 (1998), 233.

[17] V. Nemanic and T. Bogataj, Vacuum 50 (1998), 431.

[18] J.K. Fremerey, Proceedings of the 14th International Vacuum Congress, 1998, Birmingham, to be published in Vacuum 1999.

[19] K. Jousten, Vacuum 49 (1998), 359.

[20] H.F. Dylla, Vacuum 47 (1996), 647.

[21] G. Messer and N. Treitz, Proc. 7th Int. Vac. Congr. 3rd Int. Conf. Solid Surfaces, Vienna, 1977, 223.

[22] S. Komiya at al. J. Vac. Sci. Technol. 16 (1979), 689.

[23] Y. Yang, K. Saitoh and S. Tsukahara, Vacuum 46 (1995), 1371.

[24] C.W. Oatley, Brit. J. Appl. Phys. 5 (1954), 358.

[25] P.A. Redhead, J. Vac. Sci. Technol. A 14 (1996), 2599.

[26] H.F. Dylla, Vacuum 47 (1996), 647.


## BIBLIOGRAPHY


Saul Dushman, James M. Lafferty, *Scientific Foundations of Vacuum Technique*, 2nd edition, John Wiley & Sons, 1962.

James M. Lafferty, *Foundations of Vacuum Science and Technology*, John Wiley & Sons, New York, 1998.

P.A. Redhead, J.P. Hobson, E.V. Kornelsen, *The Physical Basis of Ultrahigh Vacuum*, Chapman and Hall Ltd, London, 1968. This book has recently been re-edited by the American Vacuum Society.

# DYNAMIC OUTGASSING

*Oswald Gröbner*
CERN, Geneva, Switzerland

**Abstract**
Outgassing stimulated by photons, ions and electrons created by high-energy and high-intensity particle beams in accelerators and storage rings represents a serious limitation for the design of vacuum systems. Synchrotron-radiation-induced desorption is particularly important for intense light sources and high current storage rings. Gas desorption-induced by beam generated ions was a serious limitation to the performance of the Intersecting Proton Storage Rings (ISR) at CERN in the past and is a major concern for the vacuum system design of the Large Hadron Collider.

## 1. SYNCHROTRON RADIATION-INDUCED DESORPTION

In high-energy, high-intensity accelerators and storage rings for electrons and positrons, the particle beams emit intense synchrotron radiation in a forward-directed narrow cone. This energetic photon flux produces strong outgassing from the vacuum system and thus a large dynamic pressure increase, which limits the beam lifetime in the machine and may cause increased background in the colliding-beam experiments. For the first time a significant level of synchrotron radiation will also occur with the 7-TeV proton beams in the LHC.

### 1.1 Basic equations for synchrotron radiation

The equations required to evaluate the vacuum conditions with synchrotron radiation for electrons or positrons [1] are:

Total radiated power (W):

$$P_\gamma = 88.6 \frac{E^4 I}{\rho} \tag{1}$$

where $E$ (GeV) is the beam energy, $I$ (mA) the beam current and $\rho$ (m) the bending radius. The critical energy of the S.R. spectrum (eV) is given by

$$\varepsilon_c = 2.2 \cdot 10^3 \frac{E^3}{\rho} \ . \tag{2}$$

The total photon flux ($s^{-1}$) around the machine circumference is

$$\dot{\Gamma} = 8.08 \cdot 10^{17} IE \tag{3}$$

and the linear photon flux ($m^{-1}\ s^{-1}$)

$$\frac{d\dot{\Gamma}}{ds} = 1.28 \cdot 10^{17} \frac{IE}{\rho} \ . \tag{4}$$

The gas flow due to photon induced desorption

$$Q = \eta \dot{\Gamma} \tag{5}$$

where $\eta$ is the molecular desorption yield (molecules per photon) averaged over all photon energies. The total gas load including the contribution from thermal outgassing $Q_o$ (Torr l $s^{-1}$) is

$$Q = K\eta I E + Q_o \tag{6}$$

where the constant K converts from molecules to pressure units (e.g. K = 2.8 $10^{-20}$ Torr l/molecule at 273K). The total dynamic pressure is

$$P_{dyn} = \frac{Q}{S} \,. \tag{7}$$

The instantaneous dynamic pressure increases proportionally with the beam intensity and this quantity is generally referred to as the dynamic pressure rise (Torr/mA):

$$\frac{\Delta P}{I} = \frac{Q - Q_o}{S} \tag{8}$$

which depends on the pumping speed of the vacuum system. Since the gas desorption occurs all around the bending sections of the machine, a distributed pumping system using 'linear integrated' pumps has significant advantages over individual, lumped pumps.

During machine operation, 'beam cleaning' or 'beam scrubbing' of the vacuum system effectively reduces the molecular desorption yield by several orders of magnitude. Since it would be very expensive to build a vacuum system with sufficient pumping speed to guarantee the required beam lifetime at the start-up, most synchrotron light sources and storage rings take advantage of the beam cleaning process. It is an accepted design concept that the required vacuum performance will be achieved after an initial conditioning time only.

The cleaning time or, more specifically, the integrated beam dose which is required to reduce the molecular desorption yield from its initial large value, can be estimated with good accuracy using the exponential dependence on the beam dose $D$ (mA h)

$$\eta = \eta_o D^{-a} \,. \tag{9}$$

The exponent $a$ is found to vary between 0.6 to ~1 depending on the specific machine [2–4]. The observed clean-up for LEP is shown in Fig. 1. The detailed evolution of the dynamic pressure rise in LEP depends on the combined effect of the clean-up of the vacuum system and of the saturation of the linear getter pumps (NEG). The indicated straight line with a slope of –1 corresponds to the measurements with a newly activated getter and hence a constant, maximum pumping speed.

Photon stimulated molecular desorption has been attributed to a two step process [5, 6]:

in the first step photons produce photo-electrons and secondary electrons

subsequently secondary-electrons excite strongly-bound molecules, which may desorb spontaneously. This model is strongly supported by measurements that have shown a very close correlation between gas desorption and photoelectrons produced [7].

### 1.2 Molecular-desorption yields for various vacuum chamber materials

For the design of vacuum systems the most important parameter is the desorption yield for the relevant conditions of synchrotron-radiation spectrum, angle of incidence and vacuum chamber materials. Stimulated by the need to obtain reliable input data for the design of vacuum systems,

extensive investigations have been made using the experimental system shown in Fig. 2. This set-up allows the measurement of the instantaneous desorption yield as well as its evolution with beam dose, the dependence on the synchrotron radiation spectrum (critical energy) and the angle of incidence. A detailed description of this system which was first installed on an external photon beam line at the DCI storage ring in ORSAY, France can be found in Ref. [1].

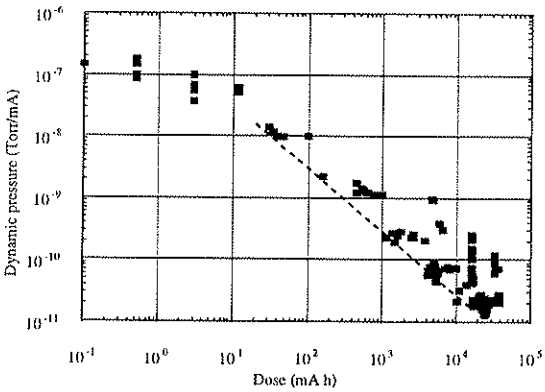



Fig. 1 Beam cleaning during the first running period of LEP.

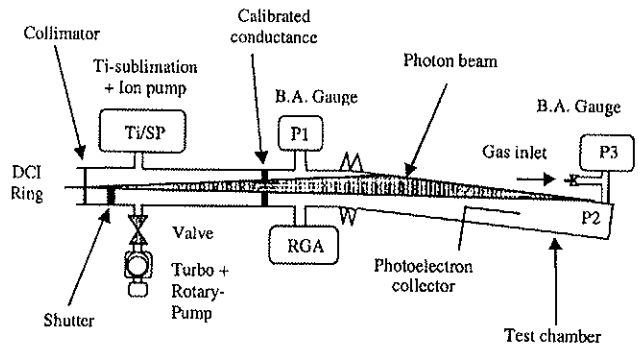



Fig. 2 Experimental system to measure synchrotron radiation-induced desorption.

The photon beam is extracted from a bending magnet and passes through a set of vertical and horizontal collimators which precisely define the total photon flux which enters the experiment. Due to the vertical divergence of synchrotron radiation the vertical collimation has been chosen sufficiently large so as not to attenuate the photon spectrum at low photon energies (e.g. below about 5 eV in the DCI set-up). To study the effect of angle of incidence, a set of bellows has been introduced so that the test chamber can be pivoted in the horizontal plane. Total and partial pressure measurements are done with a combination of calibrated total pressure gauges and a calibrated quadrupole mass spectrometer. To correlate photoelectrons with molecular desorption, an electrically biased wire electrode has been installed in some of the test chambers.

With the growing importance of reliable input data for the design of future synchrotron light sources and $e^+e^-$ storage rings, similar experiments have been installed in several other synchrotron light sources at KEK Japan, BNL USA, EPA at CERN, INP Russia and ESRF France. The results obtained with this set-up can be considered as representative for a real machine vacuum system. However, it should be noted that the measurement gives a net desorption yield since only the net amount of gas leaving a test chamber through the calibrated conductance can be observed. Furthermore, with grazing photon incidence, a fraction of the photons is reflected from the region of the primary impact and may not desorb molecules. These reflected photons may nevertheless contribute to the gas desorption when they strike the end face of the test chamber under near perpendicular incidence. For a more accurate measurement, the amount of reflected photons should be estimated and the molecular yields corrected for this effect.

### 1.3 Measured desorption yields

Measurements at DCI have been made with a critical photon energy between 0.8 and 3.75 keV and 11 mrad grazing photon incidence on a 3.5-m long test chamber. A selection of results is shown in Figs. 3 to 6 respectively referring to *in-situ* baked (between 150°C and 300°C) vacuum chambers made of aluminium alloy, untreated stainless steel and stainless steel baked under vacuum at 950°C (vacuum degassed) as a pre-treatment and finally OFHC-copper. Additional results for an unbaked stainless steel chamber may be found Ref. [3]. In order to facilitate the scaling of these data to the specific conditions in other accelerators, the beam dose has been expressed in accumulated photons per m of test chamber rather than in machine mA hours.

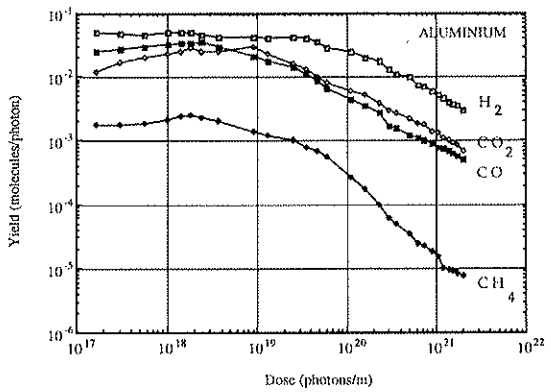



Fig. 3 Molecular desorption yields for aluminium alloy.

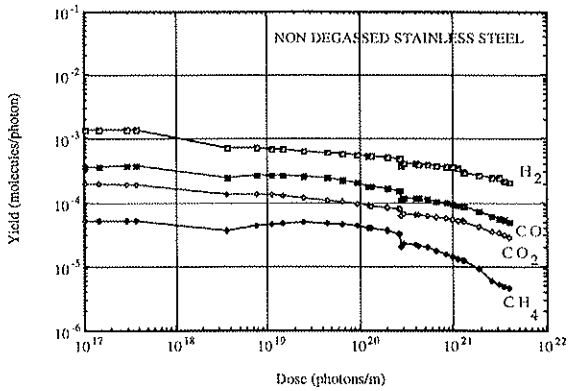



Fig. 4 Molecular desorption yields for stainless steel.

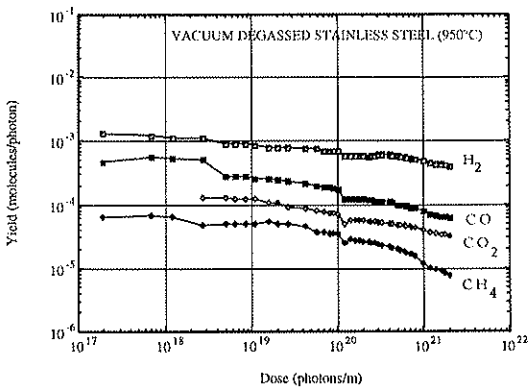



Fig. 5 Molecular desorption yields for vacuum fired stainless steel.

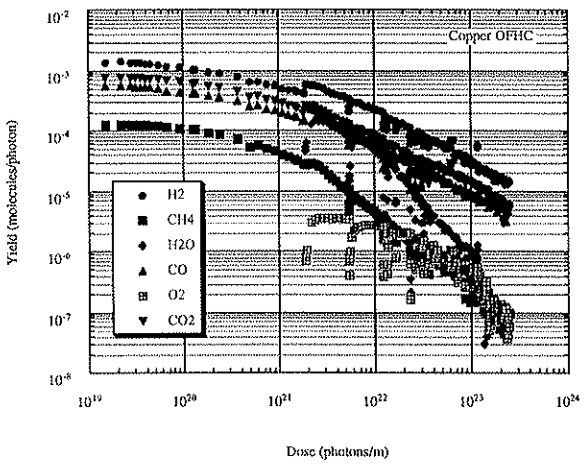



Fig. 6 Molecular desorption yields for OFHC copper.

The total duration of the photon exposures in these measurements ranges from some 24 hours to many weeks in the case of Fig. 6. Therefore, each individual plot represents many successive exposures with intermittent periods when the pressure in the test chamber has time to recover. In spite of these interruptions, the dynamic pressure evolution and the molecular desorption yields derived from it follow a steadily decreasing trend.

## 1.4 Energy dependence of the molecular-desorption yield

The dependence of the desorption yield on the critical energy of the S.R. has been measured over a range of energies and for different vacuum chambers. Basically one can expect a lower threshold at which the photon energy is insufficient to produce photoelectrons (~5 eV). With increasing energy, photons penetrate progressively deeper into the surface such that the escape probability for electrons is reduced. This effect should lead to a decreasing desorption efficiency per total number of photons. For very high photon energies, above several hundred keV and even MeV as in LEP, the Compton cross section dominates over the photo effect. Compton scattering will produce a shower of energetic recoil electrons and an increasing number of scattered secondary photons. The global effect, as seen in LEP at beam energies above 50 GeV is an increasing desorption yield. Figure 7 shows a compilation of desorption data collected from various experiments and machines and covering the range of critical photon energies from 12 eV up to several hundred keV. The data referring to a

particular machine have been obtained under the same conditions and should reflect the energy variation for this particular system while the data from different experiments may not be directly comparable and have therefore been plotted in arbitrary units. At very low energies, representative of the situation in the LHC [8], the yield increases strongly with energy.

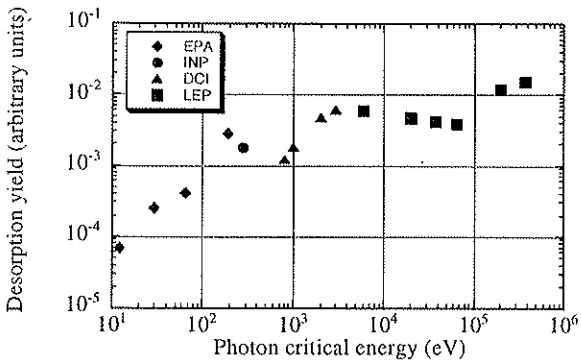



Fig. 7 Relative desorption efficiency as a function of the critical photon energy.

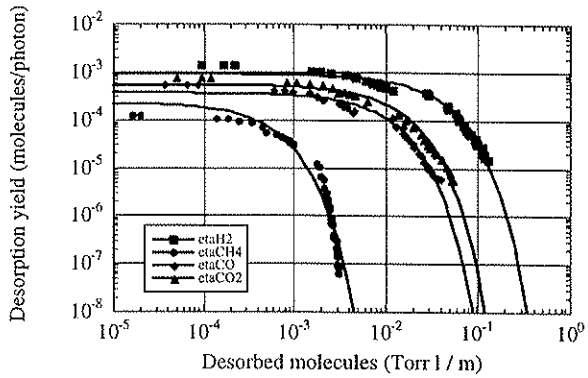



Fig. 8 Integral gas load for an OFHC copper chamber.

## 1.5 Integral desorbed-gas load

A quantity, which is particularly relevant for a getter-pumped vacuum system in order to estimate the intervals when the getter will have to be reconditioned, is the integral amount of desorbed gas during the cleaning process. This quantity can be obtained by integrating the respective cleaning curves. For the OFHC copper chamber the result of the integration is shown in Fig. 8. The observed decrease of the desorption yield $\eta$ as function of the gas load $Q$ (Torr l/m) can be described closely by the exponential expression:

$$\eta = \eta_0 \, e^{-\frac{Q}{Q_0}} \quad . \tag{10}$$

The respective values of the two parameters $\eta_0$ and $Q_0$ for different gas species are given in Table 1. The first line gives the monolayer capacity, which has been used to calculate the value of $Q_0$ in the last line.

Table 1
Parameters for OFHC copper

| | $H_2$ | $CH_4$ | CO | $CO_2$ |
|---|---|---|---|---|
| Molecules ($cm^{-2}$/monolayer) | $1.8\ 10^{15}$ | $1.2\ 10^{15}$ | $1.0\ 10^{15}$ | $9.6\ 10^{14}$ |
| $\eta_0$ (molecules/photon) | $9.2\ 10^{-4}$ | $2.3\ 10^{-4}$ | $3.7\ 10^{-4}$ | $5.5\ 10^{-4}$ |
| Q (Torr l/m) | $3.0\ 10^{-2}$ | $4.5\ 10^{-4}$ | $8.4\ 10^{-3}$ | $1.1\ 10^{-2}$ |
| $Q_0$ (monolayers) | 0.13 | $2.9\ 10^{-3}$ | $6.5\ 10^{-2}$ | $8.9\ 10^{-2}$ |

It can be seen that, e.g. for a baked copper chamber, the total amount of gas which needs to be pumped remains well within the capacity of commercial getter pumps. By comparison, the gas load for an aluminium alloy vacuum chamber is more than one order of magnitude larger but in spite of this larger load it has not caused any vacuum limitations during LEP operation [9].

### 1.6 Observations of time-dependent effects

Detailed measurements have shown that molecules desorb not only from the directly illuminated region but from the entire inner surface due to the scattered/reflected photons as well as secondary electrons. As a consequence, a vacuum chamber must be cleaned over its full perimeter and not only along a narrow directly illuminated strip.

A pronounced and persistent wall pumping effect has been observed after prolonged photon exposure, which translates into a molecular sticking probability of about $10^{-4}$. Due to this pumping effect, molecules may re-adsorb several times on the walls before leaving the system. Therefore, the system in Fig. 2 measures an effective, net yield rather than a true, intrinsic desorption rate.

Meta-stable, excited molecules accumulate and recombine on the surface producing new species, e.g. $H_2O$ and $O_2$ which have been observed after a prolonged photon exposure. This effect is illustrated in Fig. 9 where it is shown that several hours of exposure may be required until the desorption process reaches a steady state.

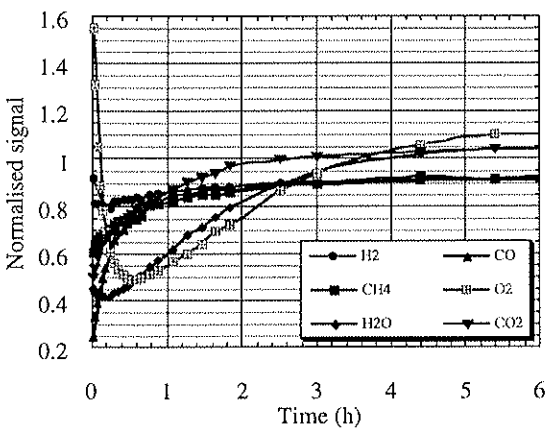



Fig. 9 Time-dependent partial pressures at the beginning of an irradiation period following an interruption of several hours.

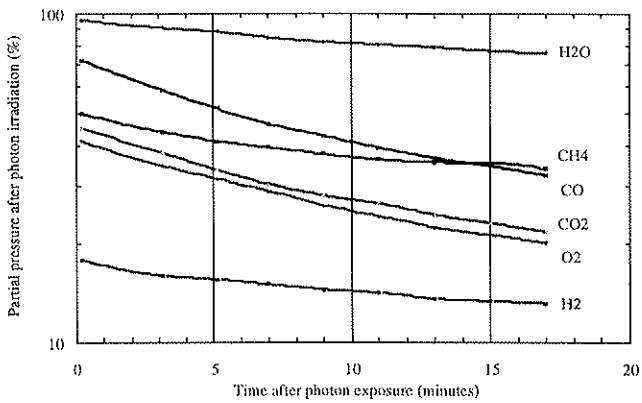



Fig. 10 Time-dependent partial pressures following an irradiation period.

The inverse process when the photon beam is stopped after a long irradiation period is shown in Fig. 10. While certain molecular species show a large instantaneous reduction, several other gases such as $H_2O$, decrease by a very small amount only and continue to evolve over a very long period of time. Indeed it may take many hours and even days until the system comes back to its pre-irradiation conditions. Based on these observations, synchrotron radiation stimulated desorption can be separated into a

- prompt component, characterised by the fact that desorption stops immediately with the photon irradiation and a

- delayed component, equivalent to an increased level of thermal desorption persisting long after the photon irradiation and characterised by the appearance of 'new' molecular species. The best illustration for this delayed component may be $O_2$ in Fig. 9. This species accumulates during the irradiation free period preceding the measurement since its initial desorption yield is by a factor of 1.6 higher than its steady state value.

Photon stimulated desorption of $H_2$ can be described very closely by a diffusion model from the bulk to the surface [10].

## 2. ION-INDUCED DESORPTION

In storage rings with intense, positively-charged beams of protons and positrons the pressure can increase due to ion-induced desorption from the walls of the beam pipe. Ions created from the residual gas are repelled by the positive space-charge potential and are accelerated towards the wall.

Ions may gain energies of several hundred eV per ampere of circulating current and are thus very effective for desorbing strongly-bound gas molecules. Since the rate of ionisation is proportional to the gas density an avalanche process may occur resulting in a continuously increasing pressure, see Fig. 11.

The total flux of desorbed gas into the vacuum system is:

$$Q = \eta\sigma \; P\frac{I}{e} + Q_o \tag{11}$$

where $\eta$ is the molecular desorption yield (molecules/ion), $\sigma$ the ionisation cross section of the residual gas molecules, $P$ the pressure, $I$ the average beam current, $e$ the unit charge and $Q_o$ is the thermal outgassing rate from the wall.

The balance of gas desorbed by the beam and removed by external pumps, $S$ is given by

$$PS = \eta\sigma\frac{I}{e}P + Q_o \tag{12}$$

and therefore

$$P = \frac{Q_o}{S - \eta\sigma\frac{I}{e}} \; . \tag{13}$$

One finds that the pressure increases with beam current and that above a critical value no equilibrium pressure exists:

$$(\eta I)_{crit} = \frac{e\,S}{\sigma} \; . \tag{14}$$

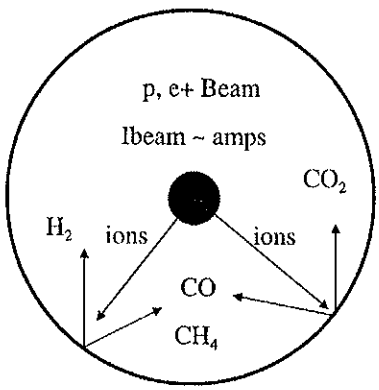



Fig. 11 Beam induced desorption mechanism.

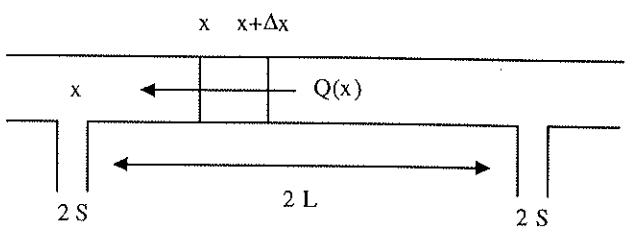



Fig. 12 Linear vacuum system with ion-induced desorption.

## 2.1 Parameters which determine vacuum stability

Ionisation cross sections for some common gas species are listed in Table 2. The values have been calculated for the beam energies in the ISR and for LHC using Ref. [11].

Table 2
Ionisation cross sections

| | Ionisation cross-section (in $10^{-18}$ $cm^2$) | |
|---|---|---|
| **Gas** | **26 GeV** | **7000 GeV** |
| $H_2$ | 0.22 | 0.37 |
| He | 0.23 | 0.38 |
| $CH_4$ | 1.2 | 2.1 |
| CO | 1.0 | 1.8 |
| Ar | 1.1 | 2.0 |
| $CO_2$ | 1.6 | 2.8 |

## 2.2 Effective pumping speed $S_{eff}$

The pumping speed which is required to compensate the desorption by ions depends strongly on the configuration of the vacuum system: For a periodically pumped, linear vacuum system with the layout as in Fig. 12, the vacuum stability can be calculated as shown below [12]. The addition of distributed pumping, e.g. by a continuous, linear getter pump or by cryopumping in a cold-bore vacuum system can be implemented as a straightforward extension of the mathematical formalism.

Gas flow $Q(x)$ [Torr $m^3$ $s^{-1}$]

Specific linear outgassing rate: $q(x)$ [Torr $m^2$ $s^{-1}$]

Specific molecular conductance of the vacuum chamber: $c$ [$m^4$ $s^{-1}$]

Pumping speed $S$ [$m^3$ $s^{-1}$]

Ion-induced desorption coefficient $b$ [$m^2$ $s^{-1}$] where

$$b = \sigma \eta \frac{I}{e} \tag{15}$$

with $I$ the beam current, $e$ the electron charge, $\sigma$ the ionisation cross section and $\eta$ the molecular desorption yield which is expressed here in molecules/ion.

In stationary conditions, the pressure is given by

$$c \frac{d^2P}{dx^2} + bP + q = 0 \ . \tag{16}$$

Boundary conditions for a uniform, periodic vacuum system with pumps (each pump has the speed $2S$) at a regular distance $2L$ are given by:

$$c \left[ \frac{dP}{dx} \right]_{x=\pm L} = \mp SP[\pm L] \ . \tag{17}$$

Three types of solutions exist for this system, which depend on the range of the parameter $b$:

For $\boldsymbol{b = 0}$, one obtains the usual parabolic pressure distribution for a periodically pumped system:

$$P(x) = q \left( \frac{2Lx - x^2}{2c} + \frac{L}{S} \right) . \tag{18}$$

For $\boldsymbol{b > 0}$, the solution has the form

$$P[x] = \frac{q}{b} \left[ \frac{\cos[\sqrt{b/c}\, x]}{\cos[\sqrt{b/c}\, L] - \frac{\sqrt{bc}}{S} \sin[\sqrt{b/c}\, L]} - 1 \right] . \tag{19}$$

A finite solution exists as long as:

$$\cos[\sqrt{b/c}\, L] - \frac{\sqrt{bc}}{S} \sin[\sqrt{b/c}\, L] > 0 \ . \tag{20}$$

The limit of vacuum stability, i.e. the largest stable value for the parameter $b$ corresponds to the first root of the transcendental equation

$$L\sqrt{b/c}\ \tan[L\sqrt{b/c}] = \frac{SL}{c} \tag{21}$$

and can be computed numerically. Nevertheless, an upper limit for $b$ exists when the vacuum system is strongly conductance limited ($S$ and $L$ very large or c very small), which can be obtained simply from the argument when the tangent goes to infinity:

$$b_{\max} = \frac{\pi^2}{4} \frac{c}{L^2} \ . \tag{22}$$

The critical beam current for this limiting condition is

$$(\eta I)_{crit} = \frac{\pi^2}{4} \frac{e\, c}{\sigma L^2} \ . \tag{23}$$

For any practical case, the critical current will be lower but any 'well designed' vacuum system should fall within 30% of this simplified condition.

Finally for $\boldsymbol{b < 0}$, the pressure decreases since the beam acts like an ion-pump — beam pumping. The pressure distribution follows a hyperbolic function and remains always stable. Numerically the 'pumping speed' of the beam can be calculated from the expression

$$S_{beam} = \frac{\sigma \kappa}{e} I \ (\mathrm{m^2 s^{-1}}) \tag{24}$$

where $\kappa$ represents the probability that an incident ion is captured at the wall. Beam pumping increases with the beam current and has been observed frequently, but not systematically, in the ISR with currents of up to 60 A [13]. The pumping depends directly on the capture probability of the ions on the wall and a rough estimate of this effect may be derived from the pumping speed of a commercial ion pump as $\kappa$~0.1 [14].

### 2.3 Molecular-desorption yield

Desorption yields which are measured in a laboratory system [15] and in a real machine where the bombarding ions are generated directly from the residual gas, are related to each other by the condition

$$\eta_m = \eta_{lab} - \kappa \ . \tag{25}$$

In the case where $\kappa > \eta_{lab}$, the effective yield is negative and the beam will act as a pump. The upper limit for $\kappa$ is unity corresponding to the situation when all ions are implanted and pumped on the wall of the vacuum system. In reality, the capture probability will be considerably less than unity and one may expect that it depends on the surface conditions, the type of ions and on the ion energy.

Ion-induced desorption yields have been measured for a number of vacuum chamber materials and for a large variety of vacuum treatments [16]. The most relevant data for designing future accelerator vacuum systems, the desorption yields for unbaked and for *in-situ* baked stainless steel chambers, are shown in Figs. 13 and 14 as a function of the ion energy [17].

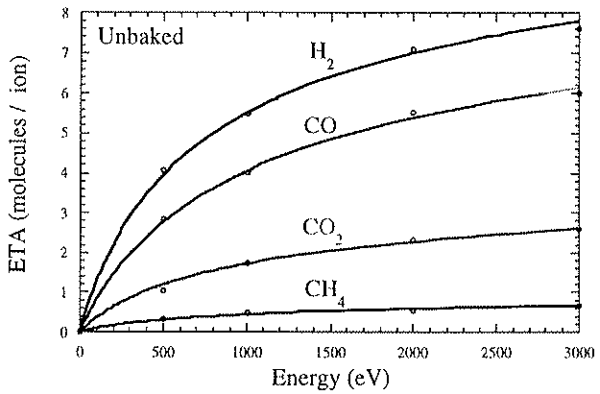



Fig. 13 Desorption yields for unbaked stainless steel as a function of incident ion energy.

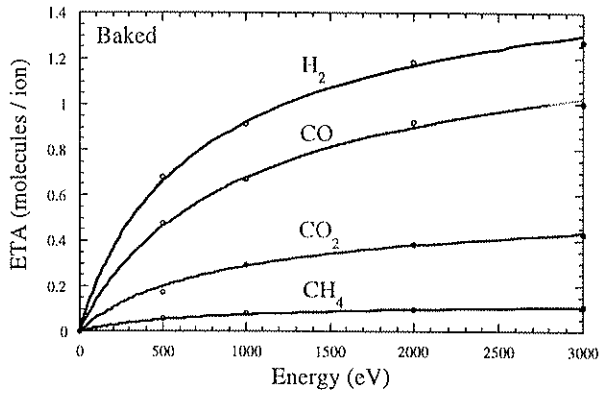



Fig. 14 Desorption yields for baked stainless steel as a function of ion energy.

The most significant permanent improvement among a variety of possible treatments has been achieved by using an argon ion glow discharge followed by an *in-situ* bakeout, see Fig. 15. A glow discharge treatment without subsequent *in-situ* bakeout of the vacuum system results only in a small, often insufficient improvement [18].

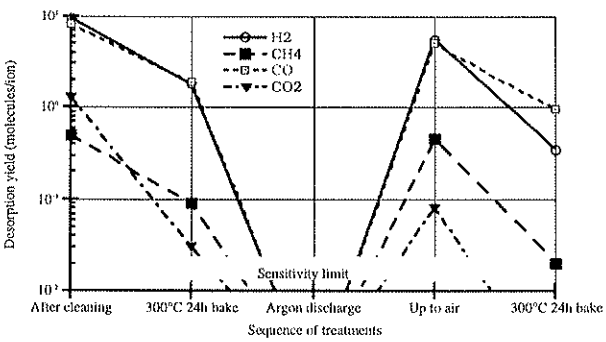



Fig. 15 Effect of various treatments on the ion-induced desorption coefficients.

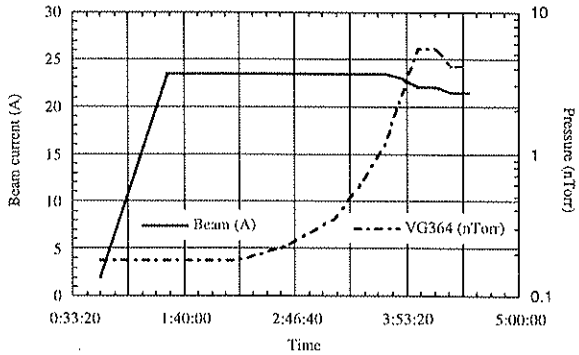



Fig. 16 Pressure instability during beam stacking in the ISR.

## 2.4 Vacuum stability in a cold bore system with cryosorbing walls (LHC)

In a cryogenic vacuum system with cold walls molecules are cryopumped with high efficiency directly onto the cold bore. The pumping speed per unit length is

$$S_{eff} = \frac{1}{4}\bar{v}\, s\, F \tag{26}$$

with $\bar{v}$ the mean molecular velocity, $s$ the sticking probability of molecules on the wall and $F$ the surface area per unit length.

With a cold beam pipe of radius $r_p$, the stability limit is

$$(\eta I)_{crit} = \frac{\pi}{2}\bar{v}\, s\, r_p \frac{e}{\sigma} \tag{27}$$

and for $s \sim 1$ the critical current should be very large, of the order of kA [19, 8]. However, this large stability limit can be offset by two factors:

Firstly, the sticking probability can be much smaller than unity

Secondly, the molecular desorption yield for condensed gas, specifically for $H_2$ accumulating on the cold bore, can become very large. Values of up to $10^4$ molecules per ion have been measured in the laboratory [20]. However, as has been demonstrated within the on-going LHC vacuum studies, cryosorbed hydrogen which is irradiated by photons will desorb with a very high efficiency. This strong recycling effect counteracts the accumulation of hydrogen on the wall of the LHC beam screen and limits the surface density and at the same time also the related desorption yield to an acceptable value. A detailed discussion of ion-induced vacuum stability in a cold vacuum system can be found in Ref. [21].

### 2.5 ISR observations and remedies for the ion-induced pressure instability

Up to now, the ISR has been the only machine where the ion-induced pressure instability has been observed. In the ISR, the pressure rise has been observed initially at very low beam currents of only a few amperes. Following many years of intensive effort this limit could be increased to about 60A by installing additional pumps, thus reducing $L$ and increasing $S$ but primarily by improving systematically the surface cleanliness of all vacuum chambers. A typical example of the build-up of an unstable pressure during beam stacking is shown in Fig. 16. A consistent observation also shown in this figure is slow build-up of the pressure over many hours following the stacking of the beam which is not expected from the simple stability condition. To understand the observed slow time response of the pressure a model, which involves the interaction of the volume gas with an adsorbed surface phase, has been developed [22] which is very similar to the more recent treatment for the LHC cold bore vacuum system [23]. A further important observation has been that while a pressure instability develops, the gas composition slowly evolves from a hydrogen-dominated rest gas to a rest gas dominated by CO [24]. Since for this gas species the desorption yield is larger (larger ion mass) and the effective pumping speed smaller (smaller vacuum conductance) compared to hydrogen, the stability limit can be significantly lower.

## 3. CONCLUSIONS

From the ISR observations the following conclusions have been drawn [25] which may be relevant also for the design of future machines:

$\eta$(CO) for 300°C baked stainless steel chambers ranges between 2 and 4.

Vacuum chambers which have been cleaned by argon-ion glow-discharge prior to installation and which have been baked *in-situ* show $\eta$ values which are reliably and permanently close to zero or even negative.

Beam pumping has been observed in many places at intermediate beam currents, particularly where the intrinsic pumping speed of the system was low and thus the relative effect of the beam larger. A safe, conservative vacuum design should, however, not be based on this transient effect.

For the LHC, it is expected that many room-temperature sections will require special measures to guarantee vacuum stability. This applies in particular to the vacuum systems for the experiments which need to be designed with adequate safety margins by making the necessary provisions for the pumping and for low desorption yields.

## REFERENCES


[1] O. Gröbner et al., Vacuum 33, No.7, 397 (1983).

[2] Ch. Falland et al., VIII Int. Vacuum Congr., Cannes, September 22-26 (1980).

[3] C. Herbeaux et. al., J. Vac. Sci. Technol. A 17(2) Mar/Apr. (1999).

[4] J-P. Bojon et al., EPAC-92, Berlin, 24-28 March (1992).

[5] G.E. Fischer, R.A. Mack, J. Vac. Sci. Technol. 2 123 (1965).

[6] M. Bernardini and L. Malter, J. Vac. Sci. Technol. 2, 130 (1965).

[7] J. Gómez-Goñi, O. Gröbner and A.G. Mathewson, J. Vac. Sci. Technol. A 12 (4), Jul/Aug (1994).

[8] O. Gröbner, Vacuum 46, No 8-10, 797 (1995).

[9] P.M. Strubin and J-P. Bojon, PAC-95, Dallas (1995).

[10] M. Andritschky et al., Vacuum, 38, No 8-10, 933 (1988).

[11] F. Rieke and W. Prepejchal, PR A6, 1507 (1972).

[12] E. Fischer and K. Zankel, CERN/ISR-VA/73-52 (1973).

[13] O. Gröbner and R. Calder, PAC 73, San Francisco, 760 (1973).

[14] J.M. Laurent, private communication, 1998.

[15] A.G. Mathewson, CERN/ISR-VA/76-5, 1st March (1976).

[16] A.G. Mathewson, Vuoto, Vol. 17, No. 3, 102, July-September (1987).

[17] Original data from A.G. Mathewson compiled by I.R. Collins.

[18] R. Veness et. al., EPAC-98, Stockholm (1998).

[19] C. Benvenuti, R. Calder and N. Hilleret, IEEE Trans. on Nucl. Sci. Vol, NS-24, No. 3, June (1977).

[20] J.C. Barnard, I. Bojko and N. Hilleret, Vacuum 47, No. 4, 347, (1996).

[21] W.C. Turner, J. Vac. Sci. Technol., A 14 (4), 2026 (1996).

[22] O. Gröbner, CERN/ISR-VA/76-25, 8th June (1976).

[23] O. Gröbner, The LHC vacuum System, these proceedings.

[24] A.G. Mathewson and P. Strubin, ISR Performance Report, ISR/VA/AM/sm, 17th August (1981).

[25] E. Fischer, ISR Performance Report ISR-VA/EF/sm, 3rd October (1975).

# CLEANING FOR VACUUM SERVICE

*R.J. Reid*
CLRC Daresbury Laboratory, Warrington, UK

**Abstract**
This paper will discuss the background to choosing an appropriate cleaning process for a vacuum vessel or component dependent on the pressure regime required. Examples of cleaning techniques and processes will be given and ways of determining whether or not an item is sufficiently clean will be discussed.

## 1. INTRODUCTION

Experience shows that in order to achieve all but the most modest levels of vacuum it is necessary to clean vacuum vessels and components in some way. This is because as far as vacuum is concerned, the world is a dirty place! In general, vessels and components will have been machined, worked in some way or another, or handled. Such processes may use greases or oils which have high outgassing rates or vapour pressures and which will remain on or in the surface. Marking pens and adhesive tapes leave residues on surfaces that can also enhance outgassing. Water, solvents and other liquids can remain embedded in cracks or pores in a surface and can outgas over long periods. All such contamination can limit the base pressures attainable in a system.

In general terms as far as vacuum is concerned, we define contamination as anything which

- prevents the vacuum system from reaching the desired base pressure
- introduces an unwanted or detrimental species into the residual gas
- modifies the surface properties of all or part of the vacuum system in an undesired way.

Thus for example, a pool of liquid mercury in a vacuum system will, at room temperature, limit the base pressure to about $2 \times 10^{-3}$ mbar, its vapour pressure. If the system were required to operate at, say, $10^{-5}$ mbar, the mercury would be a contaminant.

In an electron storage ring, the cross section for electrons scattering off a residual gas molecule increases as the square of the atomic number, Z, of the scattering species. Thus, for long beam lifetimes (i.e. low electron losses), it is important to minimise the partial pressures of high-Z species which are contaminants in the residual atmosphere.

In a system operating at $10^{-8}$ mbar, the partial pressure of hydrocarbons might be $10^{-11}$ mbar. These hydrocarbon molecules striking a mirror surface can crack or polymerise when irradiated by electrons or photons, leading to maybe graphite-like overlayers or to insulating layers. A complete monolayer might form in about 12 days, assuming that 1 in every 10 molecules impinging on the surface is cracked and sticks. The optical properties of the mirror are therefore altered by the surface becoming contaminated.

Hence there is a necessity to clean to remove actual or potential contaminants. We will define a suitable cleaning process as one which results in the residual vacuum being suitable for the task. This definition is heavily influenced by the observation that most people who use vacuum are not interested in vacuum as such but are interested in an industrial process (semiconductor chip production or metal refining for example) or a scientific experiment (such as studying chemical catalytic reactions on a surface) or operating a machine (maybe a particle accelerator). What they are interested in is having a defined, controlled atmosphere and vacuum is just the simplest way of achieving this.

In this discussion, we shall be considering in very general terms some of the factors influencing the choice of a cleaning process. This article is an overview and does not pretend to be a

comprehensive review citing the literature in detail. A (very) short and selective bibliography is given at the end and interested parties can use the references in these papers as a starting point for a more in depth survey.

## 2. CHOOSING A CLEANING PROCESS

There is no one cleaning process which is "right" for all vacuum systems, vessels or components. Some of the things which will need to be taken into account are as follows

- the level of vacuum required (rough, high, UHV, etc)
- if there is a particular performance requirement (e.g. low desorption)
- whether there is a particular contaminant (e.g. hydrocarbons) whose partial pressure must be minimised
- what materials the items are made from
- how the items are constructed
- safety
- cost.

Once these have been considered, and we will look at some of these in more detail below, then one can begin to choose a cleaning agent and the necessary processes to achieve the desired result.

It is important to realise that there are very many different "recipes" in the literature and in the folklore of vacuum. Advocates of a particular process or procedure will defend their choice with an almost religious fervour. Generally, this is because the procedure "works" for them, i.e. it meets their requirements. Often this procedure will not have been tested rigorously against other possible procedures which might be equally good if not better (in some sense) and will therefore in no way have been optimised. This is not necessarily a problem. It often simply means that the processes being used may not be as economical or convenient as they might be. Whether this is important or not depends on the individual situation.

In this article, we will look briefly at both chemical and physical cleaning processes. There are, however, some general points which should be taken into account when deciding what to use. Some of these are as follows.

- Some "cleaning" processes which are often used are applied more for cosmetic purposes to make vacuum chambers look clean by producing mirror finish surfaces and so on. Whether from a vacuum point of view they are actually cleaner is by no means to be taken for granted. Such surfaces could, for example, exhibit enhanced outgassing.
- The minimum chemical cleaning process compatible with the level of vacuum/cleanliness required should always be chosen. The less that needs to be done to a vacuum surface the better.
- Chemical cleaning is a hazardous procedure so must be done safely!
- Processes such as bead or shot blasting, grinding, scraping and mechanical polishing can leave dirt trapped in voids in the surface of materials which can then be very difficult to remove.
- Acid treatments such as pickling, passivation or electropolishing can trap acids in the surface of the material. For demanding UHV applications, a vacuum bake to 450°C is required to remove these completely.

## 3. DEPENDENCE ON THE BASE PRESSURE REQUIRED

In general the lower the base pressure required, the more rigorous the cleaning process will need to be. (Note that in general it is preferable to use the term "base pressure" rather than "ultimate pressure" which is commonly used. Base pressure refers to the lowest normal pressure attained in a

vacuum system in its working condition, whereas ultimate pressure strictly refers to the lowest pressure obtained in a standard defined system measured in a standard way.)

Figure 1 illustrates in a very schematic fashion some typical common cleaning processes which would be used in sequence. To use this figure, select the approximate pressure required and then apply all the processes in order from the top to the level corresponding to the pressure required.

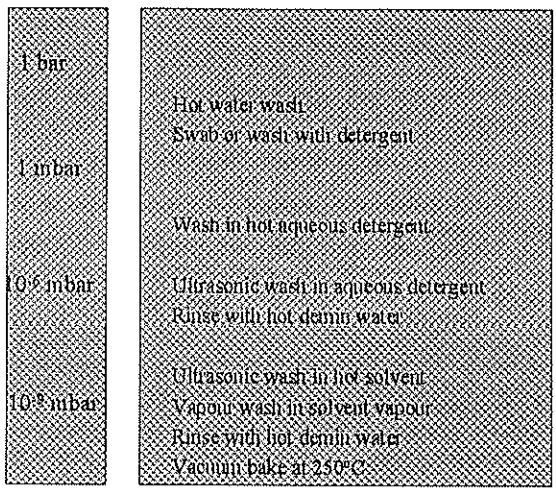



Fig. 1 Simple cleaning procedures

## 4. DEPENDENCE ON CONTAMINANTS

The user will have to determine what level of contaminants can be tolerated in any given process. It is important to distinguish between the total pressure required in the system and the partial pressures of particular species that can be allowed. As has been said, most processes use vacuum simply because it provides an easily controlled environment for the process and it is really what can and cannot be tolerated in that environment which is important.

For any given process, the user may have to determine by tests what cleaning is necessary to obtain the desired environment.

As has been discussed above, one very common requirement is to reduce hydrocarbon contamination to a minimum. This will often best be achieved by washing in a hot organic solvent such as trichloroethylene or perchloroethylene (where use of these substances, both chlorinated hydrocarbons, is permitted) followed by washing in hot clean demineralised water and a vacuum bake.

A general cleaning recipe that works in many circumstances is given in Section 7.1.

## 5. HOW CONSTRUCTION AFFECTS CLEANING

In order to clean a vacuum component to the highest standards — e.g. capable of working at UHV with very low levels of contaminant species in the residual vacuum — at the design stage careful consideration must be given to how the item is to be cleaned. In particular, crevices, blind holes, cracks, trapped volumes, etc., should be avoided as these will act as dirt and solvent traps. It can be very difficult to remove both dirt and solvent from such areas. Fortunately, good vacuum practice regarding trapped volumes will also result in a component that avoids these problems.

A good design working at lower vacuum levels will also seek to remove any such traps.

One component that must be cleaned with particular care is thin-walled edge-welded bellows often used for motions in vacuum. Care must be taken that the cleaning process does not cause particles to be left in the convolutions, since these can puncture the bellows when it is compressed.

Alkaline degreasers can be particularly prone to this as they often deposit precipitates. During cleaning, the bellows should be fully extended and a careful final wash with demineralised water and a blow dry with hot compressed air, both inside and out, must be done.

If the process uses chlorinated hydrocarbon solvents then these must be removed completely by heating, as any left behind can corrode the bellows, leading to leaks. This last point is particularly true for accelerators, where the atmosphere in accelerator tunnels can often be warm and humid. The radiation environment also promotes enhanced corrosion, leading to premature failure of the welds.

## 6. CHEMICAL CLEANING AGENTS

A chemical cleaning agent is simply any substance that might lead to removal of an unwanted contaminant. They will generally be liquids of some sort and will work either by dissolving or by reacting with the contaminant or by removing the surface layers of the substrate and hence liberating the contaminant.

Care has to be exercised that the cleaning agent itself does not introduce contaminants. This will obviously depend on the end use of the item being cleaned. For example, as noted above, hydrocarbons are often a serious contaminant in a vacuum system. However, in mineral oil sealed mechanical vacuum pumps, e.g. rotary pumps, this is not the case as much of the mechanism is of necessity bathed in hydrocarbons. Therefore a major manufacturer of such pumps uses a hydrocarbon mix to clean the piece parts for his pumps and enjoys the benefit of residual lubrication and corrosion protection on parts not normally lubricated by the pump fluid. In most cases, however, what is desired is an agent that will remove contamination from the item without recontaminating it in some way. The operator has therefore to ensure that the solvent is changed regularly before the build up of contamination in solution becomes a problem. Examples of some typical cleaning agents are shown in Table 1. The table should be read with caution and is to be regarded as a starting point only. In all cases, manufacturers data sheets should be consulted and expert technical advice sought before using agents of this type.

Table 1
Some typical cleaning agents

| Agent | Examples | Advantages | Disadvantages | Disposal |
|---|---|---|---|---|
| Water | | Cheap; readily available | Need to use demineralised for cleanliness. Not a strong solvent | To foul drain. |
| Alcohols | Ethanol, methanol, iso-propanol | Relatively cheap, readily available, quite good solvents. | Need control – affect workers, some poisonous, some flammable, stringent safety precautions. | Evaporate or controlled disposal. |
| Organic solvents | Acetone, ether, benzene | Good solvents, evaporate easily with low residue. | Either highly flammable or carcinogenic. | Usually evaporate! |
| Chlorofluoro-carbons (CFCs) | Freon™ (CFC-113) | Excellent solvents, evaporate easily with low residue. | Banned! | Strictly controlled – must not be allowed to evaporate. |
| Chlorinated hydro-carbons | Methyl chloroform; trichloroethylene ("Trike") | Excellent solvents, non toxic, low boiling point, low residue. | Methyl chloroform banned. Trike may be banned. Toxic, require stringent safety precautions. | Strictly controlled. |
| Detergents | | Aqueous solutions, non toxic, cheap and readily available, moderate solvents. | Require careful washing and drying of components. Can leave residues. | To foul drain after dilution. |
| Alkaline degreasers | Almeco™, sodium hydroxide. | Aqueous solutions, non toxic, moderate solvents. | Can leave residues and may deposit particulate precipitates. | Requires neutralisa-tion, then dilution to foul drain. |
| Citric acid | Citrinox™ | Cheap and readily available, quite good solvents. | Require careful washing and drying of components. Can leave residues. Unpleasant smell. | To foul drain after dilution. |

The majority of these agents are solvents, which dissolve the contamination present on a surface. Some rely more on a simple washing action, flushing contaminant off the surface, as in a water jet. Others, e.g. detergents, work mainly by reducing the attractive (van der Waals) force between contaminant and surface atoms. Some work by chemical action e.g. by etching a thin layer off the surface thus releasing the contaminant. The alkaline degreasers and citric acid work this way, as do, to some extent, some detergents. It is important to understand how any particular agent interacts with the materials to be cleaned in order to avoid unexpected side effects.

It should be carefully noted that, in some countries, some of these substances are either banned or their use is strictly controlled under legislation or regulation. It is therefore imperative that the operator consults the relevant authorities before implementing a process involving these agents.

## 7. CHEMICAL CLEANING OF SPECIFIC MATERIALS

Clearly, any cleaning process must not damage the component being cleaned. Most accelerator vacuum systems are made from stainless steel or aluminium, and from now on we will concentrate our discussion on matters particularly relevant to accelerators. However, contained within such vacuum systems there may well be copper, titanium, beryllium, ceramics of various sorts, and various other materials. Below we will note some cleaning processes that have proved to be satisfactory for these materials.

### 7.1 Stainless steel

The procedure described below for cleaning stainless steel is a very high specification process for the very demanding requirements of an electron storage ring where cleanliness is of paramount importance. For less demanding applications, the procedure could stop at the appropriate point in the procedure where requirements had been met.

- Remove all debris such as swarf by physical means such as blowing out with a high pressure air line, observing normal safety precautions. Remove gross contamination by washing out, swabbing or rinsing with any general purpose solvent. Scrubbing, wire
- It must be emphasised once again, however, that it is very important that the user checks that any particular cleaning agent or process is safe for the materials to be cleaned.
- brushing, grinding, filing or other mechanically abrasive methods may not be used.
- Wash in a high pressure hot water (approx. 80°C) jet, using a simple mild alkaline detergent. Switch off detergent and continue to rinse thoroughly with water until all visible traces of detergent have been eliminated.
- If necessary, remove any scaling or deposited surface films by stripping with alumina or glass beads in a water jet in a slurry blaster.
- Wash down with a high pressure hot (approx. 80°C) water jet, with no detergent, ensuring that any residual beads are washed away. Pay particular attention to any trapped areas or crevices.
- Dry using an air blower with clean dry air, hot if possible.
- Immerse completely in an ultrasonically agitated bath of clean hot stabilised trichloroethylene for at least 15 minutes, or until the item has reached the temperature of the bath, whichever is longer.
- Vapour wash in trichloroethylene vapour for at least 15 minutes, or until the item has reached the temperature of the hot vapour, whichever is longer.
- Ensure that all solvent residues have been drained off, paying particular attention to any trapped areas, blind holes etc.

- Wash down with a high pressure hot (approx. 80°C) water jet, using clean demineralised water. Detergent must not be used at this stage.
- Immerse in a bath of hot (60°C) alkaline degreaser (P3 Almeco™ P36) with ultrasonic agitation for 5 min. After removal from the bath carry out the next step of the procedure immediately.
- Wash down with a high pressure hot (approx. 80°C) water jet, using clean demineralised water. Detergent must not be used at this stage. Ensure that any particulate deposits from the alkaline bath are washed away.
- Dry using an air blower with clean dry air, hot if possible.
- Allow to cool in a dry, dust free area. Inspect the item for signs of contamination, faulty cleaning or damage.
- Vacuum bake to 250°C for 24 hours using an oil-free pumping system.
- Reduce the temperature to 200°C and carry out an internal glow discharge using a helium/10% oxygen gas mix.
- Raise the temperature to 250°C for a further 24 hours then cool to room temperature.

### 7.2 Aluminium

The CERN specification for aluminium chambers is as follows -

- Spray with high pressure jets at 60°C with a 2% solution of Almeco 29™ (an alkaline detergent).
- Repeat with a 2% solution of Amklene D Forte™.
- Rinse thoroughly with a jet of hot demineralised water.
- Dry with hot air at 80°C.

Another procedure known to give good results is:

- Immerse in sodium hydroxide (45 gm-1 of solution) at 45°C for 1 – 2 min.
- Rinse in hot demineralised water.
- Immerse in an acid bath containing nitric acid (50% v/v) and hydrofluoric acid (3% v/v).
- Rinse in hot demineralised water.
- Dry in warm air.

### 7.3 Copper

Under most circumstances, copper can be cleaned using the same procedures as for stainless steel. However, it should be noted that some of the proprietary formulations of alkaline degreasers attack copper and leave surface stains. Organic solvents are usually all right, and cleaners based on citric acid are very good for copper. Indeed, some of these are formulated specifically for this purpose. As always, thorough washing in hot demineralised water and drying in warm, dry air should be undertaken.

Copper is particularly susceptible to surface staining, fingerprints showing up very well! Under some conditions, light surface oxidation tales place which results in a visible blackish film on the surface. As long as this is not a thick, friable or flakey deposit, it will not usually be a problem in vacuum, since it thermally disassociates quite readily.

### 7.4 Beryllium

Working with beryllium is subject to stringent safety requirements and the appropriate safety authorities must always be consulted before carrying out any such work. However, provided that care is taken to ensure that no particulates are generated and that suitable precautions are taken, components may be handled safely. Impervious gloves should always be worn when handling beryllium and any skin abrasions or cuts should be covered up.

*No* stripping, cutting, machining or abrasive operations may be carried out on beryllium except in purpose built facilities. Otherwise, beryllium may be cleaned in accordance with the procedures for stainless steel.

### 7.5 Titanium

Titanium may be successfully cleaned as for stainless steel.

### 7.6 Glass

Simple detergents and hot water washing are effective for cleaning glass.

### 7.7 Ceramics

Alumina powder in a water or an isopropanol carrier may be used to remove surface marks from ceramics such as alumina or beryllia. Baking in air to the highest temperature that the material can stand or to 1000°C, whichever is lower, is very effective for removing contamination from the surface pores of the material.

## 8. PHYSICAL CLEANING TECHNIQUES

### 8.1 Blowing

A jet of compressed air or inert gas is a useful technique to remove particulates and liquids, particularly from blind holes, as well as surface dust. A laboratory or site compressed air supply may be used provided that the air supply derives from an oil-free compressor. It should be distributed via cleaned and dried air lines and there should be particulate filtration at or near the point of delivery. Alternatively, bottled gas, e.g. nitrogen, may be used, but again the delivery line should be clean and dry.

### 8.2 Bead blasting

A jet of alumina or silica or some other inert material in the form of small beads is directed onto a surface and physically strips material away. The carrier may be either air or water. The latter is gentler, but both techniques carry the risk of driving contaminants (and beads!) into the surface. Some surface damage may occur. However this is a useful technique for stripping, for example, deposited metal films from the inside of vacuum vessels.

### 8.3 "Snow" cleaning

This technique uses a jet of pellets of solid carbon dioxide that is directed onto the surface. This is a non-impact technique that is good for removing particulates and also apparently some hydrocarbon films. It is however, expensive and noisy.

Unlike bead blasting, the technique does not damage the surfaces being cleaned, since the solid $CO_2$ does not impact the surface. As the particles, usually in the form of small cylindrical pellets, approach the surface at high speed, the shock wave travelling ahead of the particle is reflected from the surface and interacts with the pellets, which sublime. The cleaning action is effected through a cavitation process in the compressed gas as it reflects from the surface.

### 8.4 Cutting and grinding

Cutting and grinding are used to remove surface layers. Both have their place provided they are done with care. Each technique usually requires a lubricant between tool and workpiece, which for vacuum use is best to be either water or water based, although a good alternative is alcohol. If possible dry cutting is best.

There is a danger of driving contaminants, cutting fluids and swarf into the surface where they can remain trapped.

### 8.5 Polishing

Polishing is a gentle form of grinding and carries the same dangers. However fine hand polishing is often required to remove surface blemishes. A useful material for this is ScotchBrite™ that is essentially fine alumina in a polymeric carrier in the form of a loose weave mat. Diamond powder wetted with alcohol and applied on a lint free pad is also acceptable. Polishing involving pastes and waxes is to be avoided

Small components are often polished in a tumble polisher in that they are immersed in a bath of small wet stones or steel pellets of various shapes which is gently stirred or tumbled.

## 9. PASSIVATION

As far as cleaning is concerned, passivation techniques are essentially ones that prevent the adsorption of contaminants into the surface of a vacuum system or prevent the permeation of gas from the bulk material into the vacuum system. By and large, these involve the formation of a barrier layer of some sort on the surface.

Some examples are

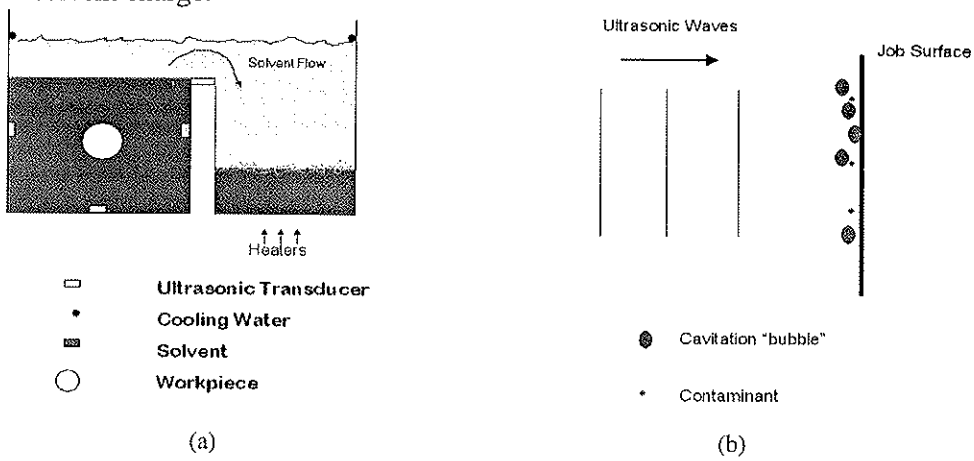
- oxide films, produced by air baking, glow discharge or other techniques.
- coatings, such as TiN or BN put on the surface by sputtering. These have not been much used except in demonstration systems and will not be discussed further here.
- active films, like getters or NEG. These are discussed in other chapters of these proceedings.

## 10. SPECIAL CLEANING TECHNIQUES

### 10.1 Ultrasonic cleaning

This is essentially a method of enhancing a chemical cleaning process. Figure 2 shows the principles involved. The figure shows a two stage cleaning plant (Fig. 2(a)). The workpiece is first suspended in the hot liquid (darker shade) in the left-hand tank. The ultrasonic transducers set up a pattern of ultrasonic waves in the liquid. As these waves reflect off the surface of the workpiece, interference takes place and a series of cavitation bubbles are generated (Fig. 2(b)). Collapse of these bubbles results in contamination particles being dragged off the surface. The contamination then either dissolves in the liquid or, if insoluble, eventually drops to the bottom of the tank as sediment. The result is an enhanced cleaning process. As jobs are cleaned, the liquid in the ultrasonic tank gradually becomes more contaminated and eventually some of this remains on the surface of the workpiece, or indeed may be transferred to it from the liquid. To ensure full cleanliness, the workpiece is withdrawn from the liquid and is suspended in the second (right-hand) tank in the solvent vapour (lighter shade). This vapour is generated by the boiling liquid solvent at the bottom of the vapour stage tank. The vapour condenses on the workpiece so that it is washed in the pure liquid solvent condensate that drops back into the boiling liquid, taking the residual (soluble) contamination with it. At the top of both baths are cooling coils that condense the vapour before it can escape from the open top of the tank. These coils are arranged so that the liquid runs back into the ultrasonic stage. This pure liquid solvent tops up the ultrasonic tank and contaminated solvent flows over the dam into the boiler. Thus a distillation process is set up with the contamination gradually being concentrated in the bottom of

the vapour stage and the liquid in the ultrasonic tank remaining relatively pure, hence enhancing the life of the solvent charge.




## 10.2 Electropolishing

Electropolishing has often been assumed to reduce outgassing by reduction of the surface area presented to a vacuum system. In practice, for any technological metal surface, the real surface area is very much larger (factors of several) than the physical area, because on a macroscopic scale there will be many pits, grooves, cracks, grain boundaries and other defects. Although electropolishing does indeed remove surface asperities and smoothes the edges of cracks, the actual benefit achieved is not very great. However, electropolishing does remove the amorphous surface layer that is formed when polishing a metal surface (the so-called Beilby layer). It replaces this layer with a relatively well ordered surface oxide layer, which may have barrier properties preventing diffusion (particularly of hydrogen) out from the bulk of the metal into the vacuum.

On the other hand, electropolishing may introduce hydrogen and other contaminants into the surface layers in significant quantities. These can subsequently diffuse out and actually increase the outgassing rate over that of the starting material. Electropolished components require a good bake (preferably at 450°C) to thoroughly outgas the surface in order to see any real benefit.

Despite many studies over a long time, it is only in some recent very careful work on aluminium prepared by various techniques that a correlation between surface roughness and outgassing rate has been demonstrated and even in these results there are some anomalies (Fig. 3).

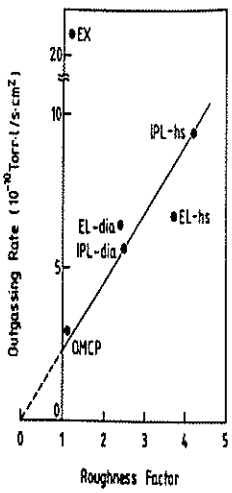



Fig. 3 Outgassing rate as a function of surface roughness.

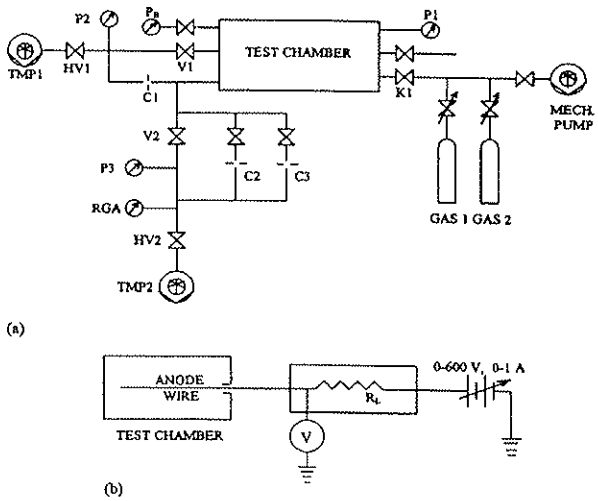



Fig. 4 Schematic of glow discharge apparatus [2]

## 10.3 Glow discharge

Glow-discharge cleaning is an effective final cleaning process which reduces outgassing and desorption rates. It achieves this essentially by three mechanisms

- The bombardment of a surface by medium energy ions (a few hundred eV) directly desorbs gas adsorbed on the surface and absorbed in the sub-surface region.
- If the discharge gas contains oxygen (or water), oxygen ions ($O_2^+$ and $O^+$) are created in the discharge. These can react with the adventitious carbon which is always present on a surface to form CO and $CO_2$ which can be pumped away. Carbon overlayers, usually graphitic or amorphous in character, seem to act as gas reservoirs on a surface.
- A good, well ordered oxide film is generated on the surface which acts as a diffusion barrier.

For accelerators, it is normal to use a simple dc glow discharge, although an ac (rf) discharge may also be used. Figure 4 shows a typical set-up for dc discharges. A gas is admitted to the vessel to be cleaned and a positive dc potential of a few hundred volts is applied to a rod or wire electrode along the axis of the vessel. If the gas pressure is of the order of a few tenths of a mbar (dependent on the actual gas being used and the precise geometry of the vessel) a discharge will be struck and a visible glow seen. The gas molecules are ionised in this discharge and are accelerated to the grounded walls of the vessel which they strike with moderate energies. It is advantageous to stream the gas along the vessel, admitting it at one end and pumping at the other. At these pressures the flow is viscous and this helps to sweep the desorbed gases from the system.

Various gas mixtures have been used. Ar/10% $O_2$ is the traditional mix and is very efficient; pure $O_2$, pure $N_2$, pure $H_2$ and He/10%$O_2$ have also been used effectively. There are a number of criteria which are used in selecting the gas. Of critical importance is whether sputtering of the metal surface is a problem or not. If there are insulators present, then sputtering can cause conducting films to be deposited over their surfaces if they are exposed to the sputtered atoms. Oxygen in the mix seems to help alleviate this by forming thermally unstable metal oxides that are removed from the surface in the discharge. Lighter species sputter less efficiently and so help obviate the problem. Also, heavier gas ions can become buried in the metal surface and can diffuse out over time. This may or may not be a problem depending on the application. Carrying out the discharge at elevated temperature (e.g. 200°C) helps minimise such accumulation. Thus the discharge can be conveniently carried out during a bake cycle.

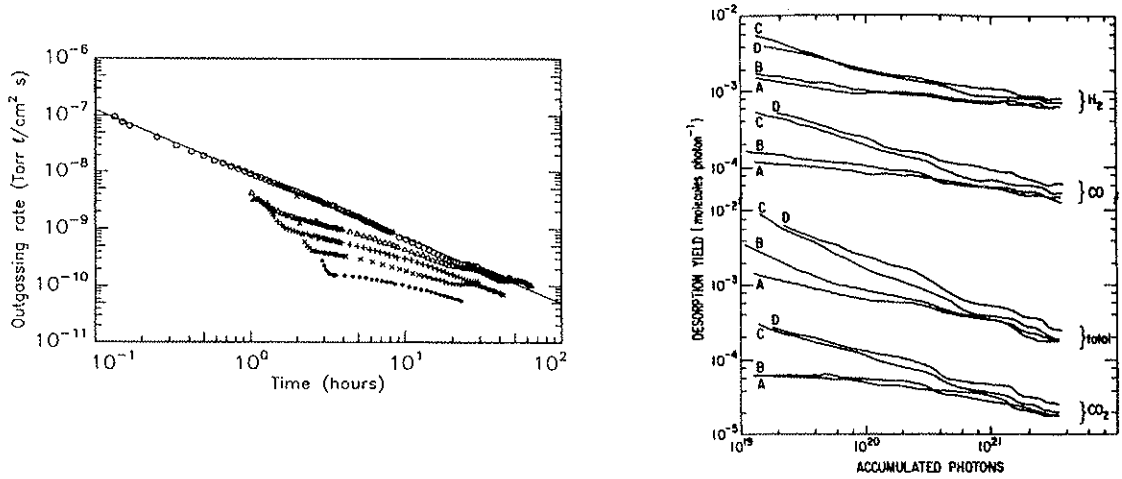
It is helpful to know when to terminate the glow discharge. If, as has been suggested above, the primary benefit is achieved by the removal of adventitious carbon by forming CO and $CO_2$, then monitoring the concentration of $C^{12}$ in the discharge exhaust gas with a residual gas analyser will give a termination point when this concentration ceases to reduce. For particle accelerators, this procedure is usually carried out as part of the pre-installation processing programme, since *in situ* glow discharge is difficult in practice.

The benefits of glow-discharge cleaning for particle accelerators are by no means fully established. For proton machines fabricated from stainless steel the benefits are well established. For electron machines where the beam space tends to be surrounded by copper, the evidence is not well in place, although it is believed to be beneficial. A definitive experiment is planned for ANKA in the near future. Figure 5 shows data from two laboratories showing the effect of glow discharge cleaning stainless steel on, respectively, outgassing rates and photodesorption yields.

## 10.4 Air baking

This procedure involves baking vacuum components at normal bake temperatures (250°C) in air. It reduces outgassing rates considerably, particularly for hydrogen. It was thought to work by inducing a barrier oxide layer, but may simply work by depleting the hydrogen from the bulk.

The technique is much favoured in the gravitational wave detector community but its value for accelerators is unproven.




(a) (b)

Fig. 5 (a) The effect of glow discharge cleaning on total outgassing rates. Between each curve (from top to bottom) is a further "dose" of *in-situ* glow discharge. (b) The effects of various cleaning treatments on partial photo-desorption yields as a function of photon dose. We simply note here that the curves labelled "A" are after a glow discharge [3].

### 10.5 Chemical "baking"

An interesting idea originating in Japan is to use a chemical which has a particular affinity for water to react with the water adsorbed on a surface in such a way that the reaction products are gases which may be pumped away.

One such chemical is dichloropropane where the reaction is

$$(CH_3)_2CCl_2 + H_2O \rightarrow (CH_3)_2C{=}O + 2HCl$$

This technique is claimed [4] to reduce the base pressure in a vacuum system by factors of 80.

However, the technique has to be carried out at a moderate temperature (80°C) so it is unclear what advantages this technique has in practice over a simple bake.

## 11. SAFETY

Safety is of paramount importance when choosing a cleaning process. For example, many processes use very hot liquids and/or large baths of chemicals.

Organic solvents, in particular, can be

- toxic (poisonous)
- flammable
- environmentally dangerous (e.g. ozone depleters).

When such solvents are used, it must be in properly designed plant suitable for the task. Vapour extraction and recovery systems must be installed. If the vapour is toxic or otherwise injurious to health, personnel should wear proper protective clothing and equipment, as well as individual monitors measuring their exposure to the vapour in the atmosphere. Alarms linked to atmospheric sampling monitors must be fitted near the cleaning plant. Breathing apparatus will be required when changing solvents and cleaning tanks.

If the solvent is flammable, then proper fire precautions will be required, including intrinsically safe electrical supplies.

When acids and alkalis are being used, neutralising pits and emergency neutralisation facilities and spill containment measures will be required.

Special buildings with controlled access may be required and rigorous operator training is essential.

## 12. COST

Cleaning to high standards can be costly. For large vessels and pieces of equipment, large cleaning baths are required, often containing many thousands of litres of expensive cleaning fluids that need to be changed regularly. Lifting and handling equipment that does not introduce contamination into the cleaning fluids, e.g. through oil drips, will be required.

The safety precautions outlined in Section 11 are expensive. It is therefore very important to clean only to the standard required and not to over specify a cleaning process as this will be uneconomic and wasteful.

## 13. HOW DO WE KNOW A SURFACE IS CLEAN?

There are two different approaches to determining whether or not a surface is clean. The more fundamental one is to perform a chemical analysis of some sort on the surface to determine what is there. Techniques are readily available to do this and the more common of these are discussed below. However, this does beg the question of what we actually mean by a clean surface. In much of surface physics and chemistry the answer is relatively straightforward since, in general terms, the experimenter will know what the "native" material is, be it a single crystal metal, an amorphous semiconductor film, or whatever. Anything else is "contamination" so the surface is "not clean". However, when dealing with the real technological surfaces of vacuum vessels or components where the macrostructure of the surface is complex (as noted above in Section 10.2) it is by no means obvious what actually should be there. For example, in a low-carbon steel, should surface carbon be regarded as part of the alloy or is it a contaminant? If there is excess chromium at the surface of a stainless steel over the bulk alloy concentration, is it a contaminant or is it an essential part of surface passivation?

Hence such analyses may raise more questions than they solve. In any case, they may or may not affect the controlled atmosphere we are seeking to obtain by using vacuum in the first place.

The second approach is more pragmatic and simply seeks to measure the properties of interest: if we want low pressures, measure outgassing rates; if we want low partial pressures of particular species in the residual atmosphere, measure them; if we want low desorption yields, measure them. To study the effects of cleaning treatments, one should measure in a systematic way how the properties of interest change. Unfortunately, few such systematic studies have been undertaken and published, so there are poor statistics on the efficacy of cleaning processes. This is why results in the literature are not definitive and why there are so many tenaciously defended but different recipes in the folklore.

## 14. SURFACE ANALYTICAL TECHNIQUES

### 14.1 Auger Electron Spectroscopy (AES)

AES is a surface sensitive technique that detects all elements except hydrogen and helium on a surface. The fundamental mechanism is illustrated in Fig. 6. An incoming electron or photon ejects an electron from a core level of an atom to create a hole. This hole is filled by an electron falling from a higher energy level. The energy thus released is transferred to a third electron (the Auger electron) which is ejected into the vacuum.

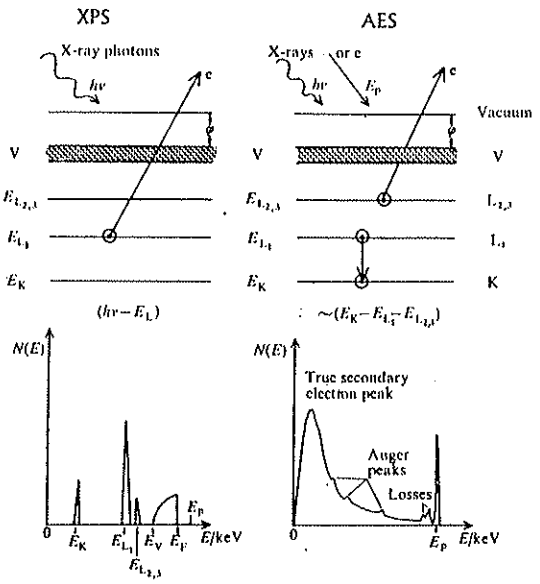



Fig. 6 Principles of two types of electron spectroscopy – XPS and AES [5]

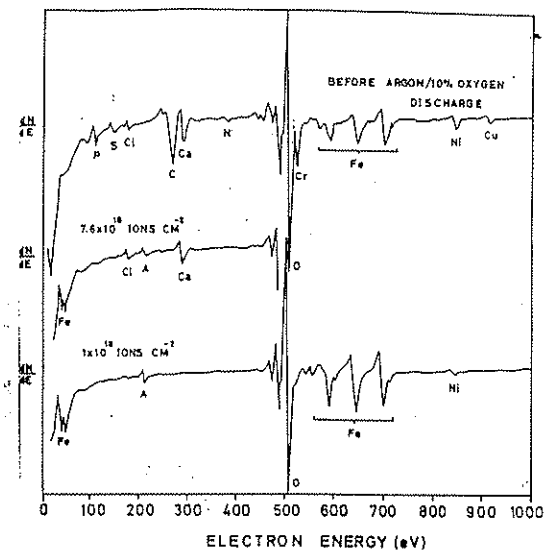



Fig. 7 Auger electron spectra from stainless steel before and after glow discharge. Note how impurities such as carbon, sulphur and chlorine disappear, whereas the native iron peaks increase. Note also the implanted argon after the discharge [6].

The energies of Auger electrons are essentially fixed and depend only on the ejecting atom. Thus measuring a spectrum of such ejected electrons from a solid, when bombarded by electrons or photons, can perform a chemical analysis of the solid. Since the Auger electrons are at relatively low energy (less than a few hundred eV) they can only escape from the surface layers of the solid without being scattered and losing energy, so the technique is surface composition sensitive. In practice the energies of the Auger electrons and the detailed shapes of the spectral lines vary slightly dependent on the chemical state of the emitting atom. Thus for example, it is perfectly straightforward to differentiate between a layer of graphite-like carbon, amorphous carbon or diamond-like carbon on a surface.

The technique is quantitative and has a surface compositional sensitivity of about 0.1%, dependent on the species concerned. Typical spectra, showing the effect of glow discharge on a stainless steel surface, are shown in Fig. 7. Note that these spectra are differentiated so that the spectral lines show up more easily against the background. AES is essentially a small-area sampling technique which cannot easily be used *in-situ* so tests are normally carried out on small coupon samples.

## 14.2 X Ray Photoelectron Spectroscopy (XPS)

This technique is similar to AES, but uses X-rays that penetrate somewhat deeper into the surface, although the emitted electrons may come only from the surface layers, dependent on their energy. The mechanism is again shown in Fig. 6. Note that here what is detected is the first ejected electron i.e. the one which leaves the core hole behind. Subsequent relaxation can then be via the Auger process, so both types of spectroscopy can be carried out simultaneously. The energies of the photoelectrons are at a fixed value below the energy of the incoming photon, the difference being characteristic of the emitting species. The technique is quantitative, in principle being more sensitive than AES, although less surface specific. It is therefore difficult to quote a sensitivity figure for surface analysis. It provides much richer information on the chemical state of the emitter than AES. Again, it is a sampling technique.

### 14.3 Secondary Ion Mass Spectroscopy (SIMS)

This technique uses a beam of ions (often argon) to knock atoms off a surface. These are then detected as ions in a mass spectrometer. Both positive and negative ions are produced and, in principle, sputtered neutrals can also be detected. It is very sensitive but produces an abundance of information that is very difficult to interpret. It is semi-quantitative. The technique is useful for detecting compositional differences between different parts of a surface e.g. across a contamination patch. It is surface sensitive, provided a low energy and low intensity incident beam is used. Prolonged exposures can drill down through the surface layers and provide information on changes in composition with depth.

### 14.4 Laser Ionisation Mass Spectroscopy (LIMA)

This technique is similar to SIMS but uses a high intensity laser beam as the primary source of excitation. Ablation rates are generally higher than for SIMS so the surface sensitivity is lower.

## 15. PHENOMENOLOGICAL METHODS

As has been noted above, these essentially rely on measuring what one needs to know in order to achieve the desired result.

For vacuum purposes, these will essentially be measurements of total and partial outgassing rates or stimulated desorption yields, using photons, electrons or ions as appropriate. All of these techniques are discussed in detail elsewhere in these proceedings.

Although from a practical point of view it is difficult to apply these techniques quantitatively, by using standard methods, preferably in a single experimental station, it is possible to produce accurate *comparisons* of the efficacy of various cleaning techniques. Hence trends can be well established. It should be noted that it is very difficult to make meaningful quantitative comparisons between work carried out using different techniques and even between work in different laboratories using the same technique.

## 16. RECONTAMINATION

Once something has been cleaned, it may or may not remain clean! Care has to be exercised to ensure that it does not become recontaminated. The precautions required will, of course, be dependent on the degree of cleanliness required.

Some points to note are

- A vacuum surface should never be touched with bare skin. Clean, lint free gloves should always be worn and arms, etc., covered up
- The operator should take care that he does not to drop bits of himself/herself into the system, e.g. by wearing a hat or hood covering up hair and possibly the mouth and nose.
- All tools should be clean. Not just the business end, but also the handles, so that they do not transfer grease onto clean gloves.
- Smoking and the use of internal combustion engines should be prohibited near an open vacuum system.
- The gas used to let up the vacuum system should be clean.
- The gas should also be dry. Figure 8 illustrates the change in outgassing rate achieved by using nitrogen containing various concentrations of water vapour to let up a vacuum system.
- Nothing should be left in the system that should not be there, e.g. tools or rags.
- Vacuum items should be stored and transported under vacuum or dry nitrogen rather than being left open to the elements after cleaning

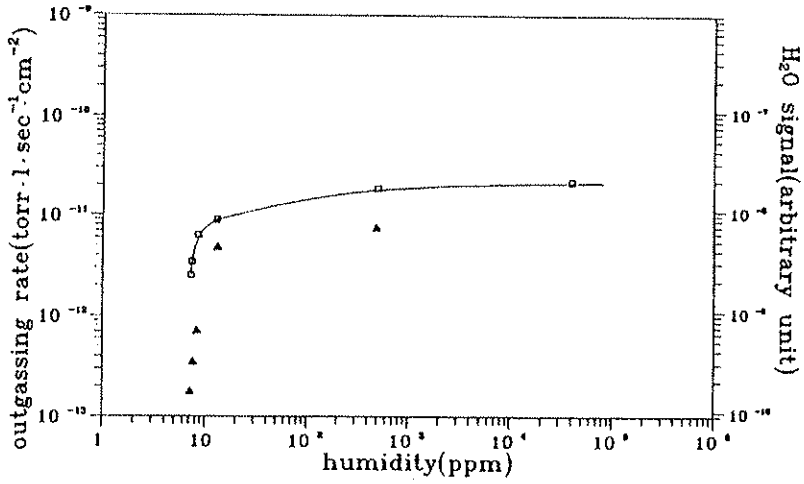



Fig. 8 The outgassing rate and water partial pressure in a vacuum system after letting up with nitrogen containing varying concentrations of water vapour [7].

## 17. SUMMARY

This paper has summarised briefly various techniques for cleaning vacuum systems and components and indicated what factors are of importance in choosing which should be used in particular circumstances. It has also discussed how one might be confident that a vacuum system is indeed clean enough for the purposes to which it might be put.

Nevertheless it must be strongly emphasised that it is only the users of any vacuum system who can determine what is required and if it has been achieved, so it is imperative that in any given situation suitable tests are carried out to establish that confidence.

## ACKNOWLEDGEMENTS

The author wishes to acknowledge the work of his colleagues at Daresbury Laboratory who operate the vacuum cleaning and processing facilities there, and in particular the work of Mr. Joseph D. Herbert who has carried out experimental measurements to establish the efficacy of many cleaning processes. He would also like to acknowledge many fruitful discussions with accelerator and other vacuum experts around the world on this topic. In particular he and the whole accelerator community owe a large debt of gratitude to the late Dr A.G. Mathewson whose work on fundamental aspects of surface cleanliness and photodesorption underpins much of our understanding in this field.

## REFERENCES

[1] Suemitsu et al. J. Vac. Sci. Technol. **A10**, 1992, 570.

[2] Li and Dylla, J. Vac. Sci. Technol. **A13**, 1995, 572.

[3] Hseuh and Cui, J. Vac. Sci. Technol. **A7**, 1989, 2418.

[4] Tatenuma et al, J. Vac. Sci. Technol. **A16**, 1998, 263.

[5] M. Prutton, Surface Physics, Clarendon Press, Oxford, 1975.

[6] A. Mathewson, CERN-LEP-VA/87-63.

[7] Liu et al, Vacuum **44**, 1995, 435.

## SELECTIVE BIBLIOGRAPHY

The following list of references is extremely selective, but will provide the interested reader with sufficient further references to pursue.


### Cleaning

J.D. Herbert, A.E. Groome, R.J. Reid, J. Vac. Sci. Technol. **A15**, 1994, 1767.

J.D. Herbert, R.J. Reid, Vacuum **47**, 1996, 693.

C. Benvenuti *et al*, Vacuum **53**, 1999, 317.

### Glow Discharge Cleaning

Li Minxu, H.F Dylla, J. Vac. Sci. Technol. **A13**, 1995, 571.

### Surface Science

J.M. Walls and R. Smith, Vacuum **45**, 1994: Special Issue "*Surface Science Techniques*".

M. Prutton, *Introduction to Surface Physics*, Oxford Science Publications, Oxford, 1994.


Procedures and specifications used at Daresbury Laboratory will be found on the World Wide Web at

<http://accelerator.dl.ac.uk/vacuum/docs.html>

These may be freely quoted and used, provided that due acknowledgement is made of their origin.

# BEAM-RESIDUAL GAS INTERACTIONS

*Søren Pape Møller*
ISA, Aarhus University, Denmark

**Abstract**
The interactions between a beam of charged particles and the residual gas will be described. In principle we should cover beams consisting of any type of charged particles. Clearly, this is an impossible task, but most types are included. Also, only the most important interactions are described. The interactions between the beam and the residual gas will lead to a reduction in the lifetime, and possibly an increase of the beam emittance. Knowing the cross section for the interaction in question, these effects can be calculated. Finally, the effects these interactions will have on the surroundings in terms of radiation and background will be mentioned briefly.

## 1. INTRODUCTION

The performance of accelerators and storage rings depends on the many components of the accelerator, and one very important component is the vacuum system. Interactions between the accelerated particles and the residual-gas atoms may degrade the beam quality. The lifetime may be reduced and/or the emittance may increase. Ideally, the aim is to obtain an average pressure so low, that the effects from the beam-residual gas interactions may be neglected altogether in comparison with other effects, for example beam-beam effects. A finite pressure can also lead to losses in transfer beam lines, but the requirements to the pressure in such beam lines are usually much more relaxed than in cyclic accelerators.

I will describe the interactions between the beam particles and the residual-gas atoms/molecules in general terms, and try to include most species. Hence both negative ions, positive ions of both low- and high-charge states, and both molecular and atomic ions are considered. Furthermore, electrons and positrons will be included. The energy range will be very wide, starting from very low energies of the order of keV and ending at very high energies in the multi-GeV region.

Emphasis is given to the physical principles, and although some exact formulae are given, crude estimates are often sufficient in order to evaluate the pressure required. Most of the vacuum related effects have been described in proceedings from previous CERN accelerator schools. Concerning electrons and positrons, see Refs. [1, 2], and for ions see [3]. As a slight addition to these treatments, the present description is more unified and formulae apply to both light and heavy projectiles.

## 2. THE CONCEPTS OF CROSS SECTIONS AND LIFETIMES

Consider a beam of intensity $I$, that is $I$ particles per cm$^2$ per second, crossing a target of thickness $dx$ with density $n$ atoms per cm$^3$. This beam will now be attenuated by the collisions, and the change in intensity $dI$ will be proportional to $I$, $n$ and $dx$

$$dI = -I\sigma n dx$$

where the constant of proportionality $\sigma$ has the dimensions of an area and is called the cross section. The cross section can be interpreted as the area of the atom inside which the considered reaction will take place; see Fig. 1. First estimates of the cross sections for the relevant processes are geometrical cross sections, which for atomic and nuclear processes would correspond to $\sigma = \pi r^2 \approx \pi(1\ \text{Å})^2 = 3\cdot10^{-16}\text{cm}^2$ and $\sigma = \pi r^2 \approx \pi(10\ \text{fm})^2 = 3\cdot10^{-24}\ \text{cm}^2$, respectively. So it is immediately seen that nuclear processes can often be neglected in comparison with atomic ones. We shall later return to more accurate estimates.

The target thickness traversed is given by $dx = vdt = \beta cdt$, where $v$ and $c$ are the velocities of the projectiles and the speed of light, respectively. The solution to the above equation is then

$$I = I_0 \exp(-t/\tau).$$

Hence the intensity will decay exponentially, with a lifetime $\tau$ given by

$$\tau = \frac{1}{v\sigma n} \quad .$$

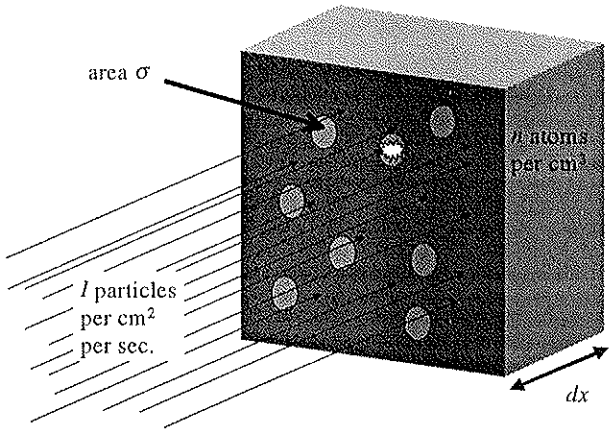



Fig. 1 Definition of cross section.

When there are more processes of comparable significance, the cross sections should be added $\sigma_{total} = \sigma_1 + \sigma_2 + ..$, or the inverse lifetimes, the decay rates, should be added $1/\tau_{total} = 1/\tau_1 + 1/\tau_2 + \ldots$ in order to find the lifetime arising from all processes. This also applies to a real residual vacuum, where several atomic species are present.

## 3. THE INTERACTION BETWEEN CHARGED PARTICLES AND RESIDUAL-GAS ATOMS

The interactions to be described subsequently depend in general on the type of projectile in question, and to some extent also on the residual-gas atom considered. The atomic number of the projectile will be designated $Z_i$ and the charge state $q$, whereas the atomic number of the rest gas atoms is called $Z_t$. The interactions will be divided into elastic processes, where no energy transfer takes place, and inelastic processes, where energy is lost by the projectile. The processes can be further classified according to the following division:

- **Interaction**
  - **Inelastic**
    - **Bremsstrahlung**
    - **Ionization energy loss**
    - **Electron capture**
    - **Electron loss**
    - **Nuclear reactions**
  - **Elastic**
    - **Single Coulomb scattering**
    - **Multiple Coulomb scattering**

Disregarding nuclear reactions and emission of bremsstrahlung, atomic collisions can be divided into two types, namely so-called electronic collisions and nuclear collisions. In the electronic collisions, the struck atoms are either excited or ionised. In the nuclear collisions, the interacting atoms recoil as a whole. The typical energy transfer is much larger for electronic than for nuclear collisions for projectile energies above a few keV. Hence, it is the electronic collisions that lead to the so-called ionisation energy loss; it is called the ionisation energy loss although atomic excitations also contribute. The nuclear collisions, on the other hand, lead to scattering of the beam particles. These scattering processes will be discussed in section 3.2 below.

### 3.1 Inelastic processes

The most important inelastic processes occurring when charged particles collide with residual-gas atoms will be described. In some special cases, ions may also interact with the black-body radiation emitted by the surrounding vacuum chambers. This interaction may lead to photo-ionisation of weakly bound negative ions [4]. This effect can only be reduced by cooling the vacuum chambers.

*3.1.1 Bremsstrahlung by electrons and positrons*

When charged particles are accelerated, they emit electromagnetic radiation, i.e. photons. This emission is only significant for light particles, and only has practical relevance for electrons (and positrons). In this respect electrons and positrons have the same radiation characteristics. When electrons and positrons penetrate matter, they are accelerated in the fields of the atomic nuclei, and electrons, and emit the so-called bremsstrahlung. This bremsstrahlung is very strong for relativistic electrons and the average energy loss due to bremsstrahlung is much larger than the ionisation energy loss to be described in the next subsection. Photons of all energies, only limited by the electrons kinetic energy $E$, can be emitted. The emission of bremsstrahlung is described by the so-called Bethe-Heitler cross section, differential in photon energy

$$\frac{d\sigma}{d\hbar\omega} = \frac{4}{3}\frac{1}{nL_0}\frac{1}{\hbar\omega}\left(1 - \frac{\hbar\omega}{E} + \frac{3}{4}\left(\frac{\hbar\omega}{E}\right)^2\right)$$

where $L_0$ is the radiation length given by

$$\frac{1}{L_0} = 4Z_t^2 \alpha n r_e^2 \ln(183 Z_t^{-1/3}).$$

Here $r_e = e^2/mc^2 = 2.82 \cdot 10^{-15}$ m is the classical electron radius, $\alpha = e^2/mc^2$ the fine structure constant and $n$ the atomic density. In the above formula contributions from scattering off the atomic nuclei have been included but a small term resulting from scattering off the electrons of the residual gas atom has been neglected. This is usually well justified due to the $Z_t^2$ front factor in the Bethe-Heitler cross section. The radiation length for $H_2$ and $CO_2$ corresponds to 7523m and 183m, respectively, at atmospheric pressure. The behaviour of this Bethe-Heitler cross section is shown in Fig. 2.

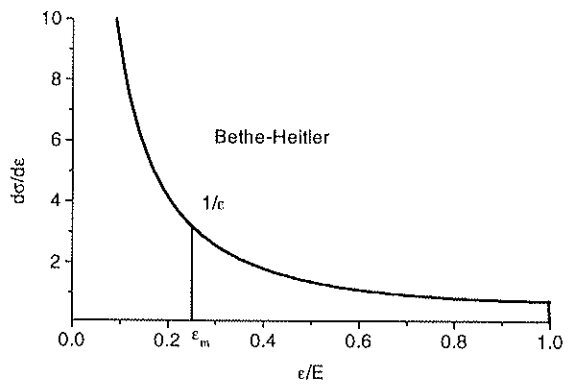



Fig. 2 Differential bremsstrahlung spectrum; the Bethe-Heitler spectrum.

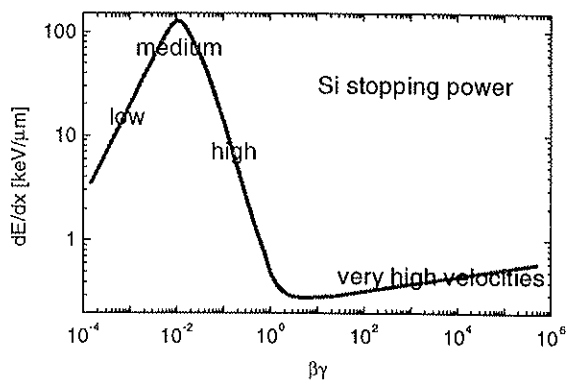



Fig. 3 Stopping power of silicon as a function of the relativistic $\beta\gamma$ factor.

When electrons are circulating in a storage ring they emit bremsstrahlung during the collisions with the residual-gas atoms, and if the photon energy is too large, the particle will be lost due to the finite momentum acceptance of the ring. This acceptance might be determined by the aperture of the vacuum chamber in a region of finite dispersion, or by the so-called dynamical aperture beyond which the transverse motion is unstable. Finally, the available RF voltage might determine the longitudinal acceptance. Hence the important cross section is the integral of the Bethe-Heitler cross section from the energy acceptance $\varepsilon_m$ to the beam energy $E$

$$\sigma_{brems} = \int_{\varepsilon_m}^{E} \frac{d\sigma}{dE} dE \approx \frac{4}{3}\frac{1}{nX_0}\left[\ln\frac{E}{\varepsilon_m} - \frac{5}{8}\right].$$

This bremsstrahlung is usually the residual-gas effect determining the lifetime of an electron beam in a storage ring. The lifetime can now be calculated and expressed in units of the radiation length, but we give here the formula for the lifetime expressed directly only in terms of the target atomic number $Z_t$

$$\frac{1}{\tau_{brems}} = \frac{16 r_e^2 Z_t^2 nc}{411} \ln\left[\frac{183}{Z_t^{1/3}}\right]\left[\ln\frac{E}{\varepsilon_m} - \frac{5}{8}\right].$$

A very strong dependence on the target atomic number is observed, which means that even a small amount of a heavy gas may have a higher influence on the lifetime than the usually dominating hydrogen content.

*3.1.2 Ionization energy loss*

The ionisation energy loss is characterised by the so-called stopping power, the energy loss per unit distance. The velocity-dependence of the stopping power for singly-charged particles is given in Fig. 3. This figure shows the stopping power of silicon, which normally is not relevant for vacuum effects. The behaviour is, however, qualitatively, and also to a large extent quantitatively independent of atomic number. Clearly, the stopping power is proportional to atomic density. Actually, the figure shows the stopping power as function of $\beta\gamma$, which is the particle momentum divided by the rest mass. As usual, $\beta = v/c$ is the particle velocity relative to the velocity of light, and $\gamma = (1-\beta^2)^{-1/2}$. This quantity $\beta\gamma$ is a very convenient parameter, which equals $\beta$ when $\beta\gamma << 1$ and $\gamma$ when $\beta\gamma >> 1$. For small velocities, the stopping power increases linearly with velocity until it reaches a maximum. This maximum occurs at a velocity corresponding to typical target electron velocities, where energy is easily transferred to the electrons. The corresponding proton energy is around 100 keV. Above the maximum, the stopping power decreases inversely proportional to the velocity squared, until a shallow minimum is reached at around $\beta\gamma \approx 1$. For even higher velocities, the stopping power has a logarithmic increase, the so-called relativistic rise. Above the maximum, the stopping power can be described by the famous Bethe-Bloch formula with corrections

$$-\frac{dE}{dx} = \frac{4\pi Z_i^2 Z_t e^4}{mv^2} n\left\{\ln\left(2mc^2\beta^2\gamma^2 / I\right) - C/Z_t - \beta^2 - \delta/2\right\}.$$

The velocity dependence described above is directly seen from this formula. The parameter describing the atomic properties of the residual-gas atoms is the so-called mean ionisation potential $I$, which is given approximately by $I = Z_t \cdot 14$ eV. The shell-correction $C/Z_t$ is usually a small correction taking into account that the atomic electrons have a finite velocity. The density effect $\delta$ is a relativistic correction saturating the distant atomic collisions at high values of $\beta\gamma$ due to polarisation of the atomic electrons. The Bethe-Bloch formula can be written in a form more suitable to estimate pressure-dependent ionisation energy-loss effects [3]

$$-\frac{dE}{dx} = K_{BB} P_t m_t Z_t q^2 \beta^{-2}\left\{11.1 - 0.9\ln Z_t + \ln(\beta^2\gamma^2)\right\}.$$

Here $P_t$ is the pressure in mbar and $m_t$ the number of atoms/molecule, and $K_{BB} = 0.0137$ eV/mbar/m.

In most cases, the ionisation energy loss is an unimportant effect masked by other effects. It can, however, be observed both for bare and almost bare nuclei at high energy [3]. Another aspect of the energy loss is that there is a spread, since it is formed by a series of discrete losses. Hence the average energy loss, and the width of the energy loss distribution, the so-called straggling, characterise the energy loss. The above two formulae describe the average energy loss. The energy-loss distribution is Gaussian at high velocities for not too thin targets.

*3.1.3 Electron capture and loss*

A positive ion can capture electrons from the residual-gas atoms during a collision. Similarly, a not fully stripped ion can loose an electron. Electron capture and loss effects cannot be described by one or a few simple formulae, since the cross section in general is a complicated function of the projectile charge $Z_i$, the projectile charge state $q$, the target atomic number $Z_t$ and the projectile velocity $\beta$, $\sigma = \sigma(Z_i,q,\beta,Z_t)$.

The charge exchange processes can be divided according to the following scheme:

$$X^{q+} + Y \rightarrow X^{(q+1)+} + Y + e$$

$$X^{-} + Y \rightarrow X + Y + e$$

$$X^{q+} + Y \rightarrow X^{(q-1)+} + Y^{+}$$

where the first two processes are electron loss by positive and negative particles, respectively, whereas the last process is electron capture. The projectile of charge $q$ is here represented by $X$ and the rest gas atom by $Y$. Clearly, multiple loss and capture processes are also possible, but they are in general much less probable than single processes.

Some scaling relations and qualitative features can, however, be given, but in general it can only be said that the relevant cross section should be found in the literature, either from measurements or from calculations. Measurements and calculations of charge exchange cross sections have been an important part of atomic physics for many decades, and the qualitative, and to a large extent also the quantitative, behaviour of the cross sections are understood today.

Qualitatively, electron-capture cross sections have a maximum at projectile velocities corresponding to the typical electron velocity of the target atom. This qualitative behaviour is given in Fig. 4. This can be easily be understood, since the captured electron has to follow the projectile, which then is the case before the collision for equivelocity particles. Similarly, for electron loss cross sections. From the same type of arguments, there exist for a given projectile with given velocity passing through a given material a so-called equilibrium charge, where the capture and loss cross sections are equal. Bohr has estimated this cross section [3] to be

$$\bar{q} \approx Z_i(1-\exp(-\frac{\beta}{\alpha Z_i^{0.67}})).$$

This equilibrium charge state only depends on the projectile velocity. From known cross sections for ions of this equilibrium charge state, the cross sections for other charge states can be estimated from simple scaling relations [3]

$$\sigma_c(q) \approx \sigma_c(\bar{q})(\tfrac{q}{\bar{q}})^a$$

$$\sigma_l(q) \approx \sigma_c(\bar{q})(\tfrac{q}{\bar{q}})^b$$

where $a \approx 4$ and $b \approx -2.3$ for charges lower than the equilibrium charge state and $a \approx 2$ and $b \approx -4$ for higher charge states. Many other scaling relations exist.

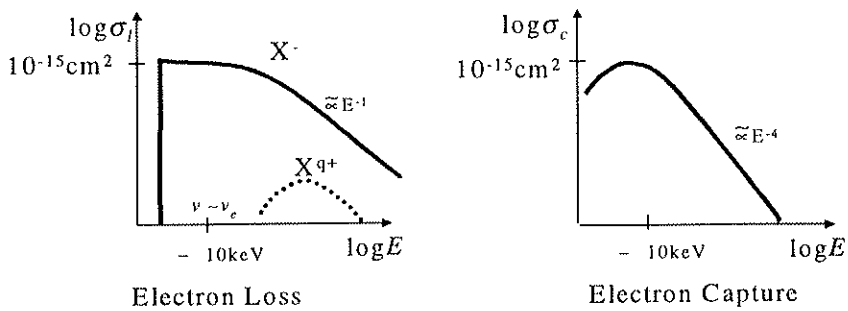



Fig. 4 Qualitative behaviour of electron loss and capture cross sections.

As an important example for a negative ion, we show in Fig. 5 the electron loss cross section for $H^-$ ions on $H_2$ molecules [5]. The cross section is seen to have a shallow maximum of around $10^{-15}$ $cm^2$ at around 100 keV, and to decrease slowly with increasing energy, slightly faster than $E^1$.

The electron capture cross section for protons, $H^+$, colliding with $H_2$ molecules is shown in Fig. 6 [5]. Again a maximum, with a value of around $10^{-15}$ $cm^2$, is observed, but with a rather fast decrease towards both lower and higher energies. The decrease for high energies is indeed very fast, roughly proportional to energy to the eighth power, $E^8$.

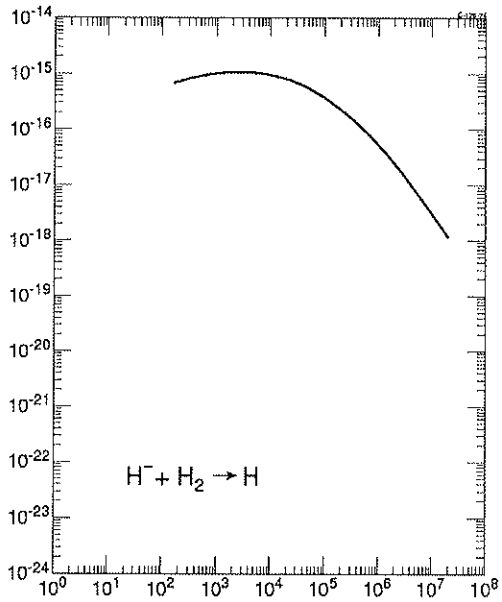



Fig. 5 Electron loss cross section for $H^-$ on $H_2$ as function of energy.

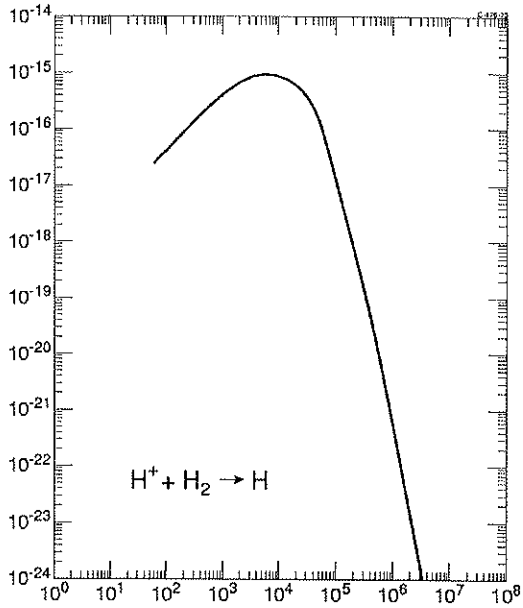



Fig 6. Electron capture cross section for $H^+$ on $H_2$ as function of energy.

Finally, as an example of an electron capture cross section for highly charged ions, we show in Fig. 7 the capture cross section as function of energy for highly charged ions colliding with argon [6]. The scaling implied by the choice of axes is observed to apply reasonably well. The decrease with energy is roughly with $E^4$. Note, that the displayed cross sections are normalised to the charge state $q$, which means that electron capture cross sections for highly charged ions can be much larger than the geometrical cross sections.

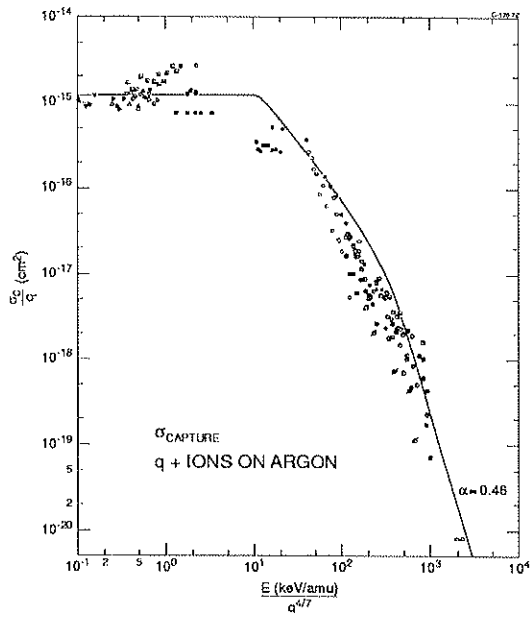



Fig. 7 Electron-capture cross section for highly-charged ions colliding with Ar atoms.

*3.1.4 Nuclear reactions*

The cross section for nuclear reactions is in general much smaller than atomic cross sections, as was also seen from the comparison between the geometrical cross sections above. There are, however, cases where a nuclear process is the dominating loss process. For example for bare nuclei at high energy, capture cross sections are very small; see above. Another important case is antiprotons, where electron capture clearly is excluded. Furthermore the energy should be so high that the elastic processes described below can be neglected.

In general, the relevant nuclear cross section has to be looked up in the literature. Some of the relevant cross sections for elementary particles can be found in [7]. The geometrical estimate of a nuclear cross section made in section 2 of around $10^{-24}$ cm$^2$ gives reasonable estimates for protons colliding with protons at beam momenta of around 100 MeV/c. At relativistic energies, above 1 GeV/c, the cross section is around a factor of 100 smaller.

## 3.2 Elastic processes

In the beginning of section 3 it was argued that the so-called nuclear collisions, where the residual-gas atom recoils as a whole, are almost elastic; i.e. there is only a very small energy transfer. There is, however, a significant scattering of the projectile associated with these collisions. If the projectile in a single-collision event is scattered outside the ring acceptance, it will be lost and lead to a reduction in the lifetime. The ring acceptance can either be physical, determined by the transverse aperture of the vacuum chambers, or it can be dynamical, determined by the boundary between stable and unstable motion. Additionally, multiple or plural scattering can lead to an increase of the beam emittance. The emittance growth can clearly be so large and fast that particles again are lost on the apertures, but a dilution of the density in phase space is in general disturbing. A slow multiple scattering can, however, be counteracted by phase-space cooling, e.g. electron or stochastic cooling or in the case of electrons by radiation damping.

*3.2.1 Single scattering*

The scattering between point-like charged particles is described by the so-called Rutherford cross section, which in the small-angle approximation reads

$$\frac{d\sigma}{d\Omega} = \left( \frac{2Z_t Z_i e^2}{m_0 c^2 \beta^2 \gamma} \right)^2 \frac{1}{\theta^4}$$

Deviations from this very simple formula exist for small angles due to electronic screening effects of the atomic nucleus and for large angles due to the finite size of the nucleus; see Fig. 8.

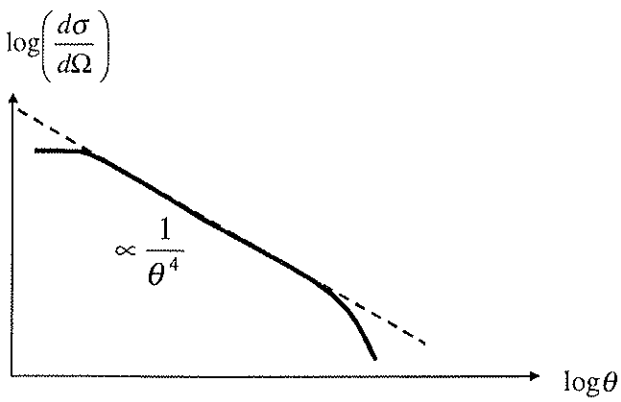



Fig. 8 Qualitative behaviour of the Rutherford scattering cross section.

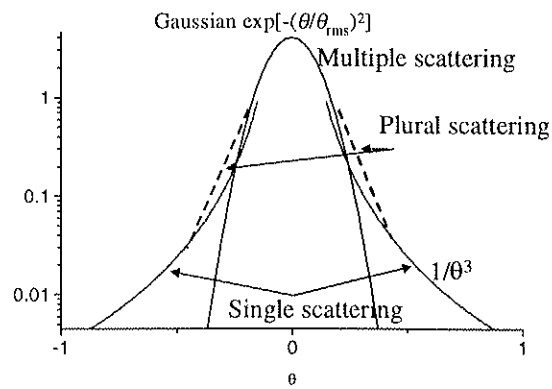



Fig. 9 Angular scattering distribution.

The relevant cross section for calculation of lifetimes is the integral of this Rutherford cross section from $\theta_0$, the minimum angle for which loss occurs, to a very large angle. This maximum angle need not be specified since the most important contribution comes from the small angles. The resulting cross section is

$$\sigma = \int_{\theta_0}^{\theta_{max}} \frac{d\sigma}{d\Omega} d\Omega = \frac{2r_p^2 (Z_i / A)^2 Z_i^2 \pi}{\beta^3 \gamma^2 \theta_0^2}$$

where $r_p$ = 1.546·10$^{-18}$ m is the classical proton radius. For electrons the projectile mass in amu is $A$ = 1/1836, which means that the formula can be written in terms of the classical electron $r_e$ = 2.82·10$^{-15}$m.

The minimum aperture gives the maximum acceptable scattering angle, and in the case of one small aperture $b$ with a betatron amplitude at this position of $\beta_u$, the angle is given by

$$\theta_0 = \frac{b}{\sqrt{\beta_u \overline{\beta}}}$$

where the betatron function $\beta$ is averaged over the whole circumference. In the case of a modern synchrotron radiation source, this aperture would be the height of the undulator vacuum chamber with smallest gap.

*3.2.2 Multiple scattering*

Many small-angle scattering events will lead to an increase of the emittance of a beam. Since this can be a relatively slow process, cooling the beam with a suitable cooling method, e.g. electron cooling or stochastic cooling can compensate it. The multiple scattering in a given thickness of material $s$ is given by

where $p$ is the momentum of the projectiles and again $L_0$ the radiation length. In a storage ring, a more

$$\sqrt{<\theta^2>} = \theta_{rms} = 14\text{mrad} Z_i \sqrt{s / L_0} / p[\text{GeV/c}]$$

interesting quantity is the emittance increase per unit time, which can be expressed by the increase in the divergence by

The increase in divergence is related to the radiation length by

$$\frac{d\varepsilon}{dt} \cong 4 < \beta_{H,V} > \frac{d<\theta^2>}{dt}.$$

From this expression, one can calculate for example the time it takes for the emittance to grow from

$$\frac{d<\theta^2>}{dt} = 2.6 \cdot 10^{-4} \frac{c}{X_0} \frac{1}{\beta^3 \gamma^2} \left( \frac{Z_i}{A_i} \right)^2 .$$

an initial to a final value

In general, there is clearly a transition between single Coulomb scattering and multiple scattering.

$$\Delta t = \frac{\varepsilon_{final} - \varepsilon_{init}}{4 < \beta_{H,V} > (d<\theta^2> / dt)}.$$

This so-called plural scattering occurs in the transition region, and the qualitative behaviour of the scattering angle distribution is given in Fig. 9.

## 4. LOCAL VERSUS AVERAGE PRESSURE

The average pressure determines in general the lifetime of a beam, and the emittance growth. There are however, exceptions to this. For example if there is a local pressure bump, the aperture determining the single-scattering lifetime calculated in subsection 3.2.1 is the aperture $(1/4+m/2)\lambda$ downstream of the pressure bump. Here $\lambda$ is the betatron wavelength and $m$ an integer. This is because a scattering will initiate a betatron oscillation, which has a maximum amplitude at these distances downstream of the scattering position.

The other important issue concerning high local pressures is the background stemming from collisions between the beam particles and the residual gas. This background can, for example, be the background in a detector in a collider, it can be the high-energy $\gamma$-radiation emerging from a straight section in a synchrotron-radiation source, or it can be the neutralised ions giving a background in a capture-experiment in a cooler synchrotron. And there are inevitably other examples, where special care has to be taken in order to minimise local pressure bumps.

## 5. CONCLUSIONS

In the present contribution to the CERN Accelerator School on Vacuum Technology the different vacuum-related processes leading to a degradation of the accelerator performance have been described. The dominating processes for different particles at different energies are outlined in Fig. 10. In several cases, the lifetimes can be estimated from the description presented, but in many cases the literature has to be consulted in order to find the relevant cross sections.

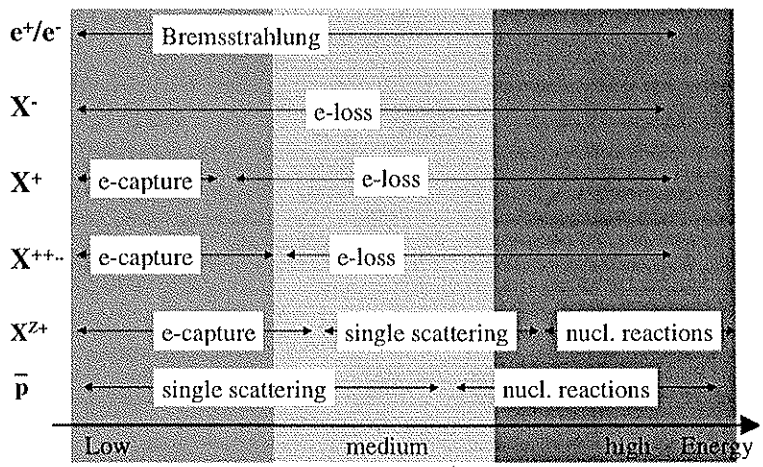



Fig. 10 The most important residual gas interactions for different projectile types in the different energy regimes.

## REFERENCES


[1] A. Wrulich, Single-beam lifetime, CERN Accelerator School, Fifth General Accelerator Physics Course, CERN 94-01, p. 409

[2] C. Bochettta, Lifetime and beam quality, CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, CERN 98-04, p. 221.

[3] B. Franzke, Interaction of stored ion beams with the residual gas, CERN Accelerator School, Fourth Advanced Accelerator Physics Course, CERN 92-01, p. 100

[4] H.K. Haugen, L.H. Andersen, T. Andersen, P. Balling, N. Hertel, P. Hvelplund, and S.P. Møller, Phys. Rev. A **46** (1992) R1.

[5] Data on collisions of hydrogen atoms and ions with atoms and molecules (I), JAERI-M, 83-013, Japan Atomic Energy Research Institute, 1983.

[6] H. Knudsen, H.K. Haugen, and P. Hvelplund, Phys. Rev. A **23** (1981) 597.

[7] R.M. Barnett, Review of Particle Physics, Phys. Rev. D **54**,1 (1996).

# ION TRAPPING, CLEARING, BEAM–ION INTERACTIONS

*A. Poncet*
CERN, Geneva, Switzerland

**Abstract**
The present lecture is based on a revised version of the one given 6 years ago at the CERN Accelerator School [1]. After a brief review of the adverse effect of neutralisation in particle storage rings, the basic topics of ion production by ionisation of the residual gas are recalled: ion production rate, natural clearing rates, ion kinematics and conditions of trapping for bunched and unbunched particle beams with positive or negative space charge. Different methods of clearing are described and their performance discussed, namely, DC clearing electrodes, empty buckets (in electron storage rings) and beam shaking. Examples of neutralisation effects and diagnostics are taken from CERN machines, such as ion clearing current measurements on electrodes, tune shifts and emittance measurements.

## INTRODUCTION

In accelerators and storage rings, ions created by the circulating particles from neutral molecules of the residual gas may be trapped in the beam-space-charge potential and may generate all sorts of ill effects: reduced beam lifetime (increased pressure), emittance growth and losses through excitation of resonances, and coherent beam instabilities. While they can occur in proton beams (e.g. CERN ISR trapping electrons), these neutralisation phenomena mainly affect machines with negative beam space charge such as electron storage rings and antiproton accumulators.

Low energy machines are more subject to ion trapping because of their small size, which leaves little space between bunches for ions to escape the beam potential, and suffer most because of their inherent high sensitivity to space-charge effects. The ion-induced space-charge tune shifts $\Delta Q$ can be unacceptably large if $\gamma$ is small (low energy) and/or the neutralisation $\eta$ is high.

## 1. NEUTRALISATION OF A BEAM: A SIMPLE DESCRIPTION

The circulating particles in a stored beam collide with residual gas molecules producing positive ions and electrons. A negatively charged beam (e.g., electrons or antiprotons) captures the ions and repels the electrons towards the vacuum chamber walls[1]. If other possible natural or artificial clearing mechanisms are not present, the neutralising ions accumulate up to the point where the remaining trapping potential is effectively zero, i.e., until the number of static neutralising particles is equal to the number of beam particles. The beam is then fully neutralised. The average neutralisation factor is defined by

$$\eta = \frac{n_i}{n_e}, \tag{1}$$

where $n_i$ is the total neutralising charge measured in units of the electronic charge and $n_e$ is the number of stored beam particles. The neutralisation is often not homogeneous along the machine azimuth $s$, and we define a local neutralisation factor by

[1] Positively charged coasting beams trap electrons, and this effect has been extensively studied in the CERN ISR (see references). Bunched proton or positron beams do not suffer from neutralisation problems because the electrons are not stably trapped (cf section 5).

$$\eta(s)=\frac{2\pi R}{n_e}\frac{dn_i}{ds}, \tag{2}$$

where $2\pi R$ *is* the machine circumference and $(dn_i/ds)$ is a local linear neutralising charge density (measured in units of electronic charge per meter)[2].

In order to get a feeling for the orders of magnitude involved in neutralisation problems, we consider a set of machine parameters corresponding to typical values for the CERN electron-positron accumulator (EPA). We disregard, for the time being, the fact that the EPA electron beam is bunched and only calculate longitudinally averaged values. Also, for the sake of simplicity, we assume a round beam with a homogeneous transverse charge distribution. The beam current is $I = 100$ mA, the energy is $E = 500$ MeV and the beam radius is 0.5 mm. The corresponding linear particle density is

$$\lambda=\frac{dn_e}{ds}=\frac{I}{e\beta c}=2\times10^9 \text{ particles/m}. \tag{3}$$

The electric field at the beam edge is obtained via Gauss's law.

$$\varepsilon=\frac{e\lambda}{2\pi\varepsilon_0 a}=1.2\times10^4 \text{ volt/m}. \tag{4}$$

The magnetic field at the edge is obtained via Ampere's law:

$$B=\frac{I\mu_0}{2\pi a}=4\times10^{-5}\text{ T}. \tag{5}$$

The total direct space-charge force on a circulating electron is

$$\vec{F}=e\left(\vec{\varepsilon}+\vec{v}\times\vec{B}\right)=\vec{F}_e+\vec{F}_m=\vec{F}_e\frac{1}{\gamma^2}, \tag{6}$$

where $\gamma$ is the total relativistic beam energy in units of the rest energy and $\vec{F}_e$ is the electrostatic force. The fact that the two forces counteract results in the so-called relativistic cancellation. With neutralisation, the electrostatic force is changed from $\vec{F}_e$ to $\vec{F}_e(1-\eta)$ and the magnetic force is unchanged. Then

$$F=F_e\left(\frac{1}{\gamma^2}-\eta\right). \tag{7}$$

The force which we have calculated at the edge of the beam is in fact proportional to the distance to the centre of the beam. The corresponding local quadrupole has a strength (with the Courant and Snyder definition [2])

$$k(s)=-\frac{1}{E}\frac{dF}{dr}. \tag{8}$$

We simplify the formulae by introducing the classical electron radius

$$r_e=\frac{e^2}{4\pi\,\varepsilon_0 m_e c^2}=2.82\times10^{-15}\text{m}. \tag{9}$$

---

[2] Some authors define an average neutralisation factor for bunched beams as the ratio of the average neutralising charge density to the bunch charge density. This is related to our definition through the bunching factor $B = (N_b L_b)/(2\pi R)$ where $N_b$ is the number of bunches and $L_b$ is the bunch length. This definition is useful when one compares the incoherent space-charge tune shift with the neutralisation induced tune shift.

and the local beam volume density

$$d_e = \frac{dn_e}{ds}\frac{1}{\pi a^2} = 2.7\times 10^5 \text{ particles/m}^3 \tag{10}$$

so that, with Eqs. (4), (7), (8), (9), and (10),

$$k(s) = -\frac{2\pi}{\gamma} r_e d_e \left(\frac{1}{\gamma^2} - \eta\right) \tag{11}$$

and the corresponding tune shift is

$$\Delta Q = \frac{1}{4\pi}\int \beta(s) k(s) ds, \tag{12}$$

where $\beta$(s) is the usual Twiss parameter and the integration is done along the machine azimuth. Instead of calculating the integral, we use average values for all quantities involved[3]:

$$\Delta Q = \frac{r_e (2\pi R)\bar{\beta}}{2\gamma} d_e \eta. \tag{13}$$

For $\bar{\beta}$ = 4 m and $2\pi R$ = 126 m, we find

$$\Delta Q \approx 2\eta \tag{14}$$

and machine performance will be limited for values of $\eta$ above a few percent.

## 2. THE IONISATION PROCESS

The circulating beam interacts with the electrons of the molecules of the residual gas and with the ions trapped in the beam. In turn the trapped ions interact with the molecules in many different ways. In the following sections, we briefly review these phenomena.

### 2.1 Transfer of energy to free electrons

An estimate can be obtained through the calculation of the electrostatic interaction between a free electron and the primary particle (Jackson [3] section 13.1). The energy transfer $E'(b)$ is a function of the impact parameter $b$ (Fig. 1).

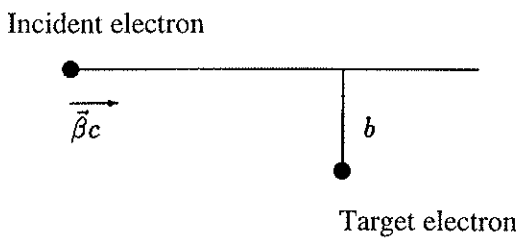



Fig. 1 The impact parameter $b$.

Equation 13.2 of Jackson, where the field at the electron is obtained by a Lorentz transformation, can be rewritten in mks units and with the classical electron radius:

$$E'(b) = \frac{2m_e c^2}{\beta^2}\frac{r_e^2}{b^2} \tag{15}$$

or

$$b^2 = r_e^2 \frac{1}{E'}\frac{2m_e c^2}{\beta^2}, \tag{16}$$

[3] We now disregard the small direct space-charge contribution.

where $\beta c$ is the velocity of the primary particle and $E'(b)$ *is* the transferred energy.

The cross-section $d\sigma$ for energy transfer between E' and E'+ dE' is

$$d\sigma = 2\pi\, b\ db \tag{17}$$

or

$$d\sigma = 2\pi \frac{m_e\, c^2}{\beta^2}\, r_e^2 \frac{dE'}{E'^2}. \tag{18}$$

One sees immediately an unphysical situation for $E' = 0$ ($b$ large, distant collisions) and $E' = \infty$ ($b$ small, close collisions). The difficulty is solved by the definition of a minimum energy $E'_{min}$ and a maximum energy $E'_{max}$.The maximum energy $E'_{max}$ can be obtained from pure kinematic considerations,[4] whereas the minimum energy requires a detailed analysis of the medium in which the interaction takes place.

The above formulae cannot be used to compute the exact ionisation cross-section, as we shall see below, but they give a fair description of the phenomenon. Detailed measurements of $d\sigma / dE$ have been made [95] and are shown in Fig. 2.

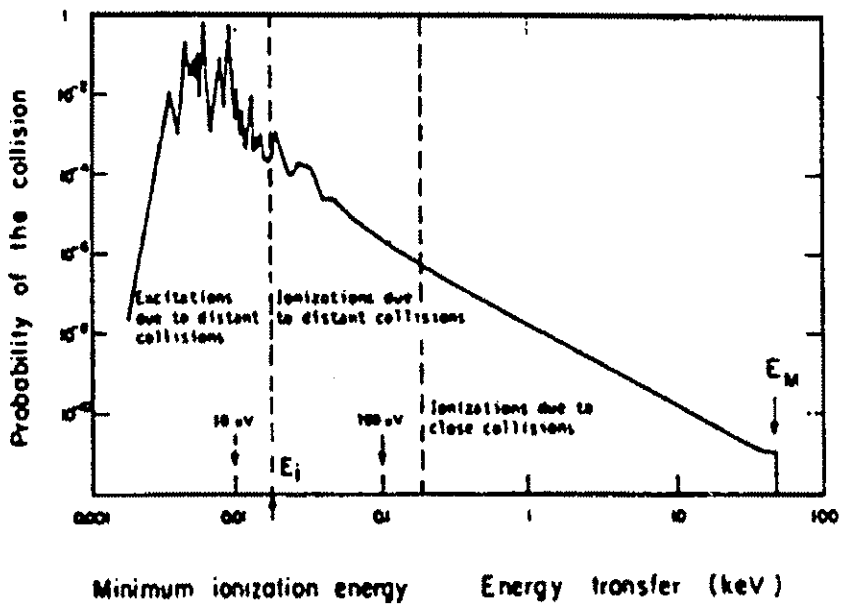



Fig. 2 Relative probability of different processes induced by fast (100 keV) electrons in water, as a function of the energy transfer in a collision [95]. The maximum kinematically allowed energy transfer, $E_{max}$ = 50 keV in this case, is also shown.

An ionisation event takes place only if the energy transferred is above the ionisation potential. Most free electrons created in such an event will be left with a rather small energy and will therefore be trapped by the beam (Fig. 2). In Ref. [96], it has been calculated that about 80% of the electrons have an energy below 45eV, the average being around a few eV. The ion energy will be smaller in the ratio of the masses and therefore negligible. These electron energies, however, should be used to calculate their drift velocities (cf. section 3) because they are several orders of magnitude higher than the thermal energy ($\simeq 10^{-2}$ eV at 300 K). This is not the case for ions. The proportion of ions not trapped because they are created with an energy larger than the potential well is negligible (less than 4% for a potential well of a few hundred volts [96]).

## 2.2 Ionisation cross-section

The ionisation cross-section depends on the molecule of the residual gas and on the velocity of the ionising particle but neither on its charge nor on its mass[5]. Measurements have been made [98] for various incident electron energies and the results were fitted to the theoretical expression by Bethe (see [4], p.45):

---

[4] For relativistic incident particles, quantum or relativistic effects further reduce the maximum energy transfer [4].
[5] Here we only consider ionisation by charged particle impact. Photo-ionisation in electron machines is analysed in ref. [97].

$$\sigma^i = 4\pi\left(\frac{\hbar}{m_e c}\right)^2\left\{M^2\left[\frac{1}{\beta_e^2}\ln\left(\frac{\beta_e^2}{1-\beta_e^2}\right)-1\right]+\frac{C}{\beta_e^2}\right\} \tag{19}$$

$$4\pi\left(\frac{\hbar}{mc}\right)^2 = 1.87\times10^{-24}\,\mathrm{m}^2\ . \tag{20}$$

The experimentally determined coefficients $C$ and $M^2$ for different molecules are shown in Table 1. Figure 3 shows a plot of the cross-sections given by the formulae above.

Table 1
Value of the $M^2$ and C constants for calculation of ionisation cross-sections

| Molecule | $M^2$ | $C$ | $Z$ | $A$ |
|---|---|---|---|---|
| $H_2$ | 0.5 | 8.1 | 2 | 2 |
| $N_2$ | 3.7 | 34.8 | 14 | 28 |
| CO | 3.7 | 35.1 | 14 | 28 |
| $O_2$ | 4.2 | 38.8 | 16 | 32 |
| $H_2O$ | 3.2 | 32.3 | 10 | 18 |
| $CO_2$ | 5.75 | 55.9 | 22 | 44 |
| $C_4H_4$ | 17.5 | 162.4 | 46 | 76 |

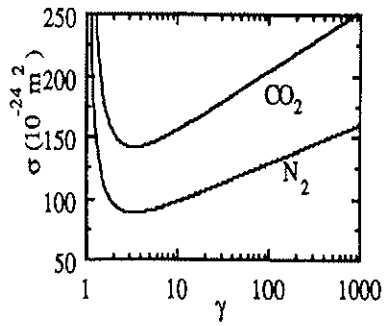



Fig. 3 Ionisation cross-section vs $\gamma$ for $CO_2$ and $N_2$

## 2.3 Ionisation rate

The time it takes for one circulating particle to create one ion is given by:

$$\tau_m = \frac{1}{d_m \sigma_m \beta c} \tag{21}$$

where $d_m$ is the molecular density ($m^{-3}$); $\tau_m$ is the ionisation cross-section for molecule $m$ ($m^2$); $\beta c$ is the velocity of the circulating beam ($m.s^{-1}$)

The molecular density $d_m$ is related to the partial pressure $P_m$ in torr by the relation (valid at 20°C).

$$d_m = 3.3\times10^{22}\,P_m \tag{22}$$

If there are several types of molecules in the residual gas then the total ionisation time $\tau_i$ is given by the relation.

$$\frac{1}{\tau_i} = \sum_m \frac{1}{\tau_m}\ .$$

In electron storage rings with typical residual pressures of 1 nTorr composed of CO and $H_2$, the ionisation time is of the order of one second or less.

## 2.4 Beam heating

Distant collisions with large impact parameter — much more probable than close ones leading to ionisation — are important, since they feed energy differentially to ions. In some circumstances (neutralisation pockets) this may be a clearing mechanism, i.e. when the trapped species get enough energy to escape the beam potential $V$:

$$R_c = \frac{1}{eV}\frac{dE}{dt} = \frac{1}{eV}\int_{b_{\min}}^{b_{\max}} \Delta E(b)\left(2\pi b\beta c N_p db\right).$$

This represents the "natural" clearing rate for a singly-charged species. The expression between brackets is the number of projectiles passing at distance $b$ to the ion target during time $dt$. $N_p$ is the projectile density of charge $Z = 1$.

To good approximation $b_{max}$ and $b_{min}$ can be chosen to have the same values as the ion and nucleus radii respectively, leading to [3]:

$$R_c = \frac{2\pi m_0 c^3 r_0^2}{\beta} \cdot \frac{N_p Z'}{eV} \cdot \ln\left(3 \cdot 10^4 \cdot Z'^{-2/3}\right) \tag{23}$$

with $m_0$ and $r_0$ being $m_e$, $r_e$ if the trapped species is an electron, $m_p$, $r_p$ for a proton; 2 $m_p$, $r_p$ for an ion of charge $Z'$.

As an example, typical clearing times for the EPA machine with $6.10^{11}$ electrons (300 mA), 1 mm beam radius, giving a beam potential of ~ 50 V, are shown in Table 2.

Table 2
EPA natural ion clearing rates

| | Clearing rate $R_c$ ($s^{-1}$) | Clearing time $\zeta_c$ (s) |
|---|---|---|
| $H^+$ | $3.10^{-3}$ | 350 |
| $H_2^+$ | $6.10^{-3}$ | 166 |
| $CO^+$ | 0.04 | 25 |
| $CO_2^+$ | 0.07 | 15 |

The process is thus slow compared to typical ionisation rates. But it may be important to explain why in some circumstances, e.g. pockets with very low gas pressure ($5.10^{-11}$ Torr, see AA case in reference [116]) fully ionised light ions can accumulate to a dangerous level, as they replace heavy ones which are chased away by beam heating.

## 2.5 Gas cooling

Seldom taken into account, gas cooling could be an important process for high pressures and long ion resting times. Charge-exchange phenomena by which a positive ion captures an electron from a gas molecule may occur at ion energies of only a few eV. The new ion is created with the primary molecule's energy, while the newly created neutral species carries away the initial ion energy. Resonant capture cross-sections between an ion and its own neutral molecule can be very high at low energy:

$$\sigma = 1.2 \times 10^{-15} \text{cm}^2 \text{ for } He^+ \text{ in He}$$

for an ion energy of 3 eV. With this cross-section, it would take only 27 seconds at $10^{-8}$ torr for an ion to capture an electron and thus leave a cool ion behind. Capture cross-sections are even higher for heavy gas species .

## 2.6 Limits on ion accumulation

In the vast majority of cases (electron storage rings with typical pressures of $10^{-9}$ torr and ionisation times of a second or less), ionisation is the dominant effect in the absence of any clearing mechanism. The neutralisation reached at equilibrium is strongly dependent on the quality of the vacuum (gas species and densities). To illustrate this, we consider the over-simplistic case of constant ionisation cross-sections. The production rates are, for singly-ionised species (density $N_i^+$):

$$\frac{d\left(N_i^+\right)}{dt} = N_m N_p \sigma_i c - N_i^+ N_p \sigma_i c = N_p \sigma_i c \left(N_m - N_i^+\right)$$

doubly ionised:

$$\frac{d\left(N_i^{++}\right)}{dt} = N_i^{\ +} N_p \sigma_i c - N_i^{\ ++} N_p \sigma_i c = N_p \sigma_i c\left(N_i^{\ +} - N_i^{++}\right)$$

etc. until ($Z'$ being the total number of electrons of the gas atom):

$$\frac{d\left(N_i^{Z'+}\right)}{dt} = N_i^{(Z'-1)+} N_p \sigma_i c.$$

In the steady state:

$$N_i^{+} = N_i^{++} = \ldots = N_i^{(Z'-1)+} \leq N_m .$$

Therefore partially ionised ions can, at most, reach the molecular density $N_m$. Only the fully ionised state $N_i^{Z'+}$ could get close to the particle density (usually much larger than the gas density) divided by the final charge state: $N_i^{Z'+} \leq N_p/Z'$ corresponding to full neutralisation of the particle beam. ("Over" neutralisation can for most cases be excluded, since the resulting potential would chase the species to the chamber wall.)

The degree of neutralisation of a particular beam can be estimated from the incoherent tune shift (see Sec. 9). Almost full neutralisation has been measured in the CERN AA when all the clearing electrodes are turned off (Fig. 4).

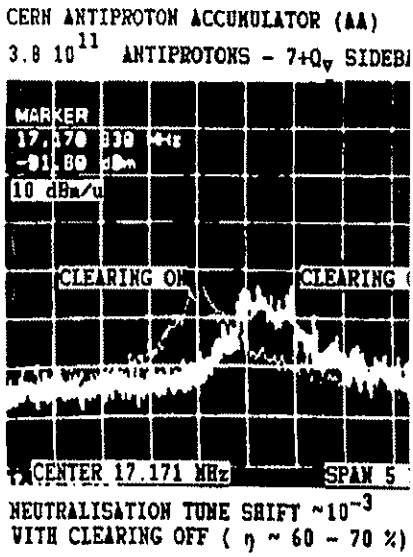



Fig. 4 AA transverse vertical Schottky scan showing the band $(7 + Q_v)f_{rev}$ with clearing electrodes turned OFF (full neutralisation) and ON (little neutralisation)

## 3. THE ION OR ELECTRON MOTION

The temperature of the molecules of the residual gas will be slightly increased by the interactions with the beam. However, the reservoir of molecules is so big that the energy in the gas will be the energy related to the temperature of the vacuum chamber walls, usually 300 K. The energy of the electrons acquired through momentum transfer from the circulating beam in the ionisation process will be of a few eV and the energy of the ions lower by the mass ratio

$$\frac{m_e}{Am_p} \tag{24}$$

where $Am_p$ is the mass of the ion, and $m_e$, the mass of the electron. The electrons and the ions created inside the beam will either be chased out or oscillate in its potential well. Their motion will be influenced by magnetic fields. The analysis of these different energies and motions is the object of this section.

### 3.1 Energy, temperature, velocity

The distribution of velocity of molecules of mass $m$ in a gas of density $d_m$ at temperature $T$ is given by the Boltzmann equation

$$dn = \frac{1}{2} d_m \sqrt{\frac{m^3}{2\pi^3 k^3 T^3}}\, e^{-\frac{m}{2kT}\left(v_x^2 + v_y^2 + v_z^2\right)} dv_x dv_y dv_z, \tag{25}$$

where $dn/dv_x dv_y dv_z$ is the number of particles per unit volume around velocities $v_x v_y v_z$ and $k = 1.4 \times 10^{-23}$ J/K or $k = 8.6 \times 10^{-5}$ eV/K is Boltzmann's constant.

One finds successive mean velocities by integration

$$\langle v \rangle = v_m = \left(\frac{8kT}{\pi m}\right)^{\frac{1}{2}}$$

$$\left(\langle v^2 \rangle\right)^{\frac{1}{2}} = v_{rms} = \left(\frac{3kT}{m}\right)^{\frac{1}{2}}$$

$$\langle |v_x| \rangle = \langle |v_y| \rangle = \langle |v_z| \rangle = v_{\|} = \left(\frac{2kT}{\pi m}\right)^{\frac{1}{2}}$$

with naturally

$$\langle v_x \rangle = \langle v_y \rangle = \langle v_z \rangle = 0\,,$$

so that

$$v_m = 2v_{\|} \quad \text{(Annex I)}.$$

The mean kinetic energy is

$$E = \left\langle \frac{1}{2} m v^2 \right\rangle = \frac{1}{2} m v_{rms}^2 = \frac{3}{2} kT. \tag{26}$$

Table 3 illustrates for different molecules the relation between energy, temperature, and velocity.

Table 3
Relation $T$(K), $E$(eV), $v_{\|}$ (ms$^{-1}$) for various molecules

$T = 300$ K, $E = \begin{cases} 6.3 \times 10^{-21} \text{ joule} \\ 3.9 \times 10^{-2} \text{ eV} \end{cases}$

| | $A$ | $v_{rms}$ | $v_{\|}$ |
|---|---|---|---|
| H | 1 | $2.7 \times 10^3$ | $3.7 \times 10^3$ |
| $H_2$ | 2 | $1.9 \times 10^3$ | $2.6 \times 10^3$ |
| $H_20$ | 18 | $6.5 \times 10^2$ | $1.0 \times 10^3$ |
| $CO/N_2$ | 28 | $5.2 \times 10^2$ | $0.7 \times 10^3$ |
| $C0_2$ | 44 | $4.1 \times 10^2$ | $0.6 \times 10^2$ |
| e- | 1/836 | $1.2 \times 10^5$ | $1.6 \times 10^5$ |

$T = 7.8 \times 10^3$ K, $E = 1$ eV

| | | | |
|---|---|---|---|
| $H_2$ | 2 | $9.8 \times 10^3$ | $13.5 \times 10^3$ |
| $N_2$ | 28 | $2.6 \times 10^3$ | $3.6 \times 10^3$ |
| e- | 1/1836 | $6.0 \times 10^3$ | $8.3 \times 10^3$ |

$T = 7.8 \times 10^3$ K, $E = 10$ eV

| | | | |
|---|---|---|---|
| $H_2$ | 2 | $3.1 \times 10^4$ | $4.3 \times 10^4$ |
| $N_2$ | 28 | $8.3 \times 10^3$ | $11.5 \times 10^3$ |
| e- | 1/1836 | $1.9 \times 10^6$ | $2.6 \times 10^4$ |

### 3.2 The electric field and the potential well

Before analysing the motion of the ions, we compute the fields which act on the ions and the electrons which have been 'just created'. In the absence of external fields, an electric field is produced by the circulating beam. This field defines a potential and the value of the potential is fixed by the fact that the vacuum chamber is at ground potential. We consider the simplified case of a circular beam in a

circular chamber, where $a$ is the radius of the beam with uniform density in real space, $r$ the radial variable and $r_0$ the vacuum chamber radius. We have seen already that the field can be calculated with Gauss's law:

$$\varepsilon_r = \begin{cases} \dfrac{e\lambda}{2\pi\varepsilon_0}\dfrac{r}{a^2} & r \le a \\ \dfrac{e\lambda}{2\pi\varepsilon_0}\dfrac{1}{r} & r \ge a \end{cases} . \tag{27}$$

The potential is obtained by integration

$$V = -\int \varepsilon_r d_r , \tag{28}$$

the constant being fixed by the condition

$$V = 0 \quad \text{for} \quad r = r_0 .$$

The result is

$$V = \frac{e\lambda}{2\pi\varepsilon_0} \begin{cases} -\dfrac{r^2}{2a^2} + \dfrac{1}{2} + \ln\left(\dfrac{r_0}{a}\right) & r \le a \\ \ln\dfrac{r_0}{r} & r \ge a . \end{cases} \tag{29}$$

Figure 5 represents the potential for our nominal beam with 100 mA circulating current, a vacuum chamber of 100 mm diameter and different beam sizes. Figure 6 represents the value of the central potential $V_0$ for different ratios $a/r_0$. The motion of ions or electrons in this field is very simple: in a beam of electrons or antiprotons, electrons are chased and arrive on the wall with an energy $eV_0$ ,ions are trapped if their transverse energy is less than $eV_0$. Since the probability of energy transfer larger than the potential well is very small, ions are always trapped (see section 5 for stability considerations in bunched beams).

The detailed calculation of the potential well for non-cylindrical geometries has been made [26]. We give in Annex II the resulting formulae to be used in practical calculations.

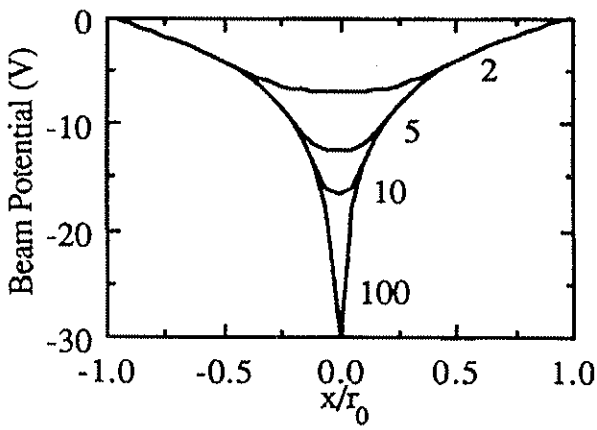



Fig. 5 Potential well for different beam sizes. The number attached to each curve is the ratio $r_0$ /a.

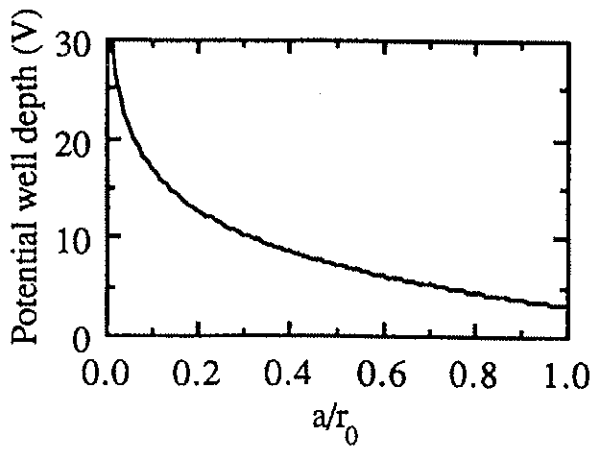



Fig. 6 Depth of the potential well vs beam size.

Figure 7 is the result of the calculation made for the ISR. These potential wells have been directly measured [34, 83].

If the vacuum chamber size or the beam size varies in a long straight section, the electron or ion will drift towards the deepest potential well. The kinetic energy gained in the process can be considerably higher than the thermal energies since the variations of potential energy can reach several tens of eV. while the thermal energies are of the order of $10^{-2}$ eV

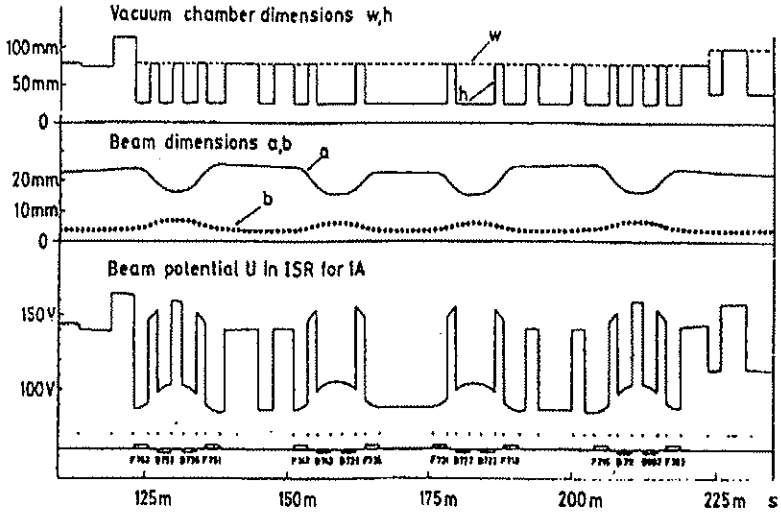



Fig. 7 Beam potential in the ISR. The position of the clearing electrodes is indicated by dots.

## 3.3 The effect of the magnetic field

The motion of a particle in a uniform magnetic field is simple, it is the well known cyclotron motion [3]. When the field has a gradient perpendicular to the field direction, a drift of the particle occurs which is called the gradient drift. If the gradient of the field is in the direction of the field there is a containment effect called the magnetic mirror. This section studies these three effects.

### *3.3.1 Cyclotron motion*

In all these problems one separates the velocity into its two components (Fig. 8):

$\vec{v}_{\parallel}$ parallel to $\vec{B}$

$\vec{v}_{\perp}$ perpendicular to $\vec{B}$

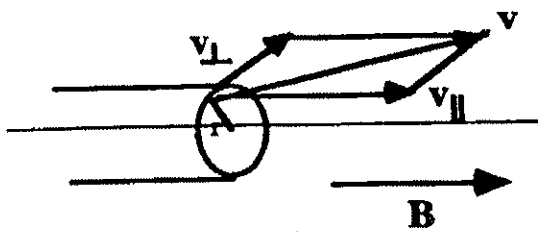



Fig. 8 $\vec{v}_{\perp}$ and $\vec{v}_{\parallel}$

with

$$v^2 = v_{\parallel}^2 + v_{\perp}^2. \tag{30}$$

If the field is uniform $\partial B / \partial r = 0$, then the velocity along the field $\vec{v}_{\parallel}$ is uniform and unchanged. The perpendicular velocity $v_{\perp}$ induces a force and therefore an acceleration.

$$m\frac{d\vec{v}_{\perp}}{dt} = e\,\vec{v}_{\perp} \times \vec{B}.$$

This is a central force perpendicular to $\vec{v}_{\perp}$ which gives a circular motion, the radius $r$ of the circle is obtained by equating the central acceleration to the centrifugal force.

$$ev_{\perp} \cdot B = m\frac{v_{\perp}^2}{r}$$

The angular frequency $\omega_c$ also called cyclotron angular frequency is

$$\omega_c = \frac{v_\perp}{r} \tag{31}$$

which gives

$$r = \frac{m v_\perp}{eB} \tag{32}$$

$$\omega_c = \frac{eB}{m} \tag{33}$$

with the remarkable result that $\omega_c$ does not depend on $v_\perp$ : in a given field the larger the velocity the longer the radius, but the frequency does not change for a given particle.

*3.3.2 Effect of a transverse gradient (the gradient drift)*

We have already seen that

$$v_\perp = r\frac{eB}{m} = r\omega_c$$

but here

$$B = B_0 + \frac{\partial B}{\partial x} \cdot x.$$

The projection of the velocity on z gives (with $x = r \cos \omega_c t$)

$$v_z = v_\perp \cos \omega_c t = r\frac{eB_0}{m}\cos \omega_c t + r\frac{e}{m}\frac{\partial B}{\partial x} \cdot r \cos \omega_c t \cdot \cos \omega_c t \ .$$

Then the mean velocity is not zero, corresponding to a drift

$$v_D = \langle v_z \rangle = \frac{1}{2} r^2 \frac{eB}{m}\frac{1}{B}\frac{\partial B}{\partial x} \ .$$

This calculation only applies if the field variation over the cyclotron motion is small, that is, if

$$r\frac{\partial B}{\partial x} \langle\langle B.$$

This effect is called the gradient drift. It can also be written:

$$v_D = \frac{1}{2\omega_c} v_\perp^2 \frac{1}{B}\frac{dB}{dx}.$$

This gradient can only be created by a curvature of the magnetic field, particles with a velocity parallel to the magnetic field $v_\parallel$ will have to follow the field lines. This curvature will give an additional drift [3] so that the final drift can be written as

$$v_D = \frac{1}{\omega_c}\left(v_\parallel^2 + \frac{1}{2}v_\perp^2\right)\frac{1}{\mathrm{B}}\frac{\partial \mathrm{B}}{\partial x}. \tag{34}$$

*3.3.3 Effect of a longitudinal gradient (the magnetic mirror)*

We assume that $B_z$ changes with $z$. This gives a set of lines of force as sketched in Fig. 9. When the particles move to the right towards higher fields, the field lines are more dense. A variation of flux through the orbit would induce an electromotive force and therefore an exchange of energy between the static magnetic field and the particles. This is not possible so the flux circled by the particle is constant. (The exact demonstration makes use of the action integral (see Jackson [3] p. 422).)

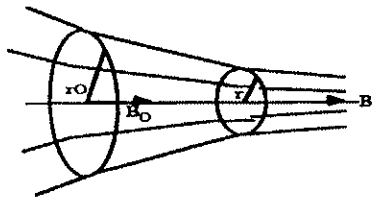



Fig. 9 The magnetic mirror.

$$\pi B r^2 = \pi B_0 r_0^2$$

or, using the equation of the cyclotronic motion,

$$\upsilon_\perp^2 = \upsilon_{\perp 0}^2 \frac{B}{B_0}.$$

Since the kinetic energy of the particle is conserved $\upsilon^2 = \upsilon_0^2$. So that (Eq. 29)

$$\upsilon_\parallel^2 = \upsilon_0^2 - \upsilon_{\perp 0}^2 \frac{B}{B_0}. \tag{35}$$

If $B$ becomes large enough, then $\upsilon_\parallel^2 = 0$. The motion of the particle is stopped. Detailed calculations show that the particle in fact spirals back.

Looking at Eq. 34 it is clear that the particle will be trapped if $\upsilon_\parallel$ can reach zero, that is if

$$\left|\frac{\upsilon_{\parallel 0}}{\upsilon_{\perp 0}}\right| < \left(\frac{B}{B_0} - 1\right)^{\frac{1}{2}}. \tag{36}$$

With an isotropic distribution of speed at the time of creation of the particle the proportion of particles trapped will be:

$$R = 1 - \frac{B_0}{B}. \tag{37}$$

## 3.4 Combined effects of $\varepsilon$ and $B$ (the cross-field drift)

We consider the magnetic field of a magnet and the electric field of the beam (Fig. 10). The electric force is: $e\vec{\varepsilon}$ and the magnetic force is: $e \cdot \vec{\upsilon}_\perp \times \vec{B}$. At equilibrium

$$\upsilon_\perp = \frac{\varepsilon}{B}.$$

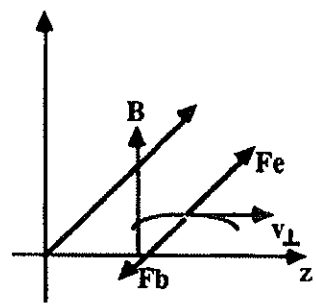



Figure 10: Cross-field drift

This equilibrium can only be attained if $\upsilon_\perp$ can reach $\varepsilon / B$ that is if

$$\frac{\varepsilon}{B} < c.$$

A rather simple analysis shows that indeed if $\varepsilon < cB$ this equilibrium is always reached. However, the time it takes to reach that equilibrium is approximately the time it takes for the field to accelerate the particle to an energy corresponding to the velocity $\upsilon_\perp$. In practice during the acceleration phase where $\upsilon_\perp \cdot B$ is very small compared to $\varepsilon$, the magnetic field can be neglected. If $\varepsilon/B > c$ this equilibrium will never be reached, the magnetic field can be neglected. In all the practical cases which will be considered, the motion can be described by a pure acceleration or a pure drift. Where $\varepsilon/B \ll c$ the equilibrium is reached in less than a $\mu$s and the transverse displacement is less than a $\mu$m.

## 4. A FEW EXAMPLES OF ION OR ELECTRON MOTION

### 4.1 Field-free section

Let us consider a straight section with enlarged vacuum chamber (Fig. 11). The potential well at the centre of the beam is

$$V_0 = \frac{\lambda}{2\pi\varepsilon_0}\left\{\ln\left(\frac{r_0}{a}\right)+\frac{1}{2}\right\} \qquad (39)$$

giving the curve of potential represented in Fig. 11 as a function of the azimuth. In this example the ions created between B and C will be trapped, the ions created between A and B will drift towards C and continue to the right with an energy between 0 and 6 eV.

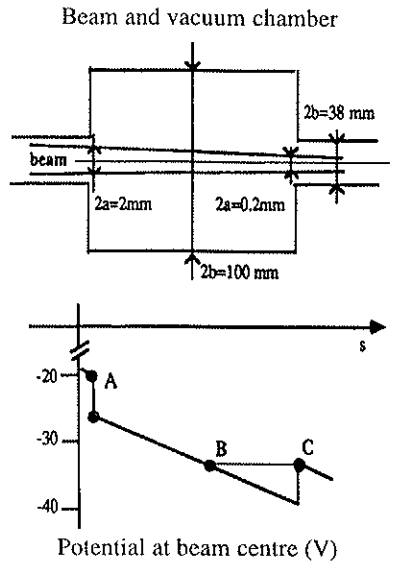



Fig. 11 Potential well in a field-free section.

### 4.2 Pure dipole field

With our nominal beam parameters (section 1), the electric field at the edge of the beam is $1.2 \times 10^4$ $Vm^{-1}$ so that the cross-field drift velocity varies from zero at the centre to $8.6 \times 10^3$ $ms^{-1}$ at the edge of the beam. The corresponding energies are $2 \times 10^{-4}$ eV for electrons and 11 eV for $N_2^+$ ions. The cyclotron radius for electrons in the centre with energy of about 1 eV is 2.4 μm (0.6 mm for ions).

Particles with the same charge as the beam are chased out of the potential well in the vertical plane. Particles with opposite charge drift towards the end of the magnet, but are contained in the potential well in the vertical plane.

### 4.3 Combined function magnet

Let us consider a magnet like the EPA magnet with a field of 1.4 T on the central orbit and a gradient of 1 T/m on the central orbit. We have seen that the typical energy of particles created in the potential well is a few electron volts, the corresponding velocities are (section 3.1):

for ions ~ $2.6 \times 10^3$ $ms^{-1}$.

for electrons ~ $6 \times 10^5$ $ms^{-1}$.

The cyclotron radii in the centre are very small. The additional gradient drift for ions or electrons of 1 eV is of the order of 0.5 $ms^{-1}$ and therefore negligible with respect to the cross-field drift.

### 4.4 The quadrupole field

If the beam is centred in a quadrupole with gradient 3 $Tm^{-1}$, the cross-field drift at equilibrium at the edge of the beam will be Eq. (37)

$$\upsilon_c = 8 \times 10^6 \text{ m s}^{-1}$$

corresponding to an energy for ions of 9 MeV. This energy would have to be provided by the potential well, which is of only a few tens of volts, so that the equilibrium required for the cross-field drift will never be reached. The electric field of the potential well dominates the ion motion.

Particles with the same charge as the circulating beam are chased toward the poles of the quadrupole. In some cases a magnetic mirror effect could ensure the containment of these particles.

### 4.5 Undulator

The field of an undulator is dipolar but with alternatively positive and negative polarity along the azimuth. The main effect in the horizontal plane is the cross-field drift, but the fact that the field is alternated gives a possibility of containment. In the vertical plane the field is concentrated in the poles so that a magnetic mirror effect can develop. This mechanism can be effective for electrons or ions depending on the geometry and the beam polarity. Since the potential well is not required to achieve containment, it is not easy to find the mechanism which limits the accumulation of ions in an undulator.

## 5. BUNCHED BEAMS

All previous studies are in principle valid only for unbunched circulating beams. In fact the bunching only introduces additional effects. In Refs. [7, 14, 57, 75], one finds a study of the stability of ions in single-beam and colliding-beam machines. We shall only summarise the results of this study in the case of single-beam machines.

Here, we have changed the definition of the neutralisation factor $\eta$ from the local definition used in Ref. [14] to the average definition given above in Eq. (1). This explains the difference between our formulae and those of Ref. [14].

At a given azimuth, an ion sees successively the focusing (or defocusing) forces induced by the bunch passage, followed by a drift time between bunches. If a vertical dipole magnetic field is applied, the horizontal transverse and longitudinal motions will be coupled, the vertical motion however being independent of the magnetic field. The horizontal motion will not be studied here [83]. It gives similar results in drift space; in a bending magnet the horizontal motion is usually stable. The present study is therefore limited to the vertical motion.

The forces induced by the passage of a bunch have been studied by several authors (e.g. [8, 14]). With the time as independent variable the bunch passage is similar to a focusing (or defocusing) lens; if the bunch is short the effect on the ion can be described by a matrix equivalent to a thin lens. The non-linear effects are not considered here.

Let $y$ and $dy = \dot{y}$ be the position and speed of the ion. Then the passage of a bunch is described by

$$\begin{pmatrix} y \\ \dot{y} \end{pmatrix}_1 = \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} y \\ \dot{y} \end{pmatrix}_0$$

with

$$\alpha = \frac{n_e}{n} \frac{4r_p}{\beta b(a+b)} \frac{1}{A} \tag{40}$$

where $n_e/n$ *is* the charge per bunch, $n$ the number of bunches, $r_p$ the classical proton radius, $a$, $b$ the beam size, horizontal resp. vertical, $A$ the atomic weight of the ion and $\beta c$ the velocity of the circulating beam.

Between bunches the ion drifts freely during the time $t = T/n$. Where $T$ is the revolution time, corresponding to the matrix transformation

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

One period of the forces applied is described by the matrix product

$$M = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix}$$

The stability is insured if the trace of the matrix satisfies

$$-2 < Tr(M) < +2$$

that is for

$$-1 < \left(1 \pm \alpha \frac{T}{2n}\right) < +1 .$$

The + sign is to be selected for cases where the beam and the ion are of the same sign. Electron beams therefore cannot accumulate electrons, and proton beams cannot accumulate positive ions, which is obvious. For the opposite sign, the requirement is:

$$1 - \alpha \frac{T}{2n} > -1 .$$

If $A.m_p$ is the mass of the ion, one can express the above formulae by the following criterion: all ionic masses larger than a critical mass $A_c$ will be accumulated.

$$A_c = \frac{n_e}{n} \frac{r_p}{n} \frac{2\pi R}{\beta \cdot b^2 \left(1 + \frac{a}{b}\right)} . \tag{41}$$

In most practical cases the critical mass varies between 0.1 and 100. This explains why electrons ($A \simeq 1/2000$) never accumulate in bunched beams of positrons or protons.

In bunched beams of electrons with a large number of bunches (e.g. synchrotron radiation sources) the critical mass can be very small ~ $10^{-2}$. Then the ion accumulation can be treated as in DC beams: the ions are too heavy to 'see' that the beam is bunched.

When the critical mass is above 44, the ions usually found in vacuum systems cannot accumulate.

Intermediate cases:

$$1 < A_c < 44$$

require a detailed analysis [14].

Note:

There is some confusion on the values to be used for $a$ and $b$ after a real measurement of particle distributions. They are related to the central density of the transverse distribution of particles. If the measured distributions were rectangular then $a$ and $b$ would be the measure of the real half width of the distributions. If the distributions are Gaussian, a simple calculation shows that

$$a = \sigma_x \sqrt{2}$$

$$b = \sigma_y \sqrt{2}$$

## 6. CLEARING MEANS

### 6.1 Missing bunches (electron storage rings)

Many small electron storage rings prone to ion trapping have partly solved their problem by introducing one or several gaps in the bunch train by not filling certain buckets at injection. Over one revolution period of a train of $p$ consecutive bunches, the motion (here vertical) of a trapped ion is the solution of:

$$\begin{bmatrix} y \\ \dot{y} \end{bmatrix}_1 = M \begin{bmatrix} y \\ \dot{y} \end{bmatrix}_0$$

with the transfer matrix

$$M = \left( \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \right)^{p} \begin{bmatrix} 1 & (h-p)t \\ 0 & 1 \end{bmatrix} \tag{42}$$

where the terms in parenthesis represent the linear kick received at the $p$ bunch passages interleaved with drifts of duration $t$, $f$ being the radiofrequency. The period of $p$ successive kicks plus drifts is terminated by the drift in the time interval $(h\text{-}p)\ t$, where $h$ is the cavity harmonic, i.e. the maximum number of bunches that the machine can handle. The Floquet's condition of stability for the ion of mass-to-charge $A$:

$$-2 < Tr\,(M) < 2$$

does not lead to a simple criterium defining which $A$'s are stable. Rather, the trace of the transfer matrix (42) is of order $p$ in $N$, the total number of circulating particles. This means that there are $p$ stable bands of ion mass-to-charge ratios for a given $N$, or that a given ion will be stable or unstable, depending on the number of beam particles. Figure 12 illustrates the conditions of linear stability for various ions in the EPA ring, as a function of the number of beam particles and consecutive bunches [89].

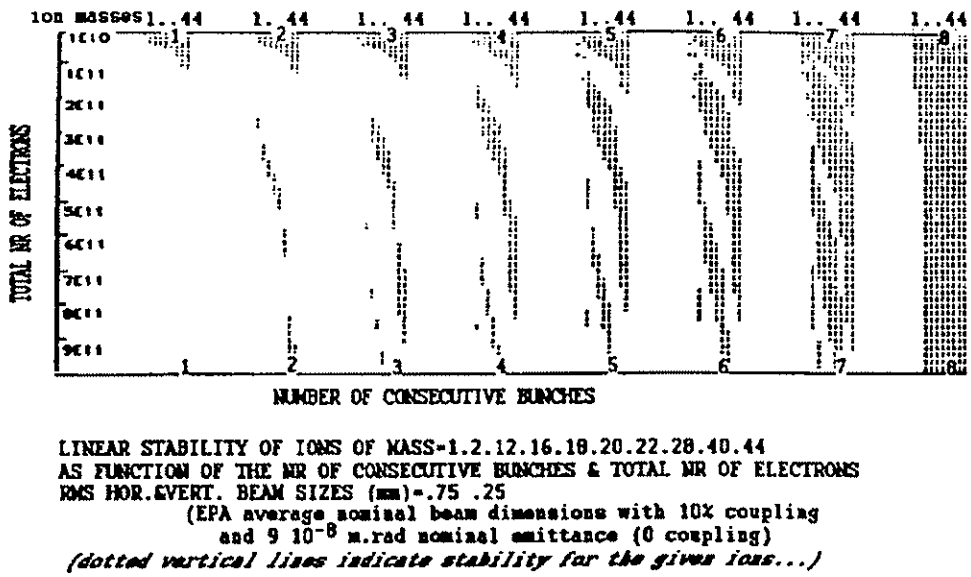



Fig. 12 Linear stability of ions in EPA as a function of total beam intensity and number of consecutive bunches for nominal beam emittances.

Although not always absent, ion trapping is indeed less severe with missing bunches. This stems from the fact that since a given trapped ion can be further ionised, it then has a good chance of falling in an unstable band and thus of being cleared.

## 6.2 Clearing electrodes

Clearing electrodes consisting of negatively polarised plates fitted into the vacuum chamber provide a transverse electric field, which diverts beam-channelled ions onto them, where they are neutralised and return into the gas phase. Figure 13 qualitatively represents the potential variation across a vacuum chamber of radius $r_c$, with an electrode on one side with a potential of $U_{ce}$.

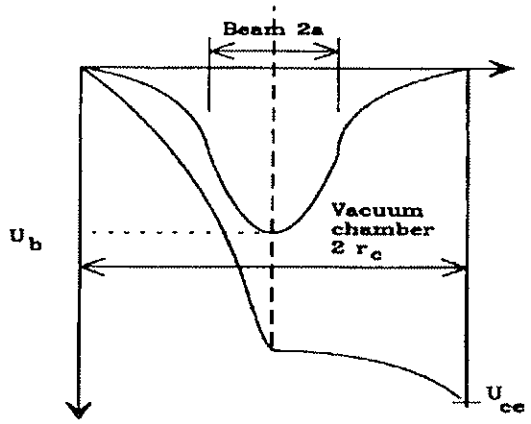



Fig. 13 Transverse potential distribution across a vacuum chamber due to the beam space charge and a clearing electrode.

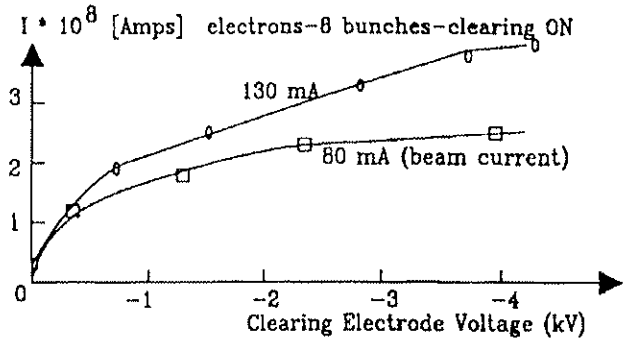



Fig. 14 Extracted ion current vs. electrode voltage in EPA.

A minimum condition for capture of the passing ion is that the transverse field provided by the electrode equals the maximum beam space charge field:

$$\frac{U_{ce}}{2r_c} > E(a) \sim \frac{I}{2\pi\beta c\varepsilon_0 a} \tag{43}$$

In fact, because ions may have — as we have seen — transverse and longitudinal velocities corresponding to energies of up to a few eV (possibly up to the field at the centre if singly ionised) and because a clearing electrode is necessarily limited in size, the field provided by the electrode must usually be larger than the calculated beam field. For instance in the EPA ring, where electrodes are of the button type with a diameter of 20 mm and installed flush to the beam, transverse clearing in field-free regions is complete with the following parameters [89]:

$$U_{ce} = 6\ \text{kV}$$

electrode field: 30 kV/m on beam axis. Beam max. field: 12 kV/m for $I$ = 0.3 A and ~ $10^{-8}$ m-rad horizontal emittance with 10% coupling.

Measurement of the electrode clearing current as a function of the applied voltage provides a verification of the required maximum to be applied for full clearing, since the current will saturate once a sufficient field is reached (Fig. 14). The number and the optimum location of electrodes are in principle dictated by the tolerable degree of residual neutralisation. In practice, even with a large number of electrodes, uncleared pockets always remain and contribute to typical residual neutralisations of a fraction to a few percent.

Few (if any) small electron storage rings exist which have reached a fully satisfactory ion-free situation, even with clearing electrodes. Perhaps one reason for this is that clearing systems have not been up to now very complete, partly owing to the fact that clearing electrodes complicate the mechanical design of the vacuum chamber and may contribute to the machine impedance. At CERN (EPA), button-type clearing electrodes presenting negligible coupling characteristics to the beam have been designed. They are made of a ceramic body, coated with a highly resistive glass layer (thick-film hybrid technology), and are terminated with a highly lossy wide-band filter as shown in Fig. 15.

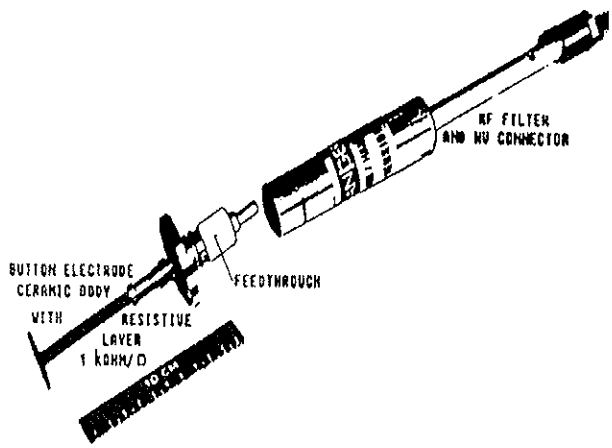



Fig. 15 EPA button-type clearing electrode.

### 6.3 Resonant transverse shaking of the beam

Neutralisation effects can be reduced when exciting vertical coherent oscillations with a transverse kicker at a given frequency [103], both in electron storage rings (bunched) and antiproton accumulators (unbunched). This technique of "RF knock out" of neutralising ions has been nicknamed "beam shaking". It has been determinant in solving ion problems in the CERN AA, where it has been studied both theoretically [103] and experimentally [102] and where it was permanently implemented, with the following parameters:

shaking: vertical
shaking frequency: 490 kHz
sideband frequency (fractional tune $q_v$):480 kHz
length of kicker electrodes: 0.6 m
kicker field: ~ 20 V/cm

Beam shaking experimental observations can be summarised as follows:

1) Beam shaking works best when applied vertically: one possible reason is that neutralisation is high in dipole fields (low ion drift velocity) where the motion along the lines of force is the only practical degree of freedom.

2) To work, beam shaking relies on the longitudinal motion of the ions. Due to changing beam dimensions, the ion "plasma frequency" spectrum is wide compared to the "knock out" frequency: ions have to "sweep" through this resonance. For this, they must be free to move longitudinally.

3) Beam shaking depends on the non-linearity of the beam space-charge field: this allows the "lock on" of the sweeping ions onto the resonance where they keep large oscillation amplitudes, thus reducing their density in the beam centre.

4) Beam shaking is efficient even with RF kicker fields of only a few 10 V/cm, provided it is applied close to a beam betatron side band whose frequency lies close to the ion plasma frequency. In this case, the beam resonant response ensures sufficiently large non-linear forces on the ion. Experimentally, it is found that for a weakly exciting RF field, shaking works best above a band ($n$+$Q$) or below a band ($n$-$Q$). This observation of asymmetry of weak resonant shaking is important in that it validates the non-linear character of the ion motion and the "lock on" conditions.

To illustrate this in a simple way, we use a quasi-linear description of the two-body resonant condition [117] for an unbunched beam. We consider only one ion species $i$, of mass-to-charge ratio $A_i$, with the following definitions:

$\Omega$: circular revolution frequency of circulating beam ($\Omega = 2\pi f_i$)

$Q_i = 2\pi f_i/\Omega$: the ion bounce number in the beam potential well

$$\Omega_i^2 Q_i^2 = \frac{2N_p r_p c^2}{\pi b(a+b)\gamma R} .$$

If $Q_v$, is the beam particle unperturbed incoherent tune and $Q_p$ the beam particle bounce number in the ion-potential well where

$$\Omega^2 Q_p^2 = \frac{2N_i r_p c^2}{\pi b(a+b)\gamma R}$$

and

$$Q = \left(Q_v^2 + Q_p^2\right)^{1/2}$$

is the perturbed beam tune, then a beam particle and an ion obey the coupled set of linear differential equations

particle:
$$\left(\frac{\partial^2}{\partial t^2} + \Omega\frac{\partial^2}{\partial\theta^2}\right) y_p + Q^2\Omega^2 y_p - Q_p{}^2\Omega^2\bar{y}_i = Fe^{i\omega t} \tag{44}$$

ion:
$$\left(\frac{d^2}{dt^2}\right) y_i + \Omega_i{}^2\Omega^2\left(y_i - \bar{y}_p\right) = 0$$

where the bar on $y$ denotes the average vertical position of each beam and the $Fe^{i\omega t}$ term is the harmonic of the external driving force close to beam ion resonance:

$$\omega_f (n \pm Q)\Omega_f Q_i \Omega \ .$$

Assuming solutions of the form:

$$y_p = \xi_p e^{i(n\theta + \omega t)}$$
$$y_i = \xi_i e^{i\omega t}$$

the ion amplitude becomes:

$$y_i = \frac{Q_i^2 \frac{F}{\Omega^2} e^{i\omega t}}{\left(x^2 - Q_i^2\right)\left((n+x)^2 - Q^2\right) - Q_p^2 Q_i^2} \tag{45}$$

with
$$x = \frac{\omega}{\Omega} \rightarrow (n \pm Q) \sim Q_i \ .$$

Therefore, shaking works when $y_i$ becomes large, i.e. when the denominator in Eq. (45) $\rightarrow 0$. But, as shown in Fig. 16, as the ions gain large amplitude, we have non-linear detuning such that $x^2 > Q_i^2$ (lock on). Therefore for $y_i$ to become large, requires that:

$$(n+x)^2 - Q^2 > 0$$

i.e.: $x < n - Q$ for excitation near a "slow-wave" beam frequency: $\omega \cong (n - Q)\Omega; n > Q$

$x > n + Q$ for a "fast-wave" frequency : $\omega \cong (n + Q)\Omega; n > Q$ .

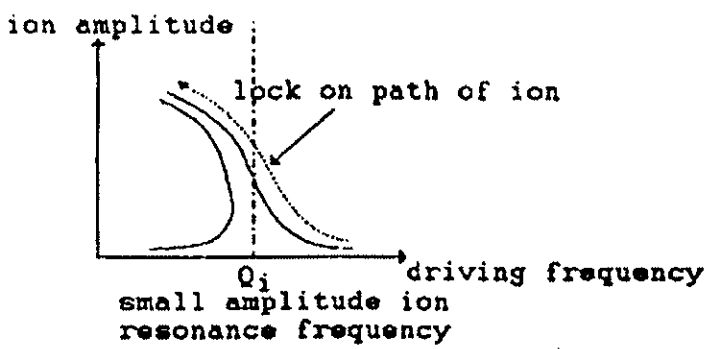



Fig. 16 Qualitative Amplitude response curve of an ion versus the driving frequency near resonance.

This asymmetry has been verified both in the CERN and Fermilab antiproton accumulators and in the EPA ring [102] [105]. Figure 17 for the Fermilab AA, shows that shaking below a "slow wave" excited by ions is more efficient than shaking above, since it suppresses the instability signal.

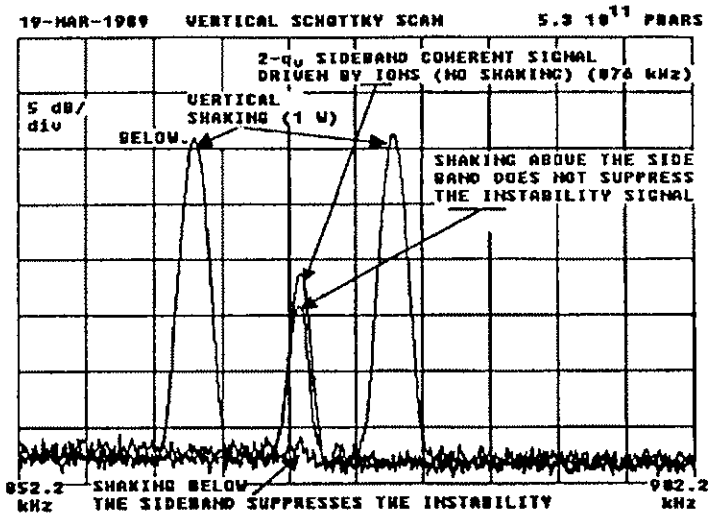



Fig. 17 Vertical Schottky scan showing the suppression of an ion-driven dipolar instability by shaking in the FNAL Antiproton Accumulator.

The analysis developed above simply hints at the reality. In fact, when one considers ions "locked" onto resonance, the amplitude of their coherent oscillation $y_i$ depends only little on the amplitude of the excitation of the oscillating external force $F$ in Eq. (44), when it is applied below or above a sideband [103]:

$$F \rightarrow F_{ef} = F \pm Q_i^2 \, \bar{y}_i, \qquad (x+n)^2 \gtrless Q^2 .$$

## 7. THE LIMIT OF ACCUMULATION

The containment of particles is in general due to the potential well. When charges start to accumulate, they diminish the depth of the potential well. The limit is reached when the density of accumulated charges is equal and opposite to the density of circulating charges so that the net resulting charge is null.

$$0 = d_i - d_p .$$

The neutralisation factor $\eta$ is then, at most, equal to 1. The beam is fully neutralised. Clearly the same limitation of neutralisation exists in bunched beams.

As we have seen, several clearing effects due to various drifts will limit the local neutralisation factor to less than 1. If required, clearing electrodes will collect electrons (or ions) and reduce the neutralisation in appropriate places. However, in some extreme cases the trapping of the ions or electrons is not due to the potential well, so that one could have local accumulation of charges to neutralisation levels higher than 1. We have seen that the undulators are probably very efficient in accumulating ions. Detailed calculations should be made in this particular case. It is not excluded that similar situations be found in the combination of end-fields of conventional magnets or quadrupoles.

## 8. THE EFFECTS OF NEUTRALISATION

There are several effects of neutralisation. The particles (ions or electrons) chased from the beam and hitting the wall with the energy of the potential well can create desorption. The ions stored in the beam modify the local pressure and therefore the life-time of stored beams. The most important effects come from the electric field of the ion (or electron) cloud. This induces a shift and a spread of

betatron oscillation frequencies and a coupling of horizontal and vertical motion. Finally, the ion cloud can interact as a whole with the beam and induce instabilities.

## 8.1 The pressure bump

The mechanism of the pressure bump is rather simple. In a beam of protons the ions are chased out of the potential well and hit the wall with enough energy to induce desorption of $\eta$ molecules per ion (here $\eta$ is the desorption factor and not the neutralisation factor). If the rate of filling the vacuum vessel with these molecules is higher than the pumping speed, the pressure will increase. The exact treatment is given in Refs. [62, 65]. It is however possible to obtain a first indication and order of magnitude by equating the pumping speeds to the desorption rate.

The equation of pumping of a vacuum vessel is in the absence of desorption [2]:

$$\frac{dP}{dt} = -\frac{SP}{V}$$

where $P$ is the pressure (torr), $S$ the pumping speed ($m^3 . s^{-1}$) and $V$ is the volume ($m^3$) . In terms of molecular density $d_m$ and linear pumping speed s, this can be written as

$$\frac{d}{dt} d_m = -\frac{s}{A} d_m \tag{46}$$

where $A$ is the cross-sectional area of the vessel (and not a molecular mass).

The rate of production of ions *is* $(I/e)d_m\sigma$ ions per second and per metre ($I/e$ is the number of protons per second and $d_m\sigma$ is the probability of ion production per metre). The desorption rate is therefore

$$\frac{d}{dt} d_m = \eta \cdot \frac{1}{A} \cdot \frac{I}{e} d_m \sigma . \tag{47}$$

Combining the two effects gives

$$\frac{d}{dt} d_m = \frac{d_m}{A}\left( \eta I \frac{\sigma}{e} - s \right) \tag{48}$$

leading to an exponential increase of density or pressure if

$$\eta I \frac{\sigma}{e} > s \ . \tag{49}$$

Ex: the ISR pressure bump at the beginning of operation

$$\eta = 4$$

$$I = 4$$

$$\frac{\sigma}{e} = 6.25 \times 10^{-4}$$

gives

$$\eta I (\sigma / e) = 10^{-2} \, m^2 \cdot s^{-1} .$$

If in some places the effective local pumping speed is reduced below 10 l $s^{-1}$ $m^{-1}$ the pressure bump can develop.

- The rise time of the pressure is of the order of $A/s$ which is, for standard vacuum chambers, a few seconds. This could explain the oscillation of the vacuum pressure observed in a slow cycling machine with high currents like the PS or the Fermilab Main Injector.

- The desorption coefficient $\eta$ depends on the energy of the ion striking the wall and therefore on the potential well depth: $V_0$. It also depends on the cleanliness of the vacuum chamber.

## 8.2 Pressure increase due to ions

The residual gas density and composition enter in the calculations of beam lifetime due to scattering of particles of the beam by the residual gas or of beam emittance growth due to multiple scattering.

The effective density to be considered is the sum of molecular and ionic densities.

We have seen (Eqs. 3 and 10) that the circulating beam density is

$$d_e = \frac{I}{e}\frac{1}{\beta c}\frac{1}{\pi a^2} . \tag{50}$$

The neutralisation factor $\eta$ gives the ion density

$$d_i = \eta d_e .$$

The molecular density is

$$d_m = 3.3 \times 10^{22} P_m .$$

If we use the data of our nominal beam

$$d_e = 2.7 \times 10^{15}\,\mathrm{m}^{-3}$$

With 10% neutralisation

$$d_i = 2.7 \times 10^{14}\,\mathrm{m}^{-3}$$

corresponding to a 'partial pressure' :

$$P_i = 8 \times 10^{-9}\ \mathrm{torr}.$$

In ultra-high vacuum systems, the pressure increase due to ions can be equal to the residual gas pressure [43, 67, 72]:

## 8.3 Tune shifts

The problem of tune shifts induced by neutralisation is similar to the problem of tune shifts induced by space charge. One assumes that the transverse distributions of ions are the same as the distributions of circulating particles[6] .The calculations are made with constant real space density [7] for a beam of width $2a$ and height $2b$. The ion density is ($\eta$ is the neutralisation factor)

$$d_i = \frac{1}{2\pi R}\frac{1}{\pi ab}\eta n_e . \tag{51}$$

The electric field of this distribution has been calculated by several authors [7]; it can be written for the horizontal plane as (mks units)

$$\frac{\partial \varepsilon}{\partial x} = \frac{e}{\varepsilon_0}\frac{d_i}{1+\frac{a}{b}} . \tag{52}$$

The local quadrupole strength as defined in [1] is

[6] This assumption is made for the sake of simplicity. In fact , calculations [19] based on a one-dimensional linear model of the ion cloud motion show that this is not true. In that model, the width of the ion distribution is much smaller than the width of the electron beam distribution.

$$k = \frac{e}{E}\frac{\partial \varepsilon}{\partial x} \tag{53}$$

where $E$ is the energy of the circulating particle.

The corresponding tune shifts are obtained using

$$\Delta Q = \frac{1}{4\pi}\int \beta(s)\, k(s)\, ds \ .$$

The tradition is to introduce the classical electron (or proton) radius

$$r_e = \frac{1}{E_0}\frac{e^2}{4\pi\,\varepsilon_0}, \tag{54}$$

where $E_0$ is the rest mass of the electron. The integration is made around the circumference of the machine where $\beta$, $a$, $b$ vary as a function of $s$. One usually replaces the quantities by their average around the machine ($\beta = R/Q$). This averaging is partly justified by the fact that the quantity $\pi a^2/\beta$ is an emittance.

$$\Delta Q_x = r_e \frac{1}{\gamma}\frac{R}{Q_x}\frac{1}{\pi a(a+b)}\eta \cdot n_e \tag{55}$$

$$\Delta Q_y = r_e \frac{1}{\gamma}\frac{R}{Q_y}\frac{1}{\pi b(a+b)}\eta \cdot n_e \ . \tag{56}$$

The values $\Delta Q_x$, and $\Delta Q_y$ are in fact tune spreads as well as tune shifts because the distribution of ions is not uniform in the beam, so that the fields are strongly non-linear. The non-linearities of the field produced by the ion cloud introduce coupling effects[7]. This is particularly visible in $e^+e^-$ machines [8, 12, 78]. These effects have been directly measured [66, 69].

## 8.4 Instabilities in proton beams

The motion of the electrons (or ions) in the potential well of the circulating beam and of the motion of the circulating beam in the electron cloud provides a feedback mechanism which can drive beam instabilities. Two such instabilities have been detected in proton machines. They received the name of ionic oscillations [110] and of electron instabilities [113]. The detailed theory of these instabilities does not have its place here. The mechanisms however are rather simple, and simplified formulae can be derived to obtain orders of magnitude.

### *8.4.1 The ionic oscillations*

The proton bunches leave ion-electron pairs after their passage. With a bunched beam the light weight electrons are so stirred up in the ionisation process that they are lost to the wall almost immediately.

The heavy positive ions, even though they are of the same polarity as the beam, will stay longer in the vacuum chamber because they are created with thermal energies and are more difficult to move [110].

The detailed calculations go through the following steps:

[7] Recent calculations [18] show that the coupling produced by an ion cloud whose transverse distribution is a replica of the distribution of a symmetrical bi-gaussian flat beam is much smaller than what is observed experimentally [89]. This suggests that, either asymmetric ion distributions are present, creating stronger *linear* coupling or that other (possibly coherent) effects play a role in the phenomenon.

- a closed pattern of oscillation of the beam around the machine is defined:

$$x(\theta,t) = a\cos(Q-n)\Omega t + n\theta$$

which implies that at a given time (e.g. $t = 0$) the centre of gravity of the beam is distributed along n oscillations around the machine and that a given particle $(\theta = \theta_0 + \Omega t)$ oscillates with the betatron frequency $Q\Omega$. For simplicity we define

$$\omega = (Q - n)\Omega$$

then

$$x = a\cos(\omega t + n\theta) \tag{57}$$

- The corresponding pattern of ions will have the same aspect but with a phase shift due to the ion motion in the beam field.

$$x_i = b_0\cos(\omega t + n\theta) + b_1\sin(\omega t + n\theta)$$

We shall skip the difficult calculation of $b_0$ and $b_1$.

- The presence of these ions will induce a force on the proton beam to be inserted in the equation of motion of the proton.

$$\ddot{x} + (Q\Omega)^2 x = K \cdot x_i \tag{58}$$

- The definition of $x$ (Eq. (52)) no longer satisfies this equation of motion. The technique is to let the two constants $a$ and $\omega$ vary slightly to take into account the small extra force introduced.

$$a \rightarrow a + \dot{a}t$$
$$\omega \rightarrow \omega + \partial\omega\,.$$

Then

$$\ddot{x} \cong -a(\omega + \partial\omega)^2\cos(\omega t + n\theta) - \dot{a}\omega\sin(\omega t + h\theta) \tag{59}$$

can be introduced in Eq. (58).

- The resolution of the cosine terms gives a negligible change in the tune shift. The resolution of the sine terms leads to the rate of rise of the instability.

$$\frac{1}{\tau_r} = \frac{\dot{a}}{a}$$

where $\tau_i$ is given as a function of $b_0$ and $b_1$.

- For $\tau_r > 0$ the amplitude is unstable. This is only obtained for $Q - n$ negative which means that only patterns with $n > Q$ will be unstable. This is a good signature of the instability.
- The final result [110] is remarkably simple for the dense beams which are common in present accelerators

$$\tau_r = \frac{8Q\gamma}{(n-Q)A}\tau_i$$

where $\tau_i$ is the ionisation time and $A$ the atomic mass number of the ion.

For example at low energy ($\gamma \sim 1$) in the CERN PS ($Q = 6.25$, $n = 7$) with a pressure of $10^{-7}$ torr of $N_2$ ($A = 28$) one finds

$$\tau_i = 10 \text{ ms } \tau_r = 20 \text{ ms}$$

*8.4.2 Electron-proton instabilities*

This instability was first observed in the ISR. The field of the electrons accumulated in a proton beam induces a tune shift $\Delta Q$. This effect can be introduced in the equation of motion of the protons [70, 113] (we have selected the vertical motion)

$$\ddot{z}_p + Q^2\Omega^2 z_p = Q_p^2\Omega(z_e - z_p) \tag{60}$$

where $z_p$ and $z_e$ are the transverse positions of the centre of gravity of the beam and of the electron cloud and $Q_p$ is given by (Section 8.3)

$$Q_p^2 = 2Q\Delta Q = \frac{2r_p}{\gamma} \cdot R \cdot \frac{1}{\pi b(a+b)} \cdot n_e \,. \tag{61}$$

In a similar way the equation of motion of the centre of gravity of the electron cloud can be written as

$$\ddot{z}_p = Q_e^2\Omega^2(z_p - z_e) \tag{62}$$

with

$$Q_e = \frac{2}{\pi}\frac{r_e R n_p}{a(a+b)}.$$

As in the previous case we can define a pattern of oscillation of the proton beam

$$z_p = A_p\, e^{i(n\theta - \omega t)}. \tag{63}$$

Note that the introduction of the complex exponential will simplify the calculation of the phase shift and of the rate of rise that was treated with real sine and cosine functions in the previous example.

The electrons have only a local oscillation:

$$z_e = A_e\, e^{i\omega t}. \tag{64}$$

The substitution of Eqs. (58) and (59) in (55) and (57) gives two homogeneous equations in $A_e$ and $A_p$. $A_e$ and $A_p$ can be eliminated with the result that the defined quantities must satisfy the resulting equation where the reduced frequency $x = (\omega/\Omega)$ has been introduced:

$$(Q_e^2 - x^2)\left[Q^2 + Q_p^2 - (n - x)^2\right] = Q_e^2 Q_p^2 \,. \tag{65}$$

In this equation $Q_e$, $Q_p$, $Q$ and $n$ are the parameters and $x$ is the unknown.

If $x$ is real it means that there exists a real frequency of oscillation of the system of two beams. The system is stable. If $x$ is complex the solutions come by pairs; one with a positive imaginary part corresponding to a damping of the pattern; one with a negative imaginary part corresponding to an anti-damping of the pattern. A detailed examination of Eq. (65) shows that for large values of $Q_p$ the solutions become complex.

The threshold value of $Q_p$ is

$$Q_p^{th} = \frac{(n - Q_e)^2 - Q^2}{2\sqrt{Q_e(n - Q_e)}} \tag{66}$$

and defines the threshold of the instability. Above this threshold the imaginary part of the complex conjugate solutions $x$ is

$$\mathrm{Im}\,\frac{\omega}{\Omega} = \frac{Q_p}{2}\sqrt{\frac{Q_e}{(n - Q_e)}}. \tag{67}$$

The growth is

$$\frac{1}{\tau_r} = \Omega\ \mathrm{Im}\,\frac{\omega}{\Omega}. \tag{68}$$

with $N_p = 6 \times 10^{13}$, $a = 3 \times 10^{-2}$ m, $b = 10^{-2}$ m, $\gamma = 16$ and $Q = 8.75$.

Then

$$Q_e = 200 \ \ (\text{Eq. (62)}).$$

The most dangerous mode $n$ will give the smaller threshold $Q_p^{th}$ This is obtained for $n$ = 209 (Eq. (66))

$$Q_p^{th} = 0.052.$$

Equation (56) gives the corresponding average neutralisation required

$$\eta = \frac{n_e}{n_p} = 2 \times 10^{-3}$$

and Eqs. (67) and (68) give the growth rate

$$\tau_r = 8T$$

where $T$ is the revolution period of the ISR

$$\tau_r = 25\ \mu\,\mathrm{s}.$$

The instability is extremely fast and the frequency observed is, in general, above the bandwidth of pick-up electrodes

$$\frac{1}{2\pi} Q_e\ \Omega = 64\ \mathrm{MHz}$$

*8.4.3 Landau damping*

In reality these instabilities are much less predictable than in the above simplified picture. The complexity comes when, instead of analysing the behaviour of the beam or of the ion cloud as a whole, one analyses the behaviour of each individual particle or ion before averaging the displacements. The fact that neutralisation forces are very non-linear, that the frequency of oscillation of electrons depends on the azimuth as well as the number of electrons trapped complicates the picture further. The calculations cannot be made on a single frequency but rather on a distribution of frequencies. It is far easier to explain why a given instability occurs in a given machine than why it does not occur in another machine.

## 8.5 Instabilities in antiproton beams

*8.5.1 Coherent instabilities*

Transverse coherent instabilities where collective oscillations of the beam centre or the beam shape grow exponentially due to the interaction with trapped ions have been identified in antiproton accumulators (coasting beams with negative space charge) at CERN and FNAL. These "two-beam" instabilities have been severe intensity limiting mechanisms for these machines, where they have been studied in detail [111].

For dipole instabilities, the theory outlined in Section 8.4.2 applies. The essential difference with proton machines trapping electrons is that the ionic oscillation frequencies in the beam potential well are much lower than with electrons (typically 1 to 2 MHz for $H^+$ and $H_2^+$, versus 20 to 80 MHz for electrons). The lowest transverse beam modes $(n - Q)\Omega$ are thus excited. These low frequency modes are the most unstable because of their low frequency spreads, resulting in loss of Landau damping.

For this reason, dipole instabilities driven by ions $(H^+, H_2^+)$ occur at very low neutralisation levels (< 0.01), lower than in proton beams. The neutralisation resulting from a single neutralisation pocket around the ring created for instance by a localised vacuum chamber enlargement, may be sufficient to drive an instability [116]. Dipole instabilities can be effectively damped by a transverse feedback system using a high sensitivity resonant pick-up tuned at the frequency of the single unstable mode [108].

Quadrupole instabilities caused by ions, whereby the beam transverse shape or envelope oscillates, have also been identified in the CERN AA. This instability manifests itself as a kind of 'breathing' mode or fast emittance blow-up phenomenon causing eventually beam loss at a quadrupolar frequency $(n - 2Q)\Omega$ which can be as low as 400 kHz in the CERN AA.

The theory for this instability, similar to the dipolar one, is also well developed [117]. The threshold neutralisation at which this instability occurs is higher than the one for a dipole instability by a factor 2 to 4 depending on the type of quadrupole instability, for which four breathing modes exist. The required frequency spreads in ionic oscillations and beam mode frequencies for Landau damping to be effective are also lower by the same factors. In the CERN AA this instability appeared only after the dipole modes had been cured by transverse feedback damping.

## 8.6 Incoherent effects

These are single-particle phenomena, by comparison to coherent effects where the beam oscillates as a whole.

Tune shifts due to the space charge field created by trapped ions cause particles to cross resonance lines in the tune diagram:

$$nQ_x + mQ_y = \text{integer} \tag{69}$$

$$n, m = 0, \pm 1, \pm 2. \tag{70}$$

Very similarly to the excitation of non-linear resonances by the beam-beam interaction in colliding beam machines, the electrostatic field of ion clouds in neutralisation pockets causes non-linear detuning, and its uneven distribution may excite very high order resonances (up to 15th order in the CERN AA) [109]. This effect, still present at low neutralisation levels of a fraction of a percent, is very detrimental to antiproton accumulators which have low transverse emittance cooling rates.

## 8.7 Instabilities in electron beams

In electron storage rings that trap ions, the emission of synchrotron radiation by the circulating particles results in high damping rates of their transverse motion. It is unlikely that high order resonances can be harmful, since their strength decreases as their order increases. Otherwise, the physics of ion-beam interactions is similar to antiproton accumulators or proton machines. Coupling resonances excited by ions yield spectacular emittance effects, in particular in the vertical plane, where normally the beam size is naturally very small [78].

Transverse coherent instabilities may induce pulsations of the beam size. This occurs when the emittance growth rate due to an ion-induced instability is larger, and the neutralisation rate smaller than the natural damping rate. As the beam inflates ions are chased away by the instability, the beam then shrinks, neutralisation builds up again and the process repeats itself in a sort of relaxation mechanism [13].

Experience on modern machines shows that ion-induced tune shifts of a few $10^{-3}$ caused by neutralisation levels of a few percent or less, already present annoying emittance effects.

## 9. DIAGNOSTICS AND PHENOMENOLOGY

As described in the preceding sections, the presence of trapped species in a beam is diagnosed qualitatively by observation of the beam behaviour: intensity limitations, reduced particle lifetime, emittance blow up, coupling and coherent beam instabilities. To assess semi-quantitatively an effective degree of neutralisation is very difficult. The experimenter has essentially three possible means at his disposal, namely tune-shift measurements, ion clearing-current recordings and Bremsstrahlung diagnostics. None of these means is always applicable or absolutely rigorous. Nevertheless they can be powerful tools in understanding the rich phenomenology of ion-trapping physics.

### 9.1 Tune-shift measurements

The incoherent space-charge tune shift expressions (Eqs. (55), (56)) provide, in principle, a means to measure the average neutralisation in a beam. This supposes the possibility of clearing the beam of its ions (clearing electrodes, beam shaking, asymmetric bunch filling) so that from measurements of the tunes $Q_x$, $Q_y$ with and without ions, the neutralisation $\eta$ can be obtained from Eqs. ((50), (51)). $\Delta Q_{xy}$ are usually extracted from frequency spectral analysis of a transverse betatron sideband (the difference signal from a pair of electrodes around a mode frequency $((n \pm Q)\Omega)$. Since the tune shift produced by ions is in this case incoherent, care must be taken to measure the incoherent tune (Schottky noise). This is an easy matter for unbunched beams like antiprotons. For electron beams, transverse coupled-bunch mode signals are more difficult to interpret precisely, particularly in cases of high neutralisation levels. Obtaining the maximum frequency shift experienced by a particle by comparison of two transverse spectra is only approximate; it pre-supposes the precise knowledge of the transverse distributions of the beam particles and of the ions. For ions, this distribution may vary from place to place around the ring and it cannot be measured. The measurement shown on Fig. 18, taken from the CERN AA, requires unfolding several spectra of the betatron oscillation amplitudes and the energy distributions of the beam particles [77]. If one supposes identical initial bi-gaussian transverse distributions for the beam particles and ions, the maximum frequency shifts $\Delta Q_{xy}$ can be obtained by multiplying the shift of the peaks of the distributions by approximately 2.5 [87].

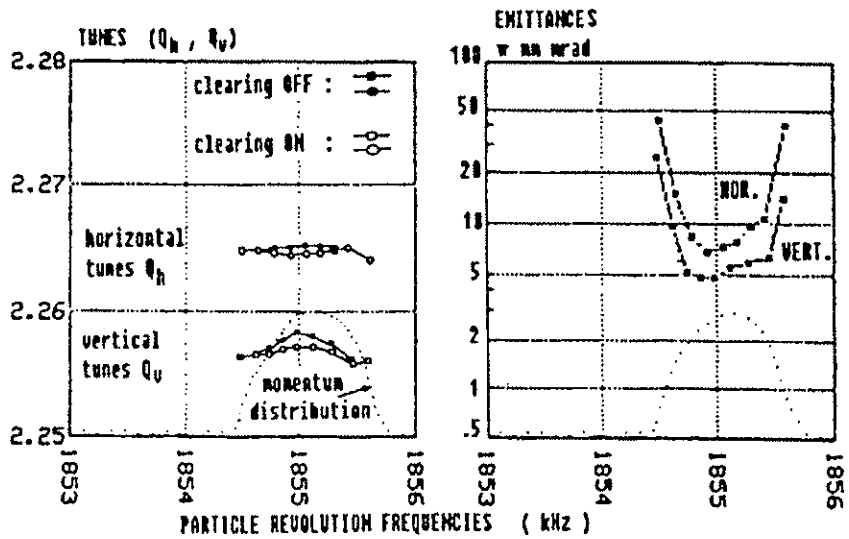



Fig. 18 Left: antiproton transverse tunes in the AA as a function of their momentum, with full (clearing off) and little (clearing on) neutralisation. Right: particle emittances. The tune shift with clearing off is consistent with an average neutralisation close to unity.

### 9.2 Clearing-current recordings

If ion clearing is procured by electrodes, and if secondary emission phenomena and direct photo-emission of electrons from these electrodes can be neglected (a serious difficulty for electron storage rings), the current drawn by electrodes is a direct measurement of the ion charge produced by

the beam. Electrodes are usually spread more or less evenly around a ring, with interspacing distances ranging from a few decimetres to several meters. Stable ions produced outside the field of action of electrodes are channelled by the beam to the nearest electrode, provided no potential barrier exists (magnet fringe fields, electrostatically charged insulators, etc.). On the other hand, unstable ions which are for instance destabilised by the bunching of the beam or shaken out of the beam potential by beam shaking, either voluntarily applied or resulting from ion-driven instabilities, do not end up on electrodes. Their absence, diagnosed as a temporary deficit on the recorded clearing current, provides a powerful means to understand the neutralisation phenomenology. The CERN AA for instance, equipped with a sophisticated ion clearing current monitoring system, has produced many interesting experimental results [88]. In machines with synchrotron radiation, clearing current measurements are rendered difficult — if not impossible - by the photo-emission of electrons produced by impinging photons on electrodes. On the CERN EPA however, a clearing electrode equipped with a shielding mask against synchrotron radiation and located at the end of a long straight section, has yielded some interesting measurements of the relative neutralisation levels for various bunch distributions (see Fig. 19), thus confirming theoretical predictions qualitatively [89].

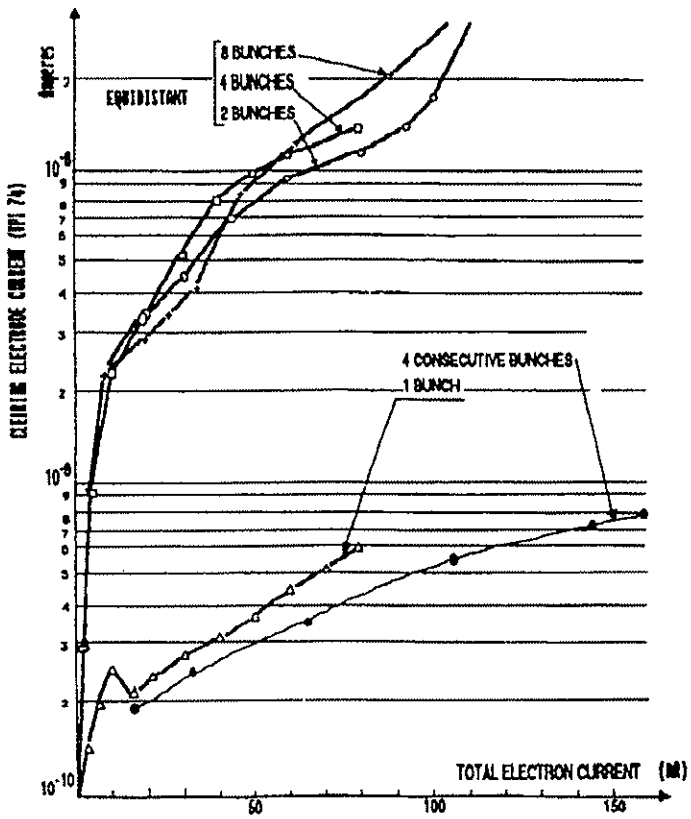



Fig. 19 Clearing current in the CERN electron accumulator for various bunch distributions. The less the neutralisation, the less the clearing current.

## 9.3 Bremsstrahlung measurements

High energy photons (gamma rays) are produced when the electrons of a circulating beam collide with the nuclei of ions trapped in the beam. By counting the number of these *bremsstrahlung* photons produced per second, one can estimate the density of target ions along the beam path seen by the detection system. This is a Cerenkov (lead-glass) or scintillator counter of dimensions large enough to contain the electromagnetic shower produced by the gammas. The whole system is aligned with the beam, looking through a metallic window in a bending magnet vacuum chamber at a straight section or a section of the beam orbit inside the dipole magnet.

The bremsstrahlung radiation covers a wide spectrum with energies extending up to the kinetic energy of the primary electron and has a very sharp collimation with a characteristic opening angle given by $1/\gamma$ where $\gamma$ is the relativistic energy of the stored beam in units of the rest energy. This allows the use of lead collimation to prevent gamma rays from other sources (mainly shower debris from lost beam electrons) from reaching the detector.

An important advantage of this technique with respect to the more conventional tune shift measurements described above is that it allows the measurement of the local neutralisation rather than the average over the whole machine. However, since many different ion species may be trapped, some assumption must be made concerning the trapped-ion composition. In fact, since the total bremsstrahlung yield is proportional to the square of the atomic number of the target atom, the count rates are proportional to a weighted average of the ion densities, each ion species contributing with a weight $Z^2$. Also, collisions with the neutral residual gas molecules produce an additional bremsstrahlung yield which one must either evaluate theoretically from the known residual gas composition or experimentally, e.g., by running the machine with positrons or comparing different machine conditions (e.g. clearing electrodes on/off).

The first observations [90] were made in the KEK Photon Factory storage ring in the form of a correlation between sudden changes in lifetime with bursts of bremsstrahlung events. Later [91], bremsstrahlung count rates were registered in synchronism with a vertical blow-up caused by the accumulation of ions. The count rates as a function of time showed a rapid decrease at the time of the blow-up, followed by a slow increase which was attributed to the creation of new ions. The time constant for this slow rise was found to be in good agreement with the expected ionisation time for CO.

In Refs. [92, 93], a vertical blow-up was artificially produced by means of an external resonant excitation and the variation of the bremsstrahlung count rates was observed in both a uniform filling and partial filling modes of operation. Since the bremsstrahlung yield from the neutral residual gas molecules is expected to be independent of the beam dimensions, a decrease in the count rates during the blow-up was interpreted as a sign of the presence of trapped ions, and the comparison between the uniform filling and the partial filling confirms the theoretical prediction that fewer ions are present in the partial filling.

In Ref. [94], a study of the longitudinal motion of trapped ions is presented. Several gamma-ray counters were aimed at successive straight sections of the UVSOR storage ring. By looking at the time response of the count rates in these counters to a pulsed high voltage applied to a clearing electrode, the authors arrive at a spread in longitudinal ion velocities substantially larger than that expected from thermal considerations only.

## REFERENCES

A considerable number of papers on the subject of neutralisation has been published. An attempt has been made to classify the large number of useful references.

[1] Y. Baconnier, A. Poncet, P.F. Tavares, "Neutralisation of Accelerator Beams by Ionisation of the Residual Gas", CERN 94-01, 1994.

[2] E.D. Courant and H. Snyder, "Theory of the alternating gradient synchrotron", Ann. Phys. (N.Y.) vol. 3, p. 1 (1958).

[3] J.D. Jackson, "Classical electrodynamics", John Wiley & Sons, New York, 1962.

[4] B. Rossi, "High energy particles", Prentice-Hall, Inc., Englewood Cligs, N.J., U.S.A., 1956.

[5] L. Spitzer, Jr., "Physics of fully ionized gases", Interscience Publishers, Inc., New York,1956.

## REVIEWS, EARLY PAPERS

[6] B. Angerth, "Review of studies on beam neutralization in storage rings", CERN AR/Int.SG/65-1 (1965).

[7] R. Jolivot, "Anneaux de Stockage. Le piegeage des ions dans ACO et leur balayage", Rapport technique 75-63/RJ- FB, Ecole Normale Superieure - Faculte des Sciences, Orsay, Laboratoire de l'Accelerateur Lineaire (1963).

[8] R.D. Kohaupt, Ion clearing mechanism in the electron-positron storage ring DORIS" , DESY MI-71/2 (1971).

[9] R. Alves Pires, J. Marriner, W. Marsh, A. Poncet, J. Rosensweig, and P. Zhou, Fermilab III Instabilities Workshop, Saint Charles, Illinois (1990).

[10] A. Poncet, "Ion trapping and clearing", CERN/MT 90-1 (ES). Course given at the the CAS, Uppsala, Sweden (October 1989) CERN 90-04 (1990).

[11] A. Poncet, "Ions and neutralization", CERN/MT 91-01. Talk given at the joint CERN/US particle accelerator school, Hilton Head Island, South Carolina, USA (November 1990) Springer Verlag, Lecture notes in Physics, n.400 (1992).

[12] D. Poteaux, "Piegeage des ions dans un anneau e+e-", Rapport technique 29-69 DP/LN Laboratoire de l'Accelerateur Lineaire, Orsay (1969).

[13] E.M. Rowe, "Trapped ion effects and their treatment in electron storage rings.", American Vacuum Society Series 5, AIP conf. proceedings No.171, p.193, New York (1988).

**TRAPPED ION DYNAMICS**

[14] Y. Baconnier and G. Brianti, "The stability of ions in bunched beam machines", CERN/SPS/80-2 (1980).

[15] C. Bernardini , "Space-charge efects in electron- synchrotrons", Nuovo Cimento, Vol X, N.5, p.804 (1988).

[16] L. Evans and D.J. Warner, "Space-charge neutralisation of intense charged particle beams: some theoretical considerations", CERN MPS/LIN 71-2 (1971).

[17] E. Fischer, "Space charges in electron storage rings and the removal of positive ions by a D.C. clearing field", PS/Int. AR/60- 14 (1960).

[18] P.F. Tavares, "Betatron coupling in ion loaded electron beams", CERN PS/92-54 (LP) (1992).

[19] P.F. Tavares, "Transverse distribution of ions trapped in an electron beam", CERN PS/92-55 (LP) (1992).

**PARTIAL FILLING, GAPS**

[20] M. Barton,"Ion trapping with asymmetric bunch filling of the NSLS VUV ring", Nucl. Instrum. Methods, A243 p.278 (1986).

[21] D. Douglas, "Ion stability in bunched electron beams", IEEE Trans. Nucl. Sci. Vol. Ns-32, No.5 p.2294 (1985).

[22] S. Sakanaka, "The stability of ions in partially filled mode operation in the electron storage ring", KEK 86-17 (1986).

**LONGITUDINAL MOTION**

[23] Y. Miyahara, K. Takayama, and G. Horikoshi, "Dynamical analysis of the longitudinal motion of trapped ions in an electron storage ring", Nucl. Instrum. Methods, A270 p.217 (1988).

[24] D. Sagan, "Some aspects of the longitudinal motion of ions in electron storage rings", Nucl. Instrum. Methods, A307 p.171 (1991).

## BEAM POTENTIAL CALCULATIONS AND MEASUREMENTS

[25] M. Bassetti and G.A. Erskine, "Closed expressions for the electrical field of a two dimensional Gaussian charge", CERN-ISR-TH/80-06 (1980).

[26] 0. Grobner and K. Hübner, "Computation of the electrostatic beam potential in vacuum chambers of rectangular cross-section", CERN/ISR-VA/75-27 (May 1975).

[27] M. Gygi-Hanney and B. Zotter, "Field strength in a bigaussian beam", LEP theory note 44 (1987).

[28] L.J. Laslett, "Potential of a uniformly charged beam with an elliptical cross-section", ERAN-44 (November 1969).

[29] L.J. Laslett, "Image field of a straight beam of elliptical cross-section", ERAN-49 (1970).

[30] R. Alves-Pires, "Beam dimensions and beam potential in the CERN Antiproton Accumulator complex", CERN PS/87-70 AA (1987).

[31] R. Alves-Pires, "Conformal mapping for two-dimensional electrostatic beam potential calculations", CERN PS/87-66 (1987).

[32] E. Regenstreif, "Potential and field created by an elliptic beam inside an infinite cylindrical vacuum chamber of circular cross-section", CERN/PS/DL 77-37 (1977).

[33] E. Regenstreif, "Potential and field created by a rectangular beam inside an infinite cylindrical vacuum chamber of circular cross section", CERN PS/DL 77-31 (1977).

[34] P. Strubin, "ISR Performance Report-Run 750, 7 July 1976, 26 GeV-Measure of the beam potential as a function of beam current, ISR-VA- ISR-VA-PS/srn (24th August, 1976).

## DUST TRAPPING

[35] P. Marin, "Microlosses of beam current in Super-ACO operated with electrons", Lure RT/90-01 (1990).

[36] P. Marin, "Longitudinal motion of a dust particle along a beam of electrons in a storage ring", Lure RT/90-06 (1990).

[37] P. Marin, "Observation of bremsstrahlung on dust particles trapped in electron beams at DCI and Super-AM, Lure RT/91-03 (1991).

[38] F. Pedersen, "Effects of highly charged, solid macroparticles in negatively charged circulating beams", CERN PS/87-25 (1987).

[39] H. Saeki, T. Momose, and H. Ishimaru, "Observations of dust trapping phenomena in the Tristan accumulation ring and a study of dust removal in a beam chamber", Rev.Sci.Instrum. Vol.62 No.4 p.874 (1991).

[40] D. Sagan, "Mass and charge measurement of trapped dust in the CESR storage rings", Cornell BN 92-12 (1992).

[41] C.J. Bochetta and A. Wrulich, "The trapping and clearing of ions in ELETTRA", Sincrotrone Trieste ST/M-88/26 (1988).

[42] E. Bozoki and H. Halama, "Ion related problems for the XLS ring", Nucl. Instrum. Methods, A307 p.156 (1991) .

[43] G. Brianti, "Upper limits of the average luminosity in the SPS p-p collider", CERN/SPS/DI/80-1 (20th February, 1980).

[44] F. Caspers, J.P. Delahaye, J.C. Godot, K. Hubner, and A. Poncet, " EPA beam-vacuum interaction and ion-clearing system", EPAC p. 1324, Rome (1988).

[45] S. Chattopadhyay, "Coherent instability and ion trapping considerations for Alladin lattices", LBL-19281 (1985).

[46] R.C. Gluckstern and A.G. Ruggiero, "Ion production and trapping in electron rings", BNL-26585 (1979).

[47] J. Herrera and B. Zotter, "Average neutralisation and transverse stability in Isabelle", BNL 50980 UC-28 (1978).

[48] T. Kasuga, "Ion clearing system for the UVSOR storage ring", Jpn. J. Appl. Phys. Vol.2 No.11, p.1711 (1986).

[49] R.Z. Liu, "Considerations on ion accumulation and clearing in Spear". SSRL ACDNote 25 (1984).

[50] F. Pedersen, A. Poncet, and L. Soby. "The CERN Antiproton Accumulator clearing system with ion current as a residual neutralization diagnostic", CERN PS/89-17 ML (1989).

[51] A. Poncet, "Ion clearing in EPA", PS/ML/Note 83-1 (1983).

[52] A. Poncet, "Quel vide pour I'EPA?", PS/ML/Note 83-3 (1983).

[53] D. Sagan, "Ion trapping in the CESR B-Factory", Cornell, CBN 91-2, (May 1990).

[54] Y. Sun, "A preliminary report on ion clearing in Spear", SSRL ACD-note 83 (may 1990).

[55] P.F. Tavares, "The ion clearing system of the UVX-2 synchrotron radiation source", Proceedings 3rd EPAC, p.1647, Berlin (1992).

[56] E.J.N. Wilson, "Does ion trapping produce a large space charge •Q in the LEP injector chain?", CERN/SPS-DI LTD (1980).

[57] Workshop on pp in the SPS. (Theoretical aspects of machine design, March 1980), SPS-pp-1, 9th May, 1980.

[58] P. Zhou and J.B. Rosenzweig, "Ion trapping in Tevatron with separated orbits", Fermilab III Instabilities Workshop, Saint Charles, Illinois (1990).

**NEUTRALISATION IN THE ISR**

[59] B. Angerth, E. Fischer, and 0. Grobner, "The clearing fields of the ISR, CERN ISR-VA/71-47 (1971).

[60] R. Calder, E. Fischer, 0. Grobner, and E. Jones, "Vacuum conditions for proton storage rings", CERN/ISR-VA/74-26 (1974).

[61] E. Fischer, "Clearing fields for the ISR", ISR-VAC/66-15, (1966).

[62] E. Fischer and K. Zankel, "The stability of the residual gas density in the ISR in presence of high intensity proton beams". CERN-ISR-VA/ 73-52 (1973).

[63] 0. Grobner, "ISR Performance Report- Clearing studies using Langmuir probe technique Run 139, 22 GeV, 20 bunches", ISR-VA/OG/ss (1972).

[64] 0. Grobner, "ISR Performance Report- Run 142 , 26 GeV, 20 bunches, Ring 1, 26 FC, Beam Decay as Function of Clearing Voltage", ISR-VA/OG/ss, (1972).

[65] 0. Grobner and R.S. Calder, "Beam induced gas desorption in the CERN intersecting storage rings", IEEE Trans.Nucl.Sci. , Vol. NS-20, No 3 p.760 (1973).

[66] 0. Grobner, "ISR Performance Report- Attempt to measure neutralisation shifts", ISR-VA/OG/srn (1974).

[67] 0. Grobner and P. Strubin, "ISR Performance Report -Decay rate due to nuclear scattering pressure as determined from the clearing currents", ISR-VA/OG/sm, llth July, 1974.

[68] 0. Grobner, "ISR Performance Report- Clearing, Pressure Bump, Beam Behaviour", ISR-VA/OG/srn (1975).

[69] 0. Grobner, "Performance study on proton- proton storage rings at several hundred GeV/c. Neutralisation and vacuum requirements in ISR". CERN/ISR-AS/74-67 (1974).

[70] 0. Grobner, "ISR Performance Report - RF Clearing, e-p lines and electron noise", ISR-VA/OG/srn (1975).

[71] 0. Grobner, "An estimate of the rate of electron removal by RF-clearing". ISR VA/OG (1975).

[72] 0. Grobner, "ISR Performance Report- -Run 904, 26 GeV, 6.12.77. Observation of beam blow-up by multiple scattering on neutralising electrons", ISR-VA/OG/srn (1978).

[73] 0. Grobner, "Ion clearing of anti-proton beams in the ISR", CERN ISR-VA/OG/srn (1978).

[74] 0. Grobner, "ISR Performance Report-Run 389, Ring 1, Dependence of clearing current on neutralisation", ISR-VA/OG/srn (1984).

[75] Technical Note - ISR Vacuum Group, "The behaviour of ions in presence of a bunched antiproton beam in the ISR", ISR-VA/EF-sm, 7th November 1978.

[76] D.G. Koshkarev, "Concerning a new way of removing electrons from an unbunched electron beam", ISR-DI/DGK (1974).

[77] S. Van der Meer, "Measurement of transverse stack emittances from Schottky scans", PS/AA/Note 84-11 (1984).

**NEUTRALISATION EXPERIMENTS**

[78] M.F. Biagini et al., "Observation of ion trapping at Adone", llth Intemational conference on high energy accelerators,p. 687, Geneva (1980).

[79] E.V. Bulyak and V.I. Kurilko, "Low frequency transverse charge oscillations in an electron storage ring", JETP Letters Vol.34, No. 9 p.471 (1981).

[80] E.V. Bulyak and V.I. Kurilko, "Space charge neutralization of a beam in high current electron storage rings", Sov. Phys. Tech. Phys. Vol.27, No.2 p.194 (1982).

[81] T.S. Chou and H.J. Halama, "Trapped ions and beam lifetime in the NSLS storage rings", Proceedings of the EPAC, p.1773, Rome (1988).

[82] R. Cappi and J.P. Riunaud, "Trapped ion effects with electron beams in the CERN PS", CERN/PS 90-17 PA (1990).

[83] 0. Grobner, "ISR Performance Report- Measurement of the self-clearing rate as function of beam current, ISR- VA/OG/sm, 27th March, 1975.

[84] H. Halama and E. Bozoki, "Ion clearing and photoelectron production in the 200 MeV SXLS ring", BNL-46471 (1991).

[85] T. Kasuga, H. Yonehara, T. Kinoshita, and M. Hasumoto, "Ion trapping effect in the UVSOR storage ring", Jpn. J. Appl. Phys. Vol.24, No.g p.1212 (1985).

[86] P. Marin, "Positron versus electron behaviour in Super- ACO", Lure RT/Anneaux/90-03 (1990).

[87] W. Marsh, "How to measure beam neutralization in the accumulator", Fermilab III Instabilities Workshop, SaintCharles, Illinois (1990).

[88] F. Pedersen et al, "The CERN Antiproton Accumulator clearing system with ion current measurements as a residual neutralisation diagnostic", CERN PS/89-17 (ML) (1989).

[89] A. Poncet, 'Trapping of ions in the EPA electron beam. Stability conditions and diagnosis", CERN PS 88-14 ML (1988).

**BREMSSTRAHLUNG EXPERIMENTS**

[90] M. Kobayashi et al, "Observations of bremsstrahlung caused by ion trapping", Proceedings of the 5th Symposium of Accelerator Science and Technology, KEK (1984).

[91] H. Kobayakawa et al, "Observation of the ion trapping phenomenon with bremsstrahlung", Nucl. Instrum. Methods, A248 p. 565 (1986).

[92] S. Sakanaka et al, "Observation of ion trapping in single bunch operation at the Photon Factory storage ring", KEK 87-168 (1987).

[93] S. Sakanaka et al, "Differences in ion trapping between uniform and partial filling", Nucl. Instrum. Methods, A256 p.184 (1987).

[94] K. Watanabe et al., "Drift velocity and pulse response of trapped ions in a circulating electron beam", Jpn. J Appl. Phys., Vol.26, No.12 , L1964 (1987).

**IONISATION, HEATING, COOLING**

[95] U. Amaldi, "Fisica delle radiazioni", Boringhieri, Torino (1971).

[96] F. Lapique and F. Piuz, "Simulation of the measurement by primary cluster counting of the energy lost by a relativistic ionizing particle in argon", Nucl. Instrum. Methods, 175 p. 297 (1980).

[97] Y. Miyahaxa, "Photo-ionization of residual gas in electron storage ring", Jpn. J. Appl. Phys., Vol. 26, No. 9, p.1544

[98] F. Rieke and W. Prepjchal, "Ionisation cross-section of gaseous atoms and molecules for high-energy electrons and positrons" , Phys.Rev. A Vol. 6, No. 4 , p.1507 (1972).

[99] F. Sauli, "Principles of operation of multiwire proportional drift chambers", CERN 77-09 (1977).

[100] K. Symon, "Are clearing electrodes really necessary?", Private Comunication (1964).

**BEAM SHAKING**

[101] E. Bozoki, "Ion shaking in the 200 MeV XLS-ring". Proceedings 3rd EPAC, p.789, Berlin (1992).

[102] J. Marriner, D. Möhl, Y. Orlov, A. Poncet and S. Van der Meer, "Experiments and practice in beam shaking", CERN/PS/89-48 AR (1989).

[103] Y. Orlov, "The suppression of transverse instabilities caused by trapped ions in the AA by shaking the $\overline{p}$ beam", CERN/PS/89-01 (1989).

[104] R. Alves-Pires, "Beam shaking for the Fermilab Antiproton Accumulator", Femilab III Instabilities Workshop, Saint Charles, Illinois (1990).

[105] A. Poncet and Y. Orlov, "EPA machine experiment note - ion shaking tests", PS/ML/Note 89-1 (1989).

[106] J.B. Rosenzweig, "Beam-ion cyclotron resonance instability", Fermilab III Instabilities Workshop, Saint Charles, Illinois (1990).

[107] P. Zhou and J.B. Rosenzweig, "Ion clearing using Cyclotron shaking", Fermilab III Instabilities Workshop, Saint Charles, Illinois (1990).

**INSTABILITIES**

[108] G. Carron et al, "Observation of transverse quadrupole instabilities in intense cooled antiproton beams in the AA", CERN/PS/89-18 (AR), Proceedings PAC, p.803, Chicago (1989).

[109] A. Dainelli, "Antiproton-positive ion transverse instabilities in the CERN AA". A tune modulated direct map simulation", CERN/PS/87-13 (AA) (1987).

[110] H.G. Hereward, "The instability of radial betatron oscillations in the CPS", CERN MPS/Int. DL 64-8 (1964).

[111] E. Jones et. al, "Transverse instabilities due to beam trapped ions and charged matter in the CERN AA", IEEE Trans.Nucl.Sci., Vol. NS -32, p. 2218 (1985).

[112] Y. Kainiya, M. Yzawa, T. Katsura, and M. Kihara, "Vertical instability caused by ion trapping in KEK-Photon factory storage ring", Proc. of 5th Symp. Acc. Sci. Tech. p.148 (1984).

[113] E. Keil and B. Zotter, "Landau-damping of coupled electron-proton oscillations", CERN-ISR-TH/71-58 (1971).

[114] D.G. Koshkarev and P.R. Zenkevich, "Resonance of coupled transverse oscillations in two circular beams", Part. Acc. Vol.3 p.1 (1972).

[115] L.J. Laslett, A. Sessler, and D. Möhl, "Transverse two-stream instability in the presence of strong species-species and image forces", Nucl. Instrum. Methods, Vol. 121 p.517 (1974).

[116] F. Pedersen, and A. Poncet, "Proton-antiproton instability in the CERN AA", CERN PS/AA/ME Note 81, 1981

[117] R. Alves-Pires et. al, "On the theory of coherent instabilities due to coupling between a dense cooled beam and charged particles from the residual gas", CERN/PS/89-14 (AR), PAC p.800 Chicago (1989).

[118] R. Alves-Pires and D. Möhl, "Landau damping of coupled quadrupole modes", PS/AR/Note 90-09 (1990).

## ANNEX I

## COMPUTATION OF THE MEAN VELOCITY IN THERMAL MOTION

By integration from $-\infty$ to $+\infty$ of the Boltzman equation (Section 4.1) over $\upsilon_x$ and $\upsilon_y$ one obtains the equation

$$\frac{dn}{d\upsilon_x} = d_m \sqrt{\frac{m}{2\pi kT}} e^{-m\frac{\upsilon_x^2}{2kT}}.$$

The mean value of $\upsilon_x$ is by definition

$$\langle \upsilon_x \rangle = \frac{1}{d_m} \int_{-\infty}^{+\infty} v_x \frac{dn}{d\upsilon_x} d\upsilon_x = 0.$$

We are interested in the mean velocity in one direction. Either:

$$\langle |\upsilon_x| \rangle 0 \geq = \frac{1}{d_m/2} \int_0^{+\infty} \upsilon_x \frac{dn}{d\upsilon_x} d\upsilon_x$$

or

$$\langle |\upsilon_x| \rangle = \frac{1}{d_m} \int_{-\infty}^{+\infty} |\upsilon_x| \frac{dn}{d\upsilon_x} d\upsilon_x$$

In both cases the result is

$$\langle |\upsilon_x| \rangle = \sqrt{\frac{2kT}{\pi \cdot m}}$$

and therefore $\langle |\upsilon_x| \rangle = \upsilon_m / 2$ and not $\upsilon_m/4$ as quoted in several papers.

## ANNEX II

## ELLIPTICAL VACUUM CHAMBER AND BEAM

The detailed calculation of the potential in a rectangular vacuum chamber of width $2w$ and height $2h$ induced by a beam with current $I$, velocity $\beta c$, height $2b$ and width $2a$ has been made [26]. The following formulae allow the detailed computation of the potential at the centre of the beam:

$$U_0 = \frac{I}{\beta c\, \varepsilon_0}$$

$$\eta_s = \frac{\pi}{2w} s$$

$$C_s = \frac{U_0}{4bw} \frac{g_s}{\eta_s}$$

$$g_s = \frac{\cos \eta_s (\omega - a) - \cos \eta_s (w + a)}{\eta_s \cdot a\left(1 - n_s^2 \frac{a^2}{\pi^2}\right)}$$

Then the potential in the centre is

$$V(0,0) = \sum_s \left(1 - \frac{\cosh[n_s(h-b)]}{\cosh \eta_s \cdot h}\right) C_s \sin \eta_s w \cdot$$

## ANNEX III

## NUMERICAL VALUE (MKS UNITS)

| | | | |
|---|---|---|---|
| $\varepsilon$ | = | $8.85 \times 10^{-12}$ | ($F \cdot m^{-1}$ or $C\ v^{-1}\ m^{-1}$) |
| $\mu_0$ | = | $4\pi \times 10^{-7}$ | ($H \cdot m^{-1}$ or $A^{-1}\ V\ s \cdot m^{-1}$) |
| $C$ | = | $3 \times 10^{8}$ | ($m\ s^{-1}$) |
| $r_p$ | = | $1.53 \times 10^{-18}$ | (m) |
| $r_e$ | = | $2.82 \times 10^{-15}$ | (m) |
| $m_p$ | = | $1.67 \times 10^{-27}$ | (kg) |
| $m_e$ | = | $9.10 \times 10^{-31}$ | (kg) |
| $m_p/e$ | = | $1.0 \times 10^{-8}$ | ($T \cdot s$ or $kg\ C^{-1}$) |
| $m_e/e$ | = | $5.7 \times 10^{-23}$ | ($T \cdot s$ or $kg\ C^{-1}$) |
| $h$ | = | $6.85 \times 10^{-16}$ | $eV \cdot s$ |
| $k$ | = | $1.4 \times 10^{-21}$ | J |

# LEAK DETECTION

*N. Hilleret*
CERN, Geneva, Switzerland

**Abstract**
Various methods used for leak detection are described as well as the instruments available for this purpose. Special emphasis is placed on the techniques used for particle accelerators.

## 1. INTRODUCTION

Leak detection is a very important step in the production of vacuum. It is needed after the production of a vacuum vessel to check that the tightness specifications are fulfilled, during and after the assembly of these vessels to locate the possible leaks created during assembly, and finally during the installation of the vessel, to guarantee that the process can be carried out under the required pressure and gas composition conditions. Hence methods of ever-increasing sensitivities have been developed to follow the ever more stringent requirements of the industry. After a summary of the various methods used to locate leaks, the most widely used leak detector will be presented with its different types. Some practical cases will then be reviewed in the context of accelerator operation.

## 2. METHODS

Depending on their size, leaks can have various effects, which can be used for their location. All methods are based on the variation of a physical property measured on one side of the vacuum vessel wall while the pressure or the nature of the gas is changed on the other side. Big leaks, involving large gas flow can generate mechanical effects, smaller leaks require finer methods. These rely on the change of the residual gas physical properties when the nature of the gas leaking into the system (tracer gas) is changed. Both categories will be reviewed hereafter. A list of the possible leak detection methods with their sensitivities can be found in Ref. [1]. A comprehensive review of the methods and apparatus for leak detection can be found in Refs. [2, 3].

### 2.1 Mechanical effects

As explained above the production of measurable mechanical effects requires a sufficient energy and hence these methods are limited to relatively large leaks. The emission of sound or the deflection of a flame can occur in the case of leaks ($10^{-2}$ to $10^{-3}$ $Pa.m^3.s^{-1}$), usually limiting the pressure in the rough vacuum domain. Ultra-sound detectors can also be used to monitor the oscillations produced by the gas in the vicinity of leaks. A more sensitive method is the formation of bubbles when water is spread on the leak, the vacuum vessel being pressurised to several bars of over pressure. The detection limit in that case can reach $10^{-5}$ $Pa.m^3.s^{-1}$ if a wetting agent is added to water and it is good practice to pressurise the vessel before immersing it in the liquid as the molecules might be unable, despite the gas pressure, to flow through the surface film because of the surface tension of the liquid. These methods have the advantages to be simple, very quick to carry out, and able to locate leaks. Their sensitivity and time constant are independent of the volume of the vessel. They apply mainly to the high-pressure region.

### 2.2 Tracer gas

In the case of small leaks, the energy of the gas flow is insufficient to generate measurable mechanical effects. In that case a greater sensitivity is obtained by relying on the variation of physical properties of the residual gas for which accurate and sensitive measurement methods are available. When the composition of the residual gas is modified by the injection in the vicinity of the leak, of a gas (the tracer gas) changing locally the air composition, these properties are altered and this alteration can be measured for determining the size and the position of a leak. The tracer gas must have the following properties [4], for the case of helium leak detection:

Be unique in the mass spectrum of the residual gas in the system and practically non-existent in the normal surrounding atmosphere.

Be readily removable from the system by pumping and should not contaminate the systems

Have a low viscosity.

Many properties of the residual gas can be used to monitor its composition changes. The most widely used are the heat conductivity, the ionisation cross section, the pumping speed and the conductance. The variation of heat conductivity is traced using a Pirani gauge and using alcohol, helium or carbon dioxide. The pressure variation on the gauge will be positive for helium and negative for alcohol or carbon dioxide. The variation in ionisation cross section can be used by monitoring the signal of an ionisation gauge and this method, very useful in accelerators, will be described in the Section 4.3. Lastly, the mass of the molecules can also be used to trace leaks and this very sensitive and widespread method is described in the next section.

## 3. HELIUM LEAK DETECTORS

### 3.1 History and principle

At the origin of the helium leak detection method was the "Manhattan Project" and the unprecedented leak-tightness requirements needed by the uranium enrichment plants. The required sensitivity needed for the leak checking led to the choice of a mass spectrometer designed by Dr. A.O.C. Nier [5] tuned on the helium mass (see the tracer gas definition above). Because of its industrial use, the material choice (originally glass) turned out to be unbearably fragile and after many complaints by the users, a new metallic version was developed and constructed. The sensitivity of the apparatus was in 1946 $\sim 10^{-7}$ Pa. $m^3.s^{-1}$ and it increased to $\sim 10^{-10}$ Pa. $m^3.s^{-1}$ by 1970. Nowadays the quoted sensitivity of the most sensitive detectors is $\sim 10^{-13}$ Pa .$m^3.s^{-1}$, a factor $10^6$ gain within 50 years.

The central piece of the helium leak detector is the cell in which the residual gas is ionised and the resulting ions accelerated and filtered in a mass spectrometer. Most of the current detectors use, as in the original design, a magnetic sector to separate the helium ions from the other gases. Permanent magnets are generally used to generate the magnetic field. The adjustment needed for the selection of the helium peak is made by varying the ion energy. A schematic layout of a helium leak detection cell is given in Fig. 1.

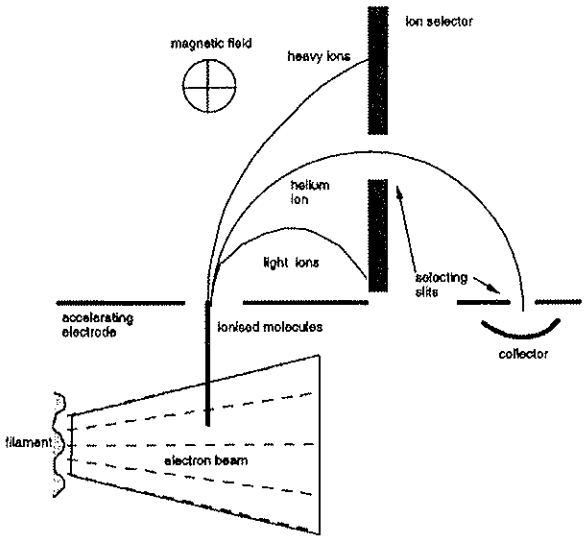



Fig. 1 Schematic layout of a leak detection cell.

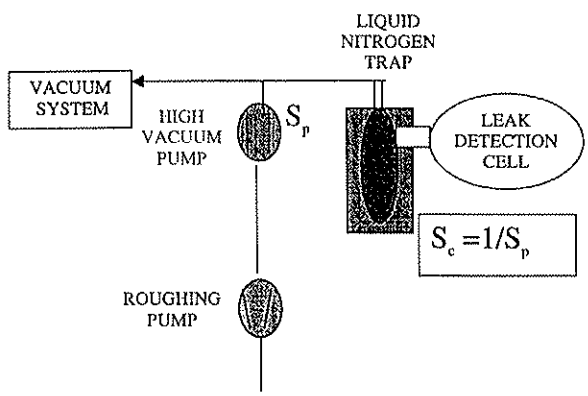



Fig. 2 The direct-flow layout.

To detect small leaks, the currents to be measured are very small: At the highest sensitivity (in the $10^{-13}$ Pa .$m^3$.$s^{-1}$ range), currents as low as femtoamperes have to be measured. This is achieved thanks to the use of an electron multiplier in the most modern detectors. If the cell of a leak detector is not much different from the original design, the pumping system has considerably changed, the original diffusion pumps now being replaced by turbomolecular pumps or dry molecular-drag pumps.

The sensitivity of the helium leak detector is given by the ratio between the helium flow through the leak and the partial pressure increase in the cell. In order to increase the sensitivity, the pumping speed of the tracer gas has to be reduced. This must be done without diminishing the pumping speed for the other gases (mainly water as leak detection usually takes place in unbaked systems) in order to keep the appropriate operating pressure for the filament emitting the ionising electrons. Selective pumping is therefore needed to provide a high pumping speed for water and a low pumping speed for helium. The various ways to achieve this will be presented in the next Section.

### 3.2 The direct-flow method

In direct-flow leak detectors, the vacuum system is connected according to Fig. 2 to the leak detection cell and to its pump.

To provide selectivity, a liquid-nitrogen trap is installed between the leak detection cell and the input flange of the detector. The high pumping speed of the trap for water allows the pressure in the cell to be lowered and, hence, leak detection to be started earlier without losing helium sensitivity. This arrangement has been very successful and such detectors were able, on small systems, to detect leaks as low as $10^{-12}$ Pa.$m^3$.$s^{-1}$. They used diffusion pumps and were sensitive to misuse such as inadequate venting leading to the oxidation of the diffusion-pump oil. Furthermore when the detector was operating, the nitrogen trap needed to be refilled periodically: an easily accessible source of liquid nitrogen was required. This trap also impeded the diffusion-pump oil-vapor to backstream to the cell. This is very important to avoid the deposition in the cell container of insulating coatings formed during the interaction of the ionising electron beam with the oil vapors. Lastly, because of the warming up time of the diffusion pump, their start-up required approximately 15 minutes and the sequence to operate them was complicated. On the other hand, the tested vacuum system was exposed to the residual gas of a trapped diffusion pump, which in these time was the most common pumping system to produce high vacuum. These detectors were used for most of the leak checks in high vacuum systems until the mid 80's. Nowadays they are being replaced by counter-flow detectors

### 3.3 The counter-flow method

The possibility of using this method for leak detection was mentioned by W. Becker in 1968 [6] and later described elsewhere [7, 8]. Since then this method has become widely adopted in the field of helium leak detection.

The method is based on the fact that the compression ratio of turbomolecular pumps and diffusion pumps increases very quickly with the mass of the pumped gas. Hence it is possible by injecting the gas from the tested vessel at the exhaust of the pump to obtain at its inlet a backstreaming flux largely enriched in lighter gases. For example, for a turbomolecular pump running at full speed the ratio between the compression ratio for helium and water vapor is $10^4$. A scheme of such a leak detector is given in Fig. 3. Although the scheme was initially proposed both for diffusion pumps and turbomolecular pumps, most of the existing counter-flow detectors use turbomolecular pumps.

A major drawback of this simple scheme is that the tested vessel is connected directly to the forepump and can be contaminated by oil vapor. Furthermore, the stability of the pumping characteristics of the forepump are very important to ensure the stability needed for accurate leak detection. Lastly the pumping speed of the forepumps used for such application is small and the time constant for the leak detection is hence greatly increased.

To remedy these problems more sophisticated commercial leak detectors have been developed using specially designed turbomolecular pumps. The vacuum system is pumped by a first turbomolecular pump ensuring clean pumping with a high pumping speed and correspondingly short time constant. A second integrated turbomolecular pump, having a common outlet flange, is

connected to the leak detection cell, Fig. 4. These advantages are also obtained by using a simple counter-flow leak detector connected to the outlet of the turbomolecular pump in a roughing station (see Fig. 5 in the chapter on mechanical pumps in these proceedings). A similar configuration can be obtained by admitting the gas at various intermediate stages of a turbomolecular pump. Depending on the size of the leak, or on the total gas flux to be evacuated, the gas can be injected in the high-pressure stages ("gross" leak configuration) or closer to the low-pressure stages for an enhanced sensitivity.

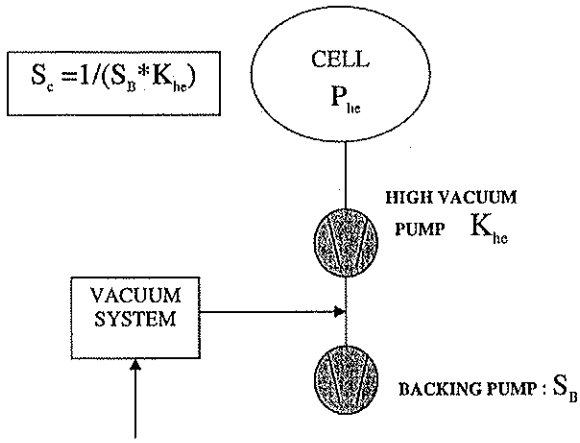



Fig. 3 Counter-flow leak detection method

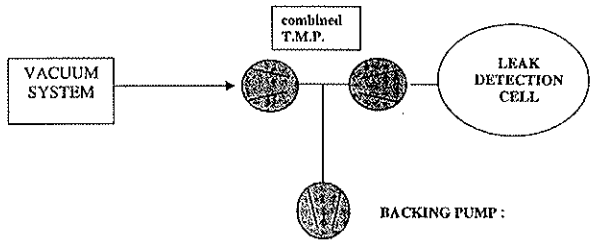



Fig. 4 Combined turbo-molecular-pump leak detector

With the development of dry pumps, new counter-flow detectors have been built using dry pumping systems to avoid contamination by the oil vapor coming from the detector pumps. Another interesting improvement is the introduction of molecular-drag pumps in place of the high-pressure stages of the turbomolecular pump. Because of the high pressure tolerable at the outlet of such pumps, leak detection is made possible with a much higher pressure in the system, thus providing the possibility of an early detection (i.e. a high pumped flow).

The advantages of counter-flow detectors are numerous: they do not need liquid nitrogen, they can be easily transported, they are more quickly put into operation, because of the inertia of the pump rotor, the filament is more protected in case of a sudden air inrush. In general they are more robust and require less maintenance, their simple mode of operation also permits easier automation and remote control. Most of the recent leak detectors use the counter-flow method. Nevertheless the direct-flow geometry offers for the case of very small leaks ($< 10^{-12}$ Pa.m$^3$.s$^{-1}$) a better stability of the signal and hence an up-to-now unsurpassed sensitivity.

### 3.4 The detector probe method (sniffer)

This method is very similar to the two preceding ones as it uses the same apparatus: a helium leak detector. In the case of the "sniffing method", the vessel itself is pressurised with the tracer gas (usually helium). Then the gas outside the vessel is tested to detect the presence of the tracer gas. This testing is made by admitting, through a needle valve or a capillary tube, the gas in a conventional helium leak detector (either direct or counter-flow) at its maximum admissible pressure. This method is very useful for locating large leaks in vessels contaminated by, or containing, helium. The detection limit of the method is close to $1 \times 10^{-1}$ Pa.m$^3$.s$^{-1}$.

This method can also be used in the case of very large leaks impeding, because of the impossibility to lower sufficiently the pressure, the use of any classical method of leak detection. In that case the "sniffer" is used to detect the presence of helium in the exhaust gas of a roughing pump while spraying helium on the leaking system.

### 3.5 Characteristics of leak detectors

To specify helium leak detectors it is important to define several characteristics. The sensitivity of the helium leak detector is often confused with the smallest detectable leak. The intrinsic sensitivity $s$ is defined as the ratio between the leakage flow $Q_l$ and the helium partial pressure $P_{He}$ in the leak detector cell:

$$S = \frac{Q_l}{P_{He}}$$

In the case of a direct-flow leak detector, with a cell pumping speed $S_{He}$ the sensitivity $S$ is:

$$S = \frac{1}{S_{He}}$$

In the case of a counter-flow leak detector, the sensitivity is:

$$S = \frac{1}{S_b \times K_{He}}$$

with $S_b$ the pumping speed of the backing pump for helium and $K_{He}$ the compression of the turbomolecular pump for helium. This sensitivity is a characteristic given by the construction of the detector, its pumping speed and compression.

The smallest detectable leak that is an important characteristic from the operational point of view is defined as the minimum detectable leak signal. It is given by the intrinsic sensitivity of the leak detector and the peak-to-peak noise of the leak signal. This noise depends on the total pressure in the cell and hence on the operating conditions. The minimum detectable leak is different when the detector is connected to a small vessel or to a 100-m long accelerator sector. To avoid any confusion the exact operating conditions for the measurement of the minimum detectable leak must be defined in the specification.

Another characteristic of interest is the long-term drift of the detector. This characteristic is especially important in the leak detection of large systems with long time constants and defines the stability in percent of the signal over a given time period.

For various reasons, the leak detectors sometimes receive large flows of helium. It is important to measure the time needed by the leak detector after such an incident to recover a low helium signal. Of course this measurement must be done on an isolated detector to eliminate the time constant introduced by the vacuum system itself. Another possible reason for a high helium background signal is the presence in the atmosphere surrounding the leak detector of a high helium concentration. This can happen in halls where large quantities of liquid helium are handled with unavoidable losses. In that case helium can penetrate the leak detector circuits by permeation through elastomer gaskets or through small internal leaks in the detector. This also causes a high background signal and must be specified and tested.

The highest operating pressure of the leak detector largely determines the speed of an intervention in an accelerator. As a matter of fact, if leak detection can be carried out with an acceptable sensitivity at pressures in excess of 10 Pa, the waiting time before starting leak detection is determined by the removal of the gas contained in the vessel. At a pressure lower than 10 Pa, the degassing of the walls plays an increasing part. In that case the pressure decay in the leak detection cell i.e. the decay of the total flow, is inversely proportional to time and cannot be influenced without increasing the value of the minimum detectable leak. On the contrary, if the operating pressure is greater than 10 Pa, larger auxiliary pumps can be used to decrease the time needed to reach the detection threshold and can be valved off during the leak detection. For these reasons, the operating pressure of the leak detector is of great importance for quick interventions in an accelerator.

The procedures and equipment to calibrate a helium leak detector are described in the Ref. [9].

## 4. PRACTICAL CASES

### 4.1 Early detection of leaks

During the operation of a vacuum system much time can be saved by logging all the relevant information concerning the pressure decay during a roughing cycle. Any deviation from the usual pump-down curve is the sign of a possible leak. If the system is pumped for the first time, the shape of the pump down curve can also indicate the possibility of a leak: in Fig. 5 the typical pressure decay of a tight and a leaky system are shown. For pressures lower than 1Pa, the normal water degassing which constitutes the main constituent of an unbaked system is inversely proportional to time (-1 slope of the pressure-versus-time curve on a log-log scale). If a leak is present in the system, the pressure will tend to a limiting pressure given by the ratio of the leak flow to the pumping speed. A similar observation can be made by isolating the vacuum system and measuring the pressure increase as a function of time. Figure 6 shows three typical curves of pressure versus time on a linear scale: a linear behaviour is an indication of a leak. The initial part of the measurement should not be considered as the normal degassing of the system is predominant in that part of the curve.

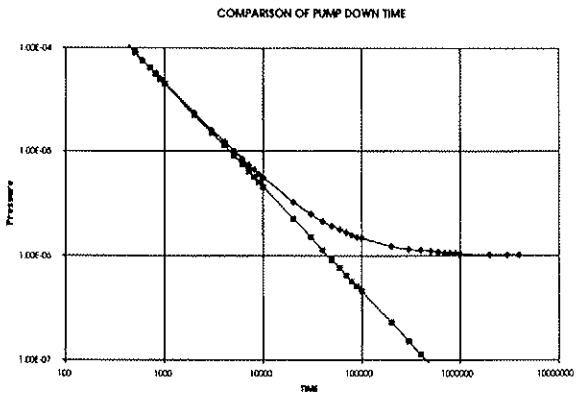



Fig. 5 Various pump-down curves

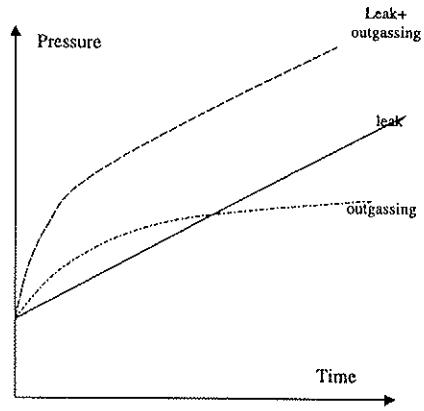



Fig. 6 Comparison of pressure increase in the presence of a leak or of degassing

If a gas analyser is present in the system, valuable information can be drawn from the residual-gas composition. On many occasions it has been observed that the presence of a peak at mass 14 (nitrogen) higher than the peaks at mass 12 or 15 (carbon and $CH_3$) after 24-hours pumping is, in an all-metal unbaked system, an indication of the possible presence of a leak. In a baked system, the mass 28 (nitrogen and carbon monoxide) is difficult to use as carbon monoxide is an important peak in baked systems. Argon (mass 40) and oxygen (mass 32) are more useful indicators of the presence of a leak although oxygen is not always visible on the mass spectrum as the pumping speed of baked stainless steel for this gas is high.

### 4.2 The use of leak detectors

In an accelerator, leak detection can be carried out in various ways: It is good practice to make an individual leak detection of all components before installing them in an accelerator since a leak is much easier to find and to repair in the laboratory than in the accelerator tunnel. This also has the advantage of reducing leak testing in the tunnel and hence the radiation dose received by the installation personnel. The leak tests can be made either by holding bags filled with helium around the mounted joints, or by blowing helium around the points to be tested. Both procedure have their advocates, the first one being longer but more systematic, the second one being faster and allowing the leaks to be located directly. In both cases, the time constant of the vacuum system must be carefully evaluated in order to spend, at each location, the time necessary for the helium signal to increase significantly. The evolution with time $t$ of the helium signal $P$, for a system of volume $V$ pumped with a pumping speed $S$ and having a leak $Q$ is given by (taking the initial pressure of tracer gas as zero):

$$P = \frac{Q}{S} \times \left(1 - e^{-St/V}\right)$$

Sixty three per cent of the full leak signal is obtained after 100 seconds if a 100-litre volume is evacuated with a 1-litre-per-second pump.

The helium leak detector is a complicated and sometime delicate instrument. It can loose partly or totally its sensitivity and hence it is of utmost importance to periodically check and readjust the sensitivity of these instruments. This is achieved by using calibrated helium leaks delivering to the instrument a known helium flow and adjusting the instrument settings (very often the ion accelerating voltage) for the maximum signal. Then, tuning the emission current sets the absolute value of the sensitivity. In most of the modern instruments, this calibration is made automatically at the start-up of the detector.

The presence of air movements is very important as they can transport the tracer gas and induce a wrong interpretation of the position of a leak. For this reason it is necessary to notice the direction of a possible air flow and to start the leak detection upstream.

### 4.3 Leak detection using total pressure gauges

This method is extremely useful in large accelerators, especially in the case of baked machines where the connection of a helium leak detector is a lengthy procedure. It is described for the case of the ISR at CERN in Ref. [10]. The method is based on the variation of the pumping speed, the gauge sensitivity and the conductance of the gas entering the leak. Because of the change of these factors, the ion current collected by an ionisation gauge is modified when the nature of the gas leaking in the system changes. The sensitivity of this method depends greatly on the configuration of the vacuum system, the types of gauges and pumps. In the forthcoming calculation, two extreme cases will be considered depending on the gauge location with respect to the pumping station: either at the pumping station or far from it. The geometry of the system studied for this example is shown in Fig. 7, the case of Bayard Alpert gauges is considered.

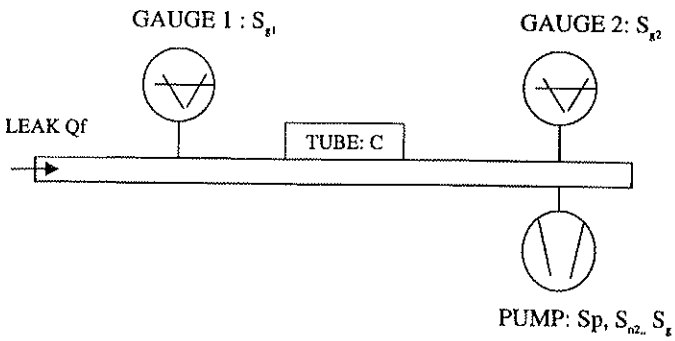



Fig. 7 Leak detection with pressure gauges

At the leak and in the accelerator tube, the ratio of the flow (in the molecular regime) for the tracer gas $Q_g$ and the air $Q_{air}$ is equal to the ratio of the conductance of the tube for these two gases:

$$\frac{Q_g}{Q_{air}} = \sqrt{\frac{29}{m_g}} = \frac{C_g}{C_{air}}$$

where the subscripts $g$ and $air$ refer respectively to the tracer gas and the air.

For the gauge 1, far from the pump, the pumping speed $S_{g1}$ is equal to the conductance $c$ of the pipe.

For the gauge 2, at the pump, the pumping speed $S_{g2}$.

At the sensors the pressure $P_{sg}$ is

$$P_g = \frac{Q_g}{S_g} .$$

The corresponding current on the collector gauge $I_g$ is:

$$i_g = i_e \times s_g \times P_g = i_e \times s_g \times \frac{Q_g}{S_g}$$

where $i_e$ is the emission current of the gauge. When changing from air to the tracer gas, the variation in collector current is (valid for the gauge 2):

$$\frac{i_g}{i_{air}} = \frac{s_g}{s_{air}} \times \frac{Q_g}{Q_{air}} \times \frac{S_{air}}{S_g} .$$

For the gauge 1 located far from the pump: $S = C$ hence:

$$\frac{i_g}{i_{air}} = \frac{s_g}{s_{air}} \times \frac{Q_g}{Q_{air}} \times \frac{C_{air}}{C_g} = \frac{s_g}{s_{air}} .$$

When changing from air to the tracer gas the ratio of the gauge readings is the ratio of the sensitivity of the gauge for nitrogen and tracer gas.

Let us consider the case of two tracer gases, helium and argon, and two types of pump, a sputter ion pump and a combination of sputter-ion pump and sublimation pump. The relevant numerical values are for the pumps and gauges in use at CERN are given in Table 1. Applied to the above equations these values give the ratios given in Table 2 for the variation of the gauge current when air is replaced by the tracer gas.

Table 1
Values of pumping speed and gauge sensitivity for different gases

| Gas | Gauge sensitivity (relative to nitrogen) | Pumping speed of ion pump | Pumping speed of ion + sub. pump |
|---|---|---|---|
| Air | 1.2 | 400 | 1400 |
| Helium | 0.15 | 100 | 100 |
| Argon | 1.4 | 100 | 100 |

Table 2
Variation ratio of the gauge reading when air is replaced by helium or argon

| Gas | Gauge far from pump | Gauge at ion pump | Gauge at ion + sub. pump |
|---|---|---|---|
| Helium | 0.18 | 2.4 | 8.4 |
| Argon | 1.7 | 5.6 | 20 |

These numbers show that depending on the position of the gauge with respect to the pump, and of the type of pumping, variations can be expected on the gauge current when air is replaced by a noble gas. The effect is especially large when the pumping relies on the chemisorption since noble gases are not pumped by this mechanism. Gauges located at the pump also show a greater effect. Argon gives an amplification of the signal under all circumstances and for this reason is preferable to helium. In the case of accelerators pumped by getter pumps, this method is especially well suited. A reactivation of the getter, or a sublimation just before detection, can greatly increase the sensitivity of the method.

### 4.4 Virtual leaks

This type of leak appears, on the pumping curve, as a real leak but cannot be located from outside. It is due to internal defects causing the retention of gas in a poorly-pumped pocket. Usual causes for

such leaks are poor design and/or poor manufacturing of the vacuum vessel. They are of course invisible by a leak detector but can be revealed by connecting a residual-gas analyser and checking the composition before and after venting with argon. After such a venting, the argon replaces the nitrogen in the virtual leak and an abnormally high argon content is present in the residual gas after the second evacuation.

### 4.5 The accumulation method

For the detection of very small leaks, an important gain can be made by leaving the unpumped system exposed to the tracer gas. This gas accumulates in the system and its concentration can be measured periodically either with a residual-gas analyser or with a helium leak detector if helium is used. Leak rates lower than $1 \times 10^{-12}$ $Pa.m^3.s^{-1}$ are measurable with this method on big volumes. For example a $1 \times 10^{-8}$ Pa partial pressure variation over a period of 48 hours in a vessel of 10 $m^3$ is produced by a leak of $8 \times 10^{-13}$ $Pa.m^3.s^{-1}$. It is possible to apply this method using air as tracer gas by checking the argon peak increase in the vessel provided that the argon degassing by welds can be neglected.

## 5. CONCLUSION

The various methods that can be used for the leak detection of accelerator systems cover the large range of possible leaks from the broken feedthrough to the tiny leaks appearing after bakeout. The vacuum specialist now has at his disposal efficient tools allying robustness with sensitivity and ease of use. Thanks to the development of molecular-drag pumps, the availability of leak detectors able to detect leaks at pressures in excess of 10 Pa is a great advantage in reducing the down-time of accelerators during emergency interventions. Despite the quality of the equipment now commercially available, leak detection remains an exercise, which, generally, is still difficult and requires well-trained technicians with a good knowledge of the vacuum system on which they intervene. Even with the best technicians using the most sophisticated equipment, emergency leak testing is always a time consuming and very expensive activity. For these reasons careful mechanical design and construction according to the rules of good vacuum practice must be applied. Preliminary tests of components must be made before installation: they are always much easier to carry out and avoid the costly installation and demounting of faulty equipment. Lastly these somewhat theoretical considerations on leak detection are a minute part of all the knowledge required to become "a subtle leak hunter" and which is only accessible "the hard way" by practice.

## ACKNOWLEDGMENTS

I would like to thank my technical colleagues at CERN for their skill in teaching me the art of leak detection.

**REFERENCES**

[1] A. Roth, Vacuum Technology, New York, Elsevier Science Publishers, 1990.

[2] J.F. O'Hanlon, A Users Guide to Vacuum Technology, New York, John Wiley & Sons, 1989.

[3] J.M. Lafferty, Foundations of Vacuum Science and Technology, New York, John Wiley & Sons, 1998.

[4] A. Nerken, J. Vac. Sci. Technol., A9, 3, 2036, 1991.

[5] A.O. Nier, C.M. Stevens, A. Hustrulid and T.A. Abott, J. Appl. Phys. 18, 30, 1947.

[6] W. Becker, Vak. Tech., 8, 203, 1968.

[7] M.H. Hablanaian and W.E. Briggs, Proc. Int. Vac., Congr. 7th, Vienna, 199, 1977.

[8] M.H. Hablanaian and W.K. Huber, Proc. Int. Vac., Congr. 7th, Vienna, 199, 1977.

[9] International Standards Organisation, 1 rue de Varembé, CP56, CH1211 Geneva 20, Switzerland, http://www.iso.ch ISO3530:1979.

[10] C. Benvenuti and J.C. Decroux, Le Vide, 162, 243, 1972.

# EXTREME HIGH VACUUM

*P.A. Redhead*
National Research Council
Ottawa, Canada

**Abstract**

Methods for producing and measuring extreme high vacuum (XHV, which is the pressure range below $10^{-10}$ Pa) and the limits on measurement and production are examined. Techniques for reducing the outgassing rates of the walls and components of XHV systems, including the gauges or RGAs, are described and some of the lowest measured outgassing rates tabulated. The lowest pressures attained in the XHV range have remained approximately constant for 35 years, explanations for this plateau in ultimate pressure are offered.

## 1. INTRODUCTION

Ultrahigh vacuum (UHV) is defined as the pressure range between $10^{-6}$ Pa (in Europe) or $10^{-7}$ Pa (in the USA) and $10^{-10}$ Pa; extreme high vacuum (XHV) is defined as the range of pressure below $10^{-10}$ Pa ($7.5 \times 10^{-13}$ Torr or $10^{-12}$ mbar). Ultrahigh vacuum is routinely achieved with commercial equipment around the world; XHV is produced and measured with difficulty in a few laboratories with special equipment. The need for UHV is widespread; the need for XHV is limited at this moment to some accelerator and storage ring facilities, to the simulation of interstellar space, in the processing of advanced semiconductor devices, and to a few other specialised applications. Research in the improvement of XHV methods, which is at the limits of vacuum science, will almost certainly improve the techniques used in the higher pressure ranges, as happened in the 1950s and 1960s with the development of UHV. Some of the gas phase parameters of XHV are shown in Table 1.

Table 1
Gas phase parameters at XHV

| Pressure (Pa) | Molecular density at 298 K | Molecular flux $N_2$ at 295 K (molec. $cm^{-2}$ $s^{-1}$) | Molecular mean free path λ* $N_2$ at 295 K | Monolayer time** |
|---|---|---|---|---|
| $10^{-10}$ | $2.5 \times 10^4$ $cm^{-3}$ | $2.9 \times 10^8$ | $3.4 \times 10^4$ km | 44 days |
| $10^{-13}$ | 25 $cm^{-3}$ | $2.9 \times 10^5$ | $3.4 \times 10^7$ km | 120 years |
| $10^{-17}$ interstellar space | 0.25 $m^{-3}$ | 29 | $2.3 \times 10^3$ astronomical units | $1.2 \times 10^6$ years |

* Mean free path for electrons is $\lambda_e = 4\sqrt{2}\lambda$, and for ions $\lambda_i = \sqrt{2}\lambda$.
**Assuming $10^{15}$ sites per $cm^2$ and a sticking probability of 0.5.

The *production* of XHV in small glass systems by cryogenic methods is not new. In the 1930s P.A. Anderson [1] was measuring the work function of barium and while working in Germany he immersed his sealed-off glass tubes in liquid hydrogen, the pressure must have been in the XHV range but it could not be measured. In the 1950s and for many years thereafter Gomer [2] immersed his sealed-off field emission and field ion macroscopes, about a litre in volume, in liquid hydrogen. These are just two examples of the early production of XHV without measurement capabilities.

In 1950 the invention of the Bayard-Alpert gauge [3] with its reduced X-ray limit made it possible to measure pressures below $10^{-6}$ Pa for the first time. By the late 1950s pressures approaching the XHV range were being achieved and measured. In 1958 Venema and Bandringa [4] built a glass system pumped by a mercury-diffusion pump which achieved a pressure of about $10^{-10}$ Pa, as

measured by a Bayard-Alpert gauge with a 25 μm diameter collector wire (to reduce its X-ray limit). Also in 1958 Hobson and Redhead [5] constructed a Pyrex glass system having a Bayard-Alpert gauge and an inverted-magnetron gauge [6] (IMG) which was pumped to about $10^{-8}$ Pa and then sealed-off. A glass finger extending from the system was immersed in liquid helium and the pressure measured by the inverted-magnetron gauge was about $10^{-10}$ Pa or less. Hobson has noted [7] that a 1-2 litre volume system pumped to UHV pressures, sealed-off, and then immersed in liquid helium will drop to a pressure of about $10^{-28}$ Pa if the residual gas is helium, this estimate was based on an extrapolation of the measured helium adsorption isotherm.

In 1962 W.D. Davis [8] used a stainless steel system containing a commercial sputter ion pump (5 l/s), a magnetic sector mass-spectrometer and a B-A gauge. With liquid nitrogen cooling of some of the components a total pressure of $4 \times 10^{-11}$ Pa was observed. In 1964 Hobson [9] made measurements on an alumino-silicate glass system (used to minimise the permeation of atmospheric helium through the glass envelope) containing a modulated B-A gauge, an inverted-magnetron pump, and a titanium getter. After pumping to $7 \times 10^{-9}$ Pa the system was sealed-off and the pressure fell to $1 \times 10^{-10}$ Pa at room temperature, when part of the system was immersed in liquid helium the pressure was $10^{-12}$ Pa. Pressure measurements were made with the modulated B-A gauge. Thus by 1964 it had been established that XHV could be achieved in small glass systems with the use of cryogenic pumps although the measurement of pressure was still difficult.

In 1977 Thompson and Hanrahan [10] took a complex surface analysis system, made from stainless steel and containing a mass spectrometer, and pumped it to UHV. It was then totally immersed in liquid helium at 4.2 K, the pressure dropped to $10^{-12}$ Pa at 30 K as the system cooled and then became immeasurable. The metal surface at 4.2 K pumped all gases (including helium) while permeation through the cold envelope was reduced to near zero.

The problem now became how to extend the techniques for producing and measuring XHV, developed for small systems with very small gas loads, to larger systems with sizeable gas loads.

This paper addresses the problems of producing and measuring XHV, it is assumed that the reader is familiar with the science and technology of the production and measurement of UHV. Review articles concerned with XHV should be consulted for further details [11-15]. Information on materials and hardware for XHV, which are not treated in this paper can be found in these reviews.

## 2. GENERAL CONSIDERATIONS

The problems in achieving XHV can be seen in a general way by examining the conservation of mass equation for a vacuum system of volume $V$, pumped by a pump of speed $S$ with a leak rate of $L$ and a total outgassing rate of $Q$. Then if $L$, $Q$, and $S$ are constant and a single gas predominates

$$V\frac{dp}{dt} = L + Q - Sp \ .$$

The total outgassing rate may be expressed as

$$Q + Q_T + Q_h + Q_d + Q_{st} \tag{2}$$

$Q_T$ is the total thermal outgassing rate from the walls and components of the system ($Q_T = q_T Af$; $q_T$ is the outgassing rate per unit area and $f$ the surface roughness factor),

$Q_h$ is the outgassing rate from heated components,

$Q_d$ is the diffusion rate through envelope ($Q_d = q_d Af$),

$Q_{st}$ is the desorption rate caused by electron, ion, or photon stimulation.

When the ultimate pressure $p_\infty$ is reached then $dp/dt = 0$ and assuming that $L = 0$ (which is essential for XHV)

$$p_\infty = \frac{Q}{S} = \frac{(q_T + q_d)Af + Q_h + Q_{st}}{S} \ . \tag{3}$$

In general the maximum pumping speed that can be applied to a vacuum system of arbitrary size is approximately proportional to the surface area ($S_{max} = kA$), this is true for all types of pump. Thus the ultimate pressure with maximum pumping speed is approximately

$$p_\infty = \text{const.}\left\{(q_T + q_d)f + \frac{Q_h + Q_{st}}{A}\right\} . \tag{4}$$

Although Eq. (4) is approximate it indicates the parameters that must be minimised to obtain ultimate pressures in the XHV range. If $Q_T + Q_d >> Q_h + Q_{st}$ then the ultimate pressure with maximum pumping speed is independent of the system volume and surface area

$$p_\infty = \text{const.}(q_T = q_d)f \ . \tag{5}$$

The above equations are a considerable simplification since they are only strictly correct if the gas density is uniform throughout the vacuum chamber; at very low pressures in the presence of adsorbing surfaces or localised pumps (e.g. gauges acting as pumps with limiting conductance to the chamber) this is not necessarily the case. In practice, most extended XHV systems tend to exhibit non-uniform pressure distributions, particularly of the chemically active gases that are readily adsorbed in a system at room temperature. The problems of pressure non-uniformity at low pressures have been considered to some extent by Da [16], Grigorev [17], and Haefer [18].

## 3. PUMPS FOR XHV

We next consider the types of pumps suitable for XHV and their limitations. These pumps fall into two categories, kinetic pumps that impart momentum to the gas molecules and remove them from the vacuum system, and capture pumps that trap gas molecules by ionic entrapment, condensation, adsorption, or chemical reaction at a surface within the vacuum system.

### 3.1 Kinetic pumps

Two types of kinetic pump are candidates for XHV, they are the diffusion pump and the turbomolecular pump (TMP). The advantage of kinetic pumps is that they can maintain a low pressure indefinitely and remove large quantities of gas permanently from the system, whereas most capture pumps require some form of regeneration after they have pumped a specific quantity of gas and do not remove the gas from the system.

Some of the early experiments to achieve XHV in the 1950s used mercury or oil diffusion pumps with liquid nitrogen traps to prevent the back-streaming of pump fluid into the UHV chamber. Diffusion pumps retain their pumping speed indefinitely as pressure is reduced and they are available with $H_2$ pumping speeds from 0.1 to over 3.5 $m^3 s^{-1}$, but the forepressure must be maintained as low as possible because hydrogen in the forevacuum can back-diffuse to some extent through the vapour jets thus reducing the effective hydrogen pumping speed. Uncontaminated pressure in the XHV range cannot be reached with a diffusion pump without extremely careful trapping to prevent any back-streaming of the diffusion pump fluid or the rotary pump oil. It has been claimed that with properly trapped oil diffusion pumps and a getter pump it is possible to achieve $10^{-12}$ Pa [19]. Diffusion pumps are not now widely used for XHV, partly as a result of difficulties in ensuring dependable, long-term trapping.

Turbomolecular pumps have considerable advantages for XHV since they can provide a completely oil-free system and, unlike diffusion pumps, most designs can be operated in any position. This is possible because the TMP can be designed to operate against a high backing pressure that can be provided by an oil-free diaphragm pump, and the rotor can be magnetically suspended to avoid lubricated bearings. The predominant gas at XHV is hydrogen, thus special steps must be taken to increase the low compression ratio for hydrogen (about $10^4$) in the TMP, the conventional TMP is only capable of achieving about $10^{-8}$ Pa. In 1990 a tandem TMP structure (with magnetic suspension) consisting of two TMPs on the same shaft was developed [20], the upper section designed for a high compression ratio and the lower portion for high throughput. This type of pump had a $H_2$ compression

ratio of $5 \times 10^8$ and a maximum backing pressure of 0.5 Torr, it achieved a pressure of $10^{-9}$ Pa. It was shown [21] in 1994 that the eddy-current losses in the magnetic suspension can cause a temperature increase in the rotor of 60°C resulting in outgassing. By coating the lower half of rotor and stator with a high emissivity layer ($SiO_2$ with emissivity of 0.9) the rotor temperature was reduced to 25°C allowing a pressure of $< 10^{-10}$ Pa to be obtained. Cho et al. [22] used a TMP with magnetic suspension backed by a small molecular drag pump followed by a dry diaphragm pump, after careful baking the final pressure was in the low $10^{-10}$ Pa range.

### 3.2 Capture pumps

Capture pumps include ion, getter, and cryopumps (and combinations thereof), and have been reviewed in some detail by Welch [23].

The main problem with a **sputter-ion pump** (SIP) at XHV is that the pumping speed decreases with pressure [24] as the discharge intensity (current per unit pressure) decreases. Re-emission of previously pumped rare gases is also a problem with SIPs. The main purpose of an ion pump in the XHV range is to provide some pumping speed for the gases that are not pumped by getter pumps (methane and the rare gases).

It was shown in the early 1960s that some speed was retained by an SIP at very low pressures, Davis [8] in 1962 achieved pressures in the $10^{-11}$ Pa range with a commercial 5 l $s^{-1}$ SIP. The rate of decrease of the pumping speed of an SIP at pressures in the UHV and XHV ranges is critically dependent on magnetic field, anode voltage, and dimensions of the pumping cell. Pumping speed is proportional to discharge intensity ($i_+/p$); $i_+/p$ increases with anode diameter and increasing the magnetic field tends to raise $i_+/p$ at low pressures.

**Getter pumps,** both evaporable and non-evaporable, are widely used in XHV, however they do not pump the rare gases and their speeds for hydrocarbons are negligible. Evaporated titanium films at room or liquid nitrogen temperatures have been used, in combination with other pumps needed to remove the rare gases and methane, down to pressures of $10^{-11}$ Pa and below. For chemically active gases (e.g. $H_2$, $H_2O$, CO, $CO_2$, $O_2$, and $N_2$) a sticking probability of about 0.5 is achieved on freshly evaporated metal films, thus pumping speeds of about 5 $ls^{-1}cm^{-2}$ can be achieved. Renewal of the film is necessary after about a monolayer of gas has been adsorbed. For titanium a capacity of about one pumped molecule to one evaporated titanium atom is possible. Freshly evaporated titanium films tend to outgas methane unless very pure titanium is used or the film is baked at 100°C for a few hours. Non-evaporable getters (NEG) are very effective in the XHV range because of their high pumping speed to hydrogen. NEGs are particularly useful in accelerators and storage rings since they can be placed very close to beam lines. Benvenuti [25] has achieved pressures as low as $5 \times 10^{-12}$ Pa in a 3 m long section of an accelerator ring using a 43.5-m long Zr-V-Fe NEG strip, a sputter ion pump (400 l $s^{-1}$), and a titanium sublimation pump.

Benvenuti has developed a method of coating the interior of a stainless steel system with a thin film (~1μm) of getter material deposited by sputtering [26]. Ti, Zr, V and their binary alloys have been used. The coating is transformed from a gas source to a pump by an in situ bakeout at temperatures of 250-300°C. Pressures as low as $10^{-11}$ Pa have been produced by this method.

Cryopumps used in the XHV range fall into two categories, a) **cryo-condensation pumps** which physisorb more than 2-3 monolayers of gas on a smooth surface of relatively small area, and b) **cryo-sorption pumps** having a porous surface of very large effective area with less than 2-3 monolayer of physisorbed gas.

For the **cryo-condensation pump** the limiting pressure is the vapour pressure of the adsorbed gas at the temperature of the surface, and the capacity of the cryo-condensation pump is essentially infinite, at least until the thickness of the adsorbed layer causes problems. Condensation coefficients of about 0.5 are typical and hence maximum pumping speeds of about 5 l $s^{-1}$ $cm^{-2}$ are possible; however, such speeds cannot be obtained in most situations because the cryosurface must be protected from room temperature radiation by suitable baffles at an intermediate temperature, the exceptions are when the complete wall of the vacuum system is a cryosurface (e.g. in the cold bore of an accelerator or storage ring with superconducting magnets). The low-pressure limit can in principle

be made infinitely low since desorption can be reduced to near zero by lowering the temperature. Table 2 shows the temperatures at which the vapour pressure of some common gases is equal to $1.3 \times 10^{-8}$ and $1.3 \times 10^{-11}$ Pa respectively[15] it can be seen that only hydrogen, neon and helium may limit the pressure in a cryo-condensation pump at 10 K to more than $10^{-11}$ Pa.

Table 2
Temperatures (K) for two selected vapour pressures of common gases

| Gas | Vapour Pressure $1.3 \times 10^{-8}$ Pa | Vapour Pressure $1.3 \times 10^{-11}$ Pa |
|---|---|---|
| $H_2$ | 3.21 | 2.67 |
| He | 0.303 | 0.250 |
| $CH_4$ | 28.2 | 24.0 |
| $H_2O$ | 130.0 | 113.0 |
| Ne | 6.47 | 5.5 |
| CO | 23.7 | 20.5 |
| $N_2$ | 21.1 | 18.1 |
| $O_2$ | 25.2 | 21.8 |
| Ar | 23.7 | 20.3 |
| $CO_2$ | 68.4 | 59.5 |
| Kr | 32.7 | 27.9 |
| Xe | 45.1 | 38.5 |

Data mainly from R.E. Honig and H.O. Hook, RCA Review, **21**, 360 (1960).

For the **cryo-sorption pump** the limiting pressure is established by an appropriate isotherm relating the equilibrium pressure to the surface coverage, for the XHV region the Dubinin-Radushkevich isotherm has been found appropriate for most gases [27]. A cryo-surface with less than monolayer coverage at 4.2 K will pump all gases to extremely low pressures.

In the preparation of highly porous surfaces for cryo-sorption pumps at XHV it is necessary to obtain good thermal contact between the porous material and the underlying cooled surface to achieve rapid thermal equilibrium. Figure 1 shows adsorption isotherms for hydrogen on several different surfaces [28] at 4.3 K, the sudden drops at pressures in the $10^{-13}$ Torr range are probably due to the X-ray limit of the extractor gauge used to measure pressure. It can be seen that the porous aluminium oxide surface (anodised aluminium with an oxide layer having small pores and a thickness of about 40 μm) has a hydrogen adsorption capacity more than $10^3$ times as great as a smooth stainless steel surface.

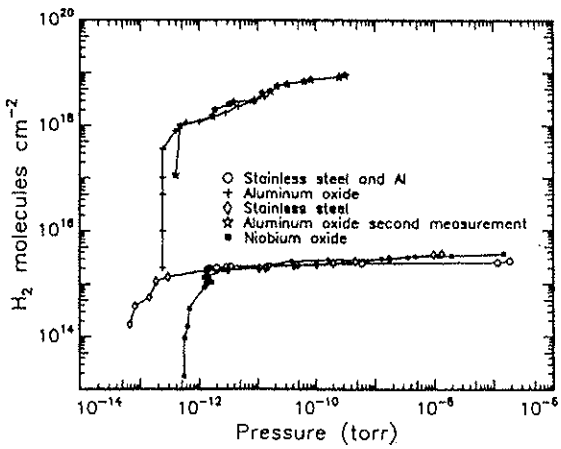



Fig. 1 Hydrogen adsorption isotherms at 4.3 K on various surfaces [28].

The choice of pump(s) for XHV depends on the application and the resources available. For relatively small systems a good choice is the magnetically suspended turbomolecular pump backed with a diaphragm pump combined with a getter pump. The use of liquid helium cryopumps, including the possibility of complete immersion, is economical where liquid helium is readily available.

For large systems the choice of pumps is dominated by costs and frequently by the difficulty of access to the vacuum chamber (as in accelerators and storage rings). Non-evaporable getters combined with sputter-ion or cryopumps are a frequent choice. Where superconducting magnets are used it is possible to use the cold bore of the beam line as a cryo-pump, this has been done in some storage rings and accelerators.

## 4. PRESSURE MEASUREMENT AT XHV

### 4.1 Ionisation gauges

The requirements for an XHV gauge is that the X-ray limit be sufficiently low and that the gauge design permits the separation of ESD ions from gas-phase ions. Table 3 lists six hot-cathode gauge designs with X-ray limits in the XHV range and which are able to separate ESD ions from gas phase ions. Only the extractor gauge is commercially available at present.

Table 3
Characteristics of gauges suitable for XHV

| Gauge | Electron Emission (A) | $P_x$** (Pa) | $P_{min}$*** (Pa) | Sensitivity (A $Pa^{-1}$) | References |
|---|---|---|---|---|---|
| Hot-cathode magnetron (Lafferty gauge) | $1 \times 10^{-6}$ | $6 \times 10^{-13}$ | $10^{-14}$ | $7 \times 10^{-4}$ [$N_2$] | [36] |
| Extractor (Leybold IE511) | $1.3 \times 10^{-3}$ | $2 \times 10^{-10}$ | $10^{-11}$ | $1 \times 10^{-4}$ [$N_2$] | [34] |
| 90• Bent-beam (Helmer gauge) | $3 \times 10^{-3}$ | $< 2 \times 10^{-12}$ | $10^{-12}$ | $1 \times 10^{-3}$ [$N_2$] | [29] |
| 180° Bent-beam* (Ion spectroscopy gauge) | $5 \times 10^{-3}$ | $< 2.5 \times 10^{-13}$ | $< 10^{-12}$ | $4 \times 10^{-4}$ [$N_2$] | [30] |
| 256.4° Bent-beam* | $1 \times 10^{-4}$ | $< 6 \times 10^{-12}$ | $4 \times 10^{-13}$ | $1.8 \times 10^{-6}$ [$H_2$] | [31] |
| Bessel box* (A-T gauge) | $3 \times 10^{-5}$ | $3.5 \times 10^{-11}$ | $10^{-12}$ | $1.8 \times 10^{-7}$ [$N_2$] | [37] |

* These gauges have electron multipliers.

** Pressure at which the ion current equals the residual current (the X-ray limit).

*** Approximate lowest measurable pressure.

Separation of ESD ions from gas phase ions is possible because of their differing energies. Electron space charge, in the grid of an ionisation gauge or the ion source of an RGA, causes a reduction of potential. Thus ESD ions released from the surface of the grid have a higher energy than gas phase ions generated within the grid.

Three of these gauges have electrostatic energy filters that separate the ESD and gas-phase ions on the basis of their initial kinetic energy, and reduce the X-ray limit by shielding the ion collector from the direct X-ray flux. These are the 90°, 180°, and 256.4° bent-beam gauges which are shown schematically in Fig. 2, the first two are also known as the Helmer gauge [29] and the ion spectroscopy gauge [30] respectively. All three gauges use electron multipliers (SEMs) and the 90° and 256.4° gauges have cylindrical deflection electrodes while the 180° gauge has hemispherical deflectors. The 256.4° gauge [31] has been tested with a cold emitter [32] (a field emission array) and its operation as a T-O-F mass spectrometer [33] will be discussed later.

Figure 3 shows the other three gauges schematically. The extractor gauge [34] can measure to less than $10^{-11}$ Pa with subtraction of the X-ray limit that can be easily measured on the operating gauge [35] this gauge does not have an SEM. It has been shown that the extractor gauge is relatively insensitive to ESD ions. The hot-cathode magnetron gauge is potentially capable of measuring to a lower pressure than any other if provided with an SEM. There is no separation of ESD and gas phase ions in this gauge but the electron current at the anode surface is so low that ESD effects are small. Its X-ray limit has been further reduced by the addition of a suppresser electrode in front of the collector [36]. Like all long-electron-path gauges it is subject to instabilities which are minimised by operation at very low electron currents. The Bessel box gauge [37] (A-T gauge) is an attractively simple design providing separation of ESD and gas phase ions.

A possible future method of measuring XHV pressures is the lifetime of charged particles in a Penning trap.

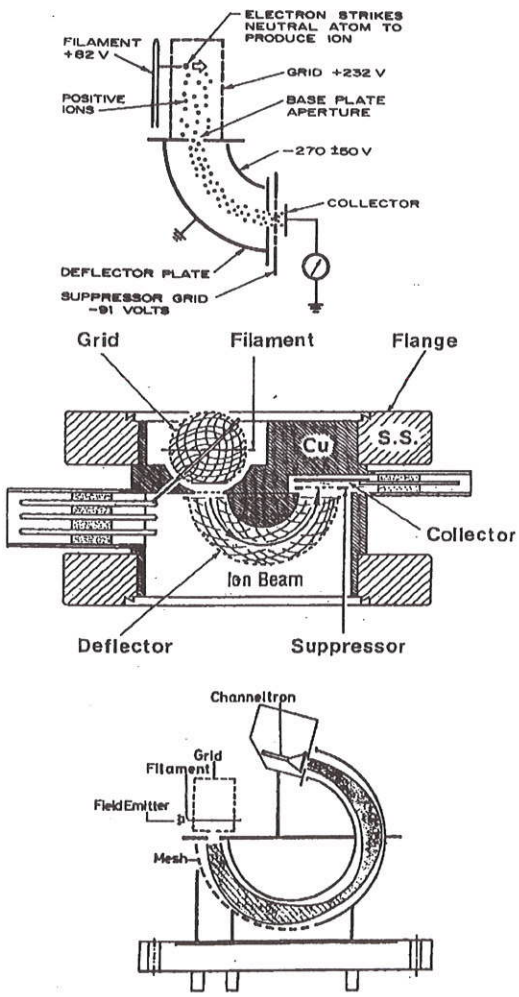



Fig. 2 Diagrams of three bent-beam gauges suitable for XHV. Top, the Helmer gauge [29] (90° bent-beam). Middle, the ion spectroscopy gauge [30] (180° bent-beam). Bottom, 256.4° bent-beam [31].

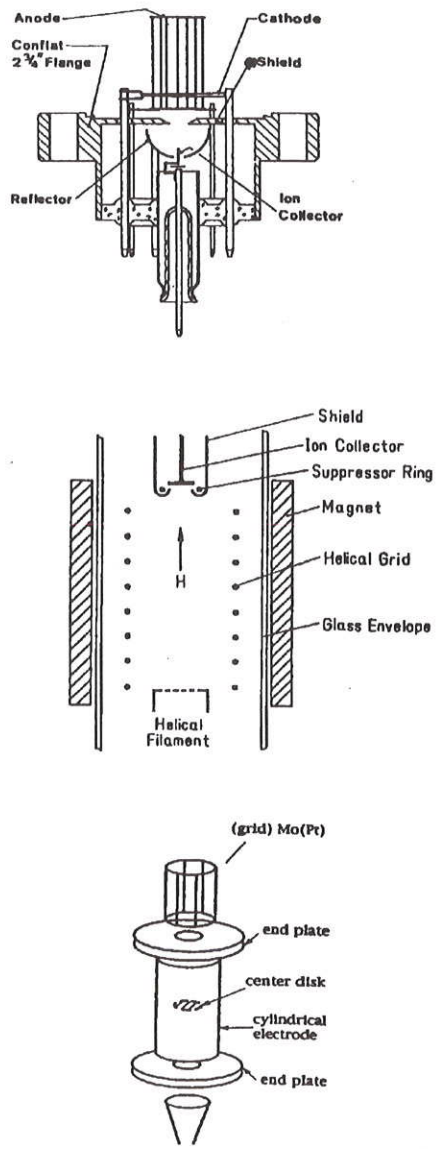



Fig. 3 Diagrams of three other gauges for XHV. Top, the extractor gauge [34]. Middle, the hot-cathode magnetron gauge with suppresser electrode [36]. Bottom, the A-T gauge [37] (Bessel box filter).

### 4.2 Processes limiting XHV measurements

The physical and chemical processes limiting the minimum measurable pressure of ionisation gauges and mass-spectrometers (i.e RGAs and leak detectors) in the XHV range include:

1. X-ray limit

2. Electron-stimulated desorption (ESD) of positive and negative ions and neutrals

3. Effects at hot cathodes

a) Increased outgassing resulting from the heating of electrodes and the envelope by radiation from the hot cathode.

b) Evaporation of neutrals and ions from the cathode.

c) Chemical reactions of gas molecules at the hot surface resulting in changes in gas composition.

4. Desorption from the gauge electrodes and walls by soft X-rays produced by the ionising electrons.

These limitations are the same as in the UHV range and have been described in detail elsewhere (see Ref. [14] and references contained therein). Here we shall only consider those effects that are particularly troublesome at XHV. The separation of ESD ions from gas phase ions in XHV gauges has been discussed above and Fig. 4 shows an example of the separation achieved in two types of bent-beam gauges, the ion spectroscopy and the Bessel box gauges. However, the ESD of neutrals from the grid of a gauge or RGA dominate the ESD problem at XHV [38]. Watanabe [39] has demonstrated that this effect can be reduced by heating the grid to about 550°C to minimise the surface coverage of adsorbed gas, see Fig. 5.

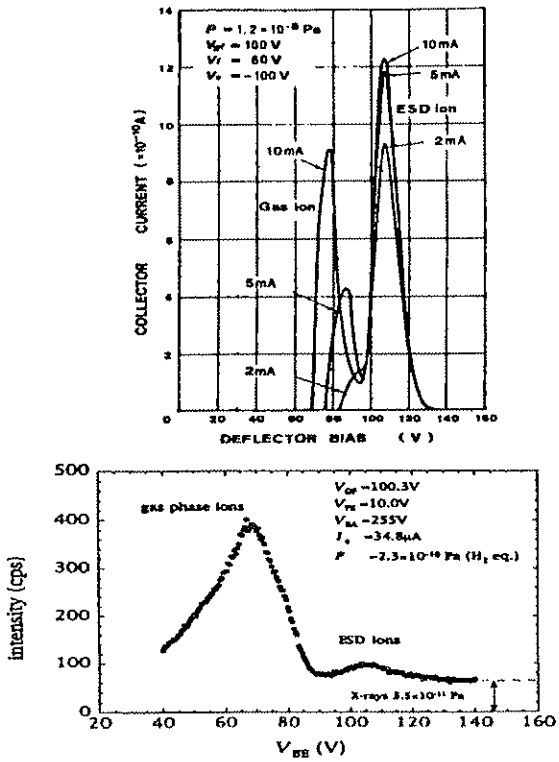



Fig. 4 Demonstration of the separation of ESD and gas phase ions in two bent-beam gauges. Top, the ion spectroscopy gauge [30]. Bottom, the Bessel box gauge [39].

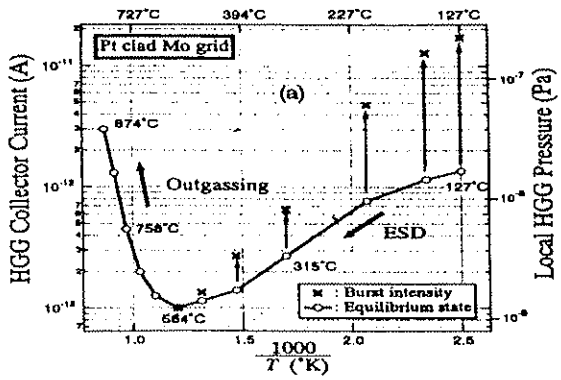



Fig. 5 Pressure, or extractor gauge ion current, vs. the reciprocal of grid temperature. The crosses indicate the pressure bursts after 30 minutes interruption of electron emission [39].

The outgassing resulting from a hot cathode, which is a particular problem at XHV, can be minimised by three methods: a) Using a low work function cathode to reduce the heating power required; thoria or yttria coatings on a refractory metal or a dispenser cathode [40] have been used successfully. b) Using a cold emitter such as an FEA, though the reliability and outgassing of these cathodes for use in vacuum gauges and RGAs is not yet assured. c) The metal envelope of the gauge or RGA can be replaced with a material of low emissivity and high conductivity to reduce the power required to heat the cathode. OFHC copper is a suitable material, a gauge envelope made of this material rather than stainless steel reduced the outgassing rate by a factor of 10 for an ionisation gauge [41] and a factor of 100 for an RGA [42].

In spite of improvements in reducing the ESD of neutrals and the outgassing due to cathode heating the outgassing from the gauge or RGA appears to be the limiting factor in attaining lower pressures in the XHV range. Desorption caused by soft X-rays may also be a limiting factor but this effect has not yet been carefully studied.

### 4.3 Residual gas analysers

In an RGA there is no practical X-ray limit since any X-ray induced photocurrent only adds to the mass independent background current. However, ESD ions are produced by electron bombardment of material adsorbed on the electron collector of the ion source, the ESD ions most frequently seen are; 1 ($H^+$), 10 and 11 ($B^+$ when $LaB_6$ cathodes are used), 16 ($O^+$), 19 ($F^+$), 23 ($Na^+$), 28 ($CO^+$), 35 and 37 ($Cl^+$), and 39 ($K^+$).

The mass-filter action of a quadrupole MS is very sensitive to the kinetic energy with which the ions are injected into the quadrupole structure. A complete separation of the two types of ions is possible either by modulating the ion accelerating voltage (typically at 125 Hz) and adjusting the mean value of the ion accelerating voltage while detecting the ion current with a lock-in amplifier [43], or by placing an electrostatic energy analyser between the ion source and the quadrupole structure [44], or between the quadrupole and the electron multiplier [45]. A time-of-flight technique has been applied to a quadrupole MS to separate ESD and gas phase ions [46]. By coupling the output of an ion-spectroscopy gauge to a quadrupole F. Watanabe has been able to completely separate the spectra of ESD ions and has demonstrated that neutral species dominate ESD at low pressures. It may be necessary to heat the grid (electron collector) of the ion source of an RGA to minimise the outgassing of neutrals by ESD.

The 256.4° bent-beam gauge has been operated as a TOF mass spectrometer [33] giving complete separation of ESD and gas phase ions, Fig. 6 shows the mass spectra obtained at a pressure of $23 \times 10^{-10}$ Pa. Only a large $H_2$ peak is seen in the gas phase spectrum, the ESD spectrum shows $H^+$, $H_2^+$, $O^+$ and other ions.

The problems of thermal outgassing from an RGA are the same as for ionisation gauges and the same methods for reduction can be used. Methods to reduce the outgassing from quadrupole RGAs have been examined by several experimenters [41, 47, 48]. Cold emitters (FEAs), to reduce heating of the ion source, have been tested in quadrupoles successfully [46] but the problem of noise and long term stability are not yet resolved.

## 5. LEAK DETECTION

It is necessary to reduce the leak rate into an XHV system to as near zero as possible, this raises the problem of the measurement of very small leak rates. The sensitivity of conventional helium leak detectors is about $10^{-12}$ Pa $m^3s^{-1}$, a higher sensitivity is desirable to test XHV systems. To achieve this increased sensitivity it is necessary to accumulate the helium, that enters the system through the leak, either in the gas or adsorbed phase.

The sensitivity of a helium leak detector can be increased by reducing the helium pumping speed to zero in a closed system and allowing the helium to accumulate in the gas phase, the pumping speed to active gases can be kept high by the use of non-evaporable getters or a cryopump operating at about 18K, neither of which will pump helium. The leak rate is determined, after calibration, by the

rate of rise of the helium ion current in the mass spectrometer (usually a quadrupole). A helium leak detector system using non-evaporable getters and the rate of rise method [49] is claimed to routinely detect leaks as small as $10^{-16}$ Pa $m^3s^{-1}$.

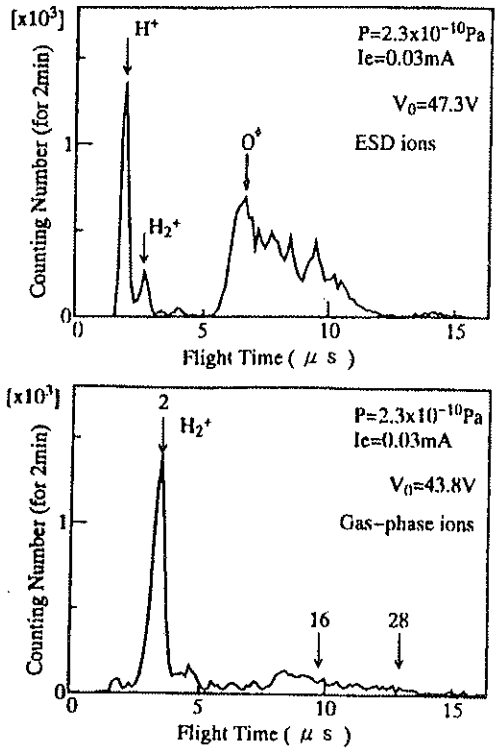



Fig. 6 The T-O-F spectra obtained with the 256.4° bent-beam gauge [33] at $2.3 \times 10^{-10}$ Pa. Top, gas phase ions. Bottom, ESD ions.

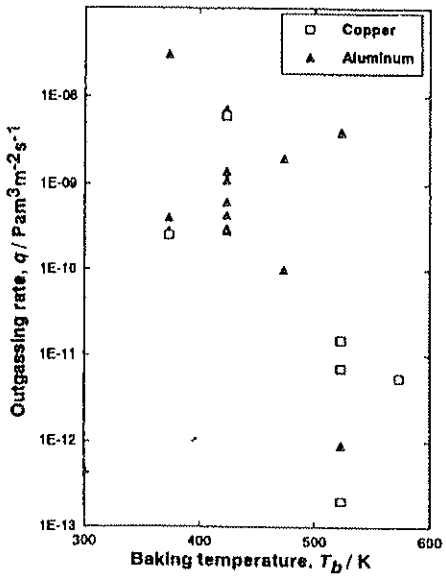



Fig. 7 Outgassing rate of copper and aluminium as a function of bakeout temperature [58].

Another method has been developed [50] in which the helium is adsorbed on a metal plate at 9 K or less in a closed system, the active gases are adsorbed on surfaces at about 20 K which may have porous carbon or non-evaporable getter material on them. The temperature of the helium adsorbing plate is raised to 10 to 20 K by an internal heater and the pressure of the desorbed helium is monitored with an RGA. This cycle is repeated with the plate cooled to 9 K or less for one minute and then heated to 10-20 K for one minute. With a one minute accumulation time at < 9 K a sensitivity of $10^{-16}$ Pa $m^3s^{-1}$ is claimed, with longer accumulation times a sensitivity of $10^{-19}$ Pa $m^3s^{-1}$ is claimed.

A special problem arises in leak detection of long pipes maintained at temperatures near 4.2 K as in cryogenic beam lines of accelerators and storage rings. Hobson and Welch have developed a theory to explain the delay in a pulse of helium gas moving down a cold pipe to a helium leak detector [51] and the predictions of the theory have been confirmed experimentally [52]. The time for a He pressure front from a $7.6 \times 10^{-6}$ Pa $m^3s^{-1}$ leak to propagate down a 75.3 m long stainless steel tube was 20 hours at 1.9 K , with a $9.3 \times 10^{-6}$ Pa $m^3s^{-1}$ leak at 4.25 K the time was 5 hours.

## 6. REDUCTION OF OUTGASSING

Reduction of the outgassing rates of the materials used in the construction of vacuum systems is essential if the XHV is to be achieved. In a metal system the residual gas is almost entirely hydrogen at XHV. Outgassing rates can be reduced by:

a) Lowering the temperature of all or part of the system.

b) High temperature vacuum firing to reduce the amount of dissolved hydrogen.

c) Baking of the vacuum system to remove water (150 to 450°C). It is important to achieve as uniform a temperature distribution as possible while baking.

d) Surface treatment to reduce surface roughness and remove porous oxides. This includes electropolishing, surface machining under special conditions, and glow-discharge cleaning.

e) Surface treatments to create oxide or other films on the surface that act as a barrier to diffusion of hydrogen from the bulk.

f) Deposition of films of low hydrogen permeability on a metal substrate, e.g. titanium nitride or boron nitride.

g) Choice of material with a low solubility for hydrogen (e.g. copper).

The general problems of reducing outgassing have been described in several texts (e.g. see Ref. [14] and references therein), here we shall only comment on a few methods that have proven useful for the extremely low rates necessary for XHV. A rough indication of the outgassing rates per unit area necessary for XHV can be obtained by assuming that the maximum applied pumping speed is proportional to the surface area of the vacuum chamber. Then if the effective entrance area of the pump(s) is $A_p$ (i.e. the area of a perfect pump where all impinging molecules are pumped) Eq. 3 becomes

$$q = kp_\infty \frac{A_p}{A} \tag{6}$$

where $k = 4.4 \times 10^2$ m$^3$ s$^{-1}$ m$^2$ for hydrogen at 295 K, and $A$ is the surface area of the vacuum chamber. Table 5 shows the outgassing rates needed to achieve $10^{-10}$ and $10^{-14}$ Pa with $A_p/A = 0.1$ and 1, the latter value being approachable when all surfaces are covered with getter material or at cryo-pumping temperatures. The outgassing rates of all the components of the system, including gauges RGAs and other active devices must be commensurate to achieve these pressures.

Table 4
Outgassing rates required to reach XHV

| $A_p/A$ | p. (Pa) | q (Pa m s$^{-1}$) |
|---|---|---|
| 0.1 | $10^{-10}$ | $4.4 \times 10^{-9}$ |
| 0.1 | $10^{-14}$ | $4.4 \times 10^{-13}$ |
| 1 | $10^{-10}$ | $4.4 \times 10^{-10}$ |
| 1 | $10^{-14}$ | $4.4 \times 10^{-12}$ |

The oxidation of stainless steel, believed to form an oxide barrier to hydrogen diffusion from the bulk, has been shown by many experimenters to reduce the outgassing rate to the $10^{-11}$ Pa m s$^{-1}$ range. Recent measurements by Bernadini [53] et al. have shown that firing stainless steel in air at about 400°C drives out most of the hydrogen in the bulk but the presence of the oxide layer does not reduce outgassing.

An alternate method of reducing the amount of hydrogen in stainless steel is to vacuum fire at 960°C. This method is more expensive than air firing and is liable to cause stress corrosion cracking if the sample is kept in the range 450 to 850°C for too long. The outgassing rates after vacuum firing observed by different experimenters are in some disagreement. Marin [54] et al. found that the outgassing rate after vacuum firing was $3 \times 10^{-11}$ Pa m s$^{-1}$ whereas the rate after air firing the same type of sample was $1 \times 10^{-12}$ Pa m s$^{-1}$. Fremery on the other hand has measured an outgassing rate of $1 \times 10^{-12}$ Pa m s$^{-1}$ after vacuum firing at 960°C.

Titanium nitride has been deposited on electrolytically polished stainless steel using a hollow cathode discharge [55] after vacuum firing twice at 430°C for 100 h and 500°C for 100 h. The outgassing rate at room temperature was lowered by a factor 10 by the TiN layer to $1 \times 10^{-13}$ Pa m s$^{-1}$

Ishimaru [56] has described an aluminium alloy system where the chamber was machined in an atmosphere of dry oxygen and argon (the EX process) after bakeout at 150°C for 24 hours the outgassing rate was about $10^{-10}$ Pa m $s^{-1}$, an ultimate pressure of $4 \times 10^{-11}$ Pa was claimed.

Koyatsu [57] has measured an outgassing rate of $10^{-12}$ Pa m $s^{-1}$ from OFHC copper after baking at 250°C (softening temperature 200°C), the same rate was obtained from chromium-copper (0.5% Cr, 99.5%Cu) after baking at 100°C (softening temperature 500°C).

Figure 7 shows the outgassing rates of copper and aluminium as a function of the bakeout temperature [58] demonstrating that higher bakeout temperatures are effective in lowering outgassing rates. Table 5 lists some of the lowest measured outgassing rates at room temperature.

Table 5
Some measured outgassing rates at room temperature suitable for XHV

| Material | Surface treatment | Outgassing rate (Pa m $s^{-1}$) | Reference |
|---|---|---|---|
| Stainless steel | Glass bead blasted + vacuum fired at 550°C /3 days + baked at 250°C/24 h. | $1.6 \times 10^{-13}[H_2]$ | [59] |
| Stainless steel (304L) | Air fired at 400°C/38h + baked at 150°C/7 days.<br>Air fired 390°C/100 h + baked at 150°C. | $1 \times 10^{-12}[H_2]$<br>$5 \times 10^{-12}[H_2]$ | [54]<br>[53] |
| Stainless steel | Vacuum fired at 960°C/25 h +baked at 180°C /6 h. | $1 \times 10^{-12}[H_2]$ | [60] |
| Aluminium | Vacuum fired at 960°C/25 h +baked at 180°C /6 . | $9 \times 10^{-13}[N_2]$ | [58] |
| Copper (OFHC) | Vacuum fired at 550°C/3 days + baked 250°C/24 h | $6 \times 10^{-14}[H_2]$ | [59] |
| Copper (OFHC) | Baked at 525°C | $2 \times 10^{-13}[N_2]$ | [58] |
| TiN on stainless steel. | Vacuum fired at 430°C/100 h and 500°C/ 100 h. | $1 \times 10^{-13}[H_2]$ | [61] |
| Aluminosilicate glass | Baked at 500°C/18 h + 600°C/2 h + 700°C/2 h/500°C/10 h. | $2.5 \times 10^{-13}[N_2]$ | [62] |

## 7. CONCLUSIONS

Methods for producing and measuring XHV in both small and large systems are well advanced although much of the necessary equipment is still not commercially available. The physical processes limiting the lowest achievable pressures are relatively well understood and methods to lower the limits are at least predictable. Two exceptions are a) the outgassing from gauges and RGAs due to the ESD of neutrals, and b) desorption in ion gauges and RGAs due to soft X-rays, where some recent work has outlined the problems but more needs to be done.

In 1964 Hobson measured a pressure of $10^{-12}$ Pa in an aluminosilicate glass system, in the last 35 years there has been no significant improvement in the lowest measured pressure. This is similar to the pressure plateau from 1916 to 1950 at $10^{-6}$ Pa as a result of the X-ray effect in hot-cathode gauges. It is probable that the present plateau also results from effects in gauges, namely the outgassing produced by the ESD of neutrals and by soft X-rays.

## REFERENCES


[1] P.A. Anderson, Phys. Rev., **47**, 958 (1935).

[2] R. Gomer, *Field Emission and Field Ionisation*, (Harvard University Press, Cambridge, MA, 1961. Reprinted A.I.P., New York, 1993.

[3] R.T. Bayard and D. Alpert, Rev. Sci. Instr., **32**, 571 (1950).

[4] A. Venema and M. Bandringa, Philips Tech. Rev., **20**, 145 (1958).

[5] J.P. Hobson and P.A. Redhead, Can. J. Phys., **36**, 271 (1958).

[6] P.A. Redhead, Trans. Nat. Symp. Vac. Technol., **5**, 148 (1958).

[7] J.P. Hobson, J. Vac. Sci. Technol., **3**, 281 (1966).

[8] W.D. Davis, Trans. Nat. Symp. Vac. Technol., **9**, 363 (1962)

[9] J.P. Hobson, J. Vac. Sci. Technol., **1**, 1 (1964).

[10] W. Thompson and S. Hanrachan, J. Vac. Sci. Technol.,**14**, 643 (1977).

[11] C. Benvenuti, *Production and Measurement of Extreme Vacua*, Proc.7th Int.Vac.Cong., p. 1 (Vienna,1977).

[12] J.P. Hobson, *Methods of Producing Utralow Vacuums and Measuring Ultralow Pressures*, Surf. and Coll. Sci., **11**, 187 (1979).

[13] C. Benvenuti, *Extreme Vacuua, Achievements and Expectations,* Physica Scripta, **T22**, 48 (1988).

[14] P.A. Redhead., *Ultrahigh and Extreme High Vacuum*, in *Foundations of Vacuum Science and Technology*, p. 625, ed. J.M. Lafferty (Wiley, New York, 1998)

[15] J.P. Hobson, *Limits of Vacuum Production and Measurement*, Proc. 9th Int. Vac. Cong., p. 35 (Madrid, 1983).

[16] D. Da and X. Da, J. Vac. Sci. Technol. A, **5**, 2484 (1987).

[17] G.I. Grigorev, K.K. Tzatzov, and I.N. Martev, Proc. 8th Int. Vacuum Congress (Cannes, 1980) vol. 2, p.291.

[18] R.A. Haefer, Vakuum Tech., **16**, 149 and 185 (1967).

[19] D.J. Santeler, J. Vac. Sci. Technol., **8**, 299 (1971).

[20] H. Enosawa, C. Urano, T. Kawashima, and M. Yamamoto, J. Vac. Sci. Technol *.A*, **8**, 2768 (1990).

[21] H. Ishimaru and H. Hisamatsu, J. Vac. Sci. Technol. A, **12**, 1695 (1994).

[22] B.Cho, S.Lee, and S.Chung, J. Vac. Sci. Technol. A, **13**, 2228 (1995).

[23] K.M. Welch, *Capture Pumping Technology* (Pergamon, Oxford 1991).

[24] T.S. Chou and D. McCafferty, J. Vac. Sci. Technol., **18**, 1148 (1981).

[25] C. Benvenuti and P. Chiggiato, Vacuum, **44**,511 (1993).

[26] C. Benvenuti, P. Chiggiato, F. Cicoira and Y. Aminot, J. Vac. Sci. Technol. A, **16**, 148 (1998).

C. Benvenuti, J.M. Cazeneuve, P. Chiggiato, F. Ciciora, A. Escudeiro Santana, V. Johanek, V. Ruzinov, J. Fraxedas, Vacuum, **53**, 219 (1999).

[27] J.P. Hobson, J. Phys. Chem., **73**, 2720 (1969).

[28] M.G. Rao, P. Kneisel, and J. Susta, Cryogenics, **34**, 377 (1994).

[29] J.C. Helmer and W.H. Hayward, Rev. Sci. Instrum., **37**, 1652 (1966).

[30] F. Watanabe, J. Vac. Sci. Technol. A, **11**, 1620 (1994).

[31] C. Oshima and A. Otaku, J. Vac. Sci. Technol. A, **12**, 3233 (1994).

[32] C. Oshima, T. Satoh and A. Otaku, Vacuum, **44**, 595 (1993).

[33] T. Satou and C. Oshima, J. Vac. Sci. Technol. A, **13**, 448 (1995).

[34] P.A. Redhead, J. Vac. Sci. Technol., **3**, 173 (1966).

[35] F. Watanabe, J. Vac. Sci. Technol. A, **9**, 2744 (1991).

[36] J.Z. Chen, C.D. Suen and Y.H. Kuo, J. Vac. Sci. Technol. A., **5**, 2373 (1987).

[37] K. Akimichi, T. Tahaka, K. Takeuchi, Y. Tuzi and I. Arakawa, Vacuum, **46**, 749 (1995).

[38] F. Watanabe, Vacuum, **52**, 333 (1999).

[39] F. Watanabe, Vacuum, **53**, 151 (1999).

[40] T. Satoh and C. Oshima, J. Vac. Soc. Jpn, **37**, 77 (1994).

[41] F. Watanabe, J. Vac. Sci. Technol. A, **11**, 1620 (1993).

[42] F. Watanabe and A. Kasai, J. Vac. Sci. Technol. A, **13**, 497 (1995).

[43] F. Watanabe and H. Ishimaru, J. Vac. Sci. Technol. A, **3**, 2192 (1985).

[44] S. Watanabe, M. Aono, and S. Kato,Vacuum , **47**, 587 (1996).

[45] S. Watanabe, H. Oyama, S. Kato, and M. Aono, Rev. Sci. Instrum., **70**, 1880 (1999).

[46] N. Ogiwara, Y. Miyo and Y. Ueda, Vacuum. **47**, 575 (1996).

[47] S. Watanabe and M. Aono, J. Vac. Sci. Technol. A, **14**, 3261 (1996).

[48] F. Watanabe, J. Vac. Sci. Technol. A, **8**, 3890 (1990).

[49] L.E. Bergquist and Y.T. Sasaki, J. Vac. Sci. Technol. A, **10**, 2650 (1992).

[50] G.R. Mynemi, US patent 5, 343, 740 (Sept. 6. 1994).

[51] J.P. Hobson and K.M. Welch, **11**, 1566 (1993)

[52] E. Whallen, J. Vac. Sci. Technol. A, **15**, 2949 (1997).

[53] .M. ernadini et al., J. Vac. Sci. Technol. A, **16**, 188 (1998).

[54] P. Marin et al., Vacuum, **49**, 309 (1998).

[55] S. Ichimura, K. Kokobun, M. Hirata, S. Tsukahara, K. Saito and Y. Ikeda, Vacuum, **53**, 294 (1999).

[56] H. Ishimaru, J. Vac. Sci. Technol. A, **7**, 2439 (1989).

[57] Y. Koyatsu, H. Miki and F. Watanabe, Vacuum, **47**, 709 (1996).

[58] Y. Ishikawa, Paper at IUVSTA Workshop on Outgassing Properties of Materials, Gräftåvallen, Sweden 1997.

[59] G. Grosse and G. Messer, Proc 8$^{th}$ Int.Vacuum Congr. p. 399 (Cannes, 1980).

[60] J.K. Fremery., Vacuum, **53**, 197 (1999)..

[61] S. Ichimura et al., Vacuum, **53**, 291 (1999).

[62] J.P. Hobson, J. Vac. Sci. Technol., **1**, 1 (1964).

# VACUUM CONTROL AND OPERATION

*D. Schmied*
ESRF, Grenoble, France

**Abstract**
The reliability of large vacuum systems is largely dependent on procedures established during installation and, later, on vacuum interventions that concern the exchange of machine parts, leak detection or bakeout. Nonetheless, vacuum equipment and its remote operation play an essential role in maintaining the reliability of the system. Therefore, as early as the design phase, particular attention should be paid to obtaining a precise structure for this system. The vacuum controls should be treated as a sub-set of the overall machine control.

## 1. VACUUM CONTROL

### 1.1 Overall layout

The successful operation of large vacuum systems must be based on distributed intelligence; this implies a reliable communication infrastructure between host computers and intelligent vacuum instrumentation.

The vacuum remote control should be designed as an open concept that allows upgrades and modifications of the system. It should be structured so that the user can access and modify different vacuum items on his own and to integrate during commissioning or intervention phases additional control parameters of bakeout or mobile pumping equipment in order to keep track of the performed interventions.

The general control structure should be split into different control levels. The vacuum instrumentation located on the lowest level is connected via dedicated control interfaces to local process computers on an intermediate level. These computers are connected to the server computers via a network, the man-machine interface being the highest level.

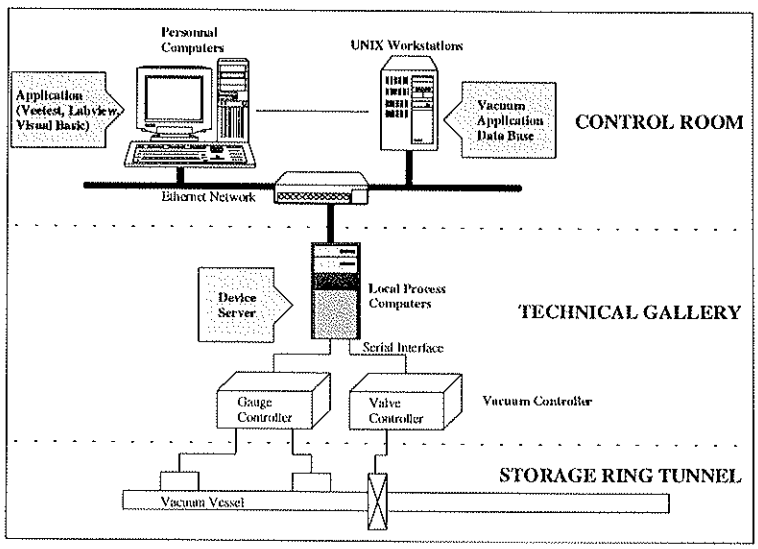



Fig. 1 Control layout

## 1.2 Vacuum equipment

Depending on the size of the installation and available resources, the vacuum control units to maintain and measure the vacuum system can be either developed in-house or bought as off-the-shelf products. Available control units can be perfectly adapted to the precise needs of the facility, including the hardware distribution, functionality and remote control interfaces. Nonetheless this implies devoting a significant amount of resources and time in order to obtain reliable operation. The more frequent solution is to use off-the-shelf products. In this way benefit from the steady development of the product is gained together with the advantage of a proven technology.

Specifications or selection criteria of vacuum equipment should include built-in microprocessor control; thus a significant amount of the specific operation and error handling is performed by the equipment itself which simplifies the necessary control software.

## 1.3 Control interface

As most commercial vacuum controllers are equipped with a microprocessor their remote operation is mainly based on a serial interface. Despite some existing electrical specifications, serial lines are not standardised and differences between the different serial interfaces are common. The RS232 transmission protocol is a one-line emission/reception standard which, due to its electrical layout (mass references), is less reliable with regard to communication errors and is limited in transmission distances (20m).

The evolution of differential signals was initiated by the RS422 serial interface. This twisted two-line emission/reception standard improved reliability as polarity changes were used instead of mass references, which may fluctuate between the connected units. The RS485 serial interface is based on the RS422 standard embedded in a multidrop. The use of this interface requires a communication protocol to be able to identify the different users. It has the disadvantage of a multidrop that, in case of a communication problem, renders the whole chain faulty.

The use of a serial interface affords the advantage of allowing a fully remote control of the vacuum equipment. However, software development can be time consuming especially if different hardware platforms are used. Special thought should be given to keeping the interface as simple as possible. The advantage of industrial standard bus systems like CAN, BITnet, G-bus are their reliable communication and a standard bus communication protocol could be used for some parts of equipment, especially for the Programmable Logic Controllers (PLC) or temperature sensors.

## 1.4 Low-level control

This equipment should be divided into different device groups consisting of pumps, gauges, valves and temperature sensors. In order to maintain identical access to the different device groups an access protocol has to be defined guaranteeing an identical interface for each member of similar device groups, independent of the hardware and the different application programs. This implies the analysis of each vacuum device concerning its functionality, capability and the properties of the different device signals/parameters [1] [2].

The local process computer contains the dedicated software (device server) to completely operate each part of the vacuum equipment. This software handles the routing of requests coming from various applications (e.g. graphical interface, historical archiving). It should be able to monitor the equipment and inform the application in case of incidents.

## 1.5 High-level control

A common man-machine interface implies that all applications must be able to access the different vacuum devices in an identical way. In order to obtain a simple and reliable control, a client/server architecture should be adopted. The device software on the local process computers are the servers which respond to requests coming from the vacuum applications. Due to the important amount of equipment an event driven data exchange between client/server, rather than polling the server on a periodical time bases, is more efficient.

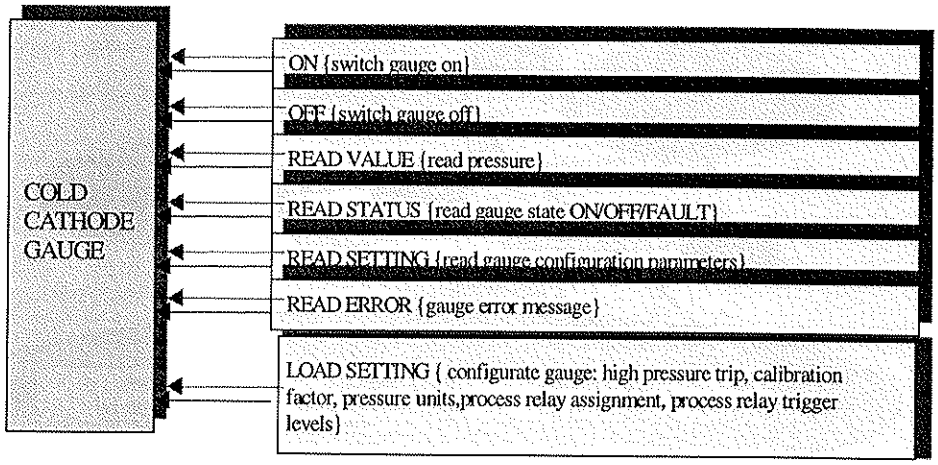



Fig. 2 Device server model

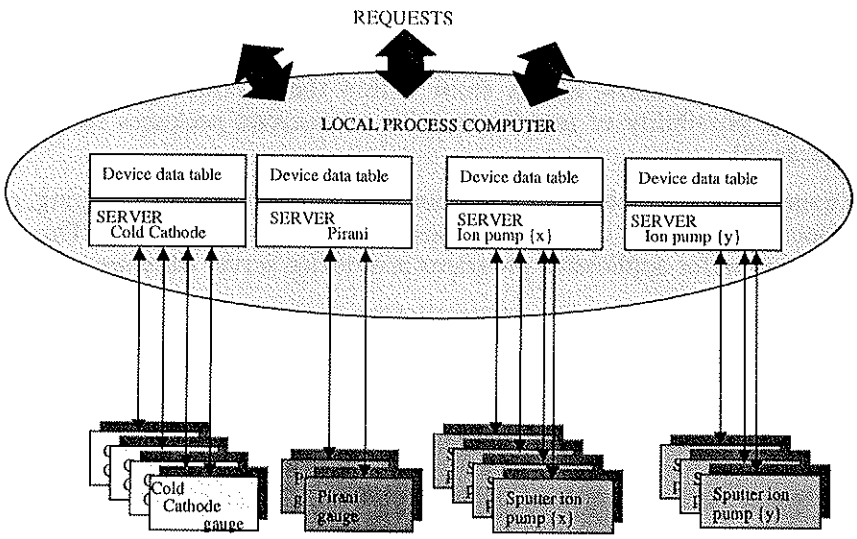



Fig. 3 Lower control level layout

Restarting application programs or dedicated device software should automatically restore the necessary configuration data from a database. This allows a dynamic update of any modification to the vacuum system with the guarantee of having a valid update on all control levels.

The main requirement of a vacuum application program is to visualise the current status of the vacuum equipment, to be able to modify it, to display all relevant physical data and to indicate errors or faulty equipment in an explicit written message. Additional information for the evaluation of the current status of machine parameters (beam current, lifetime) as well as the time stamp must be available.

Apart from the standard vacuum applications, attention should be given to providing simple development tools to create small control and analysis programs and to be able to better specify later application programs or to solve temporary problems.

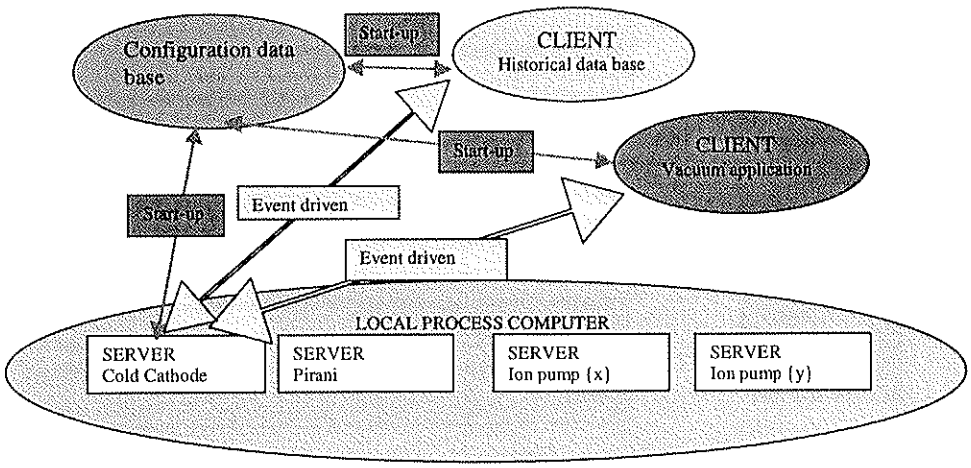



Fig. 4 Higher control level layout

## 1.6 Database

The size and complexity of control system data implies the use of databases. These reduce the difficulties of maintenance and improve understanding of the behaviour of the system. Commercial database management systems allow access from within the control application via high level languages as well as the possibility of interfacing with commercial packages such as data analyser tools or other data bases [3].

Information concerning the system configuration or measurements, data logging, cables, alarms, equipment maintenance, and failures can be integrated into a common database. This should contain all the necessary information for application programs and device server software [4]:

- ◊ The description of the vacuum system, which includes its physical layout, equipment names, and their position.
- ◊ Communication path definition between the equipment interface and the network address including the necessary communication parameters.
- ◊ Physical characteristics describing each part of the equipment (calibration, process relay assignments, set points).

The archiving and retrieval of historical data becomes an important item as time passes. It enables the identification of leaks, qualifies vacuum conditionings after interventions and enables comparison between different Insertion Device chambers. In the case of the ESRF data is processed in a common database for the machine and beamlines. Up to 3 Gbytes of vacuum data are stored over a year. Initially a commercial database was not available and the storage ring vacuum data was stored in dedicated files with the attendant difficulties of correlating vacuum data with other machine data, and the increasing complexity to extract important amounts of data.

The system should be flexible and easy to use which implies access tools to include new signals or modify the collection processes in a synchronous or asynchronous way with a maximal updating rate of the order of 1 second triggered by changes. Convenient and powerful graphic programs to display and browse the historical data must be available [5]. In addition it should be compatible for use by common spreadsheets or other specialised data analysis programs. A different feature – as important as the data logging of measurements and alarms – is the systematic recording of faulty equipment or material failures such as leaks repaired with varnish, radiation damaged cables, leakage currents on connectors, gauges, feedthroughs, etc. At the ESRF we have started an e-log, based on a database, which allows everybody to access via the Intranet from the various platforms. This enables problems to be identified, gives improved reliability and eases the maintenance burden.

## 2. INTERLOCKS

### 2.1 General concept

The choice of industrial Programmable Logic Controllers (PLCs) provides expansion capability and therefore the possibility of an easy upgrade to include changing requirements of storage ring and beamline operation during the installation and operation phases. PLCs are used extensively in industry to control sequential and/or continuous processes. These systems are designed to reliably sustain the harsh industrial environment. Compared to the classic hard-wired systems they provide more functionality but also communication mechanisms, supervision tools or archiving schemes.

At the ESRF the PLCs are connected via a proprietary RS485 interface to the local process computers. The process control part is completely covered by the industrial PLC, whereas the communication mechanisms of monitoring and control have been adopted by the standard control system. This resulted in a partial loss of important off-the-shelf features such as reliable intra-PLC communication or alarm handling. A non-negligible effort was put into the re-design and integration into the accelerator control system.

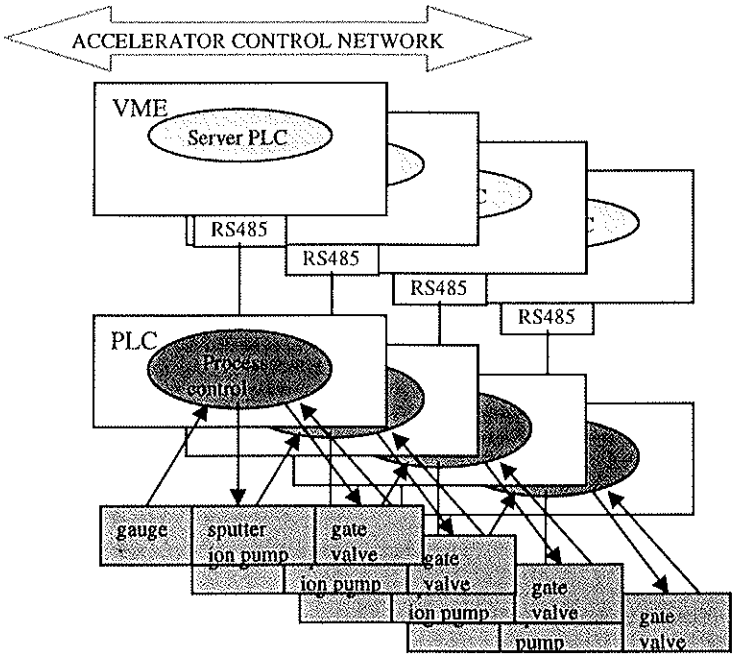



Fig. 5 ESRF layout

Possible evolution could be to use a cluster of PLCs linked together with an industrial field bus and connected to an upper control platform which acts as a supervisor. The different clients access the supervisor instead of each PLC [6]. Based on a modular system with in situ programming or via EEPROM, the PLC provides flexibility. This enables user-friendly upgrades or modifications to interlocks based on software changes, maintaining the principle of dedicated PLC programs using standard software and hardware interfaces.

### 2.2 Possible design solution

At the ESRF the same interlock strategy has been applied for the accelerators, Storage Ring (SR), Front Ends (FE) and beamlines. Depending on the size and complexity of the system several PLCs can build up the interlock system for SR or beamlines. The vacuum interlock system should be based on modules, which consist of all vacuum equipment inbetween two gate valves or windows. A module should consist at least of a Pirani gauge, Cold Cathode or Hot Cathode gauge, pump and pre-pumping valve.

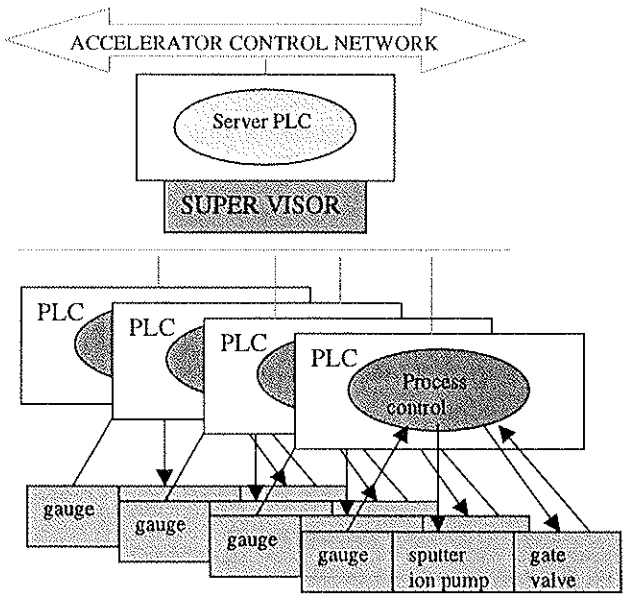



Fig. 6 PLC cluster layout

Apart from the continuous survey and processing of all connected trigger signals, which afford protection for gauges, residual-gas analysers and pumps, each PLC assures sequential processes. This concerns the operation of gate valves, shutters and absorbers related to other safety systems such as the Personnel Safety System, FE or experimental interlock signals. The remote operation via the control system allows a safe access to control vacuum items and allows quick and easy diagnostics in the case of interlocks. For the more complex use on beamlines a master/slave structure has been developed for different PLCs dedicated to cooling, safety shutter operation and beamline vacuum; and for the FE PLCs which operate the different vacuum and shutter elements of the FE.

## 3. VACUUM OPERATION

### 3.1 Total pressure measurements

#### *3.1.1 Pirani gauges*

These are chiefly used as cheap and simple high pressure gauges, necessary during interventions and for the completion of the total pressure measurements for the interlock system. The stability of these gauges is satisfactory. The use of long cables does not appear problematical with careful earth and mass connection of the gauges and controllers. These gauges remain switched on all the time. At the ESRF the Pirani gauge measurements are used for the interlock system to protect cold cathode gauges, sputter ion pumps and residual-gas analysers operated at high pressure and also during the NEG activation.

#### *3.1.2 Cold-cathode gauges*

The pressure monitoring system and interlock system at the ESRF is based on cold cathode gauges. It provides a reliable and stable operation in the dynamic pressure range of the machine ($10^{-8}$-$10^{-10}$ mbar). The gauges are easy to handle. They are mounted using CF38 elbows or cross tubes without additional photoelectron shielding. Electrical connection is by triaxial cabling because of the long cable length. Special attention is given to the electrical connection since, apart from the HV supply of the gauge, the measured currents are in the $10^{-9}$ A range, which implies careful earth/mass connections for all gauges and controllers. Occasional electrical leakage currents occur on the electrical connection due to humidity or radiation damage, or on the gauge which can be solved by high voltage

conditioning. The gauges are mainly started at about $10^{-9}$ mbar, which does not present any problems during operation.

*3.1.3 Hot-cathode gauges*

At the ESRF these gauges are only used in a few places as a reference under static vacuum conditions. Compared to the cold-cathode gauges, the use of these gauges is far more complex. In order to achieve reliable and stable measurements, several parameters need to be considered such as degassing during bakeout and regulation of emission currents. At the ESRF some problems were observed: unstable operation due to frequent filament trips; mounting on a CF63 flange with additional photoelectron shielding; a fair amount of gauge failures (short circuits, broken filaments) and sensitivity to the radiation background which required additional lead shielding.

*3.1.4 Sputter ion pumps*

Recent ion pump power supplies provide the possibility of obtaining reliable measurements of the ion pump current in the low $10^{-6}$A range. Depending on the pump size used, the limit pressure measured should be in the low $10^{-9}$ to $10^{-10}$ mbar range. This may give interesting information about the pressure distribution inbetween the gauges during operation. It will make it easier to identify local pressure bumps induced by leaks or malfunctioning of the pumps.

### 3.2 Partial pressure measurements

Residual-gas analysers (RGA) are included in the standard vacuum monitoring and data logging system to continuously follow a defined number of partial pressure measurements in order to confirm correct vacuum conditions on critical machine parts. They are also used in situ via PCs and the software to confirm correct bakeout, leak detection and RGA maintenance work. In order to achieve reliable measurements a considerable amount of work on system maintenance is necessary. Procedures have to be established to outgas filaments and to calibrate multipliers. The analysers have to be shielded against photoelectrons as well as radiation background.

### 3.3 Additional diagnostics

At the ESRF thermocouples were installed initially to control and monitor bakeout but have since been included in the standard vacuum monitoring and data logging system, as they have proved to be an interesting diagnostic tool in exotic machine operation modes to trace heating problems. Thermocouples are used since they are simple to handle and inexpensive. Special care has to be taken during their installation to assure a reliable attachment and cabling. It appears that apart from standard location on critical spots (crotch, absorber, bellows, and ceramic chamber) the system should be flexible regarding hardware and software to allow relocation/installation of thermocouples in situ and on the remote control.

### 3.4 Safety elements

*3.4.1 Gate valves*

The vacuum systems have to be divided into a sufficient number of sections limited by gate valves in order to isolate sections likely to undergo more frequent changes from sections that should remain stable. These gate valves should be equipped with an electro pneumatic control in order to close vacuum sections in case of pressure increases. This also implies the layout of the FE section and beam lines.

Several failures at the ESRF have proved that the existing interlock system can face accidental venting due to material failures (cavity window) or human errors during interventions. In all cases the incident could be limited to the zone between two pneumatic gate valves.

*3.4.2 Fast shutters, acoustic delay lines*

The UHV system of the SR has to be protected from vacuum failures in the downstream beamlines. Some of these beamlines are directly connected to the SR vacuum without any physical

limitation of a window. At the ESRF each FE has been equipped with fast shutter and acoustic delay line in order to isolate the SR vacuum system in case of an important vacuum failure. Commercial fast shutter systems are available and those installed are working well.

The propagation time of the pressure wave has to be increased by means of additional volumes or acoustic delay lines that consist of a series of baffles [7 – 9]. Tests have shown that the time taken for a wave to travel through a vacuum chamber varies according to the entry pressure, but also depends on the type of failures that vary from accidental manipulation of a valve to a window breakage. Experience gained at the ESRF shows that many vacuum accidents are not dramatic failures requiring the operation of active safety elements. Only in two cases was the fast shutter triggered and prevented the venting of SR vacuum sections. In both cases the failure happened on the first vacuum section on the beamline which is not isolated by a gate valve. In all other cases the beamline vacuum interlock system triggered the closure of the gate valves on the beamline and FE and prevented venting of the FE. Apart from the active safety elements such as fast shutters and acoustic delay lines, special attention has been given to include passive safety elements in the design of critical beamlines by using the sometimes important volumes of optical components (mirror, monochromator) linked with rather small conductance and differential pumping together with a sufficient number of gate valves to prevent the propagation up to the FE.

*3.4.3 Comments concerning windowless beamlines*

Special beamline layouts, interlock features and procedures have been developed to protect the SR vacuum against accidental venting and contamination from beam lines which do not have a physical barrier between both vacuum systems. In order to protect the SR vacuum the interlock systems of beamline, FE and SR have to be directly linked together.

◊ Accidental venting due to material failure

The mentioned FE acoustic delay line and fast shutter are the standard items. In addition a pressure gauge located on the first beam line module triggers the closure of the gate valves on the beamline and FE. An additional fast acting shutter only on beamlines protects the SR vacuum if an experimental part is placed outside a protection hutch or the experimental set-up is particularly risky (ultra-thin window).

◊ Accidental venting due to human error

The installation of a pneumatic gate valve as one of the first elements in each hutch allows systematic closure of the dedicated gate valve if personal access to the hutch is enabled.

◊ Leaks and contamination

The vacuum pressure of the beamline close to the FE has to be compatible with the SR vacuum. As mentioned earlier the FE gate valves are triggered not only by the FE gauges but also from a gauge located on the first beamline module with a large hysteresis on the trigger level.

If there is an abnormal pressure increase in the beamline the limiting gate valves are closed and cause the closure of the FE absorber. The pressure has to recover by a factor of two from the set point value in order to reopen these gate valves. In the case of an important leak which causes the pressure to rise in the first module the limiting beamline gate valves and FE gate valves will be closed. The FE gate valve will be inhibited until the pressure in the first beam line module has recovered by a factor of 100 from the set-point value.

Additional procedures have been established at the ESRF to protect the SR vacuum against contamination and for example, a partial pressure alarm system has been developed. Residual-gas analysers are installed between the FE and the limiting gate valve to the beamlines. In order to continuously protect the SR vacuum it is necessary to contain the partial pressure measurement in an alarm system connected to the standard beamline interlock system to close shutters and gate valves to isolate the vacuum sections.

◊ Differential pumping

Most beamlines use differential pumping systems to separate UHV from HV vacuum sections. In this cases the ion pump power supplies are also interlocked with the limiting gate valves.

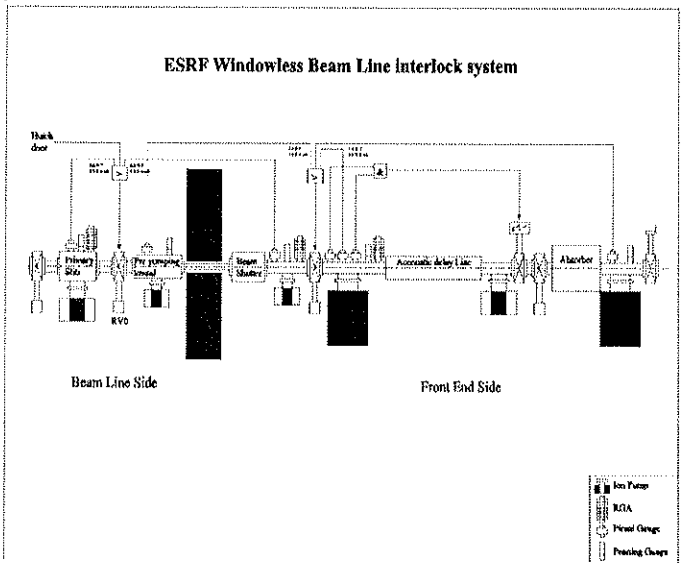



Fig.7 Interlock logic of windowless beamline

## 4. ACKNOWLEDGEMENTS

The author would like to thank the various members of the ESRF Computing Services for many fruitful discussions, and the ESRF Vacuum Group for its support.

## REFERENCES

[1] P.M. Strubin and N.N. Trofimov, First experience with control and operational models for vacuum equipment in the AD decelerator, CERN, PAC 99, New York.

[2] P.M. Strubin and N. Trofimov, Control and operational models for vacuum equipment, CERN, CERN-LHC-97-005-VAC Preprint.

[3] J. Poole, Databases for accelerators control – an operations viewpoint, CERN, PAC 95, Dallas, Proceedings IEEE, Piscataway 1995.

[4] D. Swoboda, The new vacuum control system for the SPS, AT Division, CERN, 1991 IEEE Particle Accelerator Conference, San Francisco, Proceedings IEEE, New York 1991. 9ICH3038-7 1576-1518.

[5] Y.N. Tang and J.D. Smith, Historical Data Collection, Retrieving and display in the NSLS Control System, National Synchrotron Light Source, Brookhaven National Laboratory.

[6] R. Saban, Integrating industrial and accelerator control systems, AT Division, CERN, PAC 99 Dallas Proceedings IEEE. Piscatoway 1995 (2147-2151).

[7] Y-F. Song, C-I Chen and C-N Chang, Study of the transit time of pressure propagation in an acoustic delay line, Department of Physics, National Taiwan Normal University, Taipei, Taiwan 117, Republic of China.

[8] F. Mazzoloni, Elettra front end and beamline status, Proceedings of Second Workshop on Vacuum for Future Synchrotron Light Sources.

[9] H. Betz, P. Hofbauer and A. Heuberger, Measurement of the efficiency of acoustic delay lines in view of beam lines for synchrotron radiation, J. Vac. Sci. Technol., Vol. 16, No. 3, May/June 1979.

# MATERIALS FOR ACCELERATOR VACUUM SYSTEMS

*Erhard Huttel*
ANKA, Forschungszentrum Karlsruhe, Germany

**Abstract**
Materials used in vacuum systems of accelerators are described. In the first part, general properties of materials such as the mechanical strength, thermal and electrical conductivity and transparency are discussed. In the second part the most important materials: stainless steel, aluminium, copper and ceramics are described, followed by details of the joining of different materials.

## 1. INTRODUCTION

Particle accelerators are operated in the $10^{-10}$ mbar pressure range in the case of hadron storage rings and $10^{-9}$ in the case of electron storage rings where a high gas load is caused by synchrotron radiation induced gas desorption. In order to get such a low pressure, the main criteria are a low gas desorption rate and no penetration of gases from the outside. The desorption and penetration of gases are the subject of separate lectures within these proceedings. In addition to these properties the material should fulfil some further criteria which are the subject of this lecture:

- Used as vacuum envelope, the material must withstand the atmospheric pressure. This favours a material with a high modulus of elasticity and high mechanical strength.
- Since Ultra High Vacuum (UHV) systems are baked, the materials used should have a negligible equilibrium pressure at the foreseen baking temperatures (150°C –350°C). Thus metals like zinc, magnesium or lead, or alloys containing these, which have a significant vapour pressure at relatively low temperatures, should not be used. Vapour pressures as a function of the temperature are given in the appendix.
- Obviously gases should not penetrate through the envelope of the vacuum system. There is some penetration of He and $H_2$ through glasses and some metals at higher temperature (palladium can be used as a filter for hydrogen) but for most materials discussed here, penetration is negligible. Gases can penetrate through elastomers such as Viton generally used for sealing High Vacuum (HV) systems.
- The materials should be easy available, which favours materials of general use in industrial application, for chemical plants, for building ships or aeroplanes.
- The materials should be easy to form mechanically by machining, turning, folding or deep drawing and be easy to join by welding or brazing.
- For some special applications the material should be a good thermal or electrical conductor in order to distribute the heat load or to reduce the impedance of the system. For other applications it should be non- or less-conducting in order to reduce eddy currents or for electric isolation.
- For an interaction region or for windows, the material should be most transparent for radiation, while normally it should shield radiation as much as possible.
- But increasingly the choice of the material is determined by economic criteria, it should fulfil its purpose at a minimum price.

Concerning the envelope, the most used materials are stainless steel, aluminium and copper. For small systems the material of choice is often a standard off-the-shelf solution, stainless steel. For larger

systems and special requirements other materials become interesting since the costs that are put into their development and testing often pays off.

## 2. PHYSICAL RESTRICTIONS

### 2.1 Mechanical stability

Some important material properties will now be explained but for more details the reader is referred to a text book such as Ref. [1]. The main criteria for a material used as an envelope for a vacuum system is its mechanical stability under atmospheric pressure. If a pressure or **stress** (force per area) is applied to a material it will be deformed by a certain **strain** (change in length per length). The behaviour of a material under stress is shown in Fig. 1. For small strains the deformation is elastic, it recovers when the stress is released and the strain is proportional to the stress. The factor of proportionality is called **Young's modulus** or the **modulus of elasticity**.

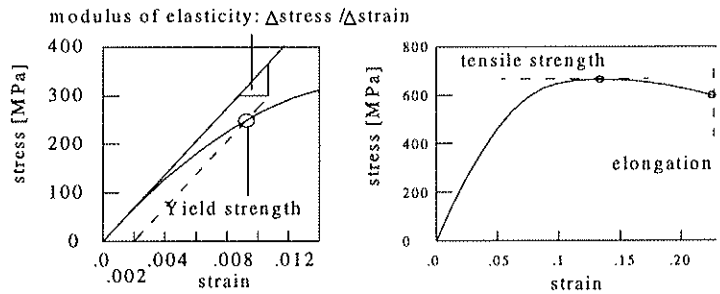
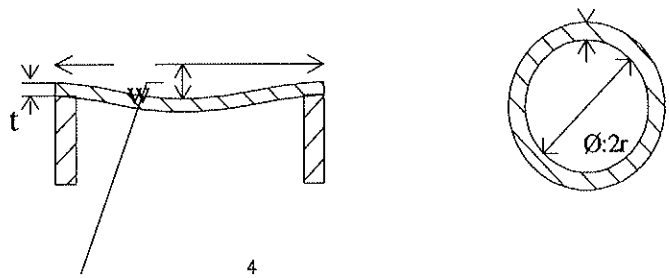



Fig. 1 Definition of modulus of elasticity, yield and tensile strength and elongation.

For large strains the deformation becomes plastic, the stress/strain relation is no longer linear. The stress value for which the stress/strain function crosses an offset of 0.002 of the linear behaviour is defined as the **yield strength**. For further increase of the strain, the stress increases until a maximum value is achieved which is defined as the **tensile strength**. For further strain, less stress is needed and the material will rupture when a maximum elongation is reached. Materials which can be deformed by several percent until rupture are called plastic or **ductile** (most metals), materials which cannot be deformed plastically are called **brittle** (ceramics). The strain at which the material ruptures is called the **elongation**.

From a mechanical point of view the critical parts of a vacuum chamber are the non-circular chambers: flat chambers like those in a dipole or an undulator, schematically shown in Fig. 2. The deflection (dy) is proportional to the fourth order of the width divided by the third order of the thickness times the modulus of elasticity. The stress is proportional to the second order of the width divided by the second order of a thickness. Obviously materials with a high modulus of elasticity are favoured but a difference in the modulus of elasticity by a factor 3 (aluminium compared to stainless steel) can be compensated by a slightly larger thickness of 1.45. Thus the design of chambers becomes important and the design has to be appropriate to the material. It should be noted, that the formulas given are only rough estimates. For realistic values of deformation and stress, finite element programs like ANSYS should be used.




deflection $\sim \frac{p w^4}{32 E t^3}$ buckling $\frac{t}{r} > (4p/E)^{1/3}$

stress $\sim \frac{p w^2}{t}$ stress $\sim \frac{pr}{t}$

Fig. 2 Deflection and stress of vacuum chambers by atmospheric pressure.

## 2.2 Hardness

The hardness of a material is another important mechanical property for the sealing of a vacuum system by a flange gasket system. It describes the resistance of a material to a local plastic deformation. For the flanges a hard material is used, mostly stainless steel. If aluminium is used for flanges an alloy with high hardness is selected and, in addition, the surface is often hardened by a special coating. For the gaskets a weak material is selected like copper, aluminium, gold, silver or indium.

There exist several measuring methods for the hardness. The Brinell hardness is measured by pressing a tungsten carbide ball (Diameter: D[mm]) with a standard load (F[kp]) into the test piece. The diameter of the imprint is measured (d[mm]) and the Brinell hardness is calculated as follows:

$$H_B := 2F/(\pi D(D-(D^2 + d^2)1/2))$$

The hardness is roughly proportional to the tensile strength $T_s$[Gpa] ~ 3.5 $H_B$.

## 2.3 Hardening

As mentioned in 2.1, a large stress applied to a material causes plastic deformation. The reason for this behaviour is explained by the structure of the metals that consist of grains with a non-perfect crystal structure. The failures or dislocations in the structure can easily move within the grains and cause their deformation. Since the movement of the dislocations is limited by the grain boundary, a small grain size increases the strength of a material.

An important possibility for increasing the strength is work hardening or cold working. If a material is cold-worked, i.e. by drawing or folding, additional dislocations are produced in the grains. The increased number of dislocations cause more friction since the dislocations hinder each other's movement. The increased strength is shown in Fig. 3 for OFHC copper as a function of the cold work. Starting from 70 M Pa for annealed copper, a yield strength of up to 300 M Pa can be achieved.

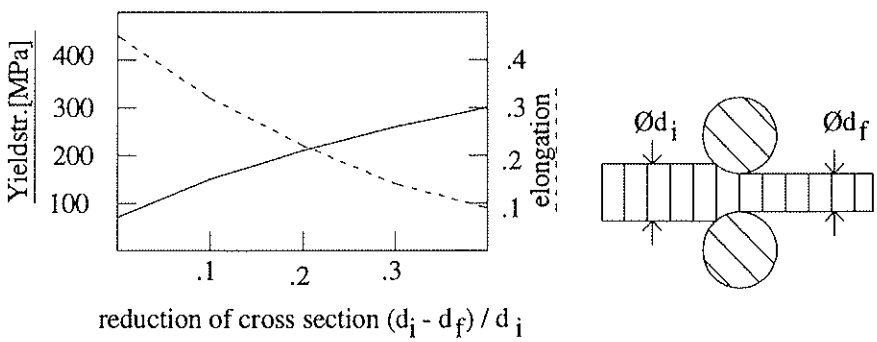



Fig. 3 Influence of cold work on yield strength and elongation.

The process of cold working can be reversed when the material is heated. At the annealing temperature all stresses in the material are released and the grain size begins to grow. This temperature is relatively low, about 1/3 of the melting temperature in [K]. The annealing temperature for some materials are given in Table 1.

Table 1
Annealing temperature for selected elements.

| Material: | Al | Al Cu Si | Cu OFHC | Stainless steel |
|---|---|---|---|---|
| Annealing temp. [°C] | 100 | 250 | 110 | 450 |

A further method for hardening some alloys is **precipitation hardening.** The strength of an alloy is generally higher since the particles introduced reduce the movement of dislocations. Some special alloys like Al Cu (5) Si (1) can be further hardened. The alloy is heated to a temperature of 400°C at which the copper is completely dissolved in the aluminium. Rapid cooling freezes this state at room temperature. (Normally the alloy consists of grains with aluminium and grains with Al-Cu.)

If the temperature is increased to 100 to 200 °C Al-Cu islands are formed within the aluminium grains. These islands constitute friction for the movement of dislocations and thus increase the strength of the alloy. The increased temperature must not be applied for too long since the strength will be reduced again. This behaviour is called over ageing. The temperature of the process as a function of the time and the yield strength as a function of the ageing time and temperature are shown in Fig. 4.

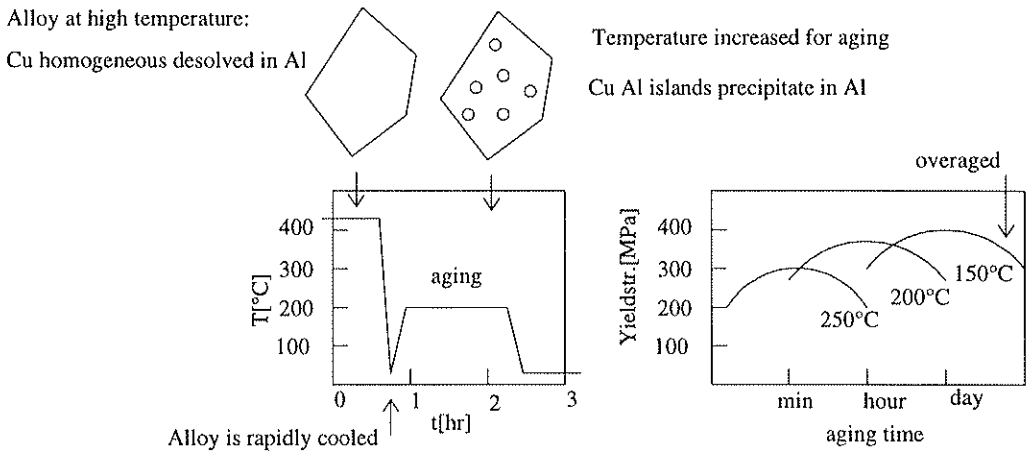



Fig. 4 Precipitation hardening i.e. the aluminium copper system.

## 2.4 Thermal conductivity

When a high heat load has to be transported materials with a high thermal conductivity are used. Thus the main lumped absorbers of synchrotron radiation sources are made from copper. When the heat load is directly absorbed on the wall of the envelope of the vacuum system aluminium (LEP) or copper (HERA, B factory) is used. For windows beryllium, beryllium-oxide or diamond are good candidates since the energy losses are small in these materials and the thermal conductivity is high. The maximum temperature at the heat source is the sum of the temperature step in the material (which increases with distance and the thermal conductivity) and the temperature step at the interface from the metal to the water (which decreases with the distance due to increased area for the heat transfer to the water). The situation of an absorber is demonstrated in Fig. 5 and the following estimates can be made:

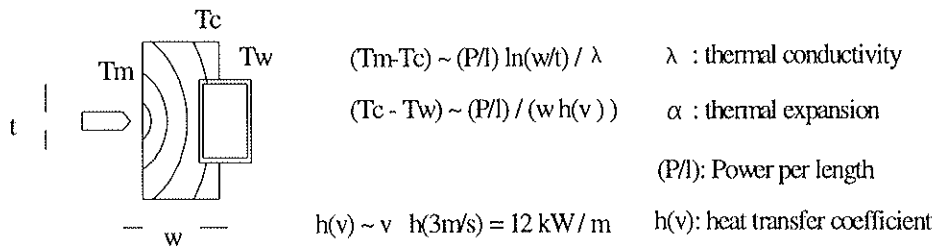



Fig. 5 Conductivity of thermal load in an absorber.

As in the case of the deformation and stress, the given formulas are only for rough estimates, the exact temperature behaviour and stress should be calculated by programs like ANSYS.

Obviously the temperature of the water duct should be kept below the boiling point of the water (the boiling temperature increases by 10 °C for each bar over-pressure). In addition the maximum temperature at the heat source should be low enough to prevent fatigue failure or even melting. The parts of the material which are at higher temperature try to expand but are hindered by the bulk of the

material, thus a stress occurs which can lead to cracks and rupture of the material. Since the stress is cyclic, fatigue failures occur at a much lower level than given by the tensile strength. The thermal stress is correlated to the strain that is given by:

$$s = \alpha \, \Delta T,$$

where $\alpha$ is the expansion coefficient of the material, $\Delta T$ the temperature difference between the bulk and the maximum temperature. This rupture depends on the number of cycles and the amount of stress from the temperature. Figure 6 shows this behaviour for copper [2] and GLIDCOP [3]. For $10^4$ cycles (the assumed lifetime of an absorber in a synchrotron radiation source) the strain should be kept below 0.002 in the case of OFHC copper.

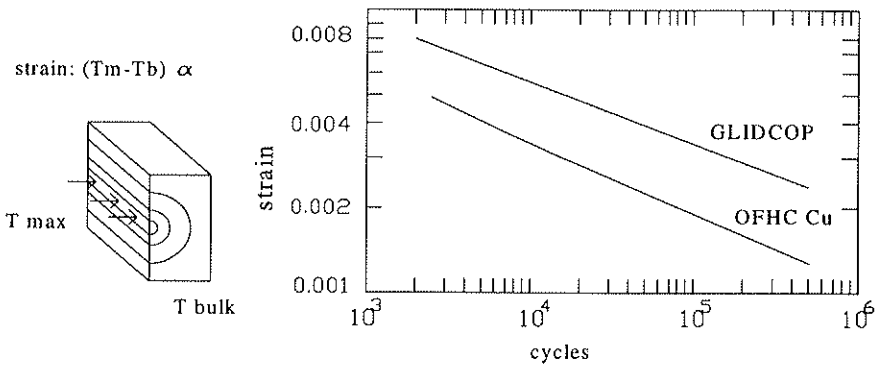



Fig. 6 Fatigue behaviour of copper.

This stress can be reduced by careful design. If the parts at higher temperature are allowed to expand (at least partially) the stress is reduced. For the crotch absorbers for ANKA and the SLS, the absorbers are split into two halves with comb-like structure and slightly separated. Each tooth can expand at the edges which reduces the stress by about a factor of two [4].

## 2.2 Electrical conductivity

Obviously high conductivity material namely copper is used wherever high currents have to be transported.

The vacuum impedance of a storage ring depends on the conductivity of the chamber material and the gap of the chamber. For insertion device chambers with extremely small gaps (5mm) the vacuum impedance can make a significant contribution to the total impedance. This is one reason for preferring aluminium as the material for the vacuum chamber of an insertion device or stainless steel with copper plating. Copper-plated stainless steel is also used for the LHC vacuum system [5] to reduce the resistive losses from the mirror currents and thus liquid helium consumption for cooling the heat load. A low-conducting steel alloy was chosen as the bulk material, since a high-conducting material (copper, aluminium) would cause destruction of the beam duct by the large forces from eddy currents when a super conducting field quenches.

In order to reduce the eddy currents, low-conductive materials are used for vacuum chambers which are situated in rapidly changing magnetic fields such as a Booster Synchrotron (operated in the 1-10 Hz range), Kickers (operated at about 1 MHz) or, as mentioned above, in the vicinity of superconducting magnets (LHC). Eddy currents cause magnetic fields opposite to the applied ones, reducing their effect. In addition they heat up the material. The dependency on frequency, conductivity and geometry is shown in Fig. 7.

The eddy currents are proportional to the electrical conductivity. Thus low-conducting materials like stainless steel, inconel or even ceramics are used, depending on the frequency of the magnetic field. The above estimates are only valid if the thickness of the material is not large compared to the skin depth which is given by: $d_{skin} = (\pi \, \mu_{rel} \, \mu_0 \, \sigma \, \omega)^{-1/2}$.

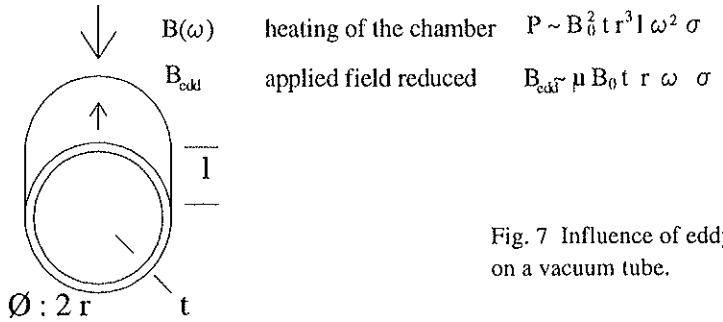



Fig. 7 Influence of eddy currents on a vacuum tube.

Kickers are interesting since their vacuum envelope is generally ceramic in order to avoid eddy currents, but a conducting layer is also needed in order to conduct the mirror currents of the particle beam, to preserve the impedance and to enclose the high field of the particle beam (500 MHz – 1GHz) to prevent heating of the ferrites. Thus a thin (about 2 μm) conducting titanium layer is sputtered onto the inner surface to shield the field. The thickness of the layer is a compromise between the rise time of the magnetic field and the eddy current losses, which favour a thin layer, and the resistive losses of the mirror currents and the need for a homogenous layer which favour a thick layer [6].

Obviously isolating materials are needed for electric feedthroughs like beam position monitors, kickers or septum magnets positioned in the vacuum, RF coupling loops etc. While plastics do their job for HV systems, especially materials like Vespel or Peek that can be baked up to 200°C, for UHV application and in radiation environment ceramics must be used.

## 2.6 Radiation penetration

At the interaction region of a collider the collision products should easily penetrate the wall of the vacuum chamber. Since most scattering processes depend on a higher order of the Z of the target material, a low Z and high mechanical strength are favoured. This was the reason that at the CERN SPS vacuum tubes made from carbon composite were developed for the 13-m long interaction region. For the shorter interaction region of the B factories and at DAFNE beryllium will be used. Beryllium is also used as windows for the synchrotron radiation in beam lines that separate the ultra-high-vacuum region of the storage ring from the user region. The left side of Fig. 8 shows the radiation lengths for selected elements [7].

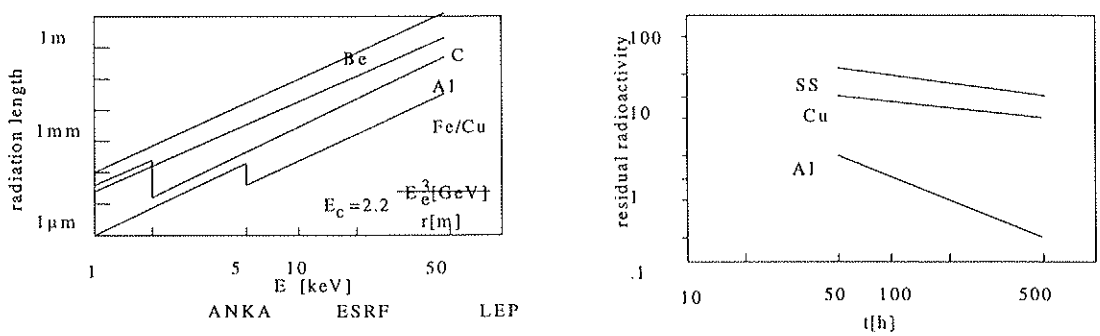



Fig. 8 Left side: Radiation length of synchrotron radiation. Right side: Residual activation of different materials.

In the normal region of a storage ring penetration of radiation is unwanted. Scattered and lost high energy particles and their reaction products cannot be shielded by the vacuum chamber. One-meter thick concrete walls or the equivalent in lead are needed. But in the case of high energy e-storage rings the emitted synchrotron radiation can also be high enough to penetrate the vacuum chamber if a light (Al) material is selected. This had been one reason for using copper (Z = 29) instead of aluminium (Z = 13) at the B factories or at HERA. At LEP a lead (Z = 82) shield was soldered onto the outside of the aluminium chamber to get a shielding effect.

### 2.7 Activation

Activation of material (by neutrons or ions) might be a nuisance for quick access to the accelerator after switching off, or by creating nuclear waste. Light materials, i.e. aluminium, (but take care of the other elements (Pb) in the alloy) are of advantage, as can be seen from Fig. 8, right side [8], the activation level is smaller and decays more rapidly. Obviously, activation also causes damage in materials, hardly in metals, but strongly in plastics which are not normally part of a UHV system. But cables for pressure gauges and electronics (often integrated into the gauges), turbo pumps and plastic or rubber water lines are indirectly part of the vacuum system and will be damaged. A list of the radiation resistance of the most used plastics is given in Table 2.

Table 2
Radiation damage of selected materials.

| Material | Polyuretan | Kapton, Peek | Polyethylene, PVC | Viton | Polyamide | Teflon |
|---|---|---|---|---|---|---|
| Dose [gray] | $10^8$ | $10^7$ | $0.5\ 10^7$ | $10^5$ | $10^4$ | $10^2$ |

## 3. SELECTED MATERIALS

### 3.1 Stainless steel

Stainless steel is the standard material for most vacuum components. Many manufactures have experience with this material. The mechanical strength (200 GPa yield strength) is sufficient for constructing dedicated vacuum systems. It is easy to weld. It is a standard material, generally used in construction work and easily available on the market. The high hardness (190 HB) protects flanges from scratches and leaks. The outgassing rates are low. Stainless steel can be baked in situ (150 to 350°C) to remove the water from the surface and it can be fired (at 900°C) to reduce the $H_2$ from the bulk of the material. Since austenitic steels are non-magnetic they can be used within magnets. A survey of the properties of stainless steel and its application to vacuum technologies is given in Ref. [9].

Steel alloys become corrosion resistant if the alloy contains more than 13% chrome, typically about 18%, since it produces a protective film of chrome oxide at the surface. The addition of nickel, typically 10%, maintains the austenitic structure (a face-centred cubic iron) at room temperature and is not magnetic.

Commonly the austenitic types 304 (Fe Cr(18)Ni(9)C(0.07)), 304L (Fe Cr(18)Ni(11)C(0.03)), 316L (Fe Cr(18)Ni(14)Mo(3)C(0.03)), 316LN (Fe Cr(18)Ni(14)Mo(3)N(0.2)C(0.03)) are used. The L stands for low carbon (less than 0.03%) which is important for conserving corrosion resistance when fired or welded. The N denotes the addition of nitrogen that increases the strength and the hardness of the steel. Within this text the chemical composition of alloys is given by the chemical elements of the alloys with its weight component in %, with the exemption of the main component. Concerning the mechanical properties austenitic steels are ductile with a relative low yield strength (200 GPa) and a high elongation (70%). The yield strength and also the hardness can be improved (300 GPa) by adding nitrogen to the alloy.

Hydrogen is the main gas component of UHV vacuum systems. While water is concentrated rather at the surface, hydrogen is dissolved in the bulk of the material (due to the production process). The hydrogen can be considerably reduced when the material is heated to 900°C. Thus the vacuum chambers of many accelerators are vacuum fired, partly before and partly after their manufacture. Since the high temperature also reduces the hardness the stronger alloy 316LN is mainly used when the material is fired.

When austenitic steels are fired or welded care must be taken to avoid the formation of carbide which can cause micro cracks and reduce the corrosion resistance. Carbon is normally completely dissolved in the alloy, but if the alloy is heated to temperatures higher than 600 °C and the cooling is later not done rapidly enough, carbon diffuses to the grain boundary to a larger extent than the chrome and the chrome content can fall below a critical limit. The formation of carbide can be avoided if the

carbon content of the alloy is low (less than 0.03%) and the critical temperature range (600 – 750°C) is passed in less than an hour [9].

Due to the production process stainless steel contains a high amount of impurities and inclusions (mostly sulphides and oxides). When the iron is cast into ingots the relatively lighter impurities collect at the surface but partly penetrate into the ingot when cooled. These impurities are the reasons for cracks and micro leaks. Forming the ingots into sheets and bars will further distribute them. The amount of impurities can be reduced considerably if the material is remoulded by an electro-slag refining process, schematically shown in Fig. 9. The impurities are collected at the top of the ingot which will be cut away. In addition the homogeneity of the steel can be improved if the ingots are hot forged. If the steel is machined to flanges etc the surface parts (3-5mm) of the bars should not be used.

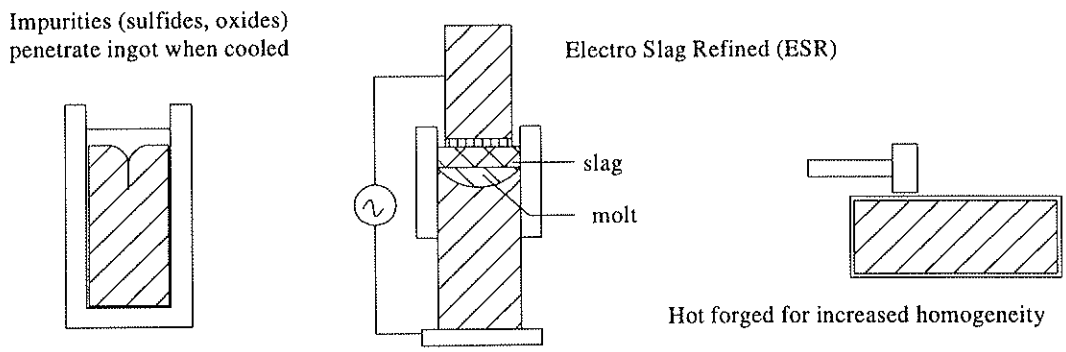



Figure 9: Quality improvements for stainless steel.

The steel quality is defined by the number of inclusions and the corn size. These items are determined by comparing a steel probe with standard probes. The corn size is classified by grain size numbers, the higher the number the smaller the grain size. The number is inversely proportional to the root of the grain size. A grain size number of 1 corresponds to a grain diameter of 0.25 mm. The inclusions are classified according to sulphides (A), oxides (B), silicates (C) and globular oxides (D). For each a severity level between 0.5 and 3 can be given. The more inclusions the higher the number. For UHV application a level of maximal 1 for inclusions (A, B, C), 1.5 for D and a grain size number of minimal 4 should be fulfilled.

Tubes and vessels are normally made from cold-rolled plates. Cold refers to the recrystallization temperature which is about 400°C for stainless steel. Cold-rolled plates have a finer surface.

While aluminium and copper are often formed by extrusion or machining from blocks, stainless steel is formed mainly by rolling, folding and deep drawing of sheets. Welding should be done without filler. Welding of long straight seams is normally done by e-beam while the flanges are welded by DC TIG (tungsten inert gas). TIG welding can lead to a distortion of the material (reduction in length and increase of the thickness close to the weld seam) due to the smaller energy density compared to e-beam welding. This distortion is in the order of the thickness of the weld and has to be taken into account for the dimensioning before welding. If the welding is not symmetric the whole vacuum chamber can be considerably deformed as experienced with the ANKA prototype vacuum chamber, when the supporting ribs were welded at the outer side of the chamber, see Fig. 10. Such a distortion can be removed by heating up the chamber at the opposite side, but has to be done by experienced workers to get acceptable results.

### 3.2 Aluminium

Aluminium has been used as the material of many electron synchrotrons (PETRA, TRISTIAN, LEP, LEAR, ALS, SRRC, APS, LEP, Pohang, Spring8). A summary of aluminium technology is given in Ref. [8]. The main reason for using aluminium has been that long beam tubes including the

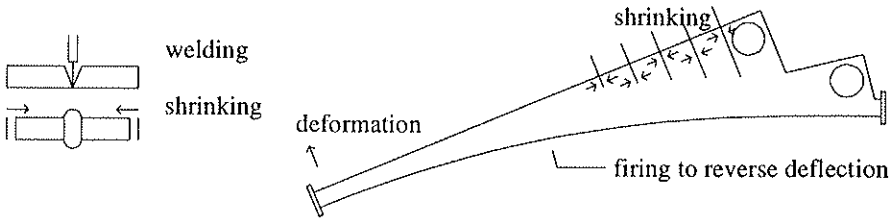



Fig.10 Weld distortion of the ANKA prototype chamber.

antechambers for distributed pumps and cooling ducts can economically be produced by extrusion. Furthermore complex chambers can be machined from solid blocks. Figure 11 shows the cross section of the LEP vacuum chamber made from extruded aluminium [10] and the cross section of the ALS vacuum chamber at a quadrupole which was machined from thick plates. The good thermal conductivity of aluminium allows an easy distribution and cooling of heat load. Thus the power density of the synchrotron radiation can be absorbed directly on the chamber wall if the power is not to high. In addition the cooling ducts can be conveniently used for mild baking (150°C) by using pressurised water. Standard parts (tubes, fittings, T and cross pieces) are now available from industry [11]. Additional advantages are that it is completely non-magnetic, has a low activation cross section and a rapid decay of radioactivity.

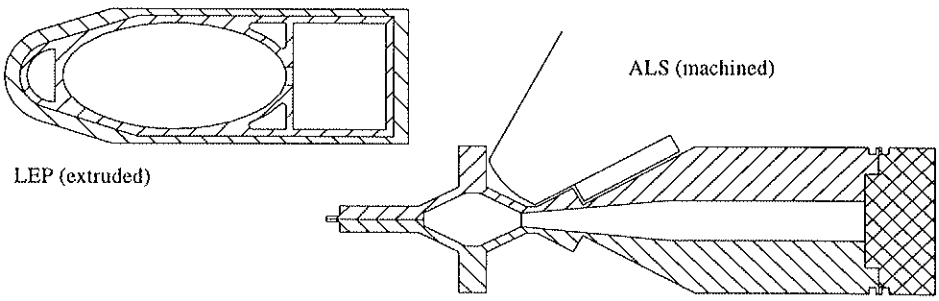



Fig. 11 Aluminium vacuum chambers of LEP and ALS.

However, aluminium has several disadvantages. Initially at least, it has an order of magnitude higher desorption yield compared to stainless steel or copper. The mechanical properties (E module, hardness) are weaker compared to stainless steel and have to be compensated by an appropriate design. Furthermore it is not so easy to make welds vacuum tight compared to stainless steel. A high heat load must be applied in order to break the thick oxide layer but the heat is rapidly diffused due to the good thermal conductivity. Thus the design of the weld should be done in such a way that only a minimum heat load is needed by doing small welds and introducing grooves close to the weld to confine the exposed area. The AC TIG method helps to remove the oxide layer more efficiently.

Both the high desorption yield and the difficulties with welding are related to the relatively thick oxide layer (15 nm) on the surface. In order to reduce this problem all extrusion and machining processes were done under argon atmosphere at KEK [8]. As a result a 3 nm low-porous oxide layer was obtained. For the same reason all machining was done with ethanol lubricant at SRRC [12]. At APS weld leaks were reduced when a filler wire extruded in argon was used [13].

For tubes the alloy 6063 (Al Mg(0.7)Si(0.4)) is mainly used, which can be extruded and is less expensive. For aluminium flanges, screws and nuts the harder 2219 Al Cu(6.3)Si(0.3) is preferred. In order to enhance the surface hardness of aluminium flanges a 2-μm TiN coating can be applied.

Connecting of aluminium chambers is partly done by using aluminium flanges, partly by using aluminium/stainless steel transitions and stainless steel CF flanges. Sealing is done with pure (1050)

aluminium Conflat gaskets or diamond-formed gaskets in combination with aluminium screws and nuts. Also the Helicoflex system is used when aluminium flanges are connected to stainless steel flanges. The variety of these techniques, shown in Fig. 12, demonstrates that the sealing of an aluminium system is not as reliable as the standard stainless steel system with copper gaskets.

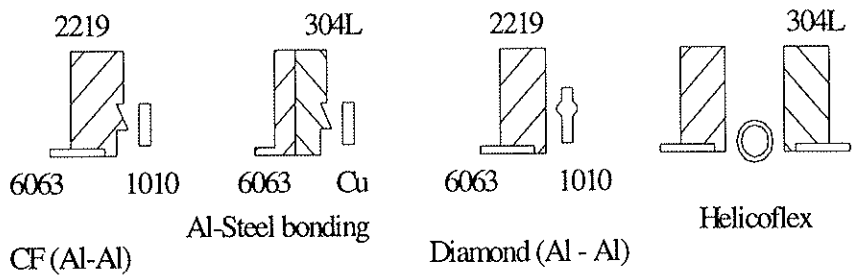



Fig.12 Sealing of aluminium vacuum systems.

### 3.3 Copper

Copper is used for the vacuum system at HERA [14] (DESY) and at the B factories [15, 16]. In nearly all accelerators it is used as an absorber for synchrotron radiation. The profile of the KEK B factory vacuum chamber is shown in Fig. 13. Normally OFHC copper is used. At HERA copper-tin (3%) bronze was used in order to have a higher stiffness and to avoid softening during brazing. Compared to aluminium large beam tubes can be produced by extrusion, but only with one duct. Therefore the antechamber for distributed pumps and the cooling duct have to be extruded separately and e-beam welded or brazed to the beam chamber. Thus the production process is more complex and expensive. Since copper vacuum systems are only used for storage rings with large radius in the dipoles, the arc sections can be formed by stretching them axially and bending them over a pre-shaped mandrel.

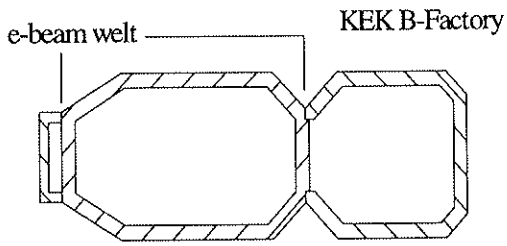



Fig. 13 Profile of the copper vacuum chamber of the KEK B factory.

Comparable to aluminium, the high thermal conductivity allows the chamber wall to be used as the absorber for the SR power. In contrast to aluminium the desorption yields are low comparable to stainless steel, albeit the differences between machined and extruded surfaces, as shown in Fig. 14.

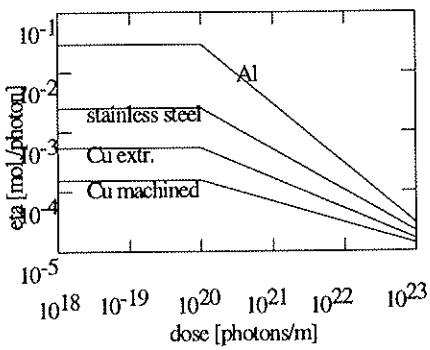



Fig.14 Desorption coefficient for aluminium, copper and stainless steel.

Of disadvantage is the fact that reliable joints can only be made by e-beam welding or brazing. At DESY different brazing techniques are used for joining copper to copper or copper to stainless steel [14]. If the brazing is done in a vacuum oven an additional cleaning is achieved. Sealing the copper vacuum system is done by using copper-to-stainless-steel transition pieces and the reliable CF system with stainless steel flanges and copper gaskets. These transitions are made by explosion bonding, roll bonding, e-beam welding or brazing.

The main absorbers of most SR sources are made from OFHC copper blocks with brazed water lines or drilled cooling ducts. Since the temperature of an absorber is not homogeneous thermal stress occurs leading to cracks and failure as described in Sec. 2.4. In order to avoid failures when the heat load is too high for OFHC copper, GLIDCOP a copper aluminium oxide Cu $Al_2O_3$(0.3) alloy which allows more stress is used [13, 17]. But GLIDCOP is rather expensive and brazing it to copper is not easy.

Copper gaskets are standard for sealing. Mainly OFHC copper is used. For baking temperatures higher then 200°C a copper-silver alloy (0.1%) should be used which has an increased recrystallization temperature and prevents flowing of the copper. In addition a silver coating can prevent sticking of the gasket.

### 3.4 Carbon-fibre composite material

Vacuum chambers made from carbon-fibre composite have been developed for the collider region of the Super Proton Synchrotron (SPS) at CERN [18]. For this region (length 13 m) the material of the vacuum tube has to be highly transparent for the collision products on their way to the detector. Therefore a material with low Z and high strength should be used. Beryllium would have been the best choice, but it is rather expensive especially for a 13-m long beam duct. As a comparable alternative a carbon-fibre tube (thickness 1 mm) with an inner aluminium layer (40 μm) was developed, having a comparable transparency to beryllium but only about 15% of the cost. The aluminium layer was needed as a barrier for gas diffusion through the composite in order to obtain a low outgassing rate and to shield the detectors electromagnetically from the high frequency field of the stored beam. The chambers were designed for 1.5 bar pressure difference. One chamber was tested to 1.95 bar before imploding.

Several techniques have been tested in co-operation with industry (Dornier System GmbH Friedrichshafen, Germany). The final technique is schematically shown in Fig. 15. For the production of the final tubes 20 μm thick 40 mm wide aluminium ribbons were wound and glued with 50% overlap around a mandrel. Liners with thickness larger than 100 μm lead to adherence problems during curing or baking due to the different thermal expansion of the carbon composite and the aluminium liner. Thin layers deform plastically without breaking the bond between the layers. The long spiral glue line did not contribute significantly to the outgassing rate. Prestressed carbon-fibre bundles (overall diameter 0.1 mm) were wound in several liners around the aluminium layer after having passed through a bath of liquid resin. The tubes were then cured at 150° – 200°C. The metallic end pieces were glued to the inner surface of the composite tube with a glue curing at room temperature, in order to avoid thermal stress.

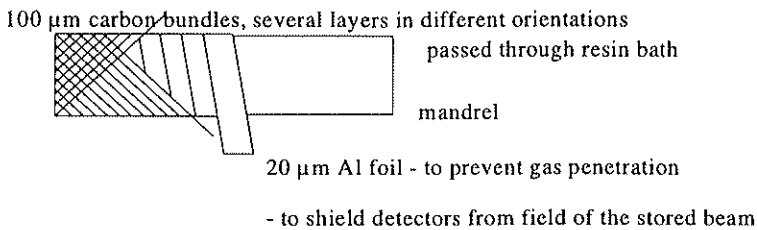



Fig. 15 Production of the CERN carbon-fibre vacuum tubes.

Chambers without an aluminium liner had a penetration rate for He ($10^{-8}$ mbar l/(s cm$^2$)) and water pressure. Chambers with an aluminium liner had a He penetration rate of $10^{-9}$ mbar l/(s cm$^2$). An outgassing rate of $10^{-12}$ mbar l/(s cm$^2$) could be obtained after two 24-h 120°C bakes.

### 3.5 Ceramics

Ceramics (mostly aluminium oxide which can be produced without porosity) are used as insulators (feedthroughs, breaks) and for vacuum chambers in rapidly changing magnetic fields (kickers) where conducting material cause unacceptably high eddy currents. Compared to glass, ceramic has the advantage of better mechanical strength. In addition the compressive strength is higher by a factor of ten compared to the tensile strength. Therefore the design of ceramic systems should only foresee compressive stress. Ceramics are formed out of a clay (oxide powder with water) which is moulded to fine particles, pressed in forms or extruded and left to dry. After drying, machining is possible. Then the parts are sintered at 2000°C. During this process the material shrinks by up to 20 % in length. The sintered parts can then only be processed by diamond tools [19].

Most ceramic components are connected to metal (flanges) by brazing. For the standard brazing procedure the ceramic surface foreseen for brazing is metalized by a MoMn or titanium hydride coating before sintering. Metals, preferably with an expansion coefficient comparable to the ceramic (nickel iron, nickel cobalt iron, titanium, molybdenum) can then be brazed to the ceramic. Copper-silver eutectic is used as filler. The interface metal can then be connected to stainless steel flanges by electron beam welding. In order to get a reliable connection different designs are used, as shown in Fig. 16. For round ceramics a thin metal is brazed from the outside like a collar. If the metal is brazed to the axis of the ceramic the metal part is sandwiched between ceramic to have a symmetric stress, or a thin transition metal is used which is more flexible.

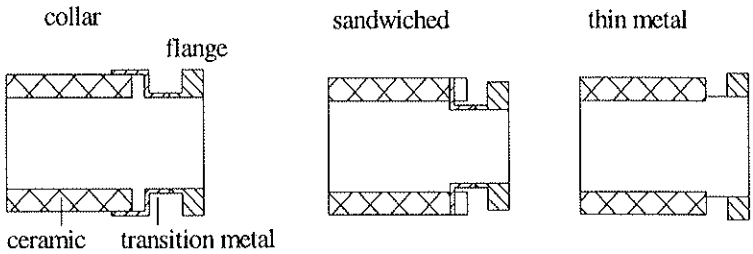



Fig. 16 Ceramic metal transition for Kicker chambers.

### 3.6 Magnetic materials

Normally, magnetic materials (ferrites, high μ iron, permanent magnets) are not used inside a UHV vacuum, since they all show a considerably higher outgassing rate compared to stainless steel. Due to the production process they might be porous and/or sensitive to corrosion. But since the performance of insertion devices increases with reduced gap some additional strength is gained by placing the magnets inside the vacuum. This is often accompanied by the installation of a high pumping speed to compensate the higher gas load.

Laminated silicon iron Fe Si(3) was used for the high field in the vacuum septum of ELETTRA [20] and BESSY II. The large surface due to the high number of thin (0.2 mm) laminations required caused considerable outgassing.

Ferrites are used for the pulsed magnetic fields found in kickers for example. Their relative permeability is about 1000 up to 10 MHz and a maximum field of 0.3 T. The general composition is (NiZn) $Fe_2O_4$. The production process is the same as for ceramics apart from the lower temperature of 1200°C for sintering [21].

Most undulators are built with permanent magnets. As material for undulators FeNbB is used for the permanent magnets and soft iron or cobalt iron for the yoke. For the production of the permanent magnets a powder of the basic materials is milled to small sizes (2μm) and then pressed into a die in the presence of a magnetic field. Then the parts are sintered at about 1100 °C. This process leads to a reduction of 20% in length. The magnets must then be magnetised in a strong field with the same direction as during the pressing process [22]. FeNbB magnets are very sensitive to corrosion and should be nickel-plated for vacuum application. Further disadvantages are the low

demagnetisation temperature of 100°C that does not allow a baking, and a low resistance to radioactivity. As an alternative, permanent magnets made from cobalt samarium can be used. They can be baked and are less sensitive to radiation damage but are more expensive and have a slightly lower remanent field.

### 3.7 Glass

Glass was the standard vacuum material in the years before the transistor replaced the electronic tube [23]. Nowadays its use is mainly restricted to windows (and some special electronic tubes (thyratrons)). Useful are silicon-oxide glasses either pure (quartz) or borosilicate (Pyrex: 81% $SiO_2$, 13% $B_2O_3$), soda glass (72% $SiO_2$, 16% $Na_2O$), or lead glasses (56% $SiO_2$, 30% $PbO_2$) if shielding of gamma radiation is required. Glass is a compound which after fusion cools to a solid without crystallisation, it has no specific melting point but its viscosity decreases until it becomes fluid. The temperature at which stresses are released is called the annealing point (Quartz: 1000°C, Pyrex: 500 °C). Borosilicate and soda glasses can be formed at about 1000 °C (working point). Quartz has a very low thermal expansion coefficient and a high working temperature range and is resistive to thermal shock. The surface desorption rate is low but helium and hydrogen can penetrate through thin glass tubes, especially quartz and to a less extent pyrex. The main disadvantage is its high fragility. Furthermore joints and transitions to metals are hard to make since the thermal expansion of the materials must be matched very closely.

### 3.8 Beryllium

Beryllium is used in vacuum systems wherever a highly transparent material is needed. At DAFNE a 0.7-m long tube with 90-mm diameter and a spherical middle part of 200-mm diameter with a thickness of 0.5 mm will be used [24]. The main application of Be at synchrotron radiation sources is for windows to separate the UHV vacuum of a storage ring from the UV or atmosphere of a beam line. A 0.5-mm thick Be foil absorbs less than 10% of 10 keV X-rays.

Beryllium has some interesting mechanical properties. It has a high modulus of elasticity and a high yield strength superior to stainless steel and a high thermal conductivity comparable to aluminium. But it is rather brittle with an elongation of 3% at room temperature. This makes it hard to form (practical only at high temperatures 700 °C), but machining and extrusion is possible [25].

Vacuum tight connection of the Be foil to stainless steel or aluminium can be achieved by electron-beam welding or brazing. The brazing can be improved (diffusion brazed) when pressure is applied to the parts to be jointed during the brazing process.

Beryllium-containing materials pose a health risk when improperly handled. Inhalation of air born beryllium particles (with a size of less than 10 μm) can cause serious lung disorders, contamination can cause skin disease, production of cancer is also possible. Thus the maximum permissible exposure limit for a working area is 2 $\mu g/m^3$. Machining is mostly done in water or in a highly-ventilated environment.

### 3.9 Further materials

Inconel, a nickel alloy Ni Cr(16)Fe(8), is used as an alternative material to stainless steel in pulsed or oscillating magnetic fields i.e. booster synchrotrons in order to reduce eddy currents, since its electric conductivity is smaller than stainless steel and the walls can be made thinner due to higher mechanical strength [26].

A titanium alloy, Ti Al(6),V(4), is used for an undulator chamber at the SRS [27]. It can be machined as easily as aluminium and has a considerably high yield strength (three times higher than stainless steel) allowing to design vacuum chambers with thin walls.

Martensitic stainless steel, AM350 Fe Ni(17)Ni(4)Mo(3), is magnetic but has higher strength than austenitic steel. It is used for bellows in regions were a significant relative permeability is allowed i.e. bellows in photon beam lines.

For the LHC beam duct a relative permeability of less than 1.003 was needed at 10 to 20 K. This could be obtained by using an iron-mangan alloy [5].

## 4. BONDING

Explosion bonding, roll bonding and inertia bonding are processes by which dissimilar materials that cannot be welded together are joined. Bringing dissimilar materials together enables one to benefit from the different properties of both materials. A thin, electrically-high-conducting material for surface currents can be combined with a low-conducting material for the envelope to reduce eddy currents. Thermal high-conducting material (copper) can be joined to stainless steel in order to disperse a heat load on the copper with the benefit of stainless steel for the vacuum system. Extruded aluminium or copper tubes can be joined to stainless steel to permit use of the more reliable CF stainless flanges.

For explosion bonding [28], schematically shown in Fig. 17, the two materials are positioned at a defined distance. The upper material is covered by an explosive which is detonated from one end. During a controlled explosion a plasma jet is formed at the impact zone of the two materials which scrubs off several layers of both materials leaving virgin material at the interface. The bond line usually has a wavy form that increases the strength of the joint. The quality of the bonding depends on the amount of explosive and the initial distance of the materials. In some cases (aluminium to stainless steel) an interlayer has to be used to get good results [29].

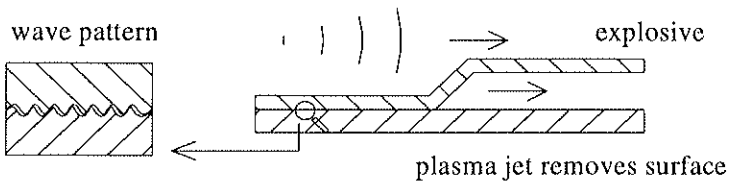



Fig.17 Explosion bonding.

Inertia bonding can be applied for joining cylindrical parts of different materials together. One part is rotated at high speed then the other non-rotating part is moved against it. Due to the high friction heat the materials are melted at the interface. With additional pressure the rotating part is decelerated and the two materials fuse.

## 5. CORROSION

A summary article about corrosion in vacuum systems can be found in Ref. [30]. Corrosion is an electrochemical process between an electrolyte and the metal surface. Thin walls, bellows, bimetallic contacts, brazings and welds and feedthroughs are mostly at danger. Atmospheric corrosion is the most usual and normally a relative humidity of more than 60% is needed. It should be noted that close to a cooling line the relative humidity can be considerably higher than in the remaining building. Furthermore the presence of radioactivity enhances corrosion considerably due to the formation of free radicals in the air. Ozone, nitrogen oxides, and subsequently acids, are formed with the moisture films on the surfaces.

Another source of corrosion is caused by spurious free chlorine ions from irradiated halogen-containing products (cable, tapes) or residuals from cleaning processes or brazing flux. Thus vacuum components should be thoroughly rinsed. Nitrogen oxides can also be formed by spurious discharges of high electric fields (i.e. feedthroughs for ion pumps, DESY has developed a specially designed plug to avoid this problem). Corrosion also occurs when materials with different electro-chemical potentials are connected (copper water pipe to an aluminium vacuum chamber). The corrosion is also enhanced by the speed of the water and impurities.

## REFERENCES

[1] W.D. Callister, Material Science and Engineering, ISBN 0-471-58128-3.

[2] K.C. Liu and C.M. Loring, Jour. Nucl. Mat. (1984) 789.

[3] K.D. Leedy, J.F. Stubbins, B.N. Singh and F.A. Garner, Jour. Nucl. Mat. 233 (1996) 547.

[4] S. Hermle, D. Einfeld, E. Huttel and G. Heidenreich, PAC, New York (1999).

[5] P. Cruikshank et. al., PAC97, Vancouver (1997) 3586.

[6] N. Terunuma, H. Nakayama and J. Urakawa, PAC, Dallas (1995) 1906.

[7] http://cindy.lbl.gov/cgi-bin/atten.pl

[8] H. Ishimaru, J. Vac. Sci. Technol., A2 (1984) 1170.

[9] C. Geyari, Vacuum 26 (1976) 287.

[10] LEP Vacuum Group, Vacuum 41 (1990) 1882.

[11] SMC Corporation, 1-016-4 Shimbashi, Minato-ku, Tokyo 105, Japan.

[12] G.Y. Hsiung, J.R. Huang, J.G. Shyy, D.J. Wang, J.R. Chen and Y.C. Liu, J. Vac. Sci. Technol., A12 (1994) 1639.

[13] J.R. Noonan, J. Gagliano, G.A. Goeppner, R.A. Rosenberg and D.R. Walters, PAC, Vancouver (1997) 3552.

[14] R. Ballion et. al., Vacuum 41 (1990) 1887.

[15] H. Hisamatsu et. al., Vacuum 47 (1996) 601.

[16] C. Perkins et. al., EPAC (1969), London, 2482.

[17] SMC Metal Products, Research Triangle Park, North Carolina, 27709-2166, USA.

[18] G. Engelmann, M. Genet and W. Wahl, J. Vac. Sci. Tech. A5 (1987) 2337.

[19] Friatec, 68229 Mannheim, Germany.

[20] R. Fabris, A. Favari and D. Tommasini, EPAC, Berlin (1992) 1460.

[21] Philips components, PO 10 63 23, 20043 Hamburg, Germany.

[22] Krupp WIDIA, Magnettechnik, PO 10 21 61, 4300 Essen, Germany.

[23] G.F. Weston, Vacuum 25 (1975) 469.

[24] A. Clozza and G. Raffone, EPAC (1999) 2491.

[25] Brush Wellman, Cleveland, Ohio 44110, USA, 517 (1997).

[26] H.C. Hseuh et. al., J. Vac. Sci. Techn. A10 (1992) 2085.

[27] N. Bliss and C.L. Dawson, PAC, Vancouver (1997) 3598.

[28] D. Shu et. al., Rev. Sci. Instrum. 66 (1995) 1783.

[29] Atlas Techn., Port Townsend, Washiongton 98368, USA, www.altlasbimetal.com

[30] J.P. Bacher and N. Hilleret, Vacuum 41 (1990) 1914.

## APPENDIX

Table 3
Material properties of selected materials [1].

| Material | Density $10^3 kg/m^3$ | Melt pt °C | Therm. cond. W/(mK) | Elect. cond. $10^6/(\Omega m)$ | exp. coeff. $10^{-6}/K$ | E-modul $10^9$ Pa | Yield strength $10^6$Pa | Tensile strength $10^6$Pa | Elonga-tion % |
|---|---|---|---|---|---|---|---|---|---|
| Al6063 | 2.8 | 507 | 192 | 29 | 23 | 72 | 250 | 300 | 12 |
| Alumina | 4.0 | 2050 | 30 | $10^{-12}/(\Omega m)$ | 8 | 393 | - | 270 | 0.1 |
| Aluminium | 2.7 | 659 | 231 | 36 | 24 | 69 | 17 | 55 | 25 |
| Beryllium | 1.9 | 1280 | 216 | 24 | 12 | 303 | 270 | 350 | 3 |
| Brass Cu-Zn | 8.5 | 915 | 120 | 16 | 20 | 110 | 75 | 303 | 68 |
| Bronze Cu-Sn | 8.8 | 880 | 62 | 8 | 18 | 110 | 152 | 380 | 70 |
| Diamond | 3.5 | | 2000 | | 1 | 1000 | - | 800 | 0.1 |
| Copper | 8.9 | 1083 | 398 | 58 | 17 | 110 | 152 | 220 | 45 |
| GlidCop-15 | 8.9 | 1083 | 365 | 54 | 17 | 130 | 330 | 420 | 25 |
| Fused silica | 2.2 | | 1.3 | $10^{-12}/(\Omega m)$ | 0.5 | 75 | - | 110 | 0.1 |
| Iron | 7.9 | 1558 | 12 | 10 | 12 | 207 | 130 | 260 | 45 |
| 304 | 7.8 | 1370 | 16 | 1.4 | 16 | 193 | 207 | 552 | 60 |
| AM350 | 7.7 | 1500 | 25 | 2 | 10 | 210 | 1000 | 1500 | |
| Magnesium | 1.7 | 650 | 122 | 17 | 27 | 45 | 41 | 41 | 14 |
| Molybdenum | 10.2 | 2610 | 142 | 19 | 4.9 | 324 | 565 | 655 | 35 |
| Nickel | 8.9 | 1454 | 80 | 12 | 13 | 207 | 138 | 483 | 40 |
| Inconel | 8.4 | 1400 | 15 | 1 | 11 | 157 | | 1200 | |
| Pyrex | 2.2 | | 1.4 | $10^{-12}/(\Omega m)$ | 3.3 | 69 | | 70 | 0.1 |
| Silver | 10.5 | 961 | 418 | 56 | 19 | 76 | 55 | 125 | 48 |
| Titanium | 4.5 | 1670 | 17 | 2 | 9 | 107 | 240 | 330 | 30 |
| TiAl(6) V(4) | 4.4 | 1600 | 6 | 0.6 | 8 | 110 | 900 | 1000 | 14 |

Table 4
Vapour pressure as a function of the temperature for different elements [23].

| Element | Melting pt | $10^{-9}$ Torr | $10^{-7}$ Torr | $10^{-5}$ Torr | $10^{-3}$ Torr |
|---|---|---|---|---|---|
| Ag | 1234 | 800 | 899 | 1030 | 1195 |
| Al | 932 | 906 | 1015 | 1155 | 1355 |
| Au | 1336 | 1020 | 1150 | 1260 | 1365 |
| Ba | 983 | 510 | 583 | 690 | 810 |
| Be | 1556 | 925 | 1035 | 1175 | 1365 |
| C | - | 1845 | 2030 | 2250 | 2520 |
| Ca | 1123 | 524 | 590 | 675 | 790 |
| Cd | 594 | 328 | 368 | 422 | 494 |
| Ce | 1077 | 1175 | 1325 | | |
| Co | 1768 | 1130 | 1265 | 1435 | 1650 |
| Cr | 2176 | 1055 | 1175 | 1335 | 1540 |
| Cs | 302 | 241 | 273 | 319 | 383 |
| Cu | 1357 | 945 | 1060 | 1215 | 1415 |
| Fe | 1809 | 1105 | 1230 | 1380 | 1595 |
| Ge | 1210 | 1030 | 1150 | | |
| Hg | 234 | 190 | 214 | 245 | 289 |
| In | 429 | 716 | 812 | 943 | 1110 |
| Ir | 2727 | 1755 | 1960 | 2070 | 2380 |
| K | 336 | 276 | 315 | 364 | 435 |
| La | 1193 | 1220 | 1375 | | |
| Mg | 923 | 432 | 487 | 560 | 655 |
| Mn | 1517 | 734 | 827 | 970 | 1125 |
| Mo | 2890 | 1770 | 1975 | 2260 | 2650 |
| Na | 371 | 328 | 370 | 431 | 511 |
| Ni | 1725 | 1145 | 1270 | 1415 | 1660 |
| Pb | 601 | 580 | 656 | 760 | 900 |
| Pd | 1823 | 1050 | 1185 | 1430 | 1660 |
| Pt | 2043 | 1480 | 1655 | 1875 | 2180 |
| Re | 3453 | 2100 | 2350 | | |
| Rh | 2239 | 1470 | 1640 | | |
| Sb | 903 | 526 | 582 | | |
| Se | 490 | 317 | 356 | | |
| Sn | 505 | 900 | 1020 | 1155 | 1365 |
| Sr | 1043 | 483 | 546 | 615 | 729 |
| Ta | 3270 | 2120 | 2370 | 2470 | 3080 |
| Th | 1968 | 1610 | 1815 | 1959 | 2272 |
| Ti | 1940 | 1265 | 1410 | | |
| W | 3650 | 2270 | 2520 | 2872 | 3280 |
| Zn | 963 | 374 | 421 | 481 | 563 |
| Zr | 2128 | 1665 | 1855 | 2110 | 2460 |

# ENGINEERING

*Lars Westerberg*
The Svedberg Laboratory, Uppsala University, Uppsala, Sweden

**Abstract**
The performance of a UHV system depends to a large extent on the effort spent in optimising the vacuum chamber as well as the components of the system. This chapter describes the design principles of a UHV system and its building blocks, joining methods, the combined action of flanges, gaskets and bolts, welding and weld preparation, valves, expansion bellows and bakeable vacuum systems.

## 1. DESIGN PRINCIPLES OF A UHV SYSTEM AND ITS BUILDING BLOCKS

### 1.1 Design criteria

In order to obtain the optimum performance at the lowest possible price it is useful to set up well-defined design criteria. Let us start with the *design pressure*. Obviously it does not pay to hunt for a pressure much better than this. On the other hand, it can be a real problem if the needed pressure is not at all achieved, or not achieved within reasonable time. Remember that it can be useful to have some margin to account for small problems. A $10^{-7}$ Pa vacuum system needs baking up to 150 °C, or, maybe, no baking at all if the outgassing is not too high and the pumps are large enough. On the other hand, $10^{-8}$ to $10^{-9}$ Pa needs baking to higher temperatures and requires that special precautions be taken to obtain sufficiently-low outgassing. The outgassing rate depends on the material properties, the surface treatment and cleaning procedures.

Let us define the needed *pump-down time*. A baked system will reach UHV conditions much faster than an unbaked system, although baking takes 24 to 36 h depending on the baking temperature. Accelerator vacuum systems with *circulating ion beams* as well as other UHV systems may have restrictions on *the rest-gas composition*. It is usually required to minimise the average mean square $Z$ of the rest gas. Therefore one should use materials with low outgassing and try to minimise the effects of *gas loads* from process gas or gas-jet targets in the vacuum system.

The *operating temperature* of the UHV system is another important factor to be taken into account. A room temperature system is quite different from a cryogenic one. Systems operated at elevated temperature or with hot cathodes have much higher outgassing than room temperature systems.

Last but not least comes the overall *budget* which can be split up into *capital investment* (purchasing of the system) and *operational costs*. The latter depends on the need for consumables and maintenance. It can be useful to include operational costs for 5 to 10 years when comparing for instance different pump alternatives.

### 1.2 Engineering formulae for vacuum systems

It is important to be aware of the conductance limitations for different shapes of vacuum chambers. Some examples for standard type tubes are given in Eqs. (1)–(3). For more complicated structures it is recommended to make Monte Carlo calculations (see Section 1.3).

The conductance, $W$ [m$^3$/s] through an orifice is given by

$$W = 36.4A\sqrt{\frac{T}{M}} \tag{1}$$

where $A$ [m$^2$] is the area of the orifice, $T$ is the temperature [$^o$K] and $M$ is the molecular weight. This is usually calculated for nitrogen. The conductance for hydrogen is 2.8 times higher. Similarly, for a long cylindrical tube where $r$ [m] is the diameter and $L$ [m] is the length

$$W = 305\frac{r^3}{L}\sqrt{\frac{T}{M}} \ . \tag{2}$$

Finally, for a long elliptic-cross-section tube where $a$ [m] and $b$ [m] are the elliptic axes

$$W = 431\frac{a^2b^2}{L\sqrt{a^2+b^2}}\sqrt{\frac{T}{M}} \ . \tag{3}$$

The effective pumping speed, $S_{\mathrm{eff}}$ [l/s] for a pump with the nominal pumping speed $S_{\mathrm{o}}$ [l/s] connected to a vacuum chamber via a tube with conductane $C$ [l/s] is given by

$$\frac{1}{S_{\mathrm{eff}}} = \frac{1}{S_{\mathrm{o}}} + \frac{1}{C} \ . \tag{4}$$

For large vacuum systems one uses either linear pumps or distributed (lumped) pumps as shown in Fig. 1.

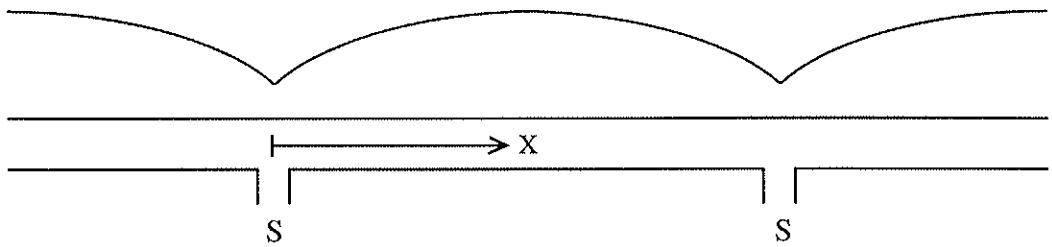



Fig. 1 Distributed pumps. Bottom: beamline with pumping ports. Top: pressure distribution in arbitrary units

Following a calculation by Gröbner [1] for the molecular flow regime we define: the distance between pumps $L$ [m], the effective pumping speed $S$ [l/s], the pressure $P$ [mbar], the specific molecular conductance $w$ [m l s$^{-1}$], the specific surface area $A$ [cm$^2$ m$^{-1}$], the specific outgassing rate $q$ [mbar l s$^{-1}$cm$^{-2}$], the gas flow $Q$ [mbar l s$^{-1}$].

$$Q(x) = -w\frac{dP}{dx} \tag{5}$$

and

$$\frac{dQ}{dx} = Aq \tag{6}$$

give

$$w\frac{d^2p}{dx^2} = -Aq \ . \tag{7}$$

$$\left.\frac{dP}{dx}\right|_{x=L/2} = 0 \tag{8}$$

and

$$P|_{x=0} = \frac{AqL}{S} \tag{9}$$

give

$$P(x) = Aq(\frac{Lx - x^2}{w} + \frac{L}{S}) \tag{10}$$

From this can be calculated the maximum pressure

$$P_{\max} = Aq(\frac{L^2}{4w} + \frac{L}{S}) \tag{112}$$

and the average pressure

$$P_{av} = \frac{1}{L}\int_0^L P(x)dx = Aq(\frac{L^2}{6w} + \frac{L}{S}) \ . \tag{12}$$

### 1.3 Computer simulations

Properly done computer simulations prior to purchasing and manufacturing of a complicated vacuum system can save money as well as a lot trouble at a later stage. A few computer programs are available for vacuum system simulations. Finite-element calculations as well as Monte Carlo calculations have been developed by Pace and Poncet [2]. Monte Carlo calculations are recommended for conductance calculations of non regular structures, where simple formulas like (1)–(3) can not be used. Ziemann [3] has developed a program for vacuum system simulations using matrix multiplications in analogy with magnetic field calculations. There is a commercial program, VacSim™ [4], which also provides design and pumpdown data.

### 1.4 Materials

Having defined the criteria mentioned above one should be able to decide which vacuum chamber material to use. The options are usually only aluminium and stainless steel. Aluminium is preferred for systems with synchrotron radiation load. It requires a special welding technique. Most vacuum components are nowadays also available in aluminium [5]. Stainless steel products are however less expensive and are available from a large number of companies. The most common qualities 304, 304L, and 316L. 316LN or 304LN are often required for their high mechanical strength and low magnetic permeability, typically 1.005. For use in magnetic field a magnetic permeability less than 1.01 is usually required.

Vacuum firing is a very effective way to reduce outgassing of stainless steel. The material is heated quickly to 950 °C and remains at this temperature for 1–2 h at a pressure lower than $10^{-3}$ Pa. It is customary to vacuum fire the completed chamber with all flanges welded on. It is recommended to use 316LN or 304LN material in vacuum fired chambers.

Titanium is an alternative material which is not often used. It is more expensive than stainless steel and more difficult to weld. When also taking into account the working cost, a Ti chamber is about 20% more expensive than a stainless steel chamber. Beryllium is often required for thin windows etc. (see Section 2.3).

### 1.5 Construction

When constructing non-standard vacuum chambers it is important to make at least simple mechanical stability calculations. For very difficult structures it is recommended to perform calculations with finite-element codes like the ANSYS™ code. An example of such calculations for a thin window is given in Ref. [6]. Standard tolerances in construction drawings are often too high for certain applications. It is therefore necessary to indicate the needed tolerances, e.g. of

flanges which should be aligned to the same axis etc. Some vacuum companies have simple construction programs available from their web site, of course only with part numbers for flanges etc from their own company.

### 1.6 Purchasing vacuum systems and vacuum components

When purchasing vacuum products it is useful to compare specifications from different manufacturers. If certain quantities are very important, ask for copies of test results of, for example, pumping capacities of H and He in a cryo pump, pumping speeds of noble gases in ion pumps, maximum magnetic field that a cryo pump or a turbo pump can accept etc. For large systems it is recommended to prepare a *call for tender* document, where the required performance of all equipment is specified. One can usually save money by placing the order split over different manufacturers, rather than requiring that one manufacturer delivers all items.

If you do not have your own workshop for the vacuum chamber manufacturing you can ask for quotations from the major vacuum companies as well as from many specialist companies and workshops. To make sure that you find competent firms ask, in your call for tender for vacuum chambers, for references to other customers who placed similar orders. If references cannot be given, be careful and insist on the production of test samples. Stipulate in the specification how welds should be done, if mechanical tolerances beyond the standard ones are necessary, what cleaning methods are allowed, to what standard the leak detection should be done at the company and/or after delivery. If the chamber will be subject to vacuum firing and/or baking, stipulate that it must be leak tight after these treatments.

## 2. TYPICAL BUILDING BLOCKS IN A UHV SYSTEM

### 2.1 Standard catalogue items

In this section the typical building blocks of a vacuum system are treated. Standard multipurpose *vacuum chambers* are available from most manufacturers. Such chambers can usually be used for auxiliary systems, while process chambers, beam line chambers and chambers for diagnostic devices etc of an accelerator usually have to be prepared specially.

***Vacuum pumps and traps*** are treated specially in other chapters of these proceedings. Here we shall only, for the sake of completeness, give an overview of available types and their specific use. *Forepumps* are either one- or two-stage *rotary vane pumps* (with oil) or ***dry pumps*** (without oil) such as *membranes, Roots or sorption pumps.*

*High vacuum pumps* operate in the pressure range from $10^{-2}$ Pa and down. *Diffusion pumps* are often used in the high vacuum (HV) range, but should avoided in UHV. Turbo pumps can be used in high vacuum as well as UHV, provided that there is a drag stage which improves the compression ratio for hydrogen. *Cryopumps* and *sputter-ion pumps* are used in HV and UHV. Some typical ***pump combinations*** for the UHV range are: sputter-ion pumps and *titanium sublimation pumps* as well as sputter-ion pumps and *non-evaporable getter pumps. Liquid He cryopumps* have long been used for the lowest pressures.

The most common *vacuum gauges* are: Pirani gauges for the fore vacuum range, Penning or cold-cathode gauges in the HV range. Special cold-cathode gauges with triax coaxial cables can measure down to $10^{-9}$ Pa. Ionisation gauges work down to $10^{-11}$ Pa. It is still a challenge to make reliable pressure measurements in the XHV range. The extractor gauge is one alternative. *Rest-gas analysers* for UHV should be vacuum fired in order to have sufficiently low self-outgassing of hydrogen.

There is a wide range of standard *electrical feedthroughs* for current and voltage available on the market. For signals there are 50 Ω feedthroughs of many types such as BNC, N, SMA etc. These feedthroughs are expensive and require much space on the flange. For projects requiring many signal feedthroughs in a close-packed array there are 15 to 50 pin D-SUB contacts. These have been available from a couple of companies for several years. The product range includes internal bakeable cable contacts for Kapton-insulated cables. On the outside one can use cheap standard contacts and cables, provided that they can be demounted during bakeout.

*In-vacuum cables* for UHV are either ceramic-bead insulated with metal inner (and outer) conductor(s), Kapton-insulated metal conductor, or hermetically-sealed cables with stainless-steel outer and a central conductor inside an insulator. The Kapton cables are available as single conductors, coaxial cables with 20 to 50 Ω characteristic impedance, and thermocouple cables. There are also Kapton-insulated band cables with 25 parallel conductors.

*Liquid feedthroughs* are available for water etc. and liquid nitrogen. *Rotational* as well as *linear feedthroughs* for UHV are usually bellows sealed. The rack-and-pinion linear feedthrough and permanent-magnet-type manipulators are used for long strokes, e. g. for sample transfer etc. All types can be motorised. Some of the commercially available motors are rather weak and are also quite expensive, probably since very few are sold. It is therefore possible to save money by home-built motorised versions.

Standard *UHV fittings* include flanges, screws, bolts, elbows and T-pieces. The standard flange steel is usually 304, while 316LN is often required for accelerator vacuum systems. *Optical windows* are available for different wavelength regimes: standard glass for 300-3000 nm, quartz for 230-4000 nm, sapphire for 150-5500 nm, and magnesium fluoride for 100-8000 nm. There are also radiation resistant windows. A shutter can be installed inside the window to protect it from evaporation onto the glass.

*Bakeout equipment* will be treated in Section 8. Since vacuum companies usually only provide heating jackets for valves and ion pumps it is necessary to go to special companies for heating tapes, heating jackets, metal heating collars, thermo elements, bakeout controllers etc. For large systems there is also a need for a computer control for the bakeout heaters. IR lamps are available as an alternative baking method, where the inside surface is heated. All areas may not see the same temperature however.

Computer controls for vacuum systems are not available from vacuum companies. See the chapter by D. Schmied in these proceedings.

### 2.2 Special parts for UHV systems, standard or non-standard items

It is recommended to use silver- or gold-plated screws in bakeable vacuum systems, since uncoated screws may seize after baking. For screws in a tapped hole it is recommended to drill a ventilation hole through the centre of the screw, see Fig. 1 in the chapter "The best laid schemes...".

There is a wide variety of insulator material which can be used in UHV systems. *Ceramics*, e.g. alumina are difficult to cut, but can be ordered specially from ceramics companies. Machinable ceramics, e.g. Macor (by Corning) or Shapal M-Soft (by Narasaki Sangyo Co) can be machined in a workshop. Other often used insulators are Vespel, PEEK and Kapton. It is recommended to make outgassing tests of such products since there are different grades. We have had problems with excessive outgassing in some deliveries.

### 2.3 Experimental equipment

There is a wide variety of evaporation sources, piezo-electric microbalances, surface analysis instruments etc. available from vacuum companies as well as from specialist companies.

In order to perform nuclear and particle physics experiments inside the UHV system of a storage ring, it is necessary to find detector materials, cables, connectors etc. which fulfil the UHV requirements. At the CELSIUS storage ring in Uppsala, Sweden a few groups are using germanium and silicon solid-state detectors as well as GSO scintillators with photo-diode readout. A report on scintillator materials for storage ring work is found in Ref. [7]. Progress reports on the CHICSi 10 μm +300 μm silicon barrel particle detector, which also includes GSO crystals are given in Refs. [8] and [9]. Microchannel plates are used for electron amplification in time-of-flight detectors for slow recoil fragments, Ref. [10]. For the latter project a ceramic thin-film circuit-board voltage divider was developed by Monolitsytem AB, Sweden. The resistors could be laser trimmed to high precision. We have vacuum tested such circuit boards with up to 19 layers.

We have tested outgassing of single- and double-sided Kapton circuit boards prepared by Xicon AB, Sweden. They were approved for use in the storage ring. Other types of Kapton multilayer circuit boards have not fulfilled our requirements.

Thin windows are needed for detection of particles (pions, protons, heavy ions) in air outside the vacuum chamber. The window should have high transparency

$$X_0 = \frac{t_{\text{eff}}}{L_R} \tag{14}$$

where $t_{\text{eff}}$ is the effective thickness of the window, taking into account if the particle enters the window normal to the surface or at an angle. $L_R$ [mm] is the radiation length. The most common materials are beryllium, carbon-fibre composites, aluminium and stainless steel. Their Youngs modulus E are 29000, 22000, 6000 and 19000 [daN/mm$^2$] and radiation length $L_R$ 353, 188, 89 and 17.5 [mm], respectively.

Beryllium is the best material. It is however expensive and dangerous to manufacture and to handle. Composite carbon fibre is also quite expensive since it has to be made to special order. Aluminium and stainless steel are the more-easy-to manufacture alternatives. For security reasons it is important to set a high safety margin. For a 0.8 mm thick and 500 mm diameter aluminium window in the form of a sphere with the pressure from the outside of the sphere, a buckling pressure of 3 bar was calculated. An experiment showed that buckling occurred at 3.1 bar [6].

## 3. JOINING METHODS

A thorough description of joining methods is found in Refs. [11] and [12]. Seals can either be demountable (such as elastomer and metal seals), or permanent seals (such as glass, glues, brazing, soldering and welding). A combination of these methods is practised at Dubna Laboratory in Russia, where two thin radial lips are sealed by welding and opened again by a cutting tool looking like a can opener. This process can be repeated several times until the welding lips have disappeared.

Historically, permanent vacuum seals and entire vacuum systems were prepared by *glass* makers. Very little of this remains today, but glass-to-meal joints can still be found in vacuum catalogues. *Glues* can be used in UHV systems provided that the outgassing is not too high. It is recommended to perform outgassing tests. In this way one can be sure that the whole process including oven heat treatment works. There is a list of glues and their outgassing properties on the NASA home page. We have used the non-conductive, conductive and heat-conductive Epotek two-component glues with the heat treatment prescribed by the manufacturer.

Permanent seals are brazing and welding [TIG (Tungsten Inert Gas), plasma welding and laser welding]. Demountable seals are O-ring seals (used on the fore vacuum side only) and various types of metal seals (Al, In, Pb, Ag, Au and Cu, with different temperature ranges). Indium and Au are used in cryogenic systems. The most common metal gaskets are Cu (OFHC). OFHC (silver plated) is used in baked systems and OFS (0.1% Ag) is a harder gasket which keeps its hardness after multiple bakeouts. Other types are Helicoflex™ and alpha C-seal™. These are spring-loaded seals with an outer seal of Al or Cu, depending on the temperature range. The VAT-seal™ is used for rectangular shapes. Ferrofluidic seals are used for rotational axes, however not in the UHV range.

### 3.1 Flange types

Small flanges (ISO-KF) are used on the fore vacuum side with elastomer seals. With a spring-loaded seal like Helicoflex™ and special clamps it is possible to use this flange type also in UHV. The available nominal widths (inside diameters) are 10, 16, 20, 25, 40 and 50 mm. ISO-K (ISO-F) are used in high vacuum systems, but not in UHV with metal seals due to lack of sealing force. Available nominal widths are 63, 100, 160, 200, 250, 320, 400, 500, 630, 800 and 1000 mm.

The Conflat™ flange is the most widely used type for UHV systems. It is available in the following nominal widths: 16, 40, 63, 100, 160, 200 and 250 mm. A few companies deliver alternative dimensions like 75 and 125 mm etc. The standard steel is 304. For vacuum firing or mounting in magnetic fields 316LN should be used. Old 250 mm flanges may not fit each other since there were different numbers of screw holes and screw diameters etc.

The Wheeler Flange™ by Varian is an alternative for very large diameters. Special flanges of the small-flange type with a large clamp chain and Helicoflex™ seal were developed for the SPS ring at CERN, Ref. [13]. Enamel-coated versions offer a cheap alternative to ceramic electrical isolation of a beam tube, at low isolation voltage though. The Pyraflat flange™ by Thermionics is a possibility for square, rectangular or other odd-shapes. It has a Conflat type Cu gasket.

## 4. COMBINED ACTION OF FLANGES, GASKETS AND BOLTS

The elastic-plastic deformation of a flange seal is illustrated in Fig. 2. At low sealing force the deflection is elastic, while at high sealing force there is a plastic flow. At higher temperature the curve is shifted downwards and the plastic deflection is increased. It is important for a flange-seal system to have adequate spring-back resulting in a large enough sealing force. This depends on the slope of the elastic curve, and the unloading due to plastic flow of the system.

For Conflat seals it is recommended to use a torque wrench and increase the torque stepwise after completing a full turn, e. g. for NW16CF 1st turn 4 Nm and 2nd turn 6.5 Nm; for NW35CF 1st turn 5 Nm, 2nd turn 10 Nm, 3rd and 4th turn 15 Nm; for NW63-200CF 1st turn 10 Nm, 2nd turn 20 Nm, 3rd and 4th turn 30 Nm.

For best performance in very large accelerator systems like the CERN ISR machine it was found that the bolts must have adequate tensile strength and yield stress. Standard bolts have class 80 (800 N/mm$^2$) tensile strength and yield stress. It was found that class 100 (1000 N/mm$^2$) screws had many less failures. The bolts are Molykote treated for better performance when baking. The nuts are glass bead blasted and Molykote treated. Washers have special high hardness. In this way the ISR ring leak rate was only 0.06 %, i.e. 1 in 1700 flanges leaked per year.

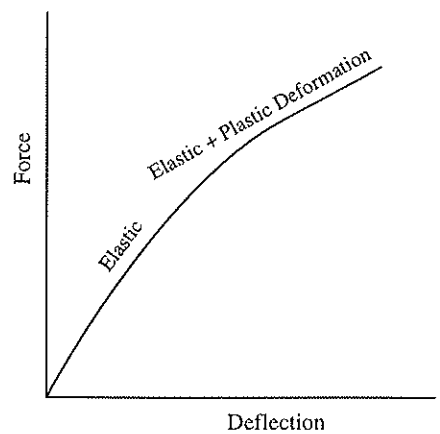



Fig. 2 Elastic and plastic deformation characteristic

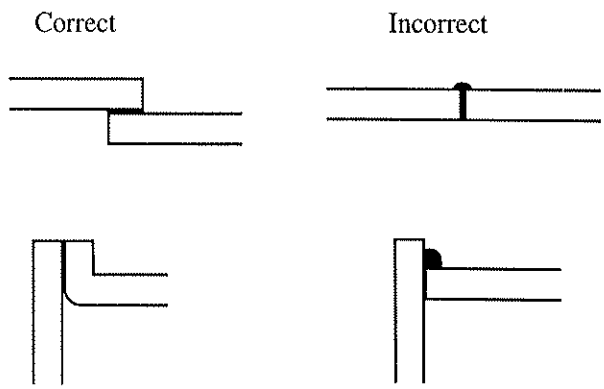



Fig. 3 Correct and incorrect brazing preparations

The outside diameter of a Conflat Cu gasket has a clearance of 0.1 mm to the flange. When the seal is partly tightened there is a radial force in the copper towards the flange. When the bolts are tightened more the seal hits the flange. When fully tightened there is a very high force back onto the sloping surface before the knife edge.

## 5. BRAZING AND BRAZING PREPARATIONS

Brazing involves a molten filler metal being drawn by capillary attraction into the space between closely adjacent surfaces. It is usually done in a hydrogen atmosphere in a dome open from below to prevent oxidation of the joints. In Fig. 3. are shown examples of correct and incorrect brazing preparations. Brazing can be done in several steps with fillers that melt at successively lower temperature.

## 6. WELDING AND WELD PREPARATION

Tungsten Inert Gas (TIG) welding is the most commonly used technique for welding vacuum components. Plasma welding is commonly used for welding thin material, e. g. bellows and thin windows. Laser welding is a new quite expensive technique which may be more used in the future. Explosion welding is used as an alternative for material which cannot be welded in any other way.

General guidelines for weld preparations:

- Cleanliness is a must.
- Use protective Ar gas, also from the outside.
- Use no filler. If necessary, use correct filler recommended and provided by the steel manufacturer.
- A vacuum chamber should always be welded from the inside. In preparing for welding it is often useful to spot weld (tack weld) from the outside to fix the structures to each other.

Examples of correct and incorrect welding preparations are shown in Fig. 4. Welding on both the inside and outside of a chamber means that the welds can not be leak detected properly. If the inside weld leaks there is a virtual leak. Some examples of welding lips for bellows are shown in Fig. 5.

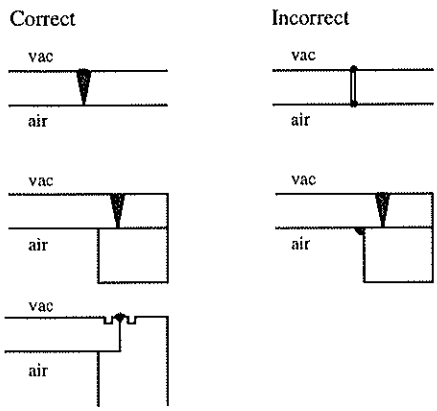



Fig. 4 Weld preparations

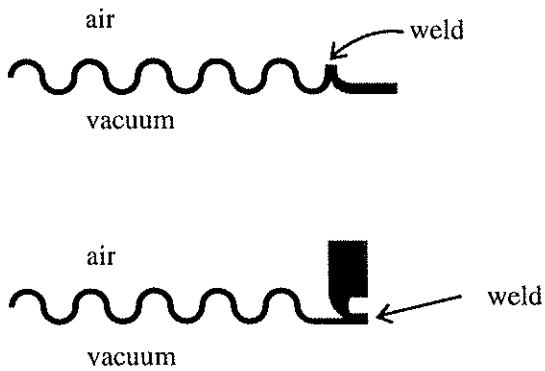



Fig. 5 Two examples of welding lips for bellows

Welding a thin window is quite delicate. The left part of Fig. 6 shows the welding preparation for a large-area, thin, stainless-steel window 450 mm ID. Welding is done from the outside. The structure to the right in Fig. 6 was used for windows from 100 to 250 mm diameter. The pieces are pressed together and are welded from the outside.

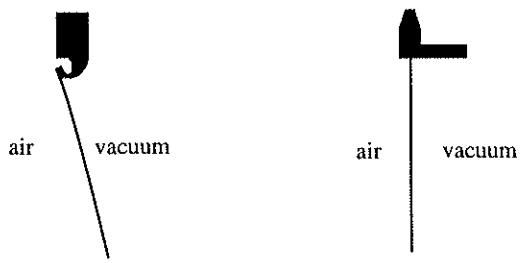



Fig. 6 Examples of thin-window welding

## 7. VALVES

Fore-vacuum and high-vacuum valves are usually O-ring sealed. These valves are made from aluminium or stainless steel. For UHV it is necessary to have all-metal valves. These are usually made from stainless steel. Aluminium all-metal valves are however also available. *Angle valves* and *straight through valves* are usually bellows sealed. When sealing a gate valve there is first a sideways movement of the sealing disc followed by a movement towards the seat.

Leak valves with sapphire crystal as a sealing surface are used for controlled inlet of gases into a UHV system e.g. or calibration of a rest-gas analyser. Calibration data on the size of the leak as a function of the number of turns should be supplied by the manufacturer.

Fast-closing valves are used for safety interlocks to protect vacuum systems from sudden pressure failures. There are leak-tight fast-closing valves and non-leak-tight fast-closing shutters. The latter type protects the system until an ordinary slow valve can be closed. The closing time is in the range 10 to 20 ms. A simple and very effective all-metal shutter is shown in Fig. 7. It has been used in the HV range, but can if made from stainless steel also be used for UHV.

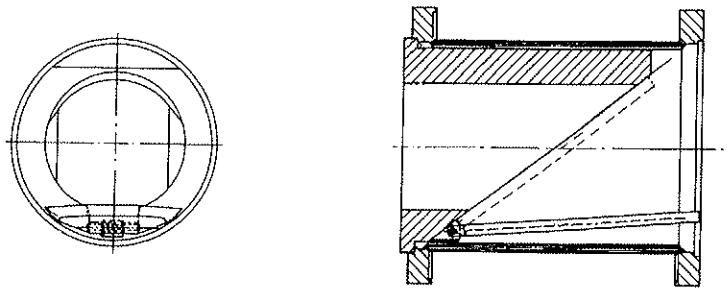


Fig. 7 A simple and very effective all-metal shutter constructed by O. Byström, TSL, Uppsala, Sweden

## 8. EXPANSION BELLOWS

The usual material for bellows is stainless steel of grades 304, 304L or 316L. Aluminium bellows are also available. Hydraulically-formed bellows have elongations of about 10%, while membrane (or diaphragm) bellows have much higher elongations, 50 to a few hundred %, depending on the type and inside and outside diameters of the membranes. For bellows which are used very frequently it is important to notice the maximum of strokes guaranteed by the manufacturer. The manufacturer can also tell how skew operation of the bellows reduces this number. Examples of welding lips for bellows are shown in Fig. 6. In many accelerators it is required that bellows assemblies have internal rf bridge shields.

## 9. BAKEABLE VACUUM SYSTEMS

All materials of a bakeable system must be compatible with the temperature cycling. There are at least 3 orders of magnitude difference in outgassing between baked and unbaked chamber material. Therefore, unbaked or low temperature baked parts will, even if their total area is only a few %, dominate the total outgassing. The most common bakeout cycles are 150 °C (to get rid of water vapour), 300 °C or 450 °C. The latter bakeout cycle can be applied to a system which is not vacuum fired in order to reduce the ougassing further. Air bakes at 400 °C have successfully been used for gravitational wave interferometer vacuum systems like LIGO. Here the stainless steel has been air baked prior to welding. A comparison outgassing and hydrogen content in air-baked and vacuum-fired stainless steel is found in Ref. [14]. Further studies, including large scale bakeout tests are going on [15].

### 9.1 Baking hardware

The standard types of heating tape are 150 and 300 W/m. There is a selection of standard lengths. The following simple formula can be useful to determine the length, $L$, of a heating tape for a vacuum chamber.

$$L = \frac{AJ}{2B} \tag{14}$$

where $A$ is the total area of a tube, $B$ is the width of the tape, $J$ is a goodness factor, which is an estimate of the fraction of the chamber which can be wound with heating tape (roughly the length of the chamber which is not covered by flange collars or other obstructions. The distance between two turns should be the same as the tape width. Never cross tapes since this can cause a short circuit. Use double tapes in magnet chambers and other hard-to-access-chambers for fast replacement in case a tape breaks down.

For good heat contact with the vacuum chamber, braze the thermocouple end onto a 10 mm × 25 mm stainless steel piece. The non magnetic type E thermocouple is recommended for chambers placed in magnetic fields. In Fig. 8 is shown how to mount the thermocouple onto a chamber. Use Kapton tape or glass fibre tape to attach the beginning and the end of the heating tape on to the chamber. Mount two thermocouples on hard-to-access chambers, e.g. in magnets. In this way a spare is easily available if the thermocouple breaks down. Continue by winding the heating tape onto the vacuum chamber as shown in Fig. 8.

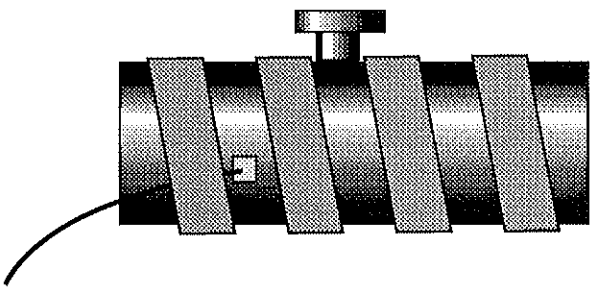


Fig. 8 Thermocouple and heating tape wound on a chamber

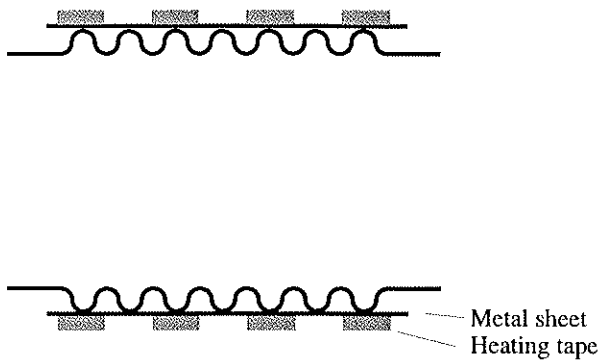



Fig. 9 Mounting of baking equipment for bellows. The outer insulation is not shown.

It is recommended to use as thick insulation as possible to reduce the power consumption. Insulation sheets can be bought from heating tape suppliers and from other companies. Use the following thickness guidelines: with 25 mm insulation the power needed for 300 °C and a bit beyond is 16 W/dm$^2$, for 15 mm 20 W/dm$^2$, for 10 mm 30 W/dm$^2$ and for 5 mm 45 W/dm$^2$. Seal the insulation on the outside by 5 cm-wide aluminium tape. For chambers where a very thin insulation is required it may be necessary to water cool a copper sheet outside the insulation to protect for example a magnet from being overheated.

Do not wind the heating tapes directly on bellows due to bad heat transfer and the risk of short circuiting. Instead, wind a thin (0.1 mm) stainless-steel sheet around the bellows and wind the heating tape on the outside as shown in Fig. 9.

Flanges can be baked by metal heating collars with the same width as the two flanges together. Use non magnetic heaters in magnetic fields. Soft heating jackets are also available for standard flanges. Bakeout jackets are commercially available for valves, ion pumps etc. For other chambers custom made bakeout jackets are an alternative to heating tape and insulation. This is also recommended where the bakeout equipment for space reasons must be demounted after bakeout.

Power controllers can be bought from many electrical companies. Make sure they are made for your thermocouple type. The controllers have usually the functions: ramp up, dwell time and ramp down. It is recommended not to ramp faster than 60 °C/h.

Large vacuum systems require more sophisticated bakeout controllers, which can handle 50 to 100 control circuits per baked sector of the vacuum system. A typical 300 °C bakeout cycle is given in Table 1.

Table 1
Bakeout cycle

| Time (h) | Temp (°C) | Action |
|---|---|---|
| 0 | Amb. | Start baking. Max. 60 °C/h. |
| 5 | 300 | Stay at full temperature for 24 h. |
| 28 | 300 | Degas sublimation pump filaments at 20A for 1 h.<br>Flash ion pumps 30s every 20 min (3–4 times). |
| 29 | 300 | Decrease temperature. Same time derivative.<br>Residual gas analysers should stay at 300 °C for 2 h. |
| 31 | 200 | Degas ionisation gauge filaments.<br>Degas sublimation pump filaments at 35A for 5 min. |
| 32 | 150 | Degas sublimation pump filaments at 40A for 2 min.<br>Degas RGA 30 min (at 200 °C).<br>Start ion pumps.<br>Close turbo pump valve. |
| 33 | 100 | Run sublimation pump filament at 48A for 1 min to pump down. |
| 35 | Amb. | Bakeout finished. |

## REFERENCES


[1] O. Gröbner, CERN Accelerator School, General Accelerator Physics, Proceedings (Eds. P. Bryant and S. Turner), Sept 1984, CERN 85-19, p 489.

[2] A. Pace and A. Poncet, Vacuum 41 (1990) 1910.

[3] V. Ziemann, SLAC/Pub/5962 (1992).

[4] VacSim™, http://www.softsim.com.

[5] H. Ishimaru, MRS Bulletin, XV (1990) 23.

[6] J. Fixdal, CERN Techn. note MT-MF/91-03, and
J. Fixdal, CERN and The Norwegian Inst. of Technology, Master Thesis, 1991.

[7] V.V. Avdeichikov, L. Bergholt, M. Guttormsen, J.E. Taylor, L. Westerberg, B. Jakobsson, W. Klamra and Yu.A. Murin, Nucl. Instr. and Meth. in Phys. Res. A349 (1994) 216.

[8] L. Evensen, V. Avdeichikov, L. Carlén, M. Guttormsen, B. Jakobsson, Y. Murin, J. Mårtensson, A. Oskarsson, A. Siwek, E.J. van Veldhuizen, L. Westerberg, T. Westgaard and H.J. Whitlow, IEEE transactions on Nucl. Sci., 44 (1997) 629.

[9] V. Avdeichikov, L. Carlén, M. Guttormsen, A. Fokin, B. Jakobsson, Yu. Murin, J. Mårtensson, A. Oskarsson, E.J. van Veldhuizen, L. Westerberg and H.J. Whitlow, Nuclear Physics A626 (1997) 439c.

[10] A. Kuznetsov, E.J. van Veldhuizen, L. Westerberg et al. (To be published).

[11] A. Roth, Vacuum Sealing Technique (AIP and Springer) 1994.

[12] A. Roth, Vacuum Technology, (North Holland) 1990.

[13] M. Ainoux, P. Fontaine and H. Wahl, CERN SPS/AMR/HW/EEK Tech. note/81-8.

[14] L. Westerberg, B. Hjörvarsson, E Wallén and A. Mathewson, Vacuum 48 (1997) 771.

[15] B. Versolatto et al. (To be published).

# THE BEST LAID SCHEMES ...

*Lars Westerberg*
The Svedberg Laboratory, Uppsala University, Uppsala, Sweden

**Abstract**
A list of recommendations for better vacuum practice is presented in this chapter.

## 1. INTRODUCTION

"The best laid schemes o'mice and men gang oft a-gley" (Rabbie Burns)

The title of this chapter refers to a poem by the Scottish poet Rabbie Burns, who when farming destroyed a mice nest with his plough. He felt so sorry for this that he wrote the poem. The conclusion we can draw is that even with the best preparations something can, and will often, go wrong because some not so obvious fact was forgotten. Much of the advice in this chapter is certainly available elsewhere in the literature, but some may be compiled here for the first time.

The topics to be covered in this chapter are recommendations concerning the choice of materials with correct properties, corrosion resistance, joining dissimilar material, rolled versus forged material, weld preparation, virtual leaks, inadequate heat treatment, magnetic welds, common errors and design mistakes. The material is presented in the form of recommendations of what to do or not to do. Clearly, deviations from these principles have caused, and will cause, trouble.

## 2. RECOMMENDATIONS

### 2.1 Design

- Make sure to match pumping speed and conductance when designing a vacuum system.
- Computer simulations are of great help in optimising the performance of a vacuum system.
- Mechanical stability calculations help to avoid such unwanted surprises as collapsed chambers after the first pumpdown.
- A chamber which can not be made stable enough against deformation during pump down can be reinforced by external bridge-like constructions.

### 2.2 Choice of materials with correct properties

- Select materials with suitable mechanical properties and good weldability.
- Use 316LN steel where low magnetic permeability and vacuum firing is required.
- Use pyrolytic carbon in UHV (density 2.2). Sintered carbon (density around 1.7) contains a lot of air and must be avoided.
- Materials with too high vapour pressure, Cd (in plated steel screws), Zn (in brass), S and Se (in 303 stainless steel) etc. should never be used in vacuum systems!
- Cast material should be used for flanges since rolled flange material can give rise to virtual leaks. See further below.
- Avoid bare insulating surfaces in an accelerator. Charged structures may deflect the beam. Sparks can cause trouble.
- Charged Vespel may carbonise on the surface when a spark occurs.
- It is possible to coat ceramic tubes by a thin layer of sputtered Au or sublimated Ti to avoid charging of the surface.

### 2.3 Corrosion resistance

- 304 series steel is not acid proof.
- Certain combinations of metals and chemicals such as fluorine solutions may give rise to corrosion.
- Corrosion of high-voltage feedthroughs (e.g. for ion pumps) in high humidity areas can be avoided by heating the feedthroughs slightly above air temperature.

### 2.4 Joining dissimilar materials

- Aluminium and stainless steel can be joined by aluminium knife-edge seals, like the LEP seal at CERN. The SMC company has a full range of aluminium Conflat-type seals and aluminium flanges, which is easily combined with standard stainless steel flanges. It is also possible to use spring-loaded seals like Helicoflex.
- Titanium/stainless steel joints can be obtained by explosion welding, performed by special companies. We have, however, bad experience for such welds since three out of four leaked. They were replaced by brazed joints Ti/Cu/stainless steel.

### 2.5 Machining

- When machining do not use silicon oils. They can not be cleaned away.
- Glass-bead blasting reduces the surface but may bury oil and dirt.
- Glass-bead blasting as well as electropolishing should be followed by vacuum firing.

### 2.6 Brazing and welding

- Virtual leaks can be created by improper weld preparations.
- Improper welding may create magnetic welds.

### 2.7 Virtual leaks

- One of the most common sources of virtual leaks is threaded holes for screws which have not been ventilated properly. Drill a 1 mm hole through the centre of the screw or, drill a hole from the bottom, or mill a narrow gap all the way down to the bottom on one side of the hole for the screw as shown in Fig. 1.
- Closed volumes can occur if welds are improperly done. See Section 6 of the Engineering chapter in these proceedings.
- Rolled instead of forged flange material can lead to virtual leaks due to silicate enclosures as shown in Fig. 2.

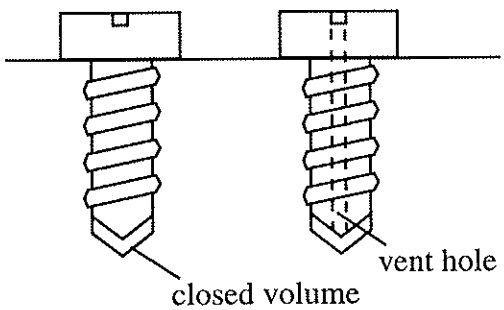



Fig 1 The screw to the right has a bored through ventilation hole to prevent virtual leaks and slow pump-down

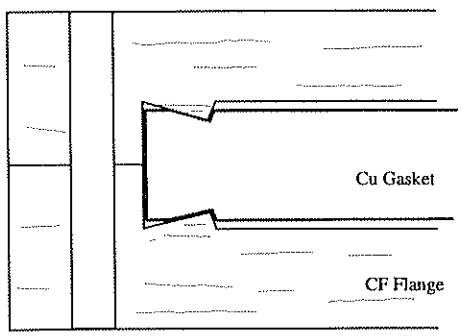



Fig. 2 Possible virtual leaks at the knife edge of a Conflat flange

### 2.8 Flanges and gaskets

- Improper torque applied to the bolts of a Conflat flange may cause a leak. Use a torque wrench and apply the recommended torque stepwise after completing a full turn of all the bolts.
- Do not re-use Conflat-type gaskets.

### 2.9 Vacuum furnace heat treatment

- It is important to have a sufficiently low pressure in the furnace, at least $10^{-3}$ Pa.
- Magnetic welds can occur if, during the heat treatment (vacuum firing), the temperature does not rise fast enough in the temperature range 500 – 800 °C.
- If a leak occurs during baking to 450 °C the chamber can be coated internally by copper oxides if OFHC copper seals are use. Use instead silver plated OFHC copper or Oxygen-Free Silver (OFS) seals

### 2.10 Baking

- Inadequate bakeout temperature will not give significantly improved vacuum.
- Non-baked areas defeat the purpose of the bakeout due to the large difference in outgassing rate between baked and unbaked stainless steel.
- Crossed heating tapes give a too-high temperature spot and will eventually cause a short circuit.
- Magnetic baking equipment will move in a magnetic field when the field is cycled and will eventually be damaged.
- Poor contact between the thermocouple and the chamber will result in excessive bakeout temperature, which can cause severe damage.
- Inadequate heat insulation gives too low temperature.
- Not degassing filaments of sublimation pumps results in excessive outgassing when the pump is operated and therefore very fast saturation of the sublimated layer.

### 2.11 Vacuum pumps

- Diaphragm or membrane pumps an cause damage when the membrane beaks down. Ask the manufacturer for the recommended interval between membrane replacements.
- Ion pumps exposed to a high noble-gas load release previously pumped gases.

### 2.12 Contamination

- Cleanliness is a must! No fingerprints. Use clean gloves! Use head cover.
- Mount equipment in a clean tent where an overpressure is established by using a sufficiently large air cleaner with HEPA filter. A movable clean bench can easily be combined with a tent. The tent is made from metal bars and clean plastic sheets. The over pressure prevents dust from entering.
- When transporting or storing vacuum chambers protect sensitive parts such as bellows from damage.
- Protect sealing surfaces by covering flanges with Al foil and protective covers after cleaning. Do not dismount until installation.
- Beware of dust particles!
- Beware of metal particles! They can stick in the seal of an all-metal gate valve resulting in expensive repairs.
- Ventilate vacuum chambers with clean nitrogen gas. Air contains water.
- During leak detection connect the leak detector between the turbo pump and the fore pump for lowest contamination risk and highest leak sensitivity. This can be done without stopping the turbo pump if two manual valves are installed in this position.

## 3. REFERENCES FOR FURTHER INFORMATION

The International Union of Vacuum Science Technique and Applications (IUVSTA) is an international organisation with 30 member national vacuum societies all over the world. It has eight divisions: Applied Surface Science, Electronic Materials & Processing, Nanometer Structures, Plasma Science and Technology, Surface Science, Thin Films, Vacuum Metallurgy and Vacuum Science. On IUVSTA's home page, http://www.vacumm.org/iuvsta you find information about its vacuum congress (IVC), conference series like ECOSS, ICTF and EVC as well as workshops and schools. On the Vacuum Science division page (http://www.vacumm.org/iuvsta/vsd/biblio.htm) you find a list of vacuum books, both historical and modern ones. There are links to national vacuum societies where you can find information about meetings, and courses they organise. Most vacuum societies publish a journal or a news letter. By joining your national vacuum society you can find very useful contacts.

*Good Luck!*

# DESIGNING ACCELERATOR VACUUM SYSTEMS

*P. Marin*
LAL, LURE, Orsay, France

**Abstract**
The reasons for achieving ultra-high vacuum are reviewed first, with a particular emphasis on lepton storage rings. The relationship between particle beams and their environment is also presented in terms of specifications. General considerations on the design of modern accelerators and their consequences are finally approached. Here and there, particular examples are given to illustrate the subject.

## 1. INTRODUCTION

Looking one or two decades back, on accelerator design and technology, one finds that vacuum is not a dominant culture in comparison with other fields such as beam optics and magnet design, or the technique of acceleration for instance.

It comes into play as a necessity that is addressed once a new machine is at a well advanced design stage. Vacuum has to adapt itself to an environment already defined. A few exceptions to such a situation could be found with the early electron-positron colliders, or the Intersecting Storage Rings at CERN which either for obvious reasons or for unforeseen ones, had to deal with ultrahigh vacuum. Progressively also came out the idea that vacuum could or should be improved on a long term operation.

To this sad picture one must add the difficulty of conducting clean and accurate experiments in the field of vacuum. Also, once a machine began to operate, beam time was usually allocated to "noble experiments" rather than to "obscure" and " unrewarding" tests.

Things began to change in the middle of the 80's and now the situation has drastically reversed in the last period with the outcome of the new $e^+e^-$ high energy colliders, the 3rd generation of synchrotron radiation sources and the future construction of LHC. The purpose of this talk is to illustrate this evolution and to describe the new trends.

## 2. THE PHYSICAL AND TECHNICAL BASIS FOR THE SPECIFICATIONS OF ACCELERATOR VACUUM SYSTEMS

Before designing and building an accelerator vacuum system, one has to recognise the various reasons for achieving ultra high vacuum and the relevance of the different technologies that will be needed. In what follows, I have listed the main problems encountered in the field of $e^+e^-$ colliders or synchrotron radiation sources. Problems somewhat similar are encountered with machines with stored proton beams. I will make a short description of each of these, thus preparing other speakers for a deeper analysis of particular cases. Most problems can be listed in five different classes.

### 2.1 The interaction of stored particles with neutrals

Whatever they are, atoms, molecules, shell electrons, eventually photons, the physical effects we are dealing with are:

- *A limitation of the lifetime of the stored particle beams* due to nuclear or electrons scattering as well as Bremsstrahlung on nuclei or shell electrons. Let us quote a few facts.

The cross section for Rutherford scattering (on nuclei) is proportional to $Z^2$, therefore argon and carbon monoxide are much more dangerous than hydrogen. Furthermore the betatron amplitude of the oscillation that results from a deflection $\theta$ from the central orbit is:

$$A = \frac{\theta}{2}\sqrt{\beta_1}\beta_2 \frac{\cos(\varphi - \pi\upsilon)}{\sin \pi\upsilon}$$

where $\beta_1$ and $\beta_2$ are the values of the betatron function respectively at the location of the scattering nuclei and at the location of observation. $\varphi$ is the so-called phase advance between these two points and $\nu$ is the betatron number. If an obstacle is set at a distance $d$ from the beam it will kill all the particles with an amplitude $A$ larger than $d$. The beam gas lifetime contribution from nuclei scattering will require an integration over the machine circumference of the Rutherford scattering cross section $\sigma(\theta)$ above the limiting $\theta$ due to the obstacle, weighted by the local pressure. The dependence of the beam gas lifetime versus the distance $d$ is of the type:

$$\tau \approx \frac{1}{d^2}$$

Usually the limitation on beam gas lifetime occurs in the vertical plane, the vacuum chamber wall being closer to the beam.

Bremstrahlung is the source of an energy loss by the particle from the beam due to a $\gamma$-ray emission. The particle loss which occurs in the horizontal plane arises from the limited energy acceptance of the machine that defines an energy window where the particle energy deviation can be accepted or even compensated.

Table 1 presents the cross section of the different processes involved when a relativistic electron interacts with residual gases.

Table 1
Interaction of ultra relativistic electrons with residual gases

- Rutherford scattering on nuclei (elastic collision)

$$d\sigma = \frac{4\, r_0^2\, Z_i^2}{\gamma_2\, \theta^4}\, d\Omega \Rightarrow \sigma_{\text{scatt.}} = \frac{2\,\pi\, r_0^2\, Z_i^2}{\gamma_2} \frac{\beta_1\, \beta_2}{d^2}$$

- Scattering on electrons (energy exchange)

$$\sigma_{\text{scatt}} = 2\pi\, r_0^2 \frac{m_0\, c^2}{\varepsilon_{\text{RF}}} Z_i$$

- Bremsstrahlung on nuclei

$$\sigma_{\text{Br.-n}} = \frac{4\, r_0^2\, Z_i^2}{137} \times \frac{4}{3} \text{Log} \frac{183}{Z_i^{1/3}} \times \text{Log}\left(\frac{E}{\varepsilon_{\text{RF}}} - \frac{5}{8}\right)$$

- Bremsstrahlung on electrons

$$\sigma_{\text{Br.-e}} = \frac{4\, r_0^2}{137} Z_i \frac{4}{3}\left[Log \frac{2E^2}{0.8\, m_0\, C^2\, \varepsilon_{\text{RF}}} - 1.4\right] Log\left(\frac{E}{\varepsilon_{\text{RF}}} - \frac{5}{8}\right)$$

- A *dispersion over the accelerator circumference of the scattered and energy degraded particles together with their secondaries* which leave the vacuum system. They are the most important contribution to the background of the physics detectors and of the sensors. Health physics is also mostly concerned by these particles.

- *The source of a beam emittance blow-up* at low energies. The last phenomenon is however rarely observed. It is particularly important at the CERN Antiproton Decelerator where the particles are injected at 3.5 GeV/C and subsequently decelerated to 100 MeV/C.

## 2.2 The interaction of the stored particles with charged particles and possibly with charged micro-objects

Here we have to consider separately negatively and positively charged beams.

- *Electron beam and positive ions*

Following the ionisation of a neutral atom or molecule, the low-energy electrons will be repelled by the beam. The positive remnant charges will be attracted by the beam. This is the so-called ion trapping [1]. The increased nuclei density lowers the beam gas lifetime and the macroscopic electric field of the ions produces extra focussing forces on the stored beam. Various adverse effects can be observed: an extra tune spread of the stored particles, an increased effect of the machine non-linear resonances and a blow up of the transverse beam cross section. Ion trapping widely observed on small electron rings was considerably weakened on large electron synchrotron radiation sources. This result is obtained from the combined effect of a beam over focussing (very flat beam cross section) together with partial circumference filling [2].

Strange effects on beam lifetime are observed when small dust particles ionised by stray photons are attracted by the electron beam. They keep traversing the beam and oscillating around it. A large reduction of the beam lifetime is seen if the micro-objects resist the high temperature increase from internal ionisation by the high-energy particles. On other occasions, a fast partial destruction of the beam occurs, the micro-object being simultaneously melted away. Such phenomena were observed both at DCI, Orsay [3] and at HERA, Hamburg.

- *Positive bunched beam interaction with ions or electrons*

While much less liable to trouble, the interaction of proton or positron beams with "vacuum" can display a number of typical effects.

The proton-positive ion avalanche was observed at the ISR [4]. Following a proton collision on residual gases, positive ions are accelerated onto the vacuum chamber wall. Under ion bombardment, neutral molecules are released thus contributing to a cumulative avalanche process with infinite asymptotic behaviour under some circumstances.

The proton (positron)-positive ion instability: a leading bunch with a vertical global betatron motion will leave its print on the positive ions it creates. These will communicate to the next coming bunch a force at the betatron frequency. The successive bunches can undergo a transverse instability above a certain pressure threshold. Seen on the CERN Proton Synchrotron in the 60's, this process was revisited by Russian physicists [5]. It was never observed however until now. Note however the problem of the fast ion instability [6].

The positron-electron cloud instability. This has been seen at KEK (photon factory) [7], at Beijing ($e^+e^-$ collider) and more recently at Argonne (APS) and at SLAC (PEP II). It is also a subject of worry for LHC. A low energy initial electron is attracted by the beam. If it crosses the orbit before the arrival of the next bunch with a sufficient energy, it will hit the chamber wall on the other side, possibly creating more than one secondary electron. In these conditions an avalanche can occur. This leads to a very strong pressure rise above a threshold current through multipactoring and possibly to beam instability.

## 2.3 The synchrotron radiation case

When bent by the dipole field of a magnet, relativistic electrons emit a characteristic radiation, the synchrotron radiation. The photon spectrum spans from infra-red to X-rays. Besides, synchrotron radiation is self collimated, very powerful, polarised and is time dependant due to the bunching of the particles. All these properties justify the tremendous interest of many researchers in various fields of science. The power emitted reaches Megawatt levels at LEP and still hundreds of kW in smaller

machines (6-8 GeV) at ESRF, APS or SPRING 8. Such an enormous beam power must be absorbed on specially cooled copper absorbers. Large fluxes of secondaries are produced, scattered or fluorescent photons and photo-electrons. These in turn produce a strong outgassing despite UHV conditions established prior to storing a beam of particles. This is the so-called Photon Stimulated Desorption. Long term outgassing will lead to surface conditioning through a progressive depletion of the protective oxide layer from its molecule content. The cleaned surface develops a very active pumping by the walls, larger in size than the installed pumping, thus modifying the initial pressure profile along the machine [8].

Two extreme cases of synchrotron radiation exist or are foreseen : the third generation of high brilliance synchrotron radiation sources and LHC. Concerning the first one, the most important problems will be addressed in § **3.3.** In the case of LHC, the protons, despite their heavy mass, have such a high energy that they emit soft energy photons. These are the sources of several problems which will be dealt with by other speakers during this course.

### 2.4 The interaction of bunched beams with the vacuum chamber wall

The image current of a bunch of particles in the wall of the vacuum chamber vessels travels together with the bunch [10]. It is the source of problems in connection with the wall conductivity and with discontinuities of the chamber cross section. The RF fields developed by the bunch moving along the machine circumference are particularly active in bellows, RF cavities, beam kickers and protrubing absorbers. These objects couple to the bunch electromagnetic field. The excitation by the beam of local parts of the vacuum chamber that can be considered as high frequency resonators is governed by the beam frequency content and the quality factor of the circuits. Depending on the damping of the RF field induced by a bunch until the arrival of the next one, different situations will be experienced. In large machines operated with different bunch patterns, wide band and narrow band excitation will therefore occur.

Besides power dissipation in the vacuum chamber walls, one observes an increase of the bunch length with current, and of the particle energy spread due to the so-called microwave instability [11]. Transverse and longitudinal excitations of the bunch are seen. The latter occurs as a bunch centre-of-gravity motion, but internal modes can also develop altering the natural Gaussian distribution of the particles. A large effort has been gradually devoted in the most recent machines to achieve a vacuum chamber cross section as smooth as possible in order to reduce the so-called chamber impedance [12]. There is a special concern for the accelerating radio-frequency cavities that from construction have many resonating modes beside the fundamental one. Great ingenuity was shown in the design of the RF cavity impedance at the frequency of the modes (HOM), with spectacular results [13]. The residual part of the phase oscillation of the bunch centre-of-gravity motion is dealt with, using damping feedback.

### 2.5 Superconductive accelerating cavities

These have been the subject of much R&D, especially during the last fifteen years. The goal here is to obtain the largest possible accelerating gradients (LEP, $e^+e^-$ linear colliders) through high quality surfaces based on elaborate construction techniques, careful surface coating and cleaning, and finally, conditioning with large pulsed RF fields [14].

Another problem encountered in RF cavities is multipactoring. Here we are mostly concerned with the multipactoring that occurs in a remote part of the cavity, the so-called energy coupler. DC voltage bias and low-secondary-emission coatings have been used successfully to overcome the trouble.

# 3. THE DESIGN OF VACUUM SYSTEMS

## 3.1 The standard technical problems

### *3.1.1 The chamber envelope*

Three different materials have been used with success in various machines for building the vessel bodies: stainless steel, aluminium and copper alloys. Concerning the connecting flanges they are mostly of the stainless steel-copper gasket-CF type, except at KEK (Japan) and more recently at Argonne and Stanford (USA) which make use of aluminium alloys.

Different arguments were used for supporting particular choices:

- Heat conductivity in relation with bake out and cooling for the evacuation of the power from the secondaries.
- Optimisation of magnet gaps, considering also the wall thickness to withstand atmospheric pressure.
- Ease of fabrication.
- Connection of vacuum vessels with reliable flange and gasket systems.
- Availability of equipment on the market: pumps, gauges, RG analysers with stainless steel CF flanges.

None of the three materials has definitely emerged as having outstanding properties in comparison with the others.

### *3.1.2 The special equipment*

This concerns the injection and beam kickers, the RF cavities and electrostatic beam separators, the current and beam position monitors, the feedback vessels.

Among these, the beam kickers with ceramic vacuum chambers deserve special attention. In order to reduce the beam electromagnetic field propagation outside the ceramic, a titanium coating, usually one micron thick, is made. A peculiar effect, investigated by A. Piwinski [15], thus occurs. Due to the lower propagation velocity of the E.M. field in the ceramic, a much higher longitudinal current density appears in comparison with the normal value in ordinary conductors. A correspondingly large joule power is developed which has to be evacuated.

### *3.1.3 The distribution of pumping in relation with the conductance of the various chamber vessels*

Montecarlo programs exist to calculate the pressure profile from static or dynamic vacuum. Experimental data on the latter are becoming available on material such as stainless steel or copper, which allow to make evaluation more realistic than in the past. Finally it should be possible to add to this a rough evaluation of the effect of the wall pumping speed at various stages of irradiation.

### *3.1.4 Pressure monitoring*

It should be emphasised that gauges and RG analysers are sensitive to several disturbing effects such as: photo-electron pick-up, photo electrons from stray photons, RF induced pick-up which requires special electrostatic shielding [16].

### *3.1.5 Three typical designs*

We have chosen to present here three typical designs of vacuum chamber vessels for electron-positron colliders or synchrotron radiation sources. They differ mostly by the disposition of the photon absorbers.

- LEP design [17], CERN, Fig. 1

It is an extruded aluminium vessel where the photons strike directly the outer side of the chamber wall. Channels with water cooling allow the power to be removed from the primary photons and from the secondaries. When run at 130 °C they can be used for vessel bake out. Note also the extra lead

shielding outside the vessel. Its sole purpose is to absorb the X-rays that escape the aluminium vessel in order to reduce the radiation level in the environment. A recess in the inner side of the chamber contains a NEG ribbon pump distributed all along the vessel.

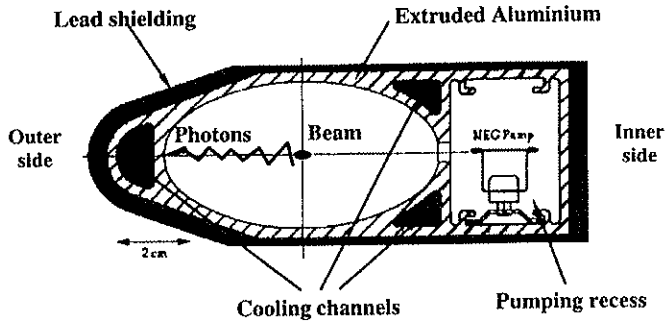



Fig. 1 LEP vacuum chamber in the arcs (cross section)

- ALS design [18], Berkeley, Fig. 2.

This is again an all aluminium vessel. In this design, full use is made, all over the circumference, of the chamber/antechamber concept. The electron beam circulates in the small particle chamber and the photons are allowed to stretch in a huge antechamber after crossing a small vertical recess. The latter decouples the particle chamber from the antechamber, from the impedance point of view. Several so-called copper crotch absorbers are put in different longitudinal locations.

- ESRF design [19], Grenoble, Fig. 3.

Here the vacuum chamber walls are made of stainless steel. Use is made of the chamber/antechamber and crotch absorber concept in the dipole vessels. In the other straight sections, longitudinal copper absorbers are brazed on the thin wall, together with outside water cooled channels. Note also the presence of two other channels with water cooling on the outside of the chamber to remove the power dissipated by the secondaries.

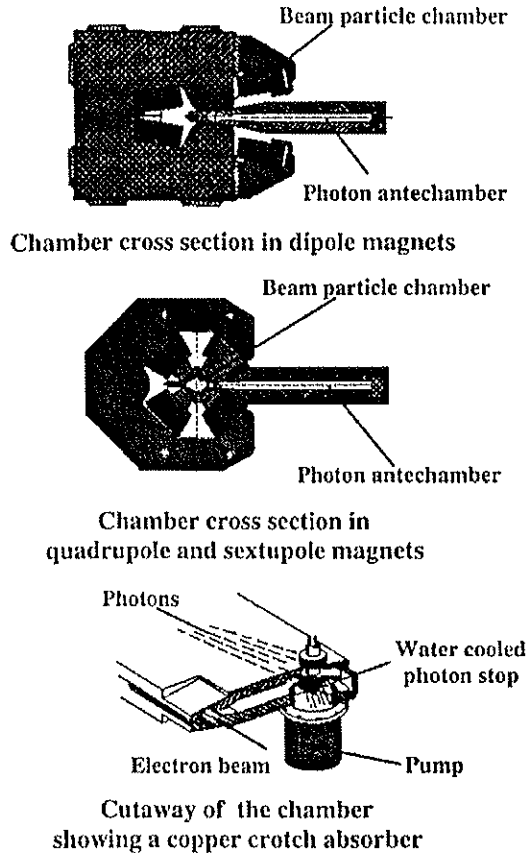



Fig. 2 ALS design. Chamber/antechamber and copper crotch absorbers

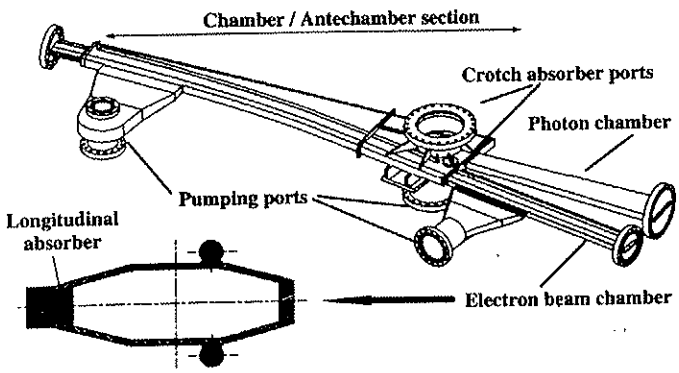



Fig. 3 ESRF mixed design

These designs have different advantages or drawbacks which would need a long discussion on comparative aspects. At the moment, all of them have given satisfaction to the users, see however § **3.3** and **3.4.**

## 3.2 Surface problems

### *3.2.1 Surface cleaning*

Various surface-cleaning techniques have been developed which are adapted to the particular metal alloy used. They have become standard and are described in the 'cleaning' chapter of these proceedings [20]. Note a problem with the use of chlorides for welding stainless steel cooling pipes outside vacuum vessels. Insufficient cleaning will produce long term corrosions.

### *3.2.2 Surface-oxide layer and bulk*

Whatever the initial preparation, there exists an oxide layer of a few tens of Å, which separates the metal bulk from vacuum. It seems that Photon Stimulated Desorption (PSD), the most powerful tool with ion bombardment, removes a number of molecules compatible with the molecule content of the outside layer. Except for hydrogen, there is no indication that one has to call for atom diffusion from the bulk to explain the photon stimulated outgassing at large photon beam doses.

### *3.2.3 Primaries and secondaries*

There is a naive idea that desorption induced by photons comes from the primary beam. However very soon after running an $e^-$ or an $e^+$ beam, degassing proceeds through the secondaries. These have several origins: the scattered primary photons which for X-rays occur mainly at glancing angle, the fluorescent photons which follow the de-excitation from the absorption of a photon by an atom, and finally the photo-electrons (and Auger electrons). The diagram of emission is very different from one type of secondaries to another. The electrons may also be deeply affected by stray magnetic fields. For a number of reasons the secondaries appear to be present practically everywhere in the vacuum vessels, see next paragraph.

### *3.2.4 Photon-stimulated desorption*

After a normal bake out, typically 24-28 hours at 150 °C, a vacuum system (surface cleaned beforehand by suitable method) reaches ultra high vacuum, well bellow $10^{-9}$ mbar, depending on the surface area and the pumping speed of the system. However, the first stored beam of relativistic electrons or positrons develops a very large pressure increase, the so-called dynamical vacuum, which is observed whatever the alloy used for the construction of the vessels.

The behaviour of the PSD has been found to be very similar for all machines. In particular, it shows a gradual conditioning with the increase of the longitudinal photon beam dose, and practically all parts of the vessels are cleaned-up. One can understand this behaviour by the fact that the parts with high photon fluxes are well conditioned, the other parts with smaller fluxes being not as well degassed. The product of the local flux and the molecular yield (molecules/photon) is therefore more or less constant. It has been observed that the static pressure of a vacuum system, including water vapour, shows a large conditioning, thanks also to stray photons.

### *3.2.5 Multipactoring and cures*

We have already dealt briefly with this subject in § **2.5.** It may occur as a two-point or a single-point effect. In the first case, two surfaces participate to the effect whereas only one participates in the second case. Electrons from secondary emission have several origins : the metal itself, the oxide layer and the molecule coverage on or within the surface layer. It is believed that multipactoring conditioning occurs from the molecule coverage which is progressively depleted with the electron bombardment. After some time, only the first two contributions dominate.

*3.2.6 Surface coatings*

Many attempts have been reported for different purposes and with more or less success. TiN coating of aluminium RF cavities has eliminated an otherwise indestructible multipactoring. Gold coating of vacuum vessels was reported to be no better than uncoated ones from the PSD point of view. New types of coating have been or are being investigated at CERN for LHC [21] and for achieving large distributed pumping speeds [22]. The problem of the long-term behaviour of such techniques is raised in view of the large-scale production of vessels.

### 3.3 New trends in accelerators design

It is not a surprise that all trends go towards increasing difficulties. Let me just briefly review the present trends before looking at their consequences in the next paragraph **3.4.**

a) For obvious reduction of construction and running cost (power bill) the goal of smaller and smaller magnet gaps, quadrupole and sextupole bores is important. In the particular case of undulator insertion devices in S.R. sources, the aim is to reach higher performances: harder monochromatic X-rays. Beam optics must follow in order to preserve beam gas lifetime and prohibitive closed orbit distorsions. However, as a result, the vacuum chamber cross section is drastically reduced.

b) Larger and larger beam currents, total or bunch current are achieved or contemplated.

c) Higher radiated power and power density are foreseen in 3rd generation synchrotron radiation sources, also from increased beam energy.

d) Higher power losses from HOM (Higher Other Modes) result also from more intense and shorter bunches.

e) Smoother vessel cross sections are required in places that were previously neglected.

f) The design of the so-called RF lined bellows, particularly sensitive to large bunch currents, has been pushed towards unprecedented but more expensive designs.

g) As a result of a), b),and c) higher density power from secondaries are expected with the adverse effects of temperature gradients which have to be limited with special cooling. This could produce vacuum vessel deformation with several consequences.

h) Finally, but for independent reasons, much tighter specifications, at the micron level, are now set for the photon and therefore the electron/positron beam stability. This difficulty cumulates with some of the above ones.

An interesting question: how far can this race go on? Before stating some physical limits to this, it is interesting to look at a certain number of consequences which we have presently to face.

### 3.4 The consequences of the new trends

a) Much tighter tolerances for fabrication are now required as a general result from most of the above new trends. We are now at the level of a small fraction of one mm, sometimes 100 microns for LHC for instance

b) The recourse to laser cutting, electron-beam welding and even to machining vacuum vessels from solid blocks are favoured in comparison with older types of fabrication.

c) Bake out is given up in favour of intense beam scrubbing. Dedicated beam-line experiments have clearly demonstrated this possibility. This has also been demonstrated recently with the operation of the new PEP II machines. The absence of bake out jackets, no matter how thin they are, leaves more free space between vacuum vessels and magnets. Air cooling of heat from secondaries is better achieved.

d) New problems for supports arise with the vacuum vessels becoming "spaghetti-like objects" with relatively heavy pumps attached to thin bodies with small moment of inertia.

e) New strategies for beam orbit and correction control are now developed in two stages. The mechanical design is of the so-called "girder type" with a number of critical items being precisely assembled on solid girders. These are aligned accurately so as to secure a closed orbit with all correctors being set at zero. This orbit must be safe from the point of view of the synchrotron radiation impinging only on copper-cooled absorbers. Final orbit corrections with very limited amplitude are achieved. They allow safe operation with large stored beams. When necessary realignment of girders is made with the help of microjacks.

f) Fast beam abort triggered from vacuum sensors distributed in all potentially dangerous sections is required.

g) Poor beam lifetime can be eased also by a regular topping-up of the beam current. The feasibility of this proposal has still to be demonstrated from the health point of view, as well as for preserving the quality of the experiments carried out by the users.

h) Finally the adverse consequences of small insertion-device cross sections can be overcome with adjustable vacuum chamber walls restoring some flexibility for the machine optics. This possibility has not been widely exploited yet, being subject to a number of problems raised in the above paragraphs.

i) Accurate codes for computing secondaries and their effects are becoming more and more necessary.

## 4. CONCLUSIONS

From vessel construction to surface problems and particle beam stability, vacuum technology for particle accelerators covers a very large number of aspects. The recent period has witnessed an evolution towards increasing complexity which brings vacuum to the forefront of accelerator technology. This situation has led to a number of experiments towards a better knowledge of the field so leading to more elaborate designs.

## REFERENCES

[1] Y. Baconnier et al., Neutralisation of accelerator beams by ionisation of the residual gas, CAS, Jyväskylä, Sept., 1992, CERN 94-01, pp. 525-564.

[2] J.-M. Filhol, Status of the ESRF, EPAC 94 Proceedings, pp. 8-14.

[3] P. Marin, Observation of Bremsstrahlung on dust particles trapped on electron beams at DCI and Super-ACO, RT/91-03, Anneaux, LURE, Orsay, France.

[4] R. Calder et al., Vacuum conditions for proton storage rings, IXth International Conference on High Energy Accelerators, Stanford 1974 pp. 70-74.

[5] D.V. Pestrikov, Collective issues related to B-factories, Nat. Lab. For High Energy Physics, Tsukuba, Feb. 17, 1993.

[6] F. Zimmerman et al., Experiments on the fast beam ion instability at the ALS, invited talk, Int. Workshop on multibunch instabilities, KEK, Tsukuba, Japan, July 15-18, 1992.

[7] K.C. Harkay, Theory and measurement of the electron cloud effect, PAC 1999 Proceedings, New York, to be published.

[8] C. Herbeaux et al., Photon stimulated desorption of an unbaked stainless steel chamber by 3.75 keV critical energy photons, J. Vac. Sci. Technol., A17 (2) March-April 1999, pp. 635-643.

[9] O. Gröbner, LHC Vacuum System, these proceedings.

[10] Physics of collective beam instabilities in high energy accelerators, A.W. Chao, John Wiley & Sons, inc. 1993, pp. 13-20.

[11] E. Shaposhnikova, Signature of microwave instability, CERN SL 99-08 HRF.

[12] See [10] A.W. Chao, pp. 67-120.

[13] H. Padamsee et al., Accelerating cavity development for the Cornell B-factory, CESR-B, IEEE, P.A.C. Proceedings 1991, pp. 786-788.

[14] M. Lung et al., High peak power processing up to 100 MV/M in various metallic samples, Proceedings of the seventh workshop on RF superconductivity, Oct. 17-20, 1995, Gif sur Yvette, France, pp. 519-522.

[15] A. Piwinski, Penetration of the field of a bunched beam through a ceramic vacuum chamber with metallic coating. IEEE, Trans. Nucl. Sc. Vol. NS 24 N°3, June 1977, pp. 1364-1367.

[16] P. Marin, Vacuum experience and experiments at Super-ACO, Conf. on vacuum design of synchrotron light sources, Argone National Laboratory, Argone Nov. 13-15, 1990.

[17] C. Benvenuti, See for instance, Ultimate pressure of the large $e^-e^+$ collider (LEP) vacuum system, 12th Int. Vac. Congress, The Hague, The Netherlands, Oct. 12-16, 1992 pp. 507-509.

[18] M. Cornacchia, The LBL 1-2 GeV Synchrotron Radiation Source, PAC 1997 Proceedings, March 16-19, Washington D.C., pp. 409-413.

[19] R. Kersevan, Status of the ESRF vacuum system, EPAC 1998 Proceedings, pp. 2178-2180.

[20] R Reid, Cleaning, these proceedings.

[21] O. Gröbner, See LHC vacuum system, these proceedings.

[22] C. Benvenuti, See Pump III (getters), these proceedings.

# HISTORY OF VACUUM DEVICES

*P.A. Redhead*
National Research Council
Ottawa, Canada

**Abstract**
An outline of the early history of vacuum devices (pumps and pressure gauges) is presented, from the first recorded experiments on vacuum in the 1640s to the start of the modern period of vacuum technology when ultrahigh vacuum was introduced in 1950.

## 1. INTRODUCTION

The first experiments to produce a vacuum were proposed in 1631 by Reneri of Leiden in correspondence with Descartes [1]. The first recorded experiment to produce a vacuum appears to have taken place in about 1641 when Gasparo Berti experimented with a water barometer [2], see Fig. 1. The lead tube AB was about 11 metres long, the tube and reservoir were filled with water and the valves G, D, and B were then closed. The valve B was then opened and the water level dropped to L, a vacuum was left in the chamber above L. The experiment was not very convincing since it was assumed that sound would not travel through a vacuum yet a bell M rung in the vacuum space could be clearly heard, the sound was probably transmitted through the supports of the bell. In 1644 Vincenzio Viviani repeated Berti's experiment using a mercury-filled glass tube which was inverted with its open end in a reservoir of mercury. The experiment was probably planned by Evangelista Toricelli who was credited by earlier historians with performing the famous experiment in 1643. The Toricellian experiments convinced most scholars at the time that a vacuum had been created. This was the time of the dual papacy and the concept of vacuum was still considered anathema, thus it was dangerous to continue experiments on vacuum in countries accepting the Roman pope so the experiments were pursued in the reformed countries and in France, where the French pope was inclined to ignore the opinions of the Roman pope.

A few years after these initial experiments on vacuum the first vacuum pumps were created by Otto von Guericke the Burgomaster of Magdeburg, this is when the history of vacuum devices begins. In this paper we examine the development of vacuum devices, in particular vacuum pumps and gauges, from 1650 to the start of modern vacuum technology in 1950 when ultrahigh vacuum was first achieved and measured.

Several histories of vacuum technology may be consulted for further details [3-11].

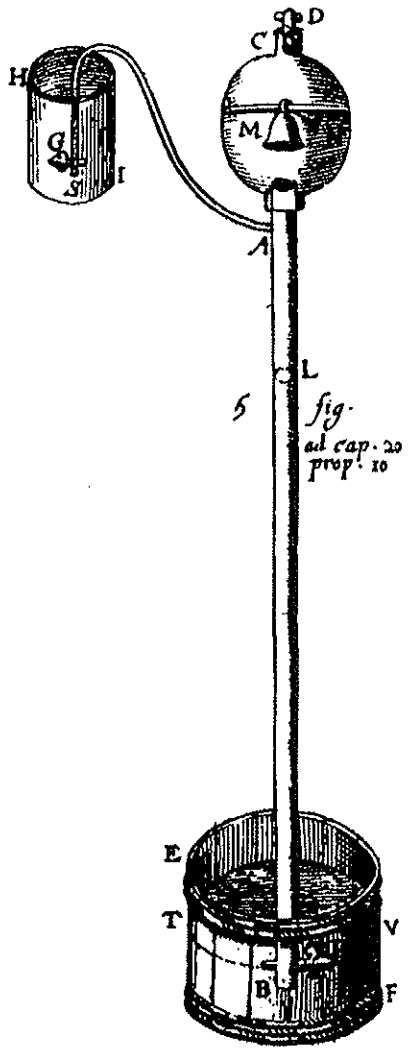



Fig. 1 The apparatus used by Gasparo Berti in about 1641 to produce a vacuum

## 2. VACUUM FOR FUN AND SCIENCE, THE FIRST TWO CENTURIES, 1650 TO 1850

The vacuum pump was one of the six instruments invented in the 17th century that had a profound impact on experimental science, the others were the pendulum clock, the telescope, the thermometer, the barometer and the microscope. Otto von Guericke started work on air-pumps in the 1640s and his work was first reported by Caspar Schott [12] in 1657. Von Guericke's first two pumps are shown in Fig. 2. These pumps were essentially water pumps being worked backwards. In his first experiment in the 1640s he filled a well-caulked wooden barrel with water and two strong men pumped the water out, as air rushed through the pores in the wood into the empty space a noise was heard. In the second experiment he used a large copper sphere attached to the pump; the water was omitted and the air pumped out directly. When most of the air was pumped out the sphere collapsed with a loud noise, von Guericke realised that this was caused by atmospheric pressure. A more accurately spherical vessel was made and the experiment was then successful.

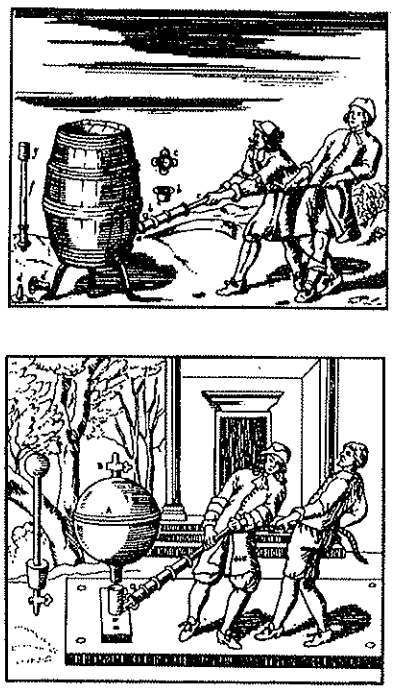


Fig. 2 The pumps of von Guericke in the 1640s. (Top) The first pump with a wooden barrel. (Bottom) The second pump with a copper sphere.

The news of von Guericke's experiments was spread through Europe by Schott's book and the next improved pump was designed by Robert Boyle [13] and built by Robert Hooke in England in 1658-9. This pump is shown in Fig. 3, it used a rack and pinion to move the piston thus greatly reducing the effort needed to operate the pump. Boyle was the first to make a measurement of vacuum with a mercury manometer placed in a bell-jar (see Fig. 4), his pump was capable of reaching a pressure of about 1/4 inch of Hg (6 Torr). Thus Boyle was the first to design a vacuum pump and combine it with a pressure gauge. The piston pump and the mercury manometer were to remain the only type of vacuum pump and pressure gauge for about two hundred years.

Fig. 3 Piston pump designed by Robert Boyle and built by Robert Hooke in 1658-9.

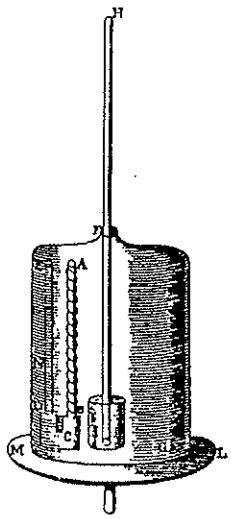


Fig. 4 The first measurement of a sub-atmospheric pressure by Robert Boyle c.1660. A beaker of mercury with a manometer tube more than 32 inches long was sealed in a bell jar and evacuated by the pump of Fig. 3.

Improvements were made to the piston pump in the next two hundred years but the basic design remained the same. The ultimate pressure achieved by these pumps decreased only slightly, the mechanical nature of the vacuum experiments of the time did not require significantly lower pressures, but the new pumps were easier to use. Hawksbee [14] built a pump in 1704 with two cylinders in which the pistons were balanced against each other as they were driven in opposite directions by the rack and pinion (see Fig. 5), this pump was capable of reaching 1.9 Torr in two minutes. At the Great Exhibition of 1851 in London the first prize for vacuum pumps was won by Newman [15] whose pump achieved a pressure of 0.5 Torr. Figure 6 shows a similar commercial pump of about 1850 that is little changed from the construction of the Hawksbee pump of 150 years before.

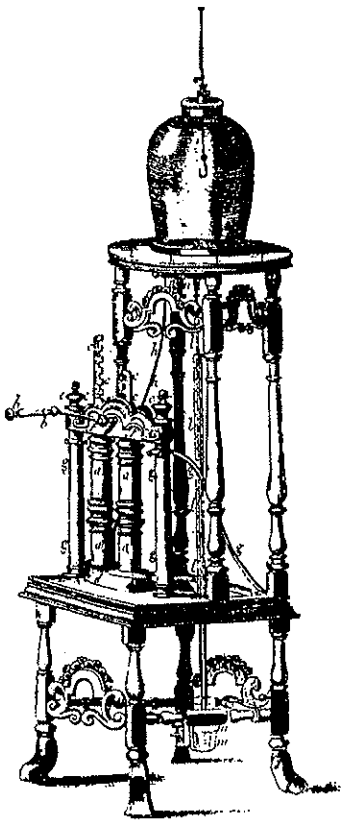


Fig. 5 The double-piston pump of Hawksbee (1704).

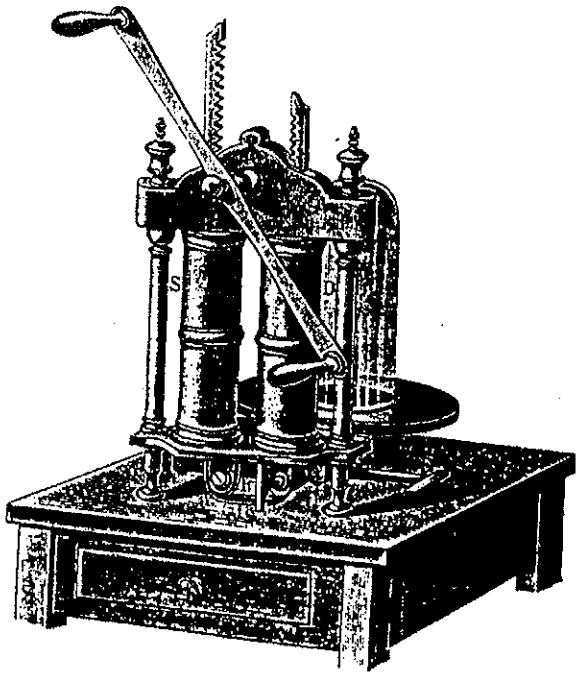


Fig. 6 A commercial double-piston pump from about 1850.

Figure 7 shows the ultimate pressure achieved by the various pump designs in the period from 1660 to 1900 and clearly indicates the limited reduction in attainable pressure in the first two hundred years. The vacuum pump made important contributions to science in its first two centuries but throughout this period it was widely used as a source of entertainment and instruction because vacuum was a novel and fascinating subject.

## 3. VACUUM FOR SCIENCE, THE GREAT LEAP FORWARD, 1850 TO 1900

The 50 years from 1850 to the turn of the century produced an intense activity in the development of vacuum technology driven initially by the needs of scientific research and later by the demands of the incandescent lamp industry. In the fifty year period the attainable pressure was reduced by six decades and vacuum equipment, though still awkward, was in use in industry. The need to thoroughly outgas the equipment and to avoid the use of rubber tubing was well established.

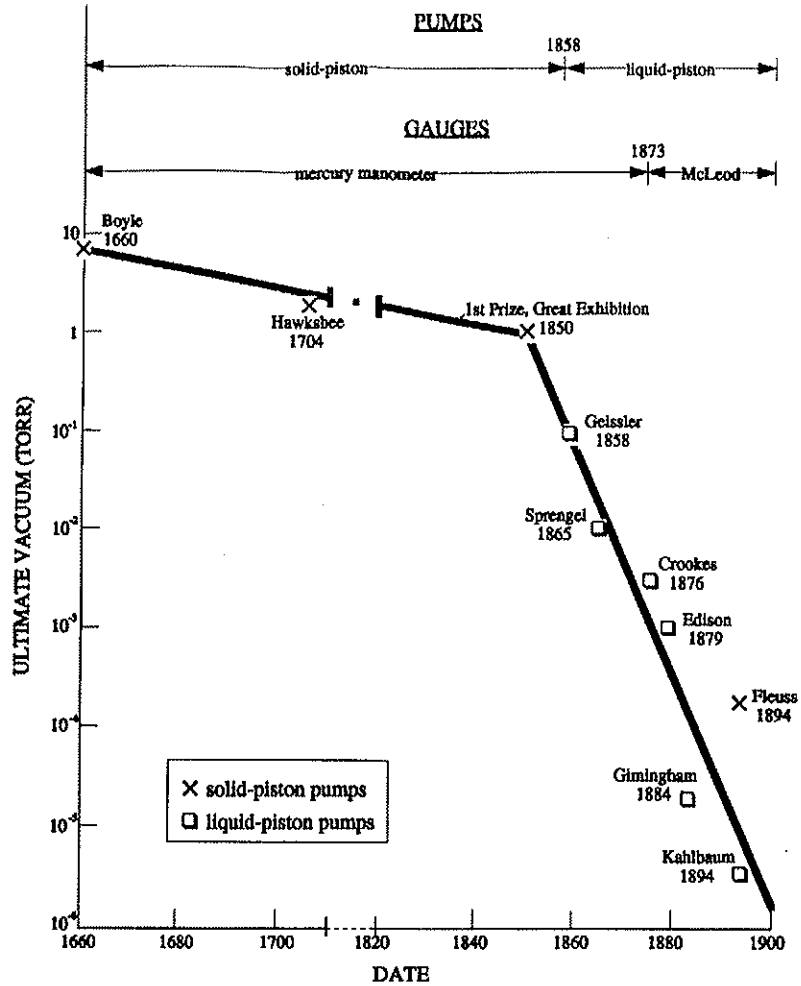



Fig. 7 The ultimate vacuum from 1660 to1900. Note the break in the time scale. The references to the experiments are; Boyle [13], Hawksbee [14], Geissler [17], Sprengel [18], Crookes [19], Edison [22], Fleuss [24], Gimingham [20], Kahlbaum [23].

The development of mercury-piston pumps was the major factor in achieving lower pressures. The first step occurred in 1854 when Julius Plücker of the University of Bonn asked Heinrich Geissler, his glass blower, to design a glass pump using a mercury piston to permit experiments on low pressure gas discharges; Geissler built a pump in 1855 capable of reaching 0.1 Torr which was first described in a pamphlet by Mayer [16] published in 1858. In 1862 Töpler [17] invented an improved form of the Geissler pump and in 1865 Sprengel [18] devised a pump in which a train of mercury droplets trapped packets of gas in a glass tube and carried the gas away.

In the 1870s William Crookes, with his assistant Charles Gimingham, attempted to achieve a vacuum "*approaching perfection*". Figure 8 shows Crookes' first pumping system [19] using a Sprengel pump. An improved version [20] of this pumping system using seven fall tubes was capable of achieving a pressure of about 2 x $10^{-5}$ Torr as measured by a McLeod gauge. Crookes greatly improved the vacuum conditions by replacing all rubber tube connections with ground glass joints and by heating the apparatus to degas it.

The vital step in pressure measurement was McLeod's invention of his vacuum gauge [21] in 1874 which was based on the compression of the gas by a mercury column to an easily measured higher pressure, and the use of Boyle's law to calculate the original pressure (see Fig. 9). This permitted pressure measurements down to the $10^{-6}$ Torr range. Since the McLeod gauge does not measure the pressure of condensable gases, such as water, the true pressure may have been higher than those reported in this period.

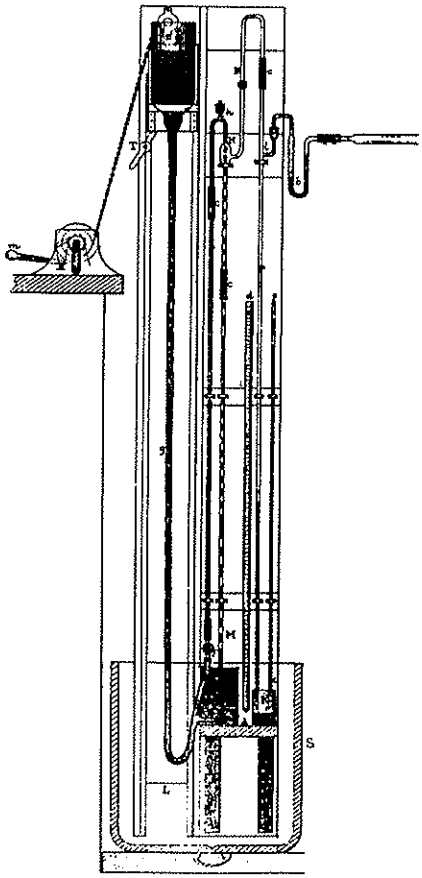


Fig. 8 Crookes' first vacuum system using a single Sprengel pump (1872).

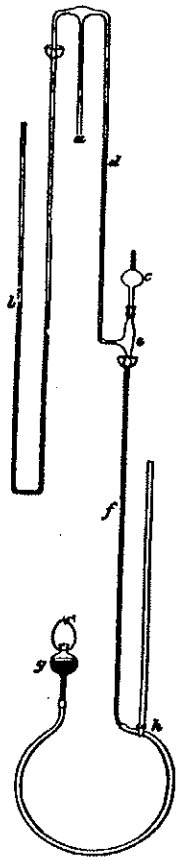


Fig. 9 The first McLeod gauge (1874).

By the end of the 1870s Crookes' vacuum techniques had moved into industry and were being used by Edison to improve the vacuum in incandescent lamps. Figure 10 shows Edison's vacuum system which contained two Sprengel pumps, a Geissler pump, and a McLeod gauge; pressures of about $10^{-3}$ Torr could be produced [22]. Edison's pumps were operated manually and were slow and clumsy for industrial use. By1894 Kahlbaum [23] was able to obtain pressures as low as 3 x $10^{-6}$ Torr with a Sprengel pump.

In the same period the solid-piston pump was greatly improved, by 1892 Fleuss had manufactured a pump with an oil-sealed piston and valves that were moved mechanically [24]. This pump was known as the Geryk pump and was capable of 2 x $10^{-4}$ Torr, it was widely used in the lamp industry to replace the manually operated mercury-piston pumps until the invention of the rotary mercury pump by Gaede in 1905. The Geryk pump and the rotary mercury pump could be motor-driven and thus had the advantage over the older liquid-piston pumps for industrial use. By 1900 pressures of the order of $10^{-6}$ Torr could be achieved and measured and the pumps for industrial vacuum systems were being motor driven.

## 4. VACUUM FOR INDUSTRY, 1900 TO 1950

The change in ultimate vacuum in the period 1900 to 1950 is shown in Fig. 11. Vacuum technology made rapid advances in the period 1900 to 1920, the two figures that dominated this period were Gaede in Germany and Langmuir in the USA. The first improvement in high vacuum pump design was the invention of the rotary mercury pump by Gaede [25] in 1905, this was a rotary mercury pump which could be motor driven (see Fig. 12). It could produce pressures in the $10^{-6}$ Torr range and was manufactured in large quantities and widely used in the lamp and vacuum-tube industries.

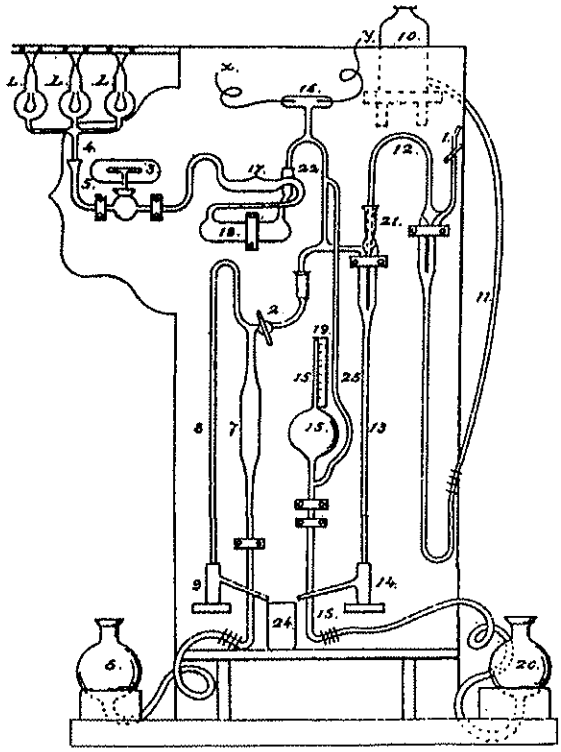


Fig. 10 Diagram of Edison's vacuum system for the production of incandescent lamps (1870s).

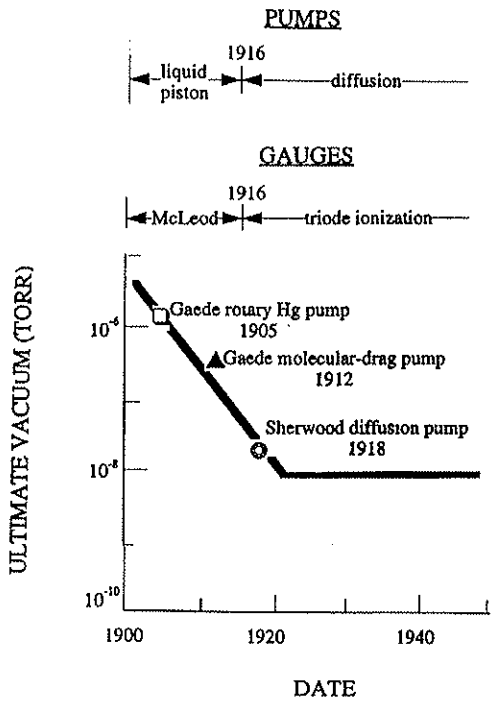



Fig. 11 The ultimate vacuum from 1900 to 1950. The references to the experiments are; Gaede (1905) [25], Gaede (1913) [27], Sherwood [34].

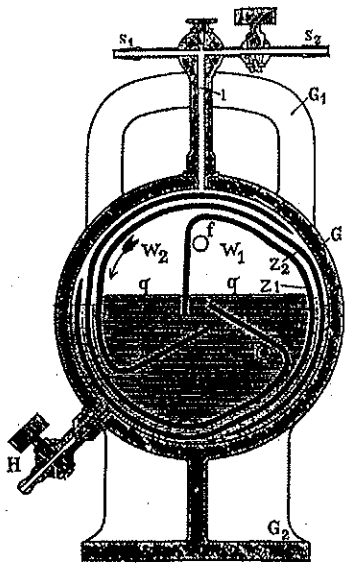


Fig. 12 Gaede's rotary mercury pump (1905).

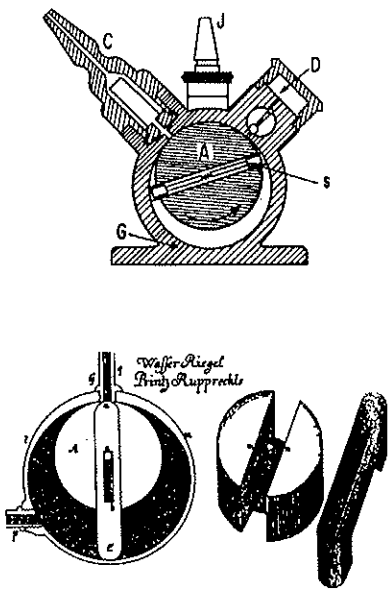


Fig. 13 Rotary oil pump designed by Gaede (1907) compared with Prince Rupert's "Water-Bolt", a water pump invented in about 1650.

In 1907 Gaede developed the rotary oil pump (see Fig. 13) primarily for use as a fore-pump for the rotary mercury pump [26], its close resemblance to the water pump invented by Prince Rupert in about 1650 is evident. The principle of this type of waterpump goes back to at least 1588 with the vane pump of Rumelli.

The next major advance in pumping methods was the invention of the molecular-drag pump by Gaede [27] in1913 (see Fig. 14). With a fore pressure of 2 x $10^{-2}$ Torr a pressure of 4 x $10^{-7}$ Torr was measured by Dushman [28] at a rotation speed of 8,000 rpm. This pump was used in the vacuum-tube industry until the arrival of the diffusion pump. Improved versions of the molecular-drag pump with higher pumping speeds were developed over the next 30 years, notably by Holweck [29] in 1923 and Seigbahn [30] in the 1920's, pressures in the $10^{-7}$ Torr range were typically obtained. Molecular-drag pumps were not much used until the 1970s because their pumping speed was low and their reliability poor as a result of the very close spacing between rotor and stator. They were ahead of their time because the technology for making high-speed rotating machinery with close spacings was still rudimentary. In recent times the molecular-drag pump has been successfully re-introduced, usually in combination with a turbo-pump.

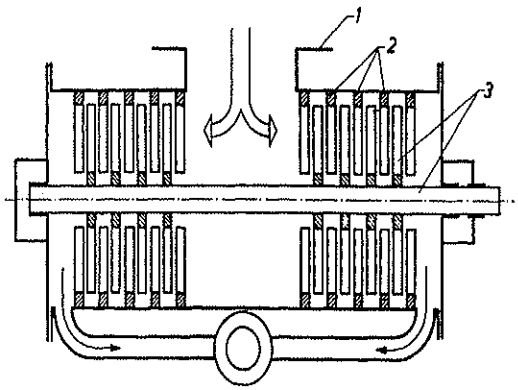



Fig. 14 Schematic diagram of Gaede's molecular-drag pump; 1 pumping port, 2 stators, 3 rotors (1913).

The mercury vapour diffusion pump was the first vacuum pump to have no moving parts. It was invented independently by Gaede [31] (see Fig. 15) and Langmuir [32] (see Fig. 16) in 1915-16. Langmuir's design with its large nozzle had much higher pumping speed and was quickly taken up for industrial use. The use of low-vapour-pressure oils rather than mercury in diffusion pumps was introduced by Burch in 1928 [33]. The diffusion pump became the most widely used high vacuum pump until the sputter-ion pump became available in 1958. Both mercury and oil diffusion pumps appeared to have an ultimate pressure of about $10^{-8}$ Torr as measured by a triode ionisation gauge, the curves of pumping speed of diffusion pumps as a function of pressure in manufacturers' literature showed the pumping speed going to zero at about $10^{-8}$ Torr (such curves were still being published as late as the 1960s). The early results using mercury diffusion pumps and ionisation gauges are typified by the work of Sherwood [34] in 1918 who measured an ultimate pressure of 2 x $10^{-8}$ Torr.

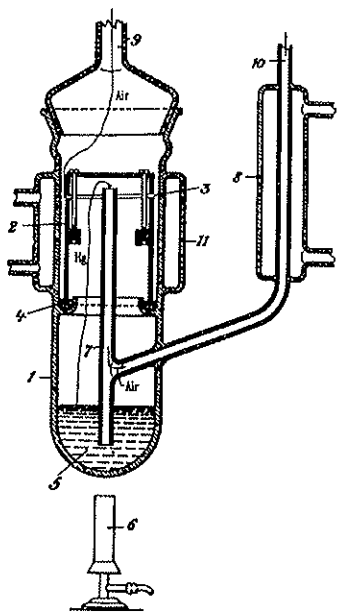



Fig. 15 Gaede's first mercury vapour diffusion pump (1915).

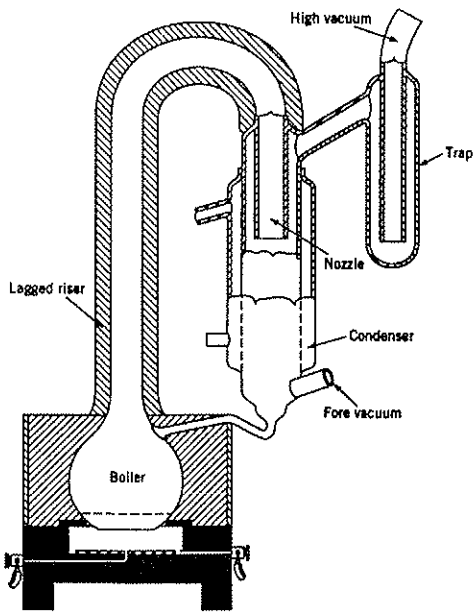



Fig. 16 Langmuir's original mercury vapour diffusion pump (1916).

Several types of pressure gauge were invented after the McLeod gauge in 1874. In 1897 Sutherland [35] invented a gauge depending on the viscosity of the gas, it could measure from 1 to $10^{-4}$ Torr. This principle was extended by Langmuir [36] in 1913 who developed a gauge using a quartz fibre which was made to oscillate in the gas, the decrement in amplitude of the oscillations was a measure of the pressure, its pressure range was similar to Sutherland's gauge. Viscosity gauges were not very widely used. In 1906 two gauges which depended on the thermal conductivity of the gas were invented covering the range from atmosphere to $10^{-4}$ Torr. The Pirani [37] gauge measured the heat loss from a hot filament to the gas by the change in resistance of the gas. The thermocouple gauge [38] measured the change in temperature of the filament with a thermocouple. In 1910 Knudsen invented the radiometer gauge [39], which with care could measure to $10^{-6}$ Torr. Only the Pirani and the thermocouple gauge are still in use. Various mechanical gauges were developed including the Bourdon and the diaphragm gauge, the only one to survive to the present day was the capacitance manometer [40] invented in 1929

Measurement of pressure in the high vacuum range prior to 1916 was difficult, the only available gauges were the Mcleod gauge and Dushman's rotating disk gauge [41] (1915), the latter was the precursor to the spinning-rotor gauge, these were both awkward to use and limited to about $10^{-7}$ Torr. In 1916 Buckley described the hot-cathode ionisation gauge [42] (von Baeyer had reported the measurement of pressure in a hot-filament triode used as an ionisation gauge [43] in 1909; however, his work was not followed up and Buckley is generally credited with the invention of the hot-cathode ionisation gauge). The triode hot-cathode ionisation gauge was almost universally used to measure high vacuum until 1950. This type of gauge had a cylindrical ion collector of large surface area surrounding the cylindrical grid and axial filament, an early example [44] is shown in Fig. 17. With the use of this type of hot-cathode ionisation gauge the measured ultimate pressure in almost all vacuum systems was about $10^{-8}$ Torr. The ultimate pressure had hit a plateau that lasted for more than 30 years.

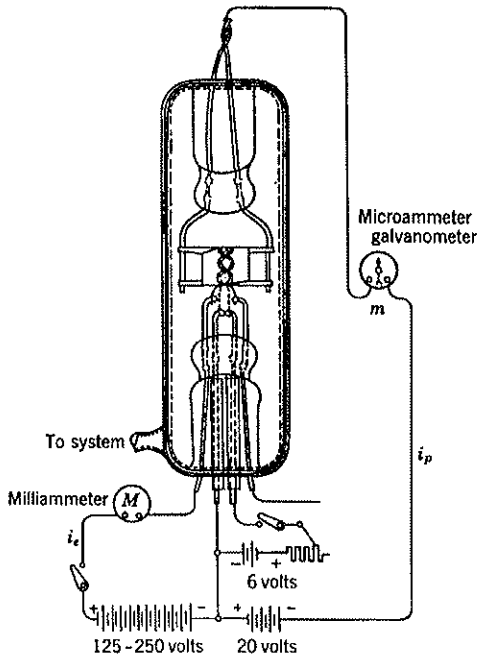



Fig. 17 Triode ionization gauge of Dushman and Found (1921)

The limit at $10^{-8}$ Torr was generally assumed to result from a failure of the pumps rather than the gauge, there was initially no suspicion that there might be a limiting process in the gauge at low pressures. The search for a method to reduce the ultimate vacuum in this period has been described in some detail [45] and will not be repeated here. Suffice to say there was considerable evidence by the late 1930s that the hot-cathode ionisation gauge was not capable of measuring below $10^{-8}$ Torr. The reason for this limit (the x-ray limit as we now know) was not understood by workers in vacuum technology although by the late 1930s the production of photoelectron emission from the grid of a high-power vacuum tube, as a result of soft x-rays produced by electron bombardment of the anode, was well understood by vacuum tube engineer (see Ref. 40 and references therein). The two communities, although closely associated, apparently did not communicate on this subject.

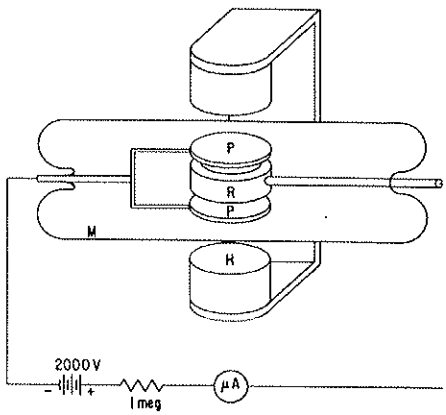



Fig. 18 Diagram of the Penning gauge, the first magnetic, cold-cathode, ionization gauge. P, cathodes, R, anode, H, magnet pole-face (1937).

The cold-cathode discharge in a magnetic field was first demonstrated by Phillips [46] in 1898 but it was not until 1937 that Penning invented the cold-cathode ionisation gauge in a magnetic field, known as the Penning or Phillips gauge [47] (see Fig. 18). This gauge was rugged and reliable but the ion current vs. pressure relation was non-linear and the discharge went out at about $10^{-7}$ Torr and thus it did not help to solve the pressure limit at $10^{-8}$ Torr. The gauge was also found to have a significant pumping speed and was the precursor of the sputter-ion pumps designed in the late 1950s. The fact that an electric discharge caused a pumping action had been known since 1858 when Plücker noted [48] "Certain gases react with the platinum cathode and the resulting compounds deposit on the walls. So we approach the absolute vacuum". In 1916 Vegard observed [49] that there was a change in the pumping speed of a cold-cathode discharge when using different materials for the cathode but not the anode. Penning noted "at high pressures a discharge with high currents should be avoided because of cathodic sputtering ... one has to take care of the fact that the discharge absorbs gas, in air about 1 litre at By 1950 the minimum measured pressure was still $10^{-8}$ Torr, the same as it had been in 1920. Some experimenters had undoubtedly achieved pressures much lower than $10^{-8}$ Torr (e.g. W.B. Nottingham [50] and P.A. Anderson [51]) but were not able to make definitive measurements of these lower pressures. It had been suggested that the hot-cathode ionisation gauge had a limit to the lowest measurable pressure set by an x-ray photocurrent but this had not been proven.

1950 is the end of our journey when the revolution in vacuum technology occurred (the Bayard-Alpert gauge [52]) which led to ultrahigh vacuum and the modern era of vacuum devices. Virtually all the new or improved vacuum devices developed since 1950 find their roots in devices developed between 1850 and 1950.

## REFERENCES


[1] Rene Descartes, Oeuvres, eds. Chas Adams and Paul Tannery, vol. 1, p. 205 (Paris, 1897).

[2] W.E.K. Middleton, The History of the Barometer (John Hopkins Press, Baltimore, 1964).

[3] E.N. da C. Andrade, The History of the Vacuum Pump, Adv. Vac. Sci. Tech., 1, 14 (Pergamon, Oxford, 1960).

[4] M.J. Sparnaay, Adventures in Vacuum (North Holland, Amsterdam, 1992).

[5] L. Fabel, Vakuum-Technik, **35**, 128 (1986).

[6] A.C. van Helden, Tractrix, **3**, 149 (1991).

[7] M.H. Hablanian, Comments on the History of Vacuum Pumps, J. Vac. Sci. Technol. A, **2**, 118 (1984).

[8] T.E. Madey, Early Applications of Vacuum, from Aristotle to Langmuir, J. Vac. Sci. Technol. A, **2**, 110 (1984).

[9] P.A. Redhead, The Measurement of Vacuum Pressures, J. Vac. Sci. Technol. A, **2**, 132 (1984).

[10] S. Shapin and S. Schaeffer, Leviathan and the Air Pump (Princeton University Press, 1985).

[11] P.A. Redhead, The Ultimate Vacuum, Vacuum, **53**, 137 (1999).

[12] Caspar Schott, Mechanica Hydraulico-pneuimatica (Wurzburg, 1657).

[13] Robert Boyle, new Experiments Physico-Mechanical, Touching the Spring of the Air and its effects, etc. (Oxford, 1660).

[14] F. Hawksbee, Physico-Mechanical Experiments on Various Subjects (London, 1709, Johnson Reprint Corp., New York, 1970).

[15] J.A. Bennett, Science at the great Exhibition (Whipple Museum, Cambridge, 1983).

[16] W.H.T. Mayer, Über das geschichtete elektrische Licht (Berlin 1858).

[17] A. Töpler, Dingler's Polytechnisches Jour., **163**, 426 (1862).

[18] H.T.P. Sprengel, J. Chem. Soc., **16**, 9 (1865).

[19] W. Crookes, Phil. Trans. Roy. Soc., **163**,277 (1873).

[20] C.H. Gimingham, J. Soc. Chem. Ind., **3**, 83 (1884).

[21] H.G. Mcleod, Phil. Mag, **48**, 110 (1874).

[22] M. Josephson, Edison (McGraw-Hill, New York, 1959).

[23] G.W.A. Kahlbaum, Wied. Ann., **53**, 199 (1894).

[24] H.A. Fleuss, German Patent, No. 76147, 13/8/1893.

[25] W. Gaede, Phys. Zeit., **6,** 758 (1905).

[26] G. Meyer, Verh. d. deutsch Phys. Ges., **10**, 753 (1907).

[27] W. Gaede, Ann. Physik, **41**, 337 (1913).

[28] S. Dushman, Phys. Rev., **5**, 224 (1915).

[29] M. Holweck, C.R. Hebd. Seances Acad. Sci., **177**, 43 (1923).

[30] M. Seigbahn, Arch. Math. Astron. Fys., **30B**, 17, (1944).

[31] W. Gaede, Ann. Phys. (Leipzig), **46**, 357 (1915).

[32] I. Langmuir, Gen. Elec. Rev., **19**, 1060 (1916); J. Franklin Inst., 182,719 (1916).

[33] C.R. Burch, Nature, **122,** 729 (1928).

[34] R.G. Sherwood, Phys. Rev., **12**, 70 (1918).

[35] W. Sutherland, Phil. Mag., 43, 83 (1887).

[36] I. Langmuir, J. Amer. Chem. Soc., **35**,107 (1913).

[37] M. Pirani, Verh. Physik Ges., **8**, 686 (1906).

[38] W. Voege, Phys. Zeit., **7**, 498 (1906).

[39] M. Knudsen, Ann. Phys. (Leipzig), **31**, 633 (1910).

[40] A.R. Olsen and L.L. Hurst, J. Amer. Chem Soc., **51**, 2378 (1929).

[41] S. Dushman, Phys. Rev., **5**, 212 (1915).

[42] O.E. Buckley, Proc. Nat. Acad. Sci., **2**, 683 (1916).

[43] O. von Baeyer, Phys. Zeit., **10**,169 (1909).

[44] S. Dushman and G.C. Found, Phys. Rev., **17**, 7 (1921).

[45] P.A. Redhead, *in* Vacuum Science and Technology, ed. P.A. Redhead, (A.I.P. Press, New York, 1994) p. 133.

[46] C.E.S. Phillips, Proc. Roy. Soc., **64**, 172 (1898).

[47] F.M. Penning, Physica, **4**, 71 (1937).

[48] J. Plucker, Poggendorf's Ann. **103**, 88 (1858); **105**, 84 (1858).

[49] L. Vegard, Ann. Phys. (Leipzig), **4**, 769 (1916).

[50] W.B. Nottingham, Phys. Rev. **49**, 78 (1936); J. Appl. Phys., **8**, 762 (1937).

[51] P.A. Andersen, Phys. Rev. **47**, 958 (1935); **49**, 320 (1936); **54**, 753 (1938); **57**, 122 (1940).

[52] R.T. Bayard and D. Alpert, Rev. Sci. Instr., **21**, 571 (1950).

# THE LHC VACUUM SYSTEM

*Oswald Gröbner, for the LHC Vacuum Group*
CERN, Geneva, Switzerland

**Abstract**
The Large Hadron Collider (LHC) project, now in the advanced construction phase at CERN, comprises two proton storage rings with colliding beams of 7-TeV energy. The machine is housed in the existing LEP tunnel with a circumference of 26.7 km and requires a bending magnetic field of 8.4 T with 14-m long superconducting magnets. The beam vacuum chambers comprise the inner 'cold bore' walls of the magnets. These magnets operate at 1.9 K, and thus serve as very good cryo-pumps. In order to reduce the cryogenic power consumption, both the heat load from synchrotron radiation emitted by the proton beams and the resistive power dissipation by the beam image currents have to be absorbed on a 'beam screen', which operates between 5 and 20 K and is inserted inside the vacuum chamber. The design of this beam screen represents a technological challenge in view of the numerous and often conflicting requirements and the very tight mechanical tolerances imposed. The synchrotron radiation produces strong outgassing from the walls. The design pressure necessary for operation must provide a beam lifetime of several days. An additional stringent requirement comes from the power deposition in the superconducting magnet coils due to protons scattered on the residual gas, which could lead to a magnet, quench and interrupt the machine operation. Cryopumping of gas on the cold surfaces provides the necessary low gas densities but it must be ensured that the vapour pressure of cryosorbed molecules, of which $H_2$ and He are the most critical species, remains within acceptable limits. In the warm straight sections of the LHC the pumping speed requirement is determined by ion-induced desorption and the resulting vacuum stability criterion.

## 1. INTRODUCTION

The Large Hadron Collider (LHC) [1] consists of a pair of superconducting storage rings with a circumference of 26.7 km. Two proton beams, of 530 mA and an energy of 7.0 TeV, circulate in opposite directions. Of the 54-km total length, the arcs account for almost 48 km and will be at 1.9 K, the temperature of the superconducting magnets. The basic layout of LHC follows the existing LEP machine, with eight long straight sections and eight bending arcs. The present experimental programme includes two high-luminosity experiments (ATLAS and CMS) as well as a heavy ion experiment (ALICE) and a B-physics experiment shown schematically in Fig.1. In this figure is also shown the position for the two beam transfer lines from the existing SPS accelerator, which will be used as the injector, the locations for the accelerating RF system, the beam dump system and the two beam cleaning insertions.

To fit the 7-TeV rings into the existing LEP tunnel special design concepts are needed for the machine components and, most specifically, superconducting magnets with 8.3 T field. The magnetic bending field of 8.3 T implies that the superconducting magnets which use commercially available NbTi superconductors must operate at a temperature well below 4.2 K. A very attractive technical solution for the LHC has been to immerse the magnets in a bath of superfluid helium at 1.9 K. A selection of the most significant machine parameters is given in Table 1.

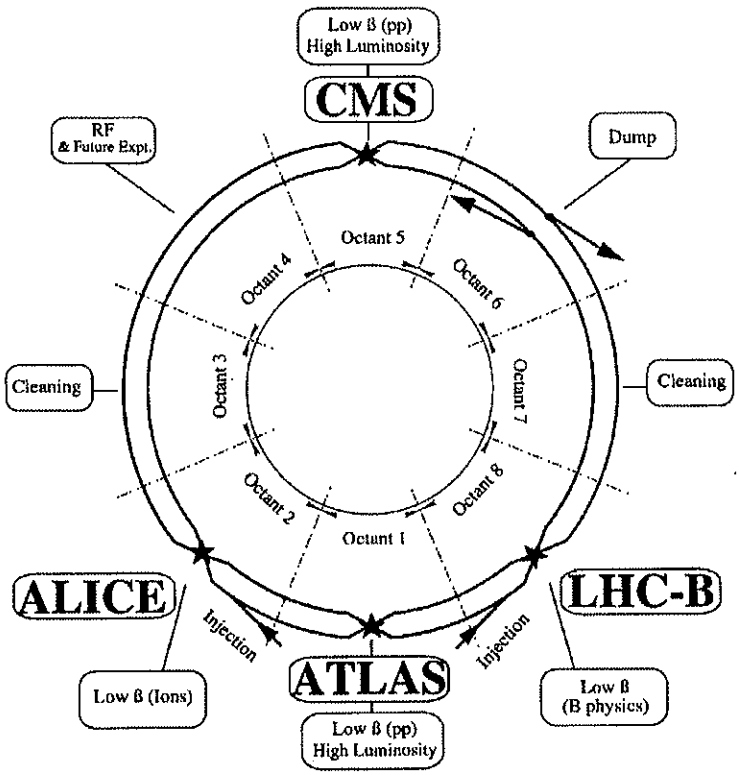



Fig. 1 General layout of LHC

Table 1
Principal machine parameters for proton-proton collisions

| | Value | Unit |
|---|---|---|
| Energy | 7 | TeV |
| Dipole field | 8.33 | T |
| Coil aperture | 56 | mm |
| Luminosity | $10^{34}$ | $s^{-1}$ $cm^{-2}$ |
| Injection energy | 450 | GeV |
| Circulating current /beam | 0.56 | A |
| Bunch spacing | 25 | ns |
| Particles/bunch | $1.1\ 10^{11}$ | |
| Stored beam energy | 350 | MJ |
| Normalised transverse emittance | 3.75 | µm |
| R.m.s. bunch length | 0.075 | m |
| Beam lifetime | 22 | h |
| Luminosity lifetime | 10 | h |
| Energy loss/turn | 6.7 | keV |
| Critical photon energy | 45 | eV |
| Linear photon flux | $1\ 10^{17}$ | $m^{-1}$ $s^{-1}$ |
| Total radiated power/beam | 3.8 | kW |

The magnets for LHC are based on a two-in-one design, where the two beam channels are incorporated into a common iron yoke in a compact single cryostat [2]. Figure 2 illustrates this design concept showing the cross section of a dipole magnet with its cryostat. A similar design has been adopted for the quadrupole magnets of the short straight sections (sss). The compact machine design which results from this construction allows the installation of the LHC machine inside the existing tunnel section of LEP and makes the necessary provision for the addition of a LEP-like machine on top of LHC at some later time.

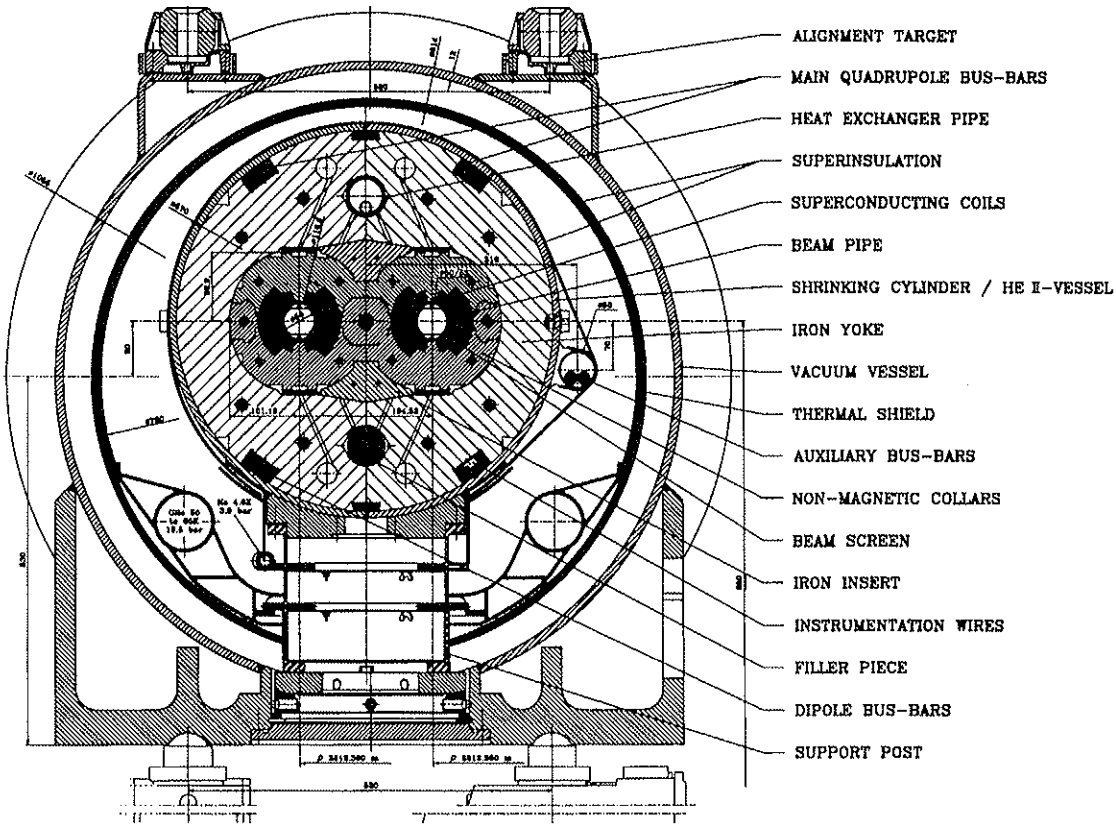



Fig. 2 Dipole magnet cross section with cryostat

In the arcs of LHC, representing more than 24 km of each ring, the machine lattice consists of a succession of regular 53.4-m long 'half-cells': three 14.2-m long twin-aperture dipole magnets, and one 6.3-m long short straight section. The remaining space is taken by various machine elements and by interconnections between magnets. In the eight long straight sections the machine layout is adapted to the specific requirements of the experiments or of the other functions and the repetitive, regular design is locally modified. In the four experimental straight sections, the beams converge into one common beam pipe, where they are tightly focused and steered into head-on collisions at the interaction points.

## 2. THE VACUUM SYSTEM

The vacuum system for LHC consists of two main parts which are kept entirely separated in terms of vacuum since they must meet rather different requirements: the cryogenic insulation vacuum necessary to avoid heat load by gas conduction requires a pressure of only about $10^{-6}$ mbar while the beam vacuum proper which must provide a good beam lifetime and low background for the experiments requires several orders of magnitude better vacuum for the operation with beam.

### 2.1 Cryogenic insulation vacuum

The cryogenic insulation vacuum for the machine consists of a nearly 3-km long continuous arc cryostat, which is, subdivided into 14 sectors each 212-m long [3]. Vacuum separation is achieved by

vacuum barriers which subdivide the arc cryostat into subsections which can be pumped, commissioned and leak checked independently. These vacuum barriers also separate the vacuum for the continuous cryogenic distribution line (QRL) from the machine cryostat as shown in Fig. 3.

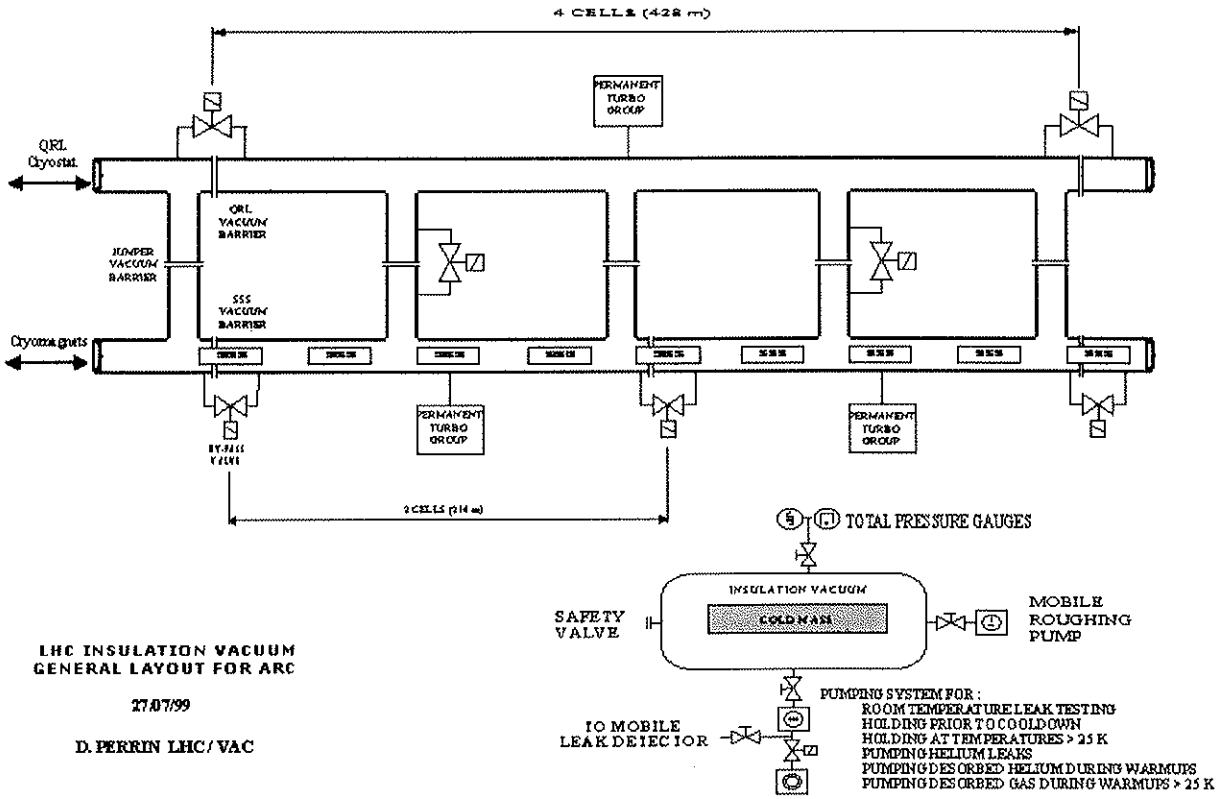



Fig. 3 Layout of cryogenic insulation vacuum

While the initial pump down can be done with conventional turbomolecular pumping groups shown in the insert of the Fig. 3, the cryopumping on the large external surface of the magnet cold mass and on the thermal screens will be entirely sufficient for maintaining the static insulation vacuum once the system has been cooled down. Provisions for any additional pumping during machine operation are required only in the case that excessive helium leaks would occur. This additional pumping can be achieved either by connecting external pumps or by installing charcoal-coated cryo-panels with a large pumping capacity for helium. Whether this option is actually needed to recover from a degraded vacuum remains to be decided on the basis of a failure analysis and based on experience with the final installation procedure and a test string of the machine.

## 2.2 Beam vacuum and beam related heat loads

The beam vacuum system has been the subject of extensive studies over the recent years and has been presented at several occasions [4–6]. The beam pipe in the centre of the superconducting magnet coils, as illustrated in Fig. , is in direct contact with the helium bath at 1.9 K (cold-bore beam pipe) so as to minimise the loss of aperture between the magnet coils and the beam stay-clear region.

At this temperature, the vacuum chamber wall becomes a very efficient cryopump with practically infinite capacity for all condensable gas species with the sole exception of helium and thus makes any external pumping superfluous and even inefficient. Unfortunately, this concept has the drawback that the cryogenic system will have to remove even very small amounts of beam induced power which, for conventional room temperature accelerators, would be insignificant. Indeed, since the removal of 1 W at 1.9 K requires nearly 1 kW of electric power, there is a strong incentive to avoid or to reduce to the bare minimum any source of heat in-leak to the cold beam pipe.

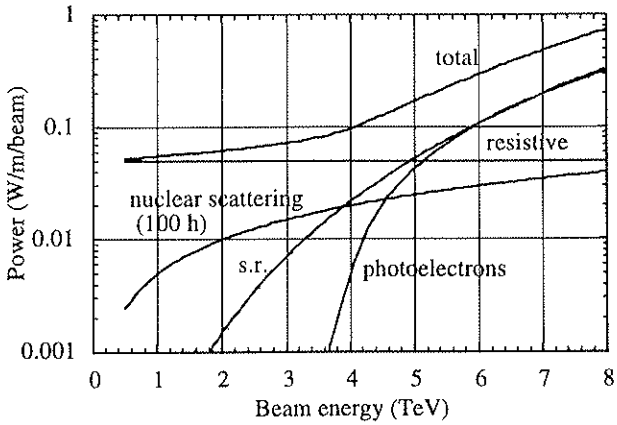



Fig. 4 Beam-induced losses in the LHC

Four distinct beam and vacuum related heat sources to the 1.9 K system can be identified as shown in Fig. 4 and these will be discussed in more detail in the following sections:

i) Synchrotron radiation

Due to the centripetal acceleration in the bending magnets the beams will emit a synchrotron radiation flux of about $10^{17}$ photons $s^{-1}$ $m^{-1}$ with a critical energy of the photon spectrum of 45 eV equivalent to a distributed linear power of 0.2 W $m^{-1}$. For LHC as a whole, this amounts to 7.6 kW.

ii) Image currents from the beam

The inner wall of the beam pipe must conduct the image currents of the beams, which may dissipate typically 0.05 W $m^{-1}$ per beam duct. This power due to image currents depends directly on the resistivity $\rho_w$ of the vacuum chamber wall material. To limit this heat load, but also to avoid resistive wall instability, the resistivity of the vacuum chamber has to be low. For the LHC this can be achieved by coating the inner surface of the stainless steel beam pipe with a thin copper layer and by profiting, furthermore, from the fact that at low temperature the resistivity of high-purity copper is strongly reduced and can be by a factor of $10^3$ better than for stainless steel.

iii) Photo-electrons and multipacting

Beam induced multipacting can arise through an oscillatory motion of a cloud of photo-electrons and of low energy secondary electrons bouncing back and forth between opposite vacuum chamber walls during successive passages of proton bunches [7]. Due to the strong electric field of the dense proton beam these electrons will be accelerated and transfer their energy to the cold wall. In the LHC the beam pipe radius $r_p$, the bunch intensity $N_b$ and the bunch spacing $L_{bb}$ are such that the wall-to-wall multipacting threshold corresponds to about 1/4 of the nominal intensity [8]. The magnitude of the heat load depends on several parameters as will be discussed in a later section but a crude estimate indicates a value of 0.2 W/m.

iv) Beam loss by nuclear scattering

Nuclear scattering of the high-energy protons on the residual gas generates an unavoidable continuous flux of high-energy particles, which are lost from the circulating beams. Even under ideal conditions there remains a small fraction of scattered protons which can not be absorbed in a controlled way by the dedicated set of collimators in the beam cleaning insertions (see Fig. 1). These particles escape from the aperture of the machine and penetrate through the cold bore into the magnets where they are lost. This flux of high-energy protons creates a shower of secondary particles, which are ultimately absorbed in the 1.9 K magnet system. The continuous heat input to the cryogenic system is directly proportional to the gas density and hence the maximum allowed value defines an upper limit of the gas density in the beam pipe. The beam lifetime $\tau$, due to nuclear scattering (the cross section for 7-TeV protons on hydrogen atoms is $\sigma \sim 5\ 10^{-30}$ $m^2$) is given by

$$\frac{1}{\tau} = c\sigma n \tag{1}$$

where $c$ is the speed of light and $n$ the gas density. The linear power load, $P$ (W/m), is proportional to the beam current $I$, the beam energy $E$ and can be expressed directly in terms of the beam-gas lifetime

$$P = \frac{IE}{c\tau} = 0.93\frac{E(\mathrm{TeV})I(\mathrm{A})}{\tau(\mathrm{h})} \tag{2}$$

For the first time in an accelerator, beam loss due to nuclear scattering on the residual gas represents a non-negligible heat load. The machine design includes therefore a nuclear scattering allowance of ~0.1 W/m for the two beams. Consistent with this requirement a beam lifetime of ~100 hours has been chosen which in turn implies that the $H_2$ density must be $\leq 1 \times 10^{15}$ molecules $m^{-3}$ (e.g. equivalent to a pressure of $\leq 1 \times 10^{-9}$ Torr at 10 K). A correspondingly lower density is required for heavier gases as is shown in Table 2 where the limits of the molecular densities for some common gas species have been listed.

Table 2
Relative cross sections and maximum gas densities for different gas species

| Gas | $S/s_{H2}$ | $N$ ($m^{-3}$) for $\tau = 100$ h |
|---|---|---|
| $H_2$ | 1 | $1\ 10^{15}$ |
| He | 1.26 | $7.8\ 10^{14}$ |
| $CH_4$ | 5.4 | $1.8\ 10^{14}$ |
| $H_2O$ | 5.4 | $1.8\ 10^{14}$ |
| CO | 7.8 | $1.2\ 10^{14}$ |
| $CO_2$ | 12.2 | $8.0\ 10^{13}$ |

Among the four heat sources shown in Fig. 4 synchrotron radiation, image currents and photo-electrons can be intercepted at a thermodynamically-more-efficient temperature of 20 K. This can be achieved by inserting in the cold bore vacuum chamber a thermally insulated 'beam screen', which is actively cooled by gaseous helium. The fourth component, due to scattered high-energy protons, cannot be intercepted by this screen and must be included in the cryogenic budget for the 1.9 K system. This effect, which has a practical importance for the first time in the LHC, defines a 'lifetime limit' of $10^{15}$ $H_2$ molecules $m^{-3}$ for the maximum average gas density in the LHC arcs. It must be noted that in the exceptional case of a limited region with a very high pressure, e.g. for a helium leak, the related beam loss may even trigger a magnet to quench.

The section of the beam screen with its two cooling tubes is shown in Fig. 5. In order not to loose the benefit of the cryopumping on the 1.9 K cold bore, the screen has narrow pumping slots along its length which enable gas molecules to leave the beam channel proper and to be adsorbed on the cold bore wall. The importance of cryosorbing gas molecules on a surface, which is screened from the effects of the beam, will be discussed in the following section.

## 3. PRIMARY DESORPTION BY SYNCHROTRON RADIATION

At 7 TeV, the critical energy of the synchrotron radiation spectrum is ~45 eV and the radiated power represents a heat load of 0.2 W/m per beam. The actively-cooled beam screen intercepts this power so that it does not load the 1.9 K system. For vacuum reasons, the screen has to be partially transparent so that molecules desorbed by the synchrotron radiation can escape from the beam channel and be pumped permanently on the cold bore surface as illustrated in Fig. 5.

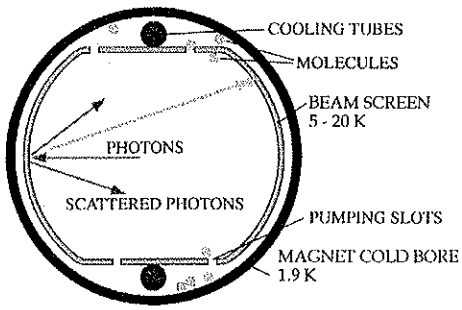



Fig. 5 Actively-cooled beam screen in the magnet cold bore

The synchrotron radiation, which strikes the copper-coated inner surface, liberates a significant amount of gas, proportional to the photon flux and to the molecular desorption yield. The molecular yield for $H_2$ from a copper surface at 10 K is typically 5 $10^{-4}$ molecules per photon while other species have considerably smaller values. Figure 6 shows the primary molecular desorption yield for copper-coated stainless steel as a function of the photon dose measured on a test chamber at 77 K, grazing photon incidence and with a critical photon energy of 50 eV.

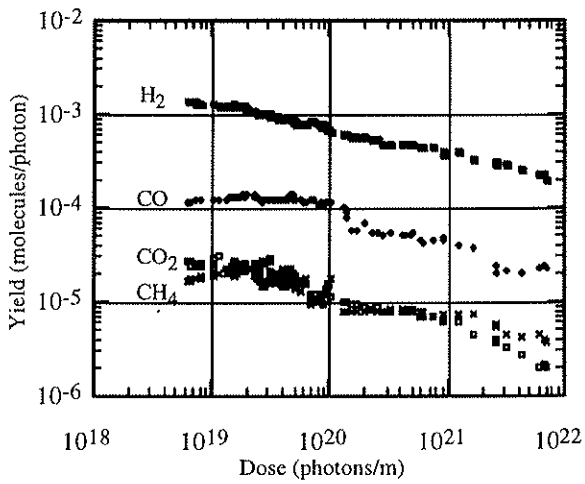



Fig. 6 Primary molecular-desorption yield versus accumulated photon dose

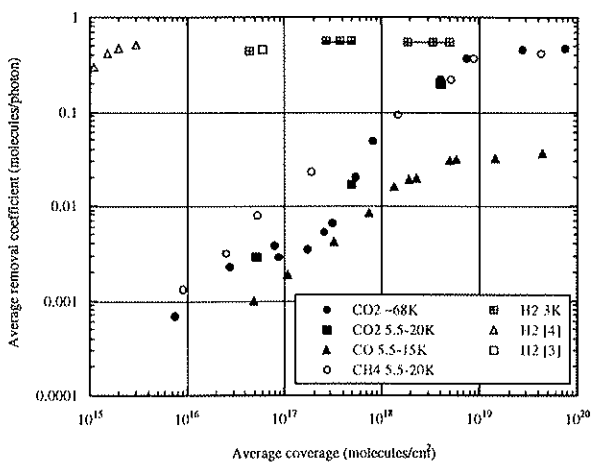



Fig. 7 Re-cycling coefficient for different gas species as a function of the coverage

## 4. VAPOUR PRESSURE AND RECYCLING OF PHYSISORBED GAS

Among the desorbed molecules, hydrogen, which has the largest molecular desorption yield, is the most critical species for which the thermal desorption is high and the sorption capacity at the temperature of the beam screen surface is rather small. Without the pumping holes, the surface of the screen would therefore not provide any useful pumping capacity since the equilibrium vapour density at 5 K for a monolayer of hydrogen exceeds by several orders of magnitude the acceptable limit [9]. With the concept of pumping holes however, hydrogen, and even more so all other gas species, will be cryosorbed on the 1.9 K surface with a negligibly-low vapour pressure and hence with a practically unlimited capacity.

Since direct and scattered/reflected photons as well as photo-electrons continuously strike the interior surface of the beam screen, cryosorbed gas molecules are efficiently re-desorbed. This 'recycling' effect increases with surface coverage and may dominate other sources of outgassing. The global result of thermal desorption and of recycling of gas implies that only a small quantity of gas can be accommodated on the inner surface of the beam screen. Thereafter any additional molecules will gradually escape through the holes and the pressure rise stabilises when the rate of production of gas molecules equals the pumping through the holes.

Experimental data for the recycling effect have been obtained from an experiment at the VEPP-2M storage ring at the Budker Institute of Nuclear Physics (BINP) in Novosibirsk, Russia [10]. In this experiment a sample can be loaded initially with a known quantity of a gas. Subsequently, the sample is exposed to a known number of photons and the quantity of molecules remaining after this exposure is measured during its warm-up. The results of this experiment are given in Fig. 7 where it can be seen that the recycling of $H_2$ by synchrotron radiation photons is a very pronounced effect already at a coverage well below a monolayer (< $10^{15}$ molecules $cm^{-2}$).

It is interesting to note that this recent result for the hydrogen re-cycling coefficient is in good agreement with observations made previously when investigating the problem of an abnormal hydrogen vapour pressure in condensation cryopumps [11]. It was observed that exposure to room temperature thermal radiation can desorb very strongly hydrogen condensed on a substrate at 2.3 K. Indeed, extrapolating from these earlier measurements to LHC conditions, the synchrotron radiation power density on a cold bore without screen (~ 0.14 W $cm^{-2}$) could increase the saturated vapour pressure to a value comparable to the lifetime limit.

The evolution of the vacuum pressure is illustrated in Fig. 8 which shows the experimental results from a test at the VEPP-2M photon beam line at BINP [12]. For better comparison, the data shown in the figure have been scaled to the photon flux in the LHC corresponding to the initial conditions of 1/10 of the nominal current. With a ~2% area of pumping holes, hydrogen saturates at ~3 $10^{13}$ $m^{-3}$, which is well below the specified lifetime limit. Warming the beam screen to 77 K while maintaining the cold-bore at ~3K, to the remove cryosorbed hydrogen demonstrates the recycling effect of cryosorbed gas on the inner wall of the beam screen. This process effectively desorbs all the hydrogen. During the second exposure hydrogen accumulates again on the screen until the gas density reaches a new saturation level.

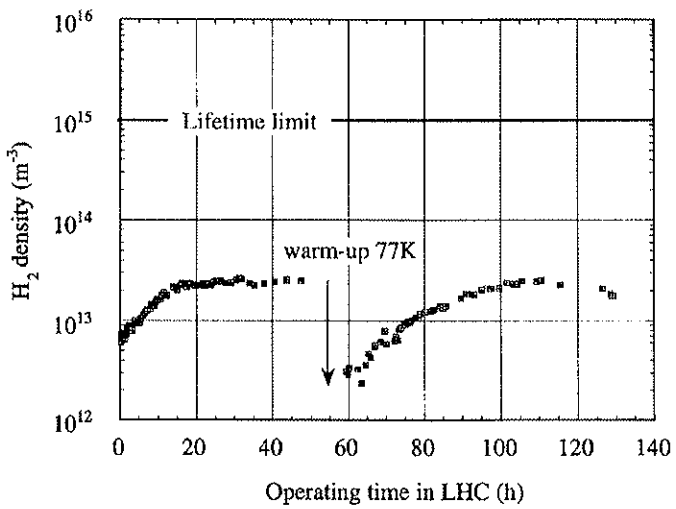



Fig. 8 Evolution of the hydrogen density during photon exposure

## 5. DYNAMIC MODEL FOR $H_2$

The dynamic vacuum behaviour for hydrogen in the cold parts of LHC with a perforated beam screen may be described very accurately by a linear model for the interaction between the volume density, $n$ ($m^{-3}$), and the surface density, $\Theta$ ($m^{-2}$), of adsorbed molecules on the wall of the beam screen. The volume and the surface gas interact according to the following set of linear equations:

$$V\frac{dn}{dt} = q - an + b\Theta \tag{3}$$

and

$$F\frac{d\Theta}{dt} = cn - b\Theta \ . \tag{4}$$

Here, $V$ is the volume and $F$ is the net wall area per unit length of the vacuum system. The parameters $a$, $b$, $c$ and $q$ represent the following physical quantities:

i) $q$ represents a source of gas, here in particular the photon induced desorption rate determined by the product of the desorption yield, $\eta$ (molecules/photon), and of the photon flux, $\dot{\Gamma}$, hence $q = \eta\ \dot{\Gamma}$ (molecules $s^{-1}m^{-1}$). In more general terms $q$ can include any source of chemisorbed gas e.g. by ion- or by electron-stimulated desorption.

ii) $a$ ($m^2\ s^{-1}$) describes the total pumping on the surface of the wall and in particular through the area of the holes in the beam screen. It can be expressed as

$$a = \frac{1}{4}\bar{v}F(s+f) \tag{5}$$

where $\bar{v}$ is the average molecular velocity, $s$ the sticking probability of the molecules on the beam screen and $f$ is the fraction of the total surface of the beam screen with pumping holes, note that $f<<1$.

iii) $b$ represents gas originating from a physisorbed surface phase, $\Theta$, by thermal- and by photon-induced desorption. In contrast to $q$, this is not a source of 'new' gas but constitutes a surface phase interacting through adsorption and desorption with the gas in the volume.

The thermal contribution to $b$ [13] may be expressed as the molecular vibration frequency $\nu o = 10^{13}\ s^{-1}$ multiplied with a Boltzmann factor

$$e^{-E/kT} \tag{6}$$

characterised by its activation energy $E$ and by the temperature $T$.

The photon-induced desorption process can be expressed as $\kappa\dot{\Gamma}$, where $\kappa$ is the re-cycling cross-section in $m^2$ per photon, thus

$$b = F\nu_o e^{-\frac{E}{kT}} + \kappa\dot{\Gamma} \ . \tag{7}$$

iv) $c$ is the rate of adsorption of gas molecules on the surface of the beam screen only. This effect is proportional to the sticking probability of the molecules, $s$,

$$c = \frac{1}{4}\bar{v}sF \ . \tag{8}$$

This dynamic model contains a reasonably small number of parameters, which, for simplicity may be assumed to be constant. In particular it implies a linear adsorption isotherm.

$$\Theta = \frac{c}{b}n = \frac{\frac{1}{4}\bar{v}sF}{F\nu_o e^{-E/kT} + \kappa\dot{\Gamma}}n \ . \tag{9}$$

Here $\Theta$ (n) depends on the re-cycling cross-section as well as on the photon flux. Therefore, the isotherm and the equilibrium surface coverage for a given volume gas density will depend on these specific conditions (i.e. on the beam current) which effectively reduce the sojourn time of any physisorbed molecules on the surface.

As shown in Fig. 9 the experimental data presented in Fig. 8, can indeed be described in a very satisfactory manner by this simple model by choosing a small number of free parameters only, see Table 3. Parameters shown in bold face have been adapted to best fit the data. The fraction of pumping slots $f$, takes into account the Clausing factor due to the wall thickness of the beam screen.

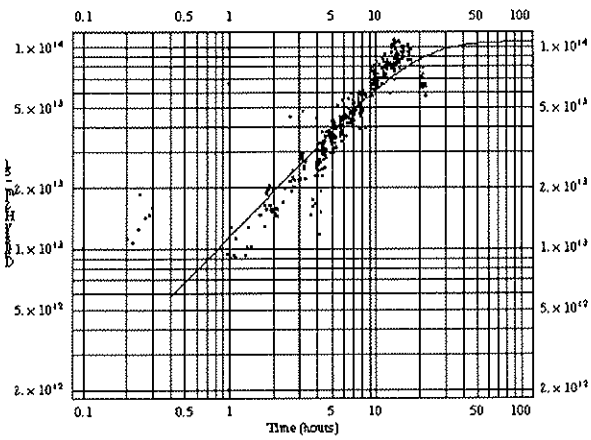



Fig. 9 Comparison of experimental data with the dynamic model for $H_2$ density ($m^{-3}$) as a function of time

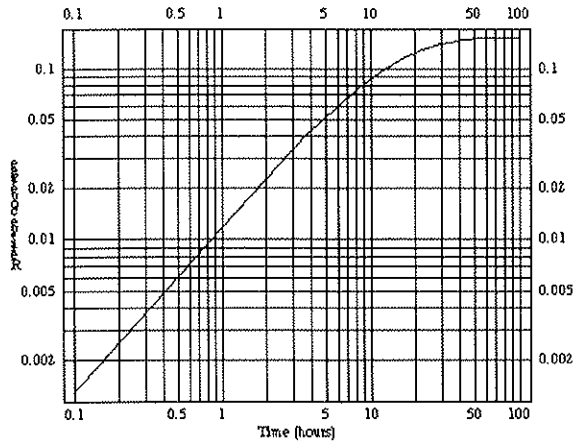



Fig. 10 Surface density of $H_2$ molecules normalised to a monolayer coverage as a function of time

Table 3
List of parameters.

| | | |
|---|---|---|
| Desorption yield | $\eta$ | $5 \times 10^{-4}$ molecules/photon |
| Screen temperature | T | 10 K |
| Linear photon flux (200 mA) | $\dot{\Gamma}$ | $3.14 \times 10^{16}$ photons/s/m |
| **Sticking probability** | $\sigma$ | 0.6 |
| Monolayer capacity | | $3 \times 10^{19}$ molecules/m$^2$ |
| **Re-cycling coefficient** | $\kappa$ | $5 \times 10^{-21}$ m$^2$ |
| **Activation energy** | $E$ | 800 cal/mole or 0.035 eV/molecule |
| Specific surface area | $F$ | 0.14 m |
| Fraction of pumping slots | $f$ | 1.28% |
| Specific volume | $V$ | $1.3 \times 10^{-3}$ m$^2$ |

The model provides in addition to the volume density also the surface density of gas, which is shown in Fig. 10. There it can be observed that the final coverage obtained remains indeed well below a monolayer as a result of the strong recycling of gas molecules.

In equilibrium conditions, when the volume and surface densities no longer change with time, the volume density is simply given by

$$n_{equ} = \frac{q}{a-c} = \frac{\eta \dot{\Gamma}}{\frac{1}{4}\bar{v} F f} \tag{10}$$

and depends solely on the primary desorption by synchrotron radiation and on the pumping through the holes in the beam screen.

## 6. ION STIMULATED DESORPTION AND PRESSURE INSTABILITY

Ionization of the residual gas molecules will produce ions, which are ejected by the positive beam potential and accelerated towards the beam screen. The final impact energy of the ions may be up to 300 eV in the arcs of the LHC. As with photons, the energetic ions are very effective in desorbing tightly-bound gas molecules, a process, which is characterised by the molecular, desorption yield, $\eta_i$ (molecules per ion). Desorbed molecules may in turn be ionised and thus increase the desorption rate. This positive feedback can lead to the so-called 'ion-induced pressure instability' known from previous experience with the Intersecting Storage Rings (ISR) at CERN [14]. As a result of this feedback mechanism the pressure increases with beam current from the initial value, $P_0$, as

$$P_I = \frac{P_0}{1 - \frac{\eta I}{(\eta I)_{crit}}} \quad . \tag{11}$$

Since the electric charge of the beam is strongly concentrated in 2835 individual, short bunches, the resulting instantaneous peak electric field amounts to several kV. The final energy of the ions at the impact on the vacuum chamber wall can reach several hundred eV depending on the beam intensity and the type of ion. Fig. 11 shows the impact energies calculated for $H_2$ He, $CH_4$, CO and $CO_2$ ions as a function of the average beam intensity. Because of the bunched structure of the beam, light ions may gain considerably more energy than heavier ones.

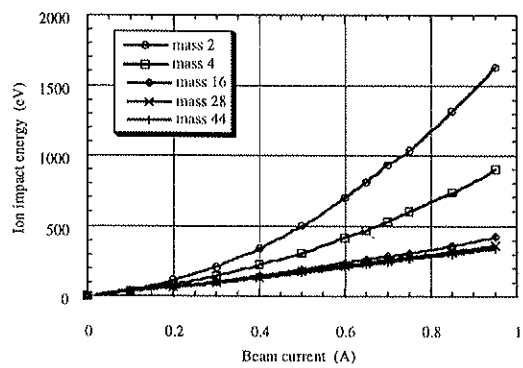



Fig. 11 Ion impact energy as a function of beam current

The pressure runaway depends critically on the local cleanliness of the surface, through the ion-induced molecular desorption yield, $\eta_i$, and on the local pumping speed. The stability limit is expressed by the condition that the product of beam current $I$ and of the molecular desorption yield must be less than a critical value given by the effective pumping speed of the system, $S_{\text{eff}}$

$$(\eta_i I)_{\text{crit}} = \frac{e}{\sigma} S_{\text{eff}} \tag{12}$$

where $e$ is the electron charge and $\sigma$ the ionisation cross section of the residual gas molecules for high energy protons. According to Eq. (12) the pressure will double at half the critical value. Ionisation cross sections for high energy protons for some common gas molecules are given in Table 4 [15].

Table 4
Ionisation cross sections for some common molecules

| | Ionization cross-section (in $10^{-18}$ cm$^2$) | |
|---|---|---|
| **Gas** | **26 GeV** | **7000 GeV** |
| $H_2$ | 0.22 | 0.37 |
| He | 0.23 | 0.38 |
| $CH_4$ | 1.2 | 2.1 |
| CO | 1.0 | 1.8 |
| Ar | 1.1 | 2.0 |
| $CO_2$ | 1.6 | 2.8 |

The effective pumping speed in the cold sections of the LHC can be very large for molecules, which condense on the wall of the beam screen. However, it has a well defined, lower limit determined by the cryosorption on the 1.9 K vacuum chamber surface through the pumping holes in the beam screen. The condition for vacuum stability can be defined by the area $f$ of pumping holes per unit length

$$(\eta_i I)_{crit} = \frac{1}{4}\frac{e}{\sigma}\bar{v} f \ . \tag{13}$$

From this equation the stability limit for the cold arc is typically $10^3$ A for hydrogen and somewhat less for the other gas species. The gradual accumulation of physisorbed gas on the wall leads to an increasing desorption yield which may reach values as large as $10^3$ $H_2$ molecules/ion at a monolayer [16]. Nevertheless, with the high effective pumping speed of the holes the cold sections with beam screens are expected to operate well inside this stability limit. In addition, as shown in Fig. 10, the equilibrium surface coverage for $H_2$ will not increase indefinitely but should remain well below a monolayer due to the continuous re-cycling effect and the pumping through the holes. Therefore, the ion-induced pressure instability can be avoided by ensuring that molecules are cryosorbed on surfaces, which are in the shadow of the beam.

More serious is the vacuum stability in the warm sections of LHC due to the limited pumping speed, which can be provided by conventional pumps, and the limited conductance of the vacuum chamber as imposed by the design of small aperture magnets. The most stringent conditions are encountered in some of the experiments due to the required small beam pipes in the central detectors and due to the locally increased ion energy. Indeed, the strong focussing of the two beams at the interaction points increases the space charge and therefore the electric force on the ions such that the impact energy increases to several keV. Measured room temperature yields for the relevant range of impact energies range from about 1 to 8 molecules/ion [17] and depend critically on the cleanliness of the surface. Special methods of surface conditioning of the beam pipes by vacuum firing, Ar-glow discharge cleaning and by *in-situ* bakeout are under study.

## 7. BEAM-INDUCED ELECTRON MULTIPACTING

Beam induced multipacting [18], which can arise through oscillatory motion of a cloud of photoelectrons and low energy secondary electrons bouncing back and forth between opposite vacuum chamber walls during successive proton bunches, represent a potential problem for the machine [19]. The key parameters for this process are:

bunch intensity and bunch spacing,
synchrotron radiation intensity,
photoelectric and secondary electron yields including the effect of the angle of incidence,
energy distribution of secondary electrons,
photon reflectivity,

beam screen shape and dimensions as well as
external electric and magnetic fields.

The average value of energy transferred to electrons may be of the order of 200 eV. Under situations where the secondary electron yield exceeds unity, the electron cloud may then grow indefinitely and, as a consequence, provoke electron-stimulated gas desorption. A first criterion for the onset of multipacting is obtained from the necessary condition that the transit time of electrons from wall to wall must be equal or less than the time between the passage of successive bunches so that the electron cloud can move in synchronism with the beam

$$N_b = \frac{r_p^2}{r_e L_{bb}} . \tag{14}$$

Here $r_p$ (m) is the beam pipe radius, $L_{bb}$ is the distance between bunches and $r_e$ the classical electron radius. In the LHC the determining parameters like the beam screen radius, the bunch intensity and the bunch spacing are such that the wall-to-wall multipacting condition is satisfied between about 1/4 to 1/3 of the nominal current [20].

The consequence for the machine operation can be an additional heat load due to the power deposited by the photo-electrons in the wall of the cold vacuum system, electron-stimulated gas desorption and finally coherent beam oscillations of the proton beam with the electron cloud leading to emittance growth and ultimately to beam loss. These effects have been studied extensively over the past years and it was concluded that an average heat load of 0.2 W/m should be included in the cryogenic budget. To reduce the electron cloud effect, one must find ways to reduce the production rate of photo-electrons and to achieve a low secondary electron yield at the internal surface of the vacuum chamber and of the beam screen [21].

A second important condition, which has to be satisfied, is that the energy transfer to the electrons is sufficiently large to obtain a secondary electron yield, which exceeds unity. Detailed numerical simulations for the nominal operating conditions have shown that in case of a secondary electron yield exceeding 1.3 to 1.4 the electron cloud could grow indefinitely until it finally saturates due to non-linear effects and space charge [22].

Secondary electron yield data for 'technical' Cu and Al surfaces measured on samples in the laboratory indicate that for both materials a unit yield is exceeded at an electron energy of less than 50 eV [23] and that the respective yield versus energy curves peak at 2.2 and 2.7 between 300 and 450 eV respectively. Therefore, to be safe against electron multipacting, it will be essential to provide a vacuum chamber surface with a lower secondary electron yield and primarily for this reason aluminium has been excluded as a material for beam pipes. Even for the more favourable materials like copper or stainless steel, it will be necessary to reduce their secondary electron yield below this critical value by *in-situ* conditioning, i.e. beam scrubbing, with synchrotron radiation.

## 8. MECHANICAL DESIGN ASPECTS

### 8.1 Beam screen

The beam screen is cooled by two longitudinal cooling tubes placed in the vertical plane above and below the beam where this is acceptable from the point of view of the required beam aperture. For mechanical and magnetic reasons, the screen must be made of a low-permeability stainless steel, approximately 1 mm thick. A detailed description of the required electric, magnetic and mechanical characteristics of the beam screen may be found in [23] and [24]. The inner wall, exposed to the electric field of the beam, is coated with a 0.05 mm layer of high-conductivity copper to keep the electric impedance low [25]. With this mechanical design an optimum compromise has to be found between many, partially conflicting requirements. For example, the good electric conductivity of the copper conflicts with the magnetic forces produced by eddy currents during a magnet quench. These forces have been evaluated at several tonnes per m and impose the mechanical strength of the screen. A list of main design parameters is given in Table 5.

Table 5
Beam screen design parameters

| | | |
|---|---|---|
| Aperture at 5 K radial/vertical | mm | 45.5 / 35.7 |
| Length | m | ~16 |
| Wall thickness | mm | ~ 1 |
| Material | | Mn alloyed stainless steel |
| Magnetic permeability | | < 1.005 |
| Temperature | K | 5 - 20 |
| Cooling tubes (number/inner diameter) | mm | 2 / 3.7 |
| Cooling capacity | W/m | 1 |
| Impedance longitudinal<br>transverse @ 3.3 kHz | Ohm | < 0.5<br>$< 10^{10}$ |
| Quench resistance @ $B\ dB/dt = 300\ T^2s^{-1}$ | | > 50 |
| Quench stress<br>deformation | MPa<br>mm | 340<br>< 0.7 |
| Copper coating @ RRR ~ 100 | mm | 0.05 |
| Pumping slots (effective area) | % | 2-4 |

### *8.1.1 Aperture*

The cross section of the beam screen inserted in the magnet cold bore has been shown in Fig. 5. Since the space inside the cold bore is very limited, a design, which maximises the aperture available for the beam, must be found. The section is the result of an extensive study which takes into account the beam requirements at all stages of operation from injection, beam cleaning, colliding beams, to beam dumping. An individual beam screen for a dipole magnet is approximately 16-m long and its nominal clearance to the cold bore is less than 0.5 mm. It will be a technical challenge to meet these very tight dimensional tolerances and the required straightness.

The beam screens in one half-cell will be cooled in series by helium with 5 K inlet and 20 K outlet temperature. The preferred method for the attachment of the cooling tubes is by laser welding to minimise deformation and to keep the heat affected zone of the welds as small as possible. After the screen has been inserted in the magnet cold bore, there should remain a small annular gap to obtain on the one hand a good centring but on the other a low thermal contact.

### *8.1.2 Magnetic properties*

The magnetic permeability of the beam screen and its components must be very low. Since any asymmetry in the distribution of magnetic material could give rise to undesirable harmonics of the magnetic guiding field of the beam. The specification imposes a magnetic permeability not exceeding 1.005 at the low temperature. This requirement eliminates more conventional types of stainless steels used in vacuum applications, e.g. 304 and 316 type steels, and imposes the use of highly stable austenitic stainless steels alloyed with Mn and N [26].

### *8.1.3 Copper coating and magnet quench*

The inner surface of the screen must be coated with high-conductivity copper to obtain good electric conductivity. The thickness chosen for this copper layer results from an optimisation of several conflicting requirements:

A sufficiently thick, high conductivity layer will reduce the power dissipated by image currents, provide a low electric impedance and thus limit the growth rate for the transverse resistive-wall beam instability. A larger power dissipation would require more or larger helium cooling channels and thus complicate the design or reduce the vertical aperture. On the other hand, a thicker copper layer will entail larger eddy currents and electro-mechanical forces during a magnet quench and hence require a mechanically stronger, i.e. thicker beam screen, which again reduces the aperture. The compromise is 0.05 mm of copper and a ~1-mm wall thickness for the stainless steel tube. The stresses induced in the beam screen during a magnet quench, results in a ~0.7-mm deflection in the horizontal plane. Among the different methods for producing the copper layer roll-bonding of copper onto stainless steel strips is the most cost effective for a series production. The coated strip is then shaped into the final profile in a continuous tube-forming machine with a single longitudinal laser weld.

### 8.2 Beam screen interconnects

The mechanical design of the cold arc vacuum system requires a large number of interconnections for the beam vacuum between the beam screens of adjacent magnets. Following the proven design of the LEP-type RF-finger contacts a new concept has been developed which takes into account the large thermal contraction (approx. 45 mm) between room temperature and operating temperature [27]. An important design criterion to be met is the requirement of installation and alignment of the long cryomagnets without the risk of damaging the delicate finger contacts even in case of lateral offsets of up to 4 mm. The principle of the design is shown in Fig. 12. To by-pass the RF bellows, the beam screen cooling tubes are joined in the insulation vacuum enclosure by a flexible tubing. With this concept the risk of a helium leak affecting the critical beam vacuum can be minimised. In the final cold position the required smooth and well-defined electric contact is established with a controlled force of the RF contact fingers.

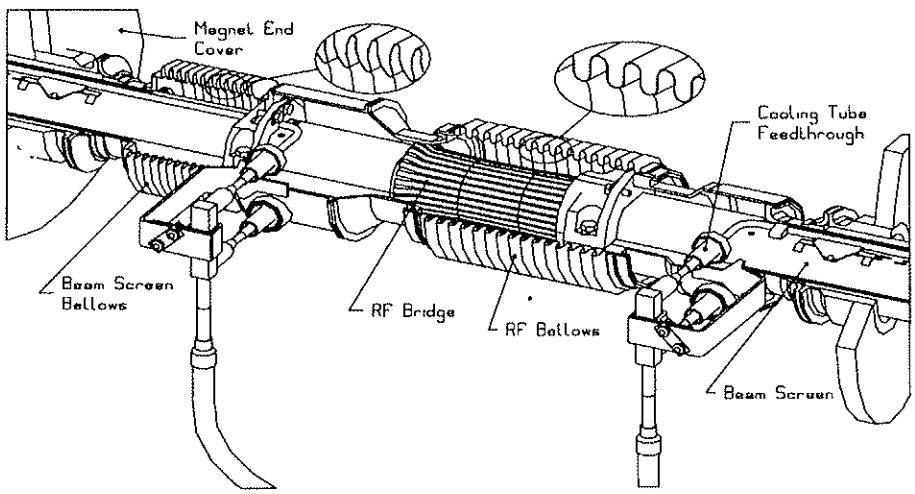



Fig. 12 Beam vacuum interconnect

## 9. SUMMARY

The design of the LHC vacuum system must satisfy a number of requirements, which depend not only on vacuum issues but also on considerations of both beam impedance and heat losses into the cryogenic system and/or the superconducting magnets. The pumping relies entirely on cryopumping of the 1.9 K cold bore surface, which offers a practically unlimited capacity for all gases except helium. The detailed design of the vacuum system is strongly influenced by the relatively large amount of synchrotron radiation, which represents a significant source of photon-stimulated outgassing, creates photoelectrons and is thus a driving mechanism for beam-induced multipacting. The beam induced heat load to the 1.9 K system has a large importance for the design and imposes by itself the presence of an actively-cooled beam screen. The intense proton bunches with their high peak

current lead to energetic ions which strike the walls and may cause ion-induced desorption and ion-induced pressure rise. Due to the cryogenic temperature of the walls, most molecular species are strongly cryosorbed and stick to the surface. This effect leads to a build-up of layers of cryosorbed gas, which in turn may desorb either thermally (vapour pressure effect) or by photon-stimulated desorption (re-cycling effect).

## ACKNOWLEDGEMENTS

The description of the LHC vacuum system presented here reflects the work of many colleagues in the LHC Vacuum Group. The important contributions, in particular from the groups LHC-CRI, EST-SM and SL-AP are also gratefully acknowledged.

## REFERENCES

[1] The LHC Study Group, The Large Hadron Collider, Conceptual Design, CERN/AC/95-05 (1995).

[2] L. Evans and Ph. Lebrun, Progress in Construction of the LHC, LHC Project Report 239 (1998).

[3] J.C. Brunet, et al., LHC-Project Report 280 and PAC-99, New York, (1999).

[4] A.G. Mathewson et al., EPAC-94, London, 1994.

[5] A.G. Mathewson et al., PAC-95, Dallas, 1995.

[6] O. Gröbner, Vacuum, 45, 767 (1995).

[7] O. Gröbner, 10th Int. Conf. On High Energy Accelerators, Protvino, July 1977.

[8] O. Gröbner, PAC-97, Vancouver 1997.

[9] E. Wallén, J. Vac. Sci. Technol. A 15(2), march/April 1997.

[10] V.V. Anashin, O.B. Malyshev, R. Calder and O. Gröbner, Vacuum 53, 269-272 (1999).

[11] C. Benvenuti, R. Calder and G. Passardi, J. Vac. Sci. Technol. 13, 6, Nov/Dec. (1976).

[12] R. Calder et al., J. Vac. Sci. Technol. A 14(4), Jul/Aug 1996.

[13] P. Redhead et al., The Physical Basis of Ultrahigh Vacuum, AIP (1993).

[14] O. Gröbner and R. Calder, IEEE Trans. Nucl. Sci., NS-20, 760, 1973.

[15] F. Rieke and W. Prepejchal, Phys. Rev. A 6, 1507 (1972).

[16] J.C. Barnard, I. Bojko and N. Hilleret, Vacuum 47, 4, 347 (1996).

[17] I.R. Collins, O. Gröbner, P. Lepeule and R. Veness, EPAC-98, Stockholm (1998).

[18] O. Gröbner, 10th Int. Conf. on High Energy Accelerators, Protvino, July 1977.

[19] F. Zimmermann, LHC Project Report 95, 27 February 1997.

[20] O. Gröbner, PAC-97, Vancouver, 1997.

[21] V. Baglin, I.R. Collins and O. Gröbner, EPAC-98, Stockholm, 1998.

[22] O. Brüning et al., PAC-99, New York (1999).

[23] A. Poncet et al., EPAC-94, London, 1994.

[24] P. Cruikshank et al., PAC-97, Vancouver, 1997.

[25] F. Ruggiero, R. Caspers and M. Morvillo, PAC-97, Vancouver, 1997.

[26] S. Sgobba and G. Hochoertler, International Congress Stainless Steel, to be published (1999).

[27] R. Veness et al., PAC-99, New York (1999).

# VACUUM SYSTEMS OF ELECTRON STORAGE RINGS

*D. Krämer*
BESSY, Berlin, Germany

**Abstract**
Designers of vacuum systems of modern high intensity electron storage rings have to carefully consider vacuum pressure effects as well as geometrical and electromagnetic constraints. This paper reviews the basic aspects of synchrotron radiation determining the dynamical vacuum pressure in the presence of beam. The processes of thermal and photon induced desorption are discussed so that a proper estimate of the gas load can be made. The relevant vacuum dependent loss mechanisms are summarised in order to calculate the beam-gas lifetime. Implicit geometrical aspects in the context of machine impedance are mentioned as they have a strong influence on beam stability.

## 1. INTRODUCTION

Beam lifetime and beam stability are of major importance to any storage ring and require a careful design of the vacuum system. The interaction of the stored particles with the molecules of the residual gas leads to particle losses, thus calculating the pressure to be expected is essential at the design stage.

The gas load and thus the resulting vacuum pressure is determined by thermal desorption and moreover by the dynamical gas load which is produced by synchrotron radiation falling on the chamber walls. The interaction of beam particles with the residual gas molecules, cause particle losses by elastic and inelastic scattering processes. The knowledge of these processes allows the beam-gas lifetime to be calculated. Nevertheless implicit requirements such as the chambers' impedances have to be taken into account by the designers to avoid beam instabilities caused by wake fields arising from the interaction of the beam self-field with the metallic vacuum-chamber boundaries.

## 2. SYNCHROTRON RADIATION ASPECTS IN ELECTRON STORAGE RINGS

It has been known for about 100 years that an accelerated "electric charge concentrated in a point" radiates energy [1]. Starting from Maxwell's equation of 1873, Liénard [1] and Larmor [2] worked out the power that is emitted by a moving charged particle at non-relativistic velocities. This energy loss of accelerated particles was of concern already at the early betatrons. [3, 4], but it took until 1947 until Elder and co-workers [5] saw a "small spot of brilliant white light" in the visible spectrum, radiated from the 70 MeV electron beam of their synchrotron. That's why this electromagnetic radiation is still called "Synchrotron Radiation" (SR).

Since the energy and the stored intensities of electron beams in storage rings have been raised to much higher values, SR has an enormous impact on the design of electron-ring accelerators. The power of the SR beams that hit the vacuum chamber in modern storage-ring-based synchrotron radiation light sources or B-meson factories can be as high as $kW \cdot m^{-1}$ causing thermal problems as well as severe desorption.

### 2.1 Basic properties of synchrotron radiation

Amongst earlier articles, Schwinger published the theory of synchrotron radiation in 1949 [6]. A more modern derivation of the theory is found in Ref. [7], where an analytical expression of the spectral and angular distribution of SR is given in terms of modified Bessel-functions $K_{n/m}$. For recent review articles with connection to accelerators see also Refs. [8, 9].

The relativistic γ-parameter of an electron with total energy $E$ is defined as

$$\gamma = \frac{E}{m_e c^2}$$

with $c$ the speed of light and $m_e$ = 511 keV/c$^2$ the electron rest mass. An electron (with charge $e$) moving in a circular accelerator will radiate a total energy $W$ per solid angle $d\Omega$ per unit frequency interval $d\omega$ per turn of

$$\frac{d^2W}{d\Omega\, d\omega} = \frac{e^2}{16\pi^3 \varepsilon_0\, c}\gamma^2 \left(\frac{\omega}{\omega_c}\right)^2 \left(1+\gamma^2\Psi^2\right)^2 \left[K_{2/3}^2(\xi) + \frac{\gamma^2\Psi^2}{1+\gamma^2\Psi^2} K_{1/3}^2(\xi)\right]. \qquad (1)$$

where $\varepsilon_0 = 8.86 \cdot 10^{-12}$ A·s·V$^{-1}$·m$^{-2}$ is the dielectric constant and the argument of the Bessel-function $\xi$ is

$$\xi = \frac{\omega}{\omega_c} \cdot \frac{(1+\gamma^2\Psi^2)^{3/2}}{2} \quad \text{with the 'critical' frequency defined as } \omega_c = \frac{3}{2} \cdot \frac{\gamma^3 c}{\rho}.$$

The terms in the square bracket of Eq. (1) describe the fact that SR is linearly polarised in the horizontal, and circularly polarised in the vertical plane. The angular divergence and thus the peak intensity of the radiated energy are functions of frequency $\omega$ and depend on the vertical observation angle $\psi$. Approximating the vertical divergence at the critical frequency $\omega = \omega_c$ by a Gaussian, the divergence (FWHM) is $\alpha \approx \gamma^{-1}$. So already at small electron energies, e.g. 500 MeV, where $\gamma \cong 1000$, the vertical opening angle of SR is of the order of 1 mrad. Thus SR is extremely well collimated with a very small vertical extension.

Integrating Eq. (1) over all frequencies, the total radiated energy per unit solid angle per electron per turn gives:

$$\frac{dW}{d\Omega} = \frac{7\, e^2 \gamma^5}{64\pi\, \varepsilon_0\, \rho} \cdot \frac{1}{\left(1+\gamma^2\Psi^2\right)^{5/2}} \left[1 + \frac{5\gamma^2\Psi^2}{7\,(1+\gamma^2\Psi^2)}\right]. \qquad (2)$$

This equation indicates that SR is of concern to electron rings rather than to current hadron machines; the $\gamma^4$-dependence of Eq. (2) results in a ~$10^{-13}$ smaller amount of SR power radiated by a proton beam (at the same energy and bending radius), as the mass ratio of proton and electron mass $m_p/m_e$ = 1836.

As energy is radiated in quanta (photons) with energies $u = h\omega/2\pi$, the distribution of the number of photons $n(u)$ emitted by a single electron to an unit energy interval $\Delta u$ per second is

$$n(u)\,\Delta u = \frac{dW}{d\omega}\,\frac{c}{2\pi\rho}\,\frac{1}{\hbar\omega}\,\Delta\omega$$

where $\hbar$ is Planck's constant and $\rho$ the electron's bending radius in the external magnetic field. Thus the flux of photons emitted per second by a single electron is:

$$\frac{dN}{dt} = \dot{N} = \int n(u)\, du = \frac{5}{2\sqrt{3}}\,\frac{e^2}{4\pi\varepsilon_0\hbar}\,\frac{\gamma}{\rho}\,. \qquad (3)$$

Equation (2) in practical units of beam energy $E$, magnetic bending field $B$ and circulating beam current $I$ for the peak power density in the horizontal plane ($\psi = 0$) becomes

$$\frac{dP}{d\Omega} = 5.42 \cdot E^4\,[\text{GeV}] \cdot B\,[\text{T}] \cdot I\,[\text{A}] \quad \text{W} \cdot \text{mrad}^{-2}$$

and integrated over all vertical angles the total power radiated by the beam into the horizontal plane is:

$$\frac{dP}{d\theta} = 4.42 \cdot E^3\,[\mathrm{GeV}] \cdot \mathrm{B}\,[\mathrm{T}] \cdot \mathrm{I}\,[\mathrm{A}] \quad \mathrm{W} \cdot \mathrm{mrad}^{-1}\,, \tag{4}$$

while from Eq. (3) the photon flux emitted into the horizontal plane is:

$$\frac{d\dot{N}}{d\theta} = 1.3 \cdot 10^{17} \cdot \mathrm{E}\,[\mathrm{GeV}] \cdot \mathrm{I}\,[\mathrm{A}] \quad \mathrm{photons} \cdot \mathrm{s}^{-1} \cdot \mathrm{mrad}^{-1}\,. \tag{5}$$

The power adsorbed by the vacuum chambers and the power distribution in the accelerator depends in detail on the vacuum chamber design but is easily estimated from geometrical considerations.

Technical problems arising from too high power loads on the chamber walls are:

- heating of the chamber walls enhancing thermal desorption;
- thermal expansion of the chambers causing current-dependent deflection of chambers and elements installed in the vacuum system, which may create position shifts of lattice elements and thus intolerable beam orbit drifts;
- photon induced desorption increasing the vacuum pressure due to a current-dependent gas load affecting the beam lifetime.

## 3. LINEAR POWER DENSITY

Equations (4) and (5) are the power and photon flux that are emitted by the beam whenever electrons are deflected in a magnetic field. In a dipole field $B$ = constant, the beam is bent in a circle with constant bending radius $\rho$. The power of SR, as well as the photon flux, are emitted tangentially to the beam path and will hit the vacuum chamber wall. Due to the small vertical dimension of the photon beam it is useful to define the 'linear power density' $p_{lin} = dP/dL$ as the vertically integrated power per unit length of chamber wall, Fig. 1. Taking small differences rather than differentials, the power radiated by the electron beam in a small angle segment $\Delta\theta$ is dissipated along a distance $\Delta L$ on the chamber walls. By this the linear power density distribution along the whole machine is:

$$p_{lin} = \frac{dP}{dL} \cong \frac{dP}{d\theta}\frac{\Delta\theta}{\Delta L}\,.$$

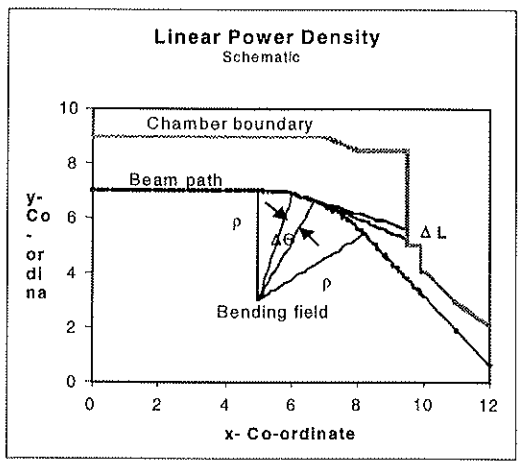



Fig. 1 Schematic of the concept of linear power density. The power $P$ radiated in the angle segment $\Delta\theta$ is spread over a distance $\Delta L$ at the vacuum system boundary. The linear power density $p_{lin}$ is the power absorbed by the chamber wall per unit length.

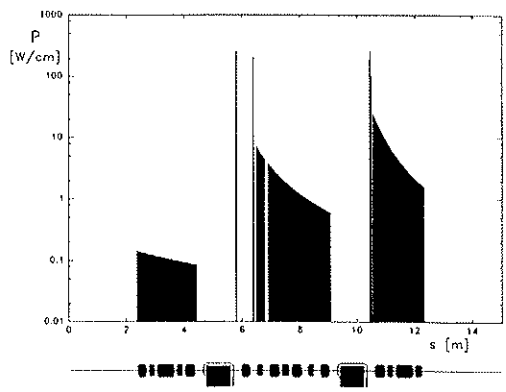



Fig. 2 Linear power density distribution along a unit cell of the BESSY II synchrotron light source (1/16 of the ring circumference). Near the exit of the bending magnets the power density is as high as 250 W/cm. Input data: energy 1.9 GeV, stored current 500 mA, bending field 1.45 T. The minor contribution of multipole magnets is neglected in the calculation. The magnetic lattice of the ring is sketched in the lower part of the graph.

It is obvious that by simple means, e.g. avoiding surfaces perpendicular to the incoming photon beam, the peak power density can be decreased by distributing the power over a larger distance $\Delta L$. Figure 2 shows a calculation of the linear power density for the 1.9 GeV BESSY II SR light source. For a stored current of 500 mA the linear power density ranges from 250 W/cm near the exit of the bending magnets to less than 1 W/cm in the straight sections. Especially at the location of high power load, a careful analysis of the expected temperatures and thermal deflection of components has to be done (using standard FEM programs) to avoid damage to the hardware components.

The same concept is used to derive the linear photon flux density $n_{lin} = \mathrm{d}(\mathrm{d}N/\mathrm{d}t)/\cdot\mathrm{d}L$ [photons·s$^{-1}$·cm$^{-1}$] on the chamber walls using Eq. (5):

$$n_{lin} = \frac{\mathrm{d}\dot{N}}{\mathrm{d}\theta}\frac{\Delta\theta}{\Delta L}. \tag{6}$$

## 4. THERMAL- AND PHOTON-INDUCED DESORPTION

For a vacuum system without beam, the gas originates from a variety of processes as permeation, diffusion, desorption/re-adsorption and the material's vapour pressure. For electron rings it is normally sufficient to consider the thermal desorption of molecules bound to the chamber-wall surfaces only. In the presence of electron beams the photon-induced desorption by synchrotron radiation gives a significant contribution to the vacuum pressure.

### 4.1 Thermal desorption

The surfaces of the vacuum vessels are always covered with some layers of molecules which are adsorbed at binding energies in the range of eV. In consequence continuous gas loads $Q_i$ consisting of gas components '$i$' like $H_2$, $CH_4$, $H_2O$, CO, $CO_2$ are desorbing, giving a total gas load $Q_o$:

$$Q_0 = \sum_i Q_i$$

The desorption rates $q_i$ [mbar·l·s$^{-1}$·cm$^{-2}$], often also termed specific outgassing rate, are the gas loads per surface area $A$: $q_i = Q_i/A$. These rates cannot be calculated by general means, as they are a property of the material and depend on its history, e.g. the way the surfaces have been cleaned [10].

According to Frenkel's equation, the gas load $Q_i$ depends on temperature $T$. $Q_o$ is the sum of all the desorbing components $i$, at a surface coverage $N_i$ [molecules·cm$^{-2}$] and binding energy $E_i$:

$$Q_0 \propto \sum_i N_i \, e^{-\frac{E_i}{kT}} .$$

Baking the vacuum system increases the temperature $T$ and enhances desorption. For clean stainless steel-samples specific outgassing rates of $10^{-12}$ mbar·l·s$^{-1}$·cm$^{-2}$ are achieved routinely after in-situ bakeout at ~300$^0$ C. In unbaked systems the outgassing rate is typically a factor of 5 to 10 higher.

Let us assume that the specific outgassing rate as function of time $q_i(t)$ is proportional to the total amount of gas $q_{o,i}$ adsorbed on the surface

$$q_i(t) = -\,\mathrm{d}q_{o,i}/\mathrm{d}t .$$

When the surface coverage is considered to change by desorption only

$$N(t) = \sum [N_i(t = 0) - (\int q_i \, \mathrm{d}t)]$$

the gas load $Q_i$ will decrease exponentially with respect to time at a characteristic time-constant $\tau_i$ for each gas species of a Frenkel type relation:

$$\tau_i = \tau_o \exp(-E_i/kT) .$$

When assuming that for a baked and an unbaked system the binding energies are the same (since the relative composition of adsorbates stays unchanged – which definitely is an over-simplification) it takes a time of approximately $2.3\tau_b$ for an unbaked system to arrive the gas load $Q_b$ of a baked one (and thus at a vacuum pressure $P_b = Q_b/S$). Here $\tau_b$ is the time it needs to reduce $Q_0$ by a factor of $1/e$ in the baked system. So, for a thermal desorption dominated vacuum system bake-out is useful to achieve low pressures in a short time.

### 4.2 Photon-induced desorption

In the presence of SR the balance of outgassing is significantly changed. Synchrotron radiation on one hand increases the temperature of the vacuum chamber, therefore increasing thermal desorption. On the other hand the SR photons hitting the vacuum chamber walls create photoelectrons which scatter on the walls, causing adsorbed molecules to desorb [11]. Thus the total gas load has to be considered as the sum of thermal and photon induced desorption

$$Q = Q_0 + \eta \dot{N} k,$$

where $Q_0$ is the thermal gas load and $\eta$ the desorption yield, e.g. the number of molecules released per incident photon, $k$ is the conversion factor from molecules to Torr l ($k = 3.1 \cdot 10^{-20}$ Torr·l·molecules$^{-1}$ at 25 °C) and $dN/dt$ the photon flux. According to Eq. (6) the photon flux is a function of position in the machine.

A variety of measurements to determine the desorption yield $\eta$ have been published, [12–14]. The desorption yield $\eta$ is decreasing with time or, to be more precise, with the accumulated photon flux, also named "photon dose". The desorping molecules follow a dependence:

$$\eta = \eta_0 D^{-\alpha},$$

where $\eta_0$ is the initial desorption rate of the order of $10^{-2}$ [molecules·photon$^{-1}$] for most gas species such as $H_2$, CO and $CO_2$ from standard materials such as stainless steel, copper and aluminium. It is obvious that $\eta_0$ is a quantity that depends on the cleaning procedure. It also depends on the history of the material. $\alpha$ determines the dependence on photon dose $D$ (e.g. the integrated photon flux [photons] or integrated electron current with respect to time [A·h] using Eqs. (4) and (5), respectively). Typical values of $\alpha$ are 0.8 to 1.2. Water, which is loosely bound by van der Waals forces, behaves somewhat differently in this context [14].

Figure 3 displays a typical measurement result of the desorption yield for an OFHC-copper surface, [15]. Starting at $\eta_0 \cong 10^{-2}$ at a dose of some $10^{20}$ photons, the yield is rapidly decreases to values below $10^{-6}$ after a dose of $10^{24}$ photons, equivalent to an integrated beam current of ~50Ah — as $10^{18}$ photons·s$^{-1}$ were hitting the sample. This well-known effect, often called "beam scrubbing", is a powerful cleaning method for vacuum surfaces in electron machines.

Calculation of the desorption distribution thus has to reflect the desorped gas load $Q(x,\eta(D))$ which is dependent on the photon flux at location $x$ as well as on the desorption yield which itself depends on the accumulated beam dose $D$. In order to determine the mean gas pressure, we have to decide on the pumping scheme, the location of pumps and their pumping speed to derive the pressure distribution.

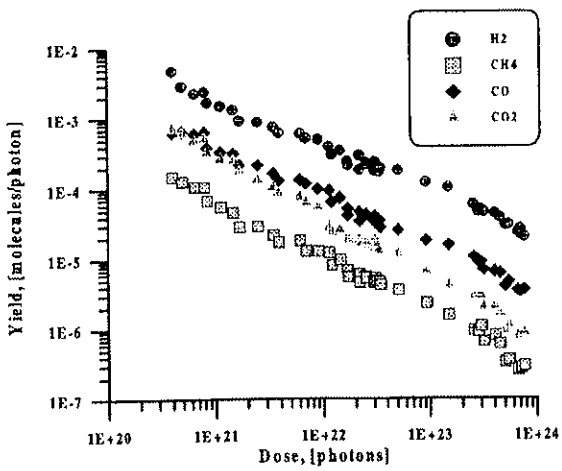



Fig. 3 Desoprtion yield vs. integrated photon dose, from Ref. [15].

## 5. PUMPING SCHEME & PRESSURE DISTRIBUTION

The magnetic elements of the accelerator set important geometrical boundaries to the layout of the vacuum system. The available apertures in the multipole magnets determine the cross section of the chambers. At the same time the flexibility for optimised positioning of vacuum pumps is limited by the presence of magnets and other elements. Thus, there is no simple way to optimise the vacuum system. Nevertheless one can derive a simple model of the vacuum system [10,16] to obtain a rough insight into the performance of the pumping system. As, indeed, major parts of the vacuum chambers do not change their shape, we consider a beam tube of constant elliptical cross section with pumps at constant distance $L$ from each other. The gas flow $Q$ [mbar·l·s$^{-1}$] gives rise to a pressure difference d$P$ due to the specific conductance w [m·l·s$^{-1}$] of the pipe. Along the direction $x$ of the pipe in linear approximation the gas load is

$$Q(x) = -\mathrm{w}\frac{\mathrm{d}P}{\mathrm{d}x} .$$

Assuming that the specific gas load $q$ [mbar·l s$^{-1}$·m$^{-2}$] per specific surface area $A$ [m] stays constant and the total surface of the pipe of length $L$ is $F = AL$ [m$^2$], then

$$\frac{\mathrm{d}Q}{\mathrm{d}x} = Aq .$$

Combining the two equations gives:

$$w\frac{\mathrm{d}^2P}{\mathrm{d}x^2} = -Aq . \tag{7}$$

With the boundary conditions that the pressure is equal to $AqL/S_0$ at the location of vacuum pumps with pumping speed $S_0$ and that due to symmetry of the problem there is a maximum at $x = L/2$:

$$\frac{\mathrm{d}P}{\mathrm{d}x}\Big|_{x=L/2} = 0 ,$$

a parabolic pressure dependence follows with the average pressure:

$$\langle P\rangle = Aq\left(\frac{L^2}{12w} + \frac{L}{S_0}\right) .$$

For an elliptical pipe of length $L$ [cm] and semi-axes $a$ [cm] and $b$ [cm], the conductance $C = w/L$ [l· s$^{-1}$] for a gas molecule of mass $M$ [amu] at temperature $T$ [K] is given by the relation 10]:

$$C = 431\cdot\left[\frac{a^2b^2}{L}\right]\cdot\frac{1}{\sqrt{a^2+b^2}}\cdot\sqrt{\frac{T}{M}} \qquad \mathrm{l\cdot s^{-1}}$$

Defining an effective pumping speed $S_{\mathrm{eff}}$, the average pressure in the section is given by:

$$\langle P\rangle = \frac{AqL}{S_{\mathrm{eff}}}$$

with

$$S_{\mathrm{eff}} = \left(\frac{L}{12w} + \frac{1}{S_0}\right)^{-1}$$

For conductance limited vacuum chambers as in the straight sections of the BESSY II storage ring the specific conductance is $w$ = 20 m·l·s$^{-1}$ for $N_2$ ($M$ = 28). The effective pumping speed as function of distance $L$ and pumping speed $S_0$ is plotted in Fig. 4. The graph immediately shows that

the optimum solution is to place as many small pumps as closely together as possible when lumped pumps are used. Irrespective of the pumping speed of the pumps, the maximum effective pumping speed is limited to12$w/L$. Realistic distances $L$ are about 2 to 3 m demanding pumps of typically less than 60 l/s pumping speed, yielding $S_{eff} \approx 20$ l/s. On the other hand the effectiveness of linearly distributed pumps all along the machine circumference is obvious as $L$ approaches zero, in this concept the maximum pumping speed $S_0$ is available at the vacuum chamber.

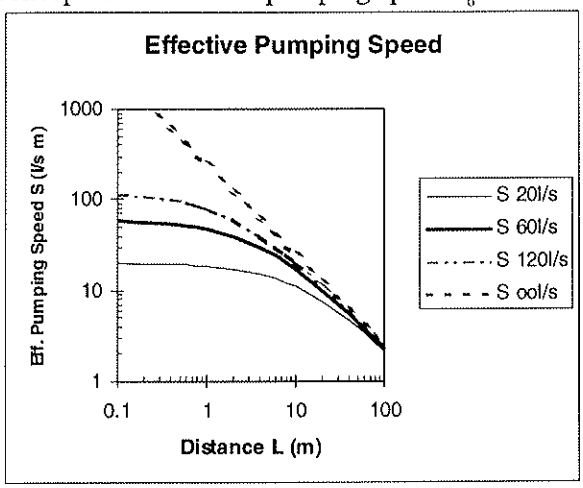



Fig. 4 Effective pumping speed of a conductance limited chamber ($w$ = 20 m·l·s$^{-1}$) as function of distance $L$. The curves are for different values of the pumping speed $S_0$ (20, 60, 120 l·s$^{-1}$ and infinite pumping speed).

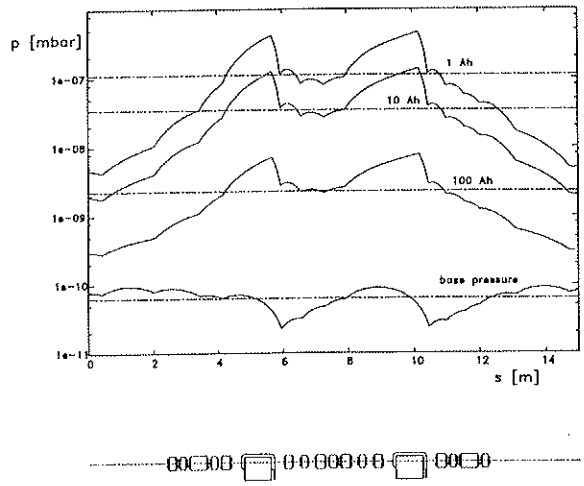



Fig. 5 Pressure distribution of a 1.9 GeV, 500 mA beam in the BESSY II storage ring as expected after accumulation of various beam doses (1, 10 and 100 A·h ) as well as the base pressure without beam. The magnetic lattice of BESSY II is sketched to scale (lower part).

In a more realistic model the assumption of constant outgassing rate has to be replaced by the position dependent $q(x)$, while the varying conductances $w(x)$ have to be considered correctly. Many different computer codes have been developed to evaluate the problem Eq. (7), ranging from spreadsheet-based macros [17], up to Monte-Carlo simulations [18, 19]. Thus the pressure profile can be approximated piecewise and/or evaluated directly taking into account the dose-dependent desorption rates. Figure 5 shows a calculation that was performed for the BESSY II storage ring. The pressure distribution is plotted for different desorption conditions (accumulated doses of 1, 10 and 100 A·h) and for thermal outgassing, i.e. without beam. To achieve the design pressure of $\langle P \rangle = 10^{-9}$ mbar, an integrated dose of ~ 200 A·h is required.

## 6. SINGLE-PARTICLE LOSS MECHANISMS

Once the vacuum pressure and the gas composition are known, scattering of the stored electrons with residual gas molecules will cause particle losses from which we will calculate the beam-gas lifetime. As the beam particles are considered to be independent from each other, beam-gas interactions are considered as incoherent effects that are treated statistically.

There are two different interactions to consider:

- Elastic scattering which leads to a transversal deflection of the beam particle. After the collision the electrons will start to perform betatron oscillations around the closed orbit. With the assumption that the beam-stay-clear aperture of the vacuum chamber is of the same size as the dynamical aperture, the particle is lost when the amplitude of the orbit oscillation exceeds the mechanical limit (mostly the vertical chamber height).
- Inelastic scattering, where a photon is emitted in the collision, changes the beam particles' energy. If the dispersive orbit exceeds the mechanical (normally the horizontal) aperture, or if the rf-acceptance limit is exceeded, the particle is lost [20, 21].

### 6.1 Coulomb scattering

Scattering of an idealised mono-energetic electron beam in the Coulomb field of the nucleus of a residual-gas atom of nuclear charge $Z$ leads to an angular deflection, depending on the impact parameter $b$, i.e. the shortest distance between the incoming particle trajectory and the point-like target. Taking into account the screening of the Coulomb field by electrons from the target-atom which is important for particles at large $b$ (equivalent to beam particles scattered into small angles) and modifying for the finite size of the nucleus, the differential cross section of the scattering process can be found in many text books. [7]. Integrating this cross section from the minimum angle $\theta_{min}$ for which a particle gets lost and maximum angle $\theta_{max} \equiv \pi$ gives:

$$\sigma = \frac{2\pi\ Z^2\ r_e^2}{\gamma^2} \frac{1}{\Theta_0^2}$$

As a consequence of the angle deflection $\theta_0$ the particle performs betatron oscillations with maximum amplitude

$$y_{\max} = \frac{1}{2\sin(\pi Q)} \cdot \sqrt{\beta(s)} \cdot \sqrt{\beta_i} \cdot \Theta_0 \ .$$

The particle hits the vacuum wall if $y \geq a$. $\beta_i$ and $\beta(s)$ denote the horizontal or vertical amplitude function of the ring at the location $i$ where the scattering takes place and $\beta(s)$ somewhere in the machine, while $a$ represents the horizontal or vertical half aperture of the chamber.

Averaging over all possible locations $i$ (replacement of $\beta_i$ by the mean beta value $<\beta>$) the cross section due to Coulomb scattering is [20, 22]:

$$\sigma_C = \frac{2\pi Z^2 r_e^2}{\gamma^2} \cdot \frac{\langle\beta\rangle \beta_{\max}}{a^2}$$

It is evident that for increasing energy this loss mechanism becomes less important.

### 6.2 Bremsstrahlung

Inelastic scattering of an electron off a nucleus causing bremsstrahlung leads to an energy loss of the circulating particle. The electron is lost when the energy deviation exceeds the rf acceptance $\varepsilon_{rf}$ of the ring. The total cross section for particle loss is [22, 23]:

$$\sigma_B = \frac{4 r_e^2 Z^2 e^2}{h\,c} \cdot \frac{4}{3} \ln \frac{183}{Z^{1/3}} \left( \ln \frac{1}{\varepsilon_{rf}} - \frac{5}{8} \right).$$

### 6.3 Elastic e-e scattering

Similar to the elastic scattering of electrons off the nucleus, the beam particles also scatter with the electrons of the atoms of the residual gas. In this process the beam electrons transfer part of their energy to the gas atom. Again if the beam particles' energy loss exceeds the rf acceptance limit, the particle is lost. The total cross section for the process is [20, 21]:

$$\sigma_{el,ee} = \frac{2\pi\ r_e^2\ Z}{\gamma} \frac{1}{\varepsilon_{rf}}$$

### 6.4 Inelastic e-e scattering

In inelastic scattering of the beam electrons with the electrons of the residual gas atoms, a photon is emitted, which carries away part of the electron energy. Particle loss occurs when the energy acceptance is exceeded. The loss cross section is given by [20, 21]:

$$\sigma_{inel,ee} = \frac{4\, r_e^2\, Z\, e^2}{h\, c} \cdot \frac{4}{3} \left[ \ln \frac{5\gamma}{2\,\varepsilon_{rf}} - 1.4 \right] \cdot \left[ \ln \frac{1}{\varepsilon_{rf}} - \frac{5}{8} \right].$$

### 6.5 The beam-gas lifetime

The number of beam particles $N$ divided by those lost per time interval $-\mathrm{d}N/\mathrm{d}t$ is the $1/e$-beam lifetime $\tau$. The relative losses equal the cross section $\sigma_i$ times the density of target atoms $n_i$ (the index $i$ corresponds to the different molecules $j$ in the residual gas composed of $k_{ij}$ atoms of type $i$ with nuclear charge of $Z_i$). The beam travels with the speed of light $c$ so the beam experiences a density of these target atoms per second of $\sum n_i c$. Thus the inverse beam lifetime is given by summation of the above cross sections:

$$\frac{1}{\tau} = -\frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}t} = c \sum_{j,i} k_{ij} n_i \left\{ \sigma_{C,i} + \sigma_{B,i} + \sigma_{el,ee,i} + \sigma_{inel,ee,i} \right\}$$

The densities $n_i$ of atoms $i$ from molecules $j$ are related to their vacuum pressure $P_i$ via Boltzmann's constant $k$ and the temperature $T$:

$$n_i = \frac{P_i}{kT}$$

At room temperature this relation in practical units is: $n_i\,[\mathrm{m}^{-3}] = 3.217 \cdot 10^{22}\, P_i\,[\mathrm{Torr}]$.

Comparing measured and calculated beam lifetimes as a function of the circulating beam intensity show good agreement at low currents, Fig. 6. For high intensity beam, i.e. high bunch currents, especially if only a few bunches are circulating discrepancies are visible. The underlying process is scattering of electrons amongst each other inside the bunch. This effect was identified first in the small Frascati ADA ring and explained by Touschek [24]. Though the "Touschek effect" is not related to the vacuum pressure at all, it is discussed in some detail here as the horizontal aperture of the vacuum chamber is affected.

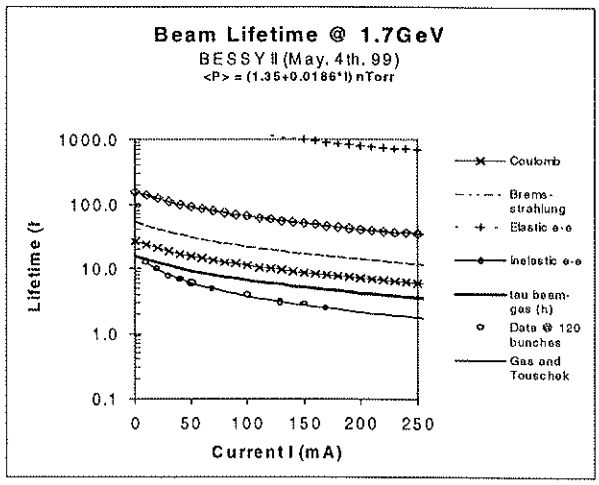



Fig. 6 Beam lifetime for the various loss mechanism as a function of the beam current in BESSY II. The calculated total beam-gas lifetime is the solid line. Solid dots represent measured data when running the current in 120 bunches. The difference of total beam-gas lifetime and measured data is due to the Touschek effect. Taking this effect into account, the measured beam lifetimes are precisely described.

### 6.6 Touschek effect

Low emittance synchrotron light sources as well as B-meson factories operate at high electron densities in the circulating bunches. The probability that electrons inside the same bunch scatter off each other is proportional to the particle density in the bunch. Thus current dependent particle losses further decrease beam lifetime with respect to beam-gas lifetime.

The scattering of particles within a bunch is called the Touschek effect. If the energy transfer in a collision is large enough the particle is lost if the momentum deviation exceeds the momentum acceptance of the machine. Furthermore if the scattering takes place in a dispersive region the amplitudes of the resulting horizontal betatron oscillations may exceed the geometrical vacuum chamber aperture. Touschek beam lifetime [20, 24, 26] is not related to the vacuum pressure but the vacuum chamber design may influence it via pure geometrical parameters. The inverse Touschek lifetime is given by:

$$\frac{1}{\tau} = \frac{N\, r_e^2\, c}{8\pi\, \sigma_x\, \sigma_z\, \sigma_s\, \gamma^2\, \varepsilon_{rf}^3} \cdot D(\varepsilon)$$

with

$$\varepsilon = \frac{\varepsilon_{rf}^2}{2\gamma^2} \frac{\beta_x^2}{\sigma_x}$$

for a flat beam and the $D$-function defined by

$$D(\varepsilon) = \sqrt{\varepsilon} \left\langle -\frac{3}{2} e^{-\varepsilon} + \frac{\varepsilon}{2} \int_{\varepsilon}^{\infty} \frac{\ln u}{u\, e^{u}} du + \frac{1}{2}(3\varepsilon - \varepsilon \ln \varepsilon + 2) \cdot \left\{ \int_{\varepsilon}^{\infty} \frac{e^{-u}}{u} du \right\} \right\rangle$$

It may be worthwhile to point out that the Touschek loss rate depends on the square of the number of electrons in a bunch as two electrons are involved in the collision, $dN/dt = -aN^2$. Therefore the total beam lifetime does not show an exponential decay with time $t$ as in case of the beam-gas events. The development of current $I(t)$ with time is:

$$I(t) = \frac{I_0}{1 + I_0 \cdot a \cdot t}.$$

Depending on the machine momentum acceptance, Touschek-scattered electrons require a certain minimum beam-stay-clear aperture. As an example Fig. 7 shows the theoretical predicted Touschek lifetime at BESSY II as a function of horizontal chamber half-apertures at different momentum acceptance $\Delta p/p$ of the ring. The Monte-Carlo approach [27] is in excellent agreement with the analytical approach used in the ZAP program [28]. Thus the vacuum chamber design has to provide an aperture wide enough to allow safe operation.

Especially for low energies and high intensity storage rings one requires a more detailed analysis of multiple-small-angle Coulomb scattering (multiple Touschek scattering) [29]. The effect of intra-beam scattering has been seen in hadron and heavy-ion machines especially when phase-space-cooling techniques are used [30]. However, little time has been spent on this problem in the case of electron storage rings.

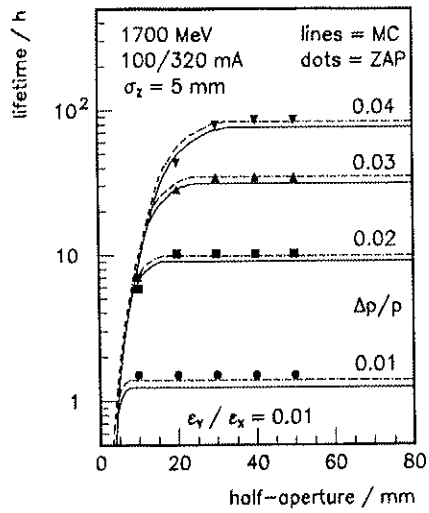



Fig. 7 Calculated Touschek lifetime for the BESSY II Synchrotron Light Source as a function of horizontal aperture and machine-momentum acceptance $\Delta p/p$. A 100 mA beam spread in 320 bunches with bunch length $\sigma$ = 5 mm is assumed. The dots are lifetimes obtained using the program ZAP, the solid (dashed) lines are relativistic (non-relativistic) Monte Carlo results [27].

## 7. IMPLICIT REQUIREMENTS FOR THE VACUUM SYSTEM

The beam particles generate an electromagnetic "self"-field according to Maxwell's equations. The boundary conditions that are given by the geometry and electromagnetic properties of the vacuum chamber materials strongly influence the solution of the differential equations. An analytical solution of Maxwell's equation taking into account all details of an actual machine environment is definitely impossible. Nevertheless, in order to get information on instabilities generated in the beam in the presence of the electromagnetic fields, the concept of impedances was introduced. In analogy to Ohm's law, longitudinal and transversal impedances $Z_{||}$ and $Z_{\perp}$ are defined [31, 32] by:

$$\left[\vec{E}_{//} + \vec{v} \times \vec{B}\right]_{//}(t,\Theta) = -\frac{1}{2\pi R} \int_{\omega=-\infty}^{\omega=\infty} Z_{//}\, S_{//}(\omega,\Theta)\, e^{j} \quad d\omega$$

and

$$\left[\vec{E} + \vec{v} \times \vec{B}\right]_{T}(t,\Theta) = -\frac{j\beta_0}{2\pi R} \int_{\omega=-\infty}^{\omega=\infty} Z_{\perp}\, S_{T}(\omega,\Theta)\, e^{j\ t}\, d\omega$$

where $S_{//}$ is the beam current and $S_T$ the beam signal while E and B are the electrical field and induction, respectively. The impedances are considered as the origin of longitudinal and transversal beam instabilities. The impedances are complex function of frequency $\omega$.

It is useful to distinguish between single-bunch and multi-bunch operation in the accelerator. Single bunches are affected, with respect to stability, by broadband impedances associated with short-range wake fields. The longitudinal broadband impedance may cause bunch lengthening, while the transverse impedance allows the current-dependent growth rates of instabilities (i.e. head tail and strong head tail instability) to be derived. If these growth rates are faster than the "natural" damping times due to the emission of SR, the beam will be unstable.

Similarly, when trains of bunches are circulating in the ring (multi-bunch operation), the long-range wake field and their associated longitudinal and transverse impedances can generate longitudinal and/or transverse coupled-bunch oscillations (instabilities). Only in very special cases can analytical formulae for the impedance be derived, i.e. the resistive-wall impedance of a long circular tube.

### 7.1 Resistive-wall impedance

The finite resistivity of the vacuum chamber is the origin of a longitudinal and transversal impedance (restive-wall impedance). For a long circular pipe of radius $b$ the impedances are given by the analytical expressions [32, 33]:

$$Z_{\perp}(\omega) = \frac{(1-i)}{b^3} \cdot Z_0\, R\, \delta$$

and

$$\frac{Z_{//}(\omega)}{n} = \frac{(1-i)}{2b} \cdot Z_0\, \delta$$

with the chamber length $C = 2\pi \cdot R$ equal to the machine circumference. $\rho$ is the resistivity of the wall material with skin depth $\delta$ given by:

$$\delta = \sqrt{\frac{2\rho}{\omega \cdot \mu_0}}\ .$$

where $\mu_0 = 1.25 \cdot 10^{-6}$ Vs $A^{-1}$ $m^{-2}$ is the permeability of the vacuum, and $Z_0 = \mu_0 \cdot c = 377\ \Omega$ the impedance of vacuum. As it is common to use $Z/n$ rather than $Z$, we also will use the relation:

$$\frac{Z(\omega)}{n} = Z(\omega) \cdot \frac{\omega_0}{\omega},$$

where $\omega_0 = 2 \cdot \pi \cdot f_0$ with $f_0$ the revolution frequency of the beam in the ring.

The longitudinal impedance gives rise to a growth rate of longitudinal coupled bunch modes for a beam of $n$ bunches and bunch-current $I$. If the inverse growth rate is larger than the natural damping time, the beam will be damped, otherwise the bunches start longitudinal oscillations.

Transverse coupled bunch oscillations are excited when the growth rate is smaller than the transverse radiation damping time. Fortunately – as the resistive-wall growth rate depends on chromaticity and working point of the machine – changes to the excitation of sextupole magnets in the ring or selection of a new tune can help to damp the instability. Nevertheless the $b^{-3}$ scaling of the transverse impedance is a problem for modern SR light sources where a considerable fraction of the machine circumference is equipped with narrow vertical chambers for small gap undulators.

## 7.2 Impedance of the vacuum system in general

Designers of vacuum chambers are confronted with the problem of including into the machine various items such as bellows, flanges, pumping ports and tapers for changes of the chamber cross section. To evaluate their contribution to the machine impedance each item needs to be considered separately, using well known codes such as MAFIA [34] and TBCI [35]. In the time domain the wake function $W(t)$ generated by the bunch with charge distribution $I(t)$ is calculated from Maxwell's equations. The coupling impedance $Z(\omega)$ then is the complex power spectrum of the wake function normalised to the charge distribution in the frequency domain:

$$Z(\omega) = \frac{\int_{-\infty}^{\infty} W(t)\, e^{-i\omega t}\, \mathrm{d}t}{I(\omega)}.$$

Figure 8 shows an example using a Gaussian charge distribution $I(\omega) = q_1 \exp[-{}^1/_2(\omega\tau)^2]$ with the resulting wake function [36]. The geometry under consideration in the example is that of one of the 112 beam position monitors in the BESSY II storage ring. They give a significant contribution to the longitudinal impedance with a prominent resonance at $\omega = 45$ GHz.

Summing up all impedances at BESSY II gave a total longitudinal coupling impedance nearly constant over the frequency range $\omega = 0 - 140$ GHz for the BESSY II ring with $|Z/n| \cong 0.3\ \Omega$, Fig. 9.

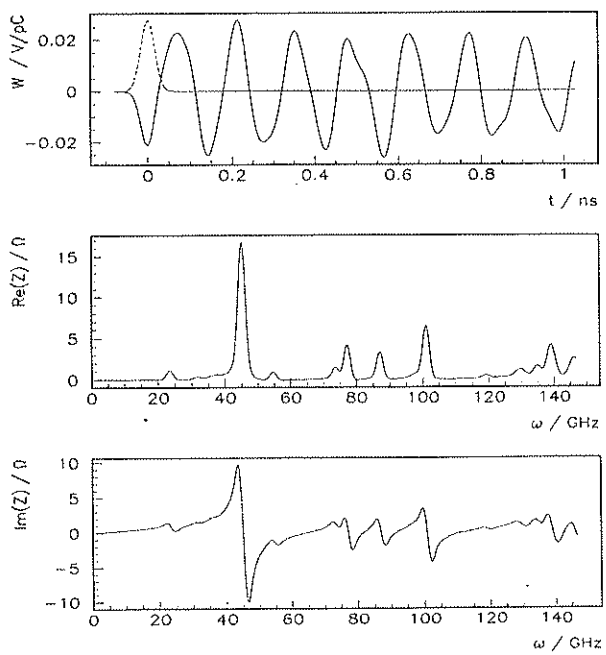



Fig. 8 Upper part: Beam charge distribution $i(t)$ and wake function $W(t)$ caused by a beam position monitor as function of time. The two lower graphs show the real and imaginary part of the longitudinal impedance [36].

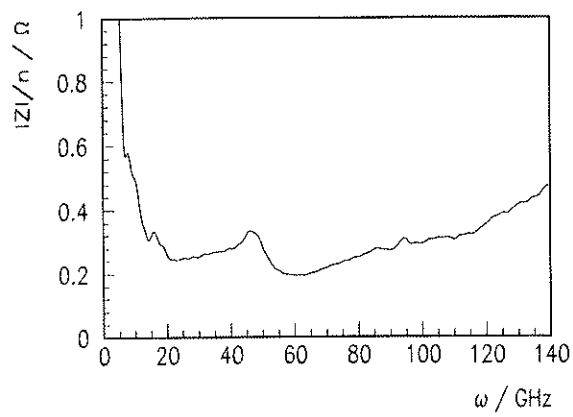



Fig. 9 Calculation of the longitudinal coupling impedance for the BESSY II storage ring [36].

## 8. ACCELERATOR VACUUM SYSTEMS IN ROUTINE OPERATION

Already during accelerator commissioning any error in the design and production of the vacuum hardware will show up. Effects range from incorrect vacuum gauge pressure readings, due to the sensors being installed too close to the fringe field of magnets, to enhanced desorption rates of heavy contaminants caused by inappropriate cleaning methods. Obstacles in the chambers as a result of manufacturing or cleaning errors as well as melted rf liners in bellows sometimes reduce the vacuum chamber aperture to the point that no beam can be injected. In such cases, the vacuum system has to be re-opened to cure the problem. Macroscopic particles such as "dust" degrade the beam lifetime in some electron machines in an unpredictable way. Finding the origin of the problem and its cure is a very difficult job.

## 9. CONCLUSION

There is a wide range of technical details that have to be considered by the vacuum designers. Whatever solutions are selected they have to allow correct operation of the machine. Thus vacuum design of today's accelerators is a challenging task.

## ACKNOWLEDGEMENT

L. Schulz*, J. Feikes and S. Khan from BESSY are thanked for helpful suggestions, comments and providing figures for the preparation of this paper.

(* present address: PSI, Villigen, Switzerland)

## REFERENCES

[1] A. Liénard, L'Éclairage Électrique, 16 (1898) 5.

[2] J. Larmor, Phil. Mag. 5 (1897) 503.

[3] D. Iwanenko et. al., Phys. Rev. 65 (1944) 343.

[4] J.P. Blewett, Phys. Rev. 69 (1946) 87.

[5] F.R. Elder, Phys. Rev. 71 (1947) 829.

[6] J. Schwinger, Phys. Rev. 75 (1949) 1912.

[7] J.D. Jackson, Classical Electrodynamics, John Wiley & Sons, New York (1962).

[8] K. Hübner, CAS 1989, CERN 90-03 (1990) 24.

[9] R.P. Walker in CAS 1992, CERN 94-01 (1994) 437.

[10] A.G. Mathewson, CAS 1992, CERN 94-01 (1994) 717.

[11] O. Gröbner et al., EPAC, Berlin (1992) 132.

[12] O. Gröbner et al., Vacuum 33 (1983) 397.

[13] H.J. Halama et al., Vacuum 42 (3) (1991) 185.

[14] C. Herbeaux et al., LURE RT/97-03 (1997).

[15] V. Anashin et al, EPAC 98 (Stockholm) 2163.

[16] A.I. Ivanenko et al., EPAC London (1994) 2523.

[17] M. Nordby, EPAC, London (1994) 2515.

[18] A. Poncet, CERN, program Pressure Distribution 2.β; http://www.cern.ch.

[19] R. Kersevan, Cornell University, program MOLFLOW.

[20] J. LeDuff, Nucl. Instr. Meth. A239 (1985) 83.

[21] A. Wrulich, CAS 1992, CERN 94-01 (1994) 409.

[22] J. Haissinski, Thesis, LAL, Orsay (1965).

[23] H.A. Bethe et. al., Proc. R. Soc. A146 (1934) 83.

[24] C. Bernardini et. al., Phys. Rev. Lett. 10 (1963).

[25] H. Bruck, Accélérateurs Circulaires de Particles, Presse Universitaires de France, Paris (1966) and translation into English LANL, LA-TR-72-10 Rev. (1972).

[26] H. Bizek, PAC, Vancouver (1997) 1496.

[27] S. Khan, EPAC, London (1994) 1192.

[28] M.S. Zisman et. al., ZAP Users's Manual, LBL-21270 UC-28 (1986).

[29] A. Piwinski, Proc. of the Int. Conf. on High Energy Acc. (1974) 105.

[30] M. Martini, CERN PS/AA/84-7 and CERN PS/AA/84-9 (1984).

[31] A. Chao, Physics of Collective Beam Instabilities in High Energy Accelerators, John Wiley & Sons, New York (1993).

[32] J.L. Laclare, CAS 1992, CERN 94-01 (1994) 385.

[33] N. Nakamura et. al., PAC, Vancouver (1997) 1756.

[34] MAFIA User's Guide, CST, Darmstadt.

[35] T. Weiland, IEEE Trans. Nucl. Sci., NS-32 (1985) 2738.

[36] S. Khan, PAC, Vancouver (1997) 1703.

# QUALITY ISSUES IN VACUUM

*R.J. Reid*
CLRC Daresbury Laboratory, Warrington, UK

**Abstract**
This paper takes a brief and somewhat subjective look at some of the issues which arise when one begins to think about the concept of "Quality" as applied particularly to vacuum system specification.

## 1. WHAT IS "QUALITY"?

It is perhaps unfortunate that when the word "quality" is used in relation to some object, component, system or whatever, it is often assumed to mean that it is of the highest possible standard or that it is in some way state-of-the-art. The word therefore carries connotations of great care, superb workmanship and great expense. It is in some way used to convey a sense of "goodness".

However, to use the word in this way is misleading. A much more appropriate definition of quality is fitness for purpose. In other words, if a job is a quality job then it will fulfil what it is meant to fulfil but it will not exceed those requirements. The standard of the job will be sufficient and not excessively high.

Quality in the sense that is going to be assumed from now on means meeting the needs and expectations of the user, that is fitness for the purpose of the user. Difficulty arises in that users may not know what they actually require. A user will often tend to over-specify just to be safe, but is then unwilling to pay the necessary costs. It is therefore imperative that users' requirements are carefully checked by system designers.

## 2. QUALITY SYSTEMS

It is salutary to note that most organizations and commercial undertakings now work under so-called Quality Systems such as, for example, ISO 9001 but that in the main, these systems actually say nothing about quality in reference to standards of work *per se*. They are primarily documentation and tracing systems which will detail how a job has been handled, although, where appropriate, they will at some stage refer to the adherence of a job to its technical specification.

Quality systems are, as stated, primarily documentation systems and indeed usually involve a great deal of documentation. In the case of a manufacturing enterprise for example, all processes – handling, working, treatment, packing, etc. – from goods inward to dispatch are spelt out in detail and the appropriate tracing documentation and certification required at each stage laid down.

Such documentation systems have a great potential for becoming fossilized, but when used properly they provide the necessary information required to carry out audits of the process, to help the system to evolve by, for example, reducing costs without sacrificing standards, and to improve in some way the whole process.

## 3. VACUUM SYSTEM QUALITY

There are two main aspects to be borne in mind when thinking about the quality of a vacuum system, *viz.*, system design and system performance.

### 3.1 Vacuum system design

To some extent this is the straightforward part of the process since most competently designed vacuum systems will be designed and built to well understood engineering standards. Factors such as size, shape, tolerances, stability, motions, heat dissipation, etc. can all be specified in detail. When the system is built, all of these can be accurately checked by standard techniques. It is sometimes the case

that in vacuum systems for very particular purposes, the state of the art may not be good enough, but that is a different story.

### 3.2 Vacuum system performance

Specifying the performance of a vacuum system may also be thought at first to be straightforward, but in practice, this is not the case.

Firstly, it is important to note that the majority of users of vacuum systems are not really interested in vacuum at all. It is simply that they wish to work in a controlled atmosphere and that vacuum happens to be the easiest such controlled atmosphere to use. Examples of such uses are particle accelerators, where beam scattering should be minimized; metal refiners and semiconductor manufacturers where contamination of the product is an issue, etc.

## 4. SPECIFICATION OF A VACUUM SYSTEM

### 4.1 General

Fundamentally there are two different methods of specifying the performance of a vacuum system: specification by process and specification by performance. The former involves the designer, who in this case will probably either be the user or closely associated with the user, spelling out in detail every aspect of the vacuum system, confident that provided the manufacturer sticks to procedures, the desired result will be achieved. The latter approach simply states what the vacuum system will do and leaves the manufacturer to decide how to meet this performance, which will, of course, be checked.

### 4.2 Specification by process

This approach puts the responsibility for achieving the desired quality firmly on the system designer. This is fine so long as the system in fact meets its requirements, but gives rise to possible conflicts if the performance falls short of requirements. Has the manufacturer not adhered to procedures properly? Is the system under-pumped? The designer accepts the costs of the design process, so manufacturing costs should be less. However the purchaser must be prepared to supervise the manufacturing process closely and check frequently that everything is being done as required.

The designer will require an intimate knowledge and understanding of things like choice of materials, manufacturing techniques, cleaning processes, handling procedures, outgassing rates, pumping and gauging, etc.

Traditionally, accelerator builders have used this approach, but tend to muddy the waters slightly by also placing performance specifications on manufacturers.

Most of the risk here is borne by the designer/purchaser.

### 4.3 Specification by performance

Here, the user will specify the way in which the system is to perform with respect to important operational parameters which might include the required pressure, pump down times, residual gases, outgassing rates, etc., and may include the elimination of unwanted side effects such as particulate production and cross contamination.

The responsibilities for ensuring that the performance is achieved are here quite clear. They lie firmly with the manufacturer who will design the whole system mechanically, specify the pumping and gauging, etc. However the purchaser must be prepared to accept the costs of doing this and most manufacturers will also build in contingency costs for rectification work if the desired performance is not achieved.

A potential problem lies in how the performance is to be measured and this must be clearly understood before an order is placed. Performance testing is discussed in the next section.

This approach has traditionally been used by industrial purchasers, where in-house understanding of vacuum may be quite low. In practice a modified version of this approach will often

be used with the purchaser carrying out part of the mechanical design work, leaving the manufacturer to do the vacuum bits.

Most of the risk here is borne by the manufacturer but the customer pays for this.

## 5. VACUUM SYSTEM PERFORMANCE AND ITS MEASUREMENT

There are many factors affecting vacuum system performance and its measurement, many of which have been dealt with in detail in this course so are simply listed here for completeness. The specifier, the designer and the purchaser's inspector will all require a knowledge of these if the finished system is to be of suitable quality. The limitations apply equally to whichever approach to specification as discussed above is used: the approach simply determines who carries the responsibility.

*Factors which affect performance*

- Materials
  - Mechanical Properties
    - Changes induced by heat, vacuum, by what the system is used for
    - Techniques
  - Working, joining
  - Outgassing rates
    - What influences outgassing
    - How is it measured
    - Are "book" values to be trusted
- Cleaning
  - What is to be cleaned
  - How and why
- Handling

*Factors which affect measurement*

- Required pressure
  - Measurement
    - Gauge calibration and stability
  - Pumping speed and stability
  - Uncertainties in measurement
- Residual gases
  - RGA calibration and stability
  - Uncertainties in measurement

## 6. REAL-WORLD VACUUM GAUGING

It is useful to note that the measurement of pressure in a vacuum system is usually a key factor in any specification. It is therefore of vital importance to know to what extent measurements of pressure can be trusted. Much of the published values for gauge sensitivities, stabilities, etc., come, understandably, from standards laboratories where gauges are used in near ideal circumstances.

Things are quite different in the real world of vacuum pressure measurements and not much is really known about the uncertainties inherent in such measurements. For example, there is usually little time allowed for taking such measurements, gauges are not allowed to outgas and stabilize and the system may not be in equilibrium. The personnel taking the measurements have limited vacuum knowledge and skills and are indeed often not really interested in the subject. They may have little experience on which to base an assessment of whether or not a reading is sensible. There is no appreciation of the effects of the state of the vacuum system (e.g. an external pump set) on measurements and how the past history of the system can affect readings.

The problem is compounded when it comes to using residual gas analyzers. Calibration of such instruments is in fact difficult and fundamentally not well defined. Judgement is required in interpreting the spectra. Most modern instruments are software driven and the real data is hidden from the user. The effects of interpretation algorithms are often not specified.

However, this is not to counsel that there is no point in doing such measurements, simply that a degree of caution needs to be applied in interpreting results

## 7. SUMMARY

This paper has discussed two basic approaches to the application of quality thinking to the specification of vacuum systems in which the specifier either accepts control of the manufacturing process and the risk involved or specifies what is required and passes the risk to the manufacturer but pays the price.

In each case, both specifier and manufacturer must clearly understand who is accepting the risk and where responsibility lies. They must also clearly understand from the start how the assessment of the completed system is to be achieved and the limitations of the processes involved.

# LIST OF PARTICIPANTS

| | |
|---|---|
| ANDERSEN, E. | Pfeiffer Vacuum Scandinavia, Taastrup, Denmark |
| AULENBACHER, K. | University of Mainz, Germany |
| BAGLIN, V. | CERN, Geneva, Switzerland |
| BATTISTI, A. | INFN-LNF, Frascati, Italy |
| BOJON, J-P. | CERN, Geneva, Switzerland |
| BOURQUIN, P. | CERN, Geneva, Switzerland |
| BUENAVENTURA, A. | CERN, Geneva, Switzerland |
| BURMANN, P. | DANFYSIK AS, Jyllinge, Denmark |
| BURTIN, G. | CERN, Geneva, Switzerland |
| BUTIN, F. | CERN, Geneva, Switzerland |
| CALATRONI, S. | CERN, Geneva, Switzerland |
| CHANG, S-H. | Synchrotron Radiation Research Centre, Hsinchu, Taiwan |
| COSTA, E. | INFN-LNS, Catania, Italy |
| COSTA PINTO, P. | CERN, Geneva, Switzerland |
| DE GIORGI, M. | INFN-LNF, Frascati, Italy |
| DILLON, S. | Elekta Oncology Systems Ltd, Crawley, UK |
| DONG, H. | IHEP, Beijing, P.R. China |
| FAVRE, G. | CERN, Geneva, Switzerland |
| FEDER, P. | ESRF, Grenoble, France |
| GADDI, A. | INFN, Pisa, Italy |
| GUPTA, A.D. | Variable Energy Cyclotron Centre, Calcutta, India |
| HANSEN, J. | CERN, Geneva, Switzerland |
| HENRIST, B. | CERN, Geneva, Switzerland |
| HERBERT, J. | Daresbury Lab., Warrington, UK |
| HERMLE, S. | Forschungszentrum Karlsruhe Gmbh, Karlsruhe, Germany |
| HOEG, P.L. | DANFYSIK AS, Jyllinge, Denmark |
| HOURICAN, M. | CERN, Geneva, Switzerland |
| IVERSEN, K. | University of Aarhus, Aarhus, Denmark |
| JAHNKE, P. | The Svedberg Lab., Uppsala, Sweden |
| JENNINGER, B. | CERN, Geneva, Switzerland |
| JIMINEZ, J.M. | CERN, Geneva, Switzerland |
| KIVER, A. | IHEP, Protvino, Russia |
| KNASTER REFOLIO, J.R. | CERN, Geneva, Switzerland |
| KOOPMAN, J. | CERN, Geneva, Switzerland |
| LATIONS, D. | Mass. Institute of Technology, Middleton, MA, USA |
| LAUGIER, I. | CERN, Geneva, Switzerland |
| LEE, J.C. | SRRC, Hsinchu, Taiwan |
| LOLLO, V. | INFN-LNF, Frascati, Italy |
| LOWE, D. | University of Saskatchewan, Saskatoon, Canada |
| LOZANO BERNAL, M.P. | CERN, Geneva, Switzerland |
| MADSEN, N. | CERN, Geneva, Switzerland |
| MAHNER, E. | CERN, Geneva, Switzerland |
| MARLETTA, R. | INFN-LNS, Catania, Italy |
| MARQUE, S. | CERN, Geneva, Switzerland |
| MATHOT, S. | CERN, Geneva, Switzerland |
| MIRALLES, L. | IFAE, Barcelona, Spain |
| MORAJA, M. | SAES GETTERS s.p.A., Lainate, Italy |
| MURDOCH, G. | RAL, Didcot, UK |
| NEUPERT, H. | CERN, Geneva, Switzerland |
| PERKINS, C. | Elekta Oncology Systems Ltd, Crawley, UK |
| PRODROMIDES, A. | CERN, Geneva, Switzerland |
| RAU, M. | GSI, Darmstadt, Germany |
| ROMMELUERE, P. | LURE, Bât. 209A, Orsay, France |

| | |
|---|---|
| ROSSI, A. | CERN, Geneva, Switzerland |
| ROWE, A. | Fermilab, Batavia, IL, USA |
| SARTA, G. | INFN-LNS, Catania, Italy |
| SCHEUERLEIN, C. | CERN, Geneva, Switzerland |
| SCHNEIDER, G. | CERN, Geneva, Switzerland |
| SENSOLINI ARRA, G. | INFN-LNF, Frascati, Italy |
| SHUKLA, S.K. | CAT, Indore, India |
| SKOV, H. | DANFYSIK AS, Jyllinge, Denmark |
| SORCHETTI, R. | INFN-LNF, Frascati, Italy |
| SØRENSEN, E.M. | Manne Siegbahnlaboratoriet, Stockholm, Sweden |
| STRUBIN, P. | CERN, Geneva, Switzerland |
| TABORELLI, M. | CERN, Geneva, Switzerland |
| TANGUAY, M. | Mass. Institute of Technology, Middleton, MA, USA |
| TOFT, V. | University of Aarhus, Aarhus, Denmark |
| TONGU, H. | Kyoto University, Kyoto, Japan |
| VACCAREZZA, C. | INFN-LNF, Frascati, Italy |
| VALENTE, A-M. | CERN, Geneva, Switzerland |
| VANT HOF, C.A. | CERN, Geneva, Switzerland |
| VENESS, R. | CERN, Geneva, Switzerland |
| VIGNOLI, C. | INFN-PAVIA, Pavia, Italy |
| VOLLENBERG, W. | CERN, Geneva, Switzerland |
| WELSCH, C. | Johann Wolfgang Göthe-Univ., Frankfurt, Germany |
| WESTON, T. | Daresbury Lab., Warrington, UK |
| WEVERS, I. | CERN, Geneva, Switzerland |
| XIONG, S. | Beijing Synchrotron Radiation Lab., Beijing, P.R. China |
| YANDON, J. | TRIUMF, Vancouver, Canada |
| ZAPFE, K. | DESY, Hamburg, Germany |

## List of CERN Reports published in 1999

CERN 99–01
CERN. Geneva
Russenschuck, S [ed]
Proceedings, ROXIE: Routine for the optimization of magnet X-sections, inverse field calculation and coil end design, CERN, 16–18 March 1999
CERN, 12 April 1999. – 258 p

CERN 99–02
CERN. Geneva
Autin, B; Blondel, A; Ellis, J [eds]
Prospective study of muon storage rings at CERN
CERN, 30 April 1999. – 144 p

CERN 99–03
CERN. Geneva
D'Agostini, G
Bayesian reasoning in high-energy physics: Principles and applications
CERN, 19 July 1999. – 184 p

CERN 99–04
CERN. Geneva
Ellis, N; March-Russel, J [eds]
Proceedings, 1998 European School of High-Energy Physics, St. Andrews, Scotland, 23 August – 5 September 1998
CERN, 28 July 1999. – 472 p

CERN 99–05
CERN. Geneva
Turner, S [ed]
Proceedings, CAS – CERN Accelerator School; Vacuum Technology, Snekersten, Denmark 28 May–3 June 1998
CERN, 19 August 1999. – 340 p

**CERN 99–05**
**ISSN 0007–8328**
**ISBN 92–9083–149–9**