CERN–2007–003
11 June 2007

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

CAS CERN Accelerator School

Vacuum in accelerators

Platja d'Aro, Spain
16 – 24 May 2006

Proceedings

Editor: D. Brandt

GENEVA
2007

CERN–320 copies printed–June 2007

Abstract

These proceedings present the lectures given at the twentieth specialized course organized by the CERN Accelerator School (CAS), the topic being Vacuum in Accelerators. The course was held in Platja d'Aro, Spain, from 16 to 24 May 2006. A similar course took place in Snekersten, Denmark, in 1999, with proceedings published as CERN 99-05. After an interval of seven years, the aim of this course was to present a review of the actual state of the art and to highlight the latest developments in the field.

The lectures start with a general overview of vacuum, accelerators and cryogenics followed by a more detailed review of the basic principles concerning thermal and non-thermal outgassing. More specialized lectures are then proposed on gas dynamics, on the interaction of energetic particles with matter, and on beam–gas collisions. The production and measurement of vacuum is addressed in subsequent lectures on pumps and vacuum gauges which present a detailed view of the materials currently used in accelerators. The next part of the course is dedicated to the construction and operation of vacuum systems: design rules, leak detection, sealing, cleaning, and materials. Finally, lectures of general interest are presented on topics like industrial application of vacuum, history of vacuum, cold vacuum systems, ITER, and future challenges for extremely high vacuum and accelerators.

Foreword

The aim of the CERN Accelerator School (CAS) to collect, preserve and disseminate the knowledge accumulated in the world's accelerator laboratories applies not only to general accelerator physics, but also to related sub-systems, equipment, and technologies. This wider aim is achieved by means of specialized courses. For 2006, the topic of the course was Vacuum in Accelerators and it was held at the Park Hotel San Jorge, Platja d'Aro, Spain, from 16 to 24 May 2006.

Vacuum in Accelerators had already been treated in the framework of CAS courses, namely in 1999 in Snekersten, Denmark. However, after seven years and considering the enormous developments achieved in the field, it was unanimously felt that there was a real need to present an updated version of the previous course.

The present course was elaborated in collaboration with the ALBA Synchrotron Light Facility (Consortium CELLS) in Barcelona. In particular, the outstanding contribution of the Programme coordinators Dr N. Hilleret (CERN) and Dr L. Miralles (CELLS) and the enthusiastic contribution of the Local Organizing Committee composed of Lluis Miralles, Laura Campos and Yolanda Ruiz, were most invaluable.

It is also important to underline that the Local Organizing Committee successfully contributed to identifying sponsors from Industry to provide financial support and offer scholarships to highly deserving young students, who would otherwise not have been able to attend the school.

As always, the backing of the CERN management, the guidance of the CAS Advisory and Programme Committees, the attention to detail of the Local Organizing Committee and also the management and staff of the Park Hotel San Jorge ensured that the school was held under optimum conditions.

Very special thanks must go to the lecturers for the enormous task of preparing, presenting, and writing up their topics.

Finally, the enthusiasm of the participants who came from 24 different countries was convincing proof of the usefulness and success of the course.

It is my pleasure and my privilege to thank most sincerely all those persons who helped in making the course a success, including the team of the CERN Scientific Text Processing Service for their dedication and commitment to the production of this document.

Daniel Brandt
CERN Accelerator School

PROGRAMME
VACUUM IN ACCELERATORS (16-24 May 2006, Platja D'Aro, Spain)

Time	Wednesday 17 May	Thursday 18 May	Friday 19 May	Saturday 20 May	Sunday 21 May	Monday 22 May	Tuesday 23 May	Wednesday 24 May
08:30 09:30	Opening Talk The ALBA Project L. Miralles	Thermal Outgassing P. Chiggiato	Non-thermal Outgassing N. .Hilleret		Partial Pressure Gauges G. Peter	Analytical & Numerical Tools for Vacuum Systems R. Kersevan	Cold/Sticky Systems V. Baglin	Buses to Airport
09:30 10 :30	Introduction toVacuum F. Dylla	General Design Rules C. Hauviller	Materials S. Sgobba	**E X C U R S I O N**	Beam-Vacuum Interaction I Lifetime, Cross-sections, Activation P. Grafström	Beam-Vacuum Interaction II Lifetime, Ion-Electron, Trapping, Dust, E-Cloud O. Gröbner	Iter Vacuum System M. Wykes	
	COFFEE	**COFFEE**	**COFFEE**		**COFFEE**	**COFFEE**	**COFFEE**	
11:00 12:00	Introduction to Accelerators D. Brandt	Mechanical Pumps A. Chew	Ion Pumps M. Audi		Leak Detection K. Zapfe	Getter Pumps C. Benvenuti	Control Systems P. Strubin	
	LUNCH	**LUNCH**	**LUNCH**		**LUNCH**	**LUNCH**	**LUNCH**	
15:00 16:00	Physics of Outgassing P. Chiggiato	Gauges for Fine and High Vacuum K. Jousten	UHV Gauges K. Jousten		Cryo Pumps C. Day	Surface Properties and Cleaning M. Taborelli	Vacuum Aspects in Synchrotron Light Sources R. Reid	
16:00 17:00	Gas Dynamics F. Sharipov	Water Outgassing F. Dylla	Interaction with Matter J. Schou		Sealing Technology K. Sonderegger	Use of NEG Pumps and Coatings in Large Vacuum Systems F. Mazzolini	Misconceptions and other Sources of Problems N. Hilleret	
	T E A	**T E A**	**T E A**		**T E A**	**T E A**	**T E A**	
17:30 18:30	Introduction to Cryogenics P. Lebrun	Industrial Applications of Vacuum D. Mueller	History of Vacuum Technology for Physics Research F. Dylla		Large Systems Commissioning K. Zapfe	Challenges for XHV F. Dylla	Vacuum Challenges for Future Machines R. Reid	
	Welcome Drink	**DINNER**	**DINNER**	**DINNER**	**DINNER**	**Special Dinner**	**DINNER**	

Contents

Foreword
D. Brandt v
Rarefied gas dynamics and its applications to vacuum technology
F. Sharipov 1
Introduction to cryogenics
P. Lebrun 15
Design rules for vacuum chambers
C. Hauviller 31
Mechanical vacuum pumps
A.D. Chew 43
Gauges for fine and high vacuum
K. Jousten 65
Non-thermal outgassing
N. Hilleret 87
Materials for high vacuum technology: an overview
S. Sgobba 117
Ultrahigh vacuum gauges
K. Jousten 145
Slowing-down processes, energy deposition, sputtering and desorption in ion and electron interactions with solids
J. Schou 169
The development of ultrahigh and extreme high vacuum technology for physics research
H.F. Dylla 179
Partial pressure gauges
G.J. Peter and N. Müller 195
Lifetime, cross-sections and activation
P. Grafström 213
Leak detection
K. Zapfe 227
Basics and applications of cryopumps
C. Day 241
Commissioning of large vacuum systems
K. Zapfe 275
Analytical and numerical tools for vacuum systems
R. Kersevan 285
Getter pumping
C. Benvenuti 313
Cleaning and surface properties
M. Taborelli 321
The use of NEG pumps and coatings in large vacuum systems: experience and limitations
F. Mazzolini 341
Cold/sticky systems
V. Baglin 351
Vacuum controls and interlocks
P. Strubin 369

Vacuum aspects of synchrotron light sources
R.J. Reid 389
Vacuum challenges for future machines
R.J. Reid 401
List of Participants 407

Rarefied gas dynamics and its applications to vacuum technology

F. Sharipov
Universidade Federal do Paraná, Curitiba, 81531-990, Brazil

Abstract
Basic concepts of rarefied gas dynamics are given in a concise form. Some problems of rarefied gas flows are considered, namely, calculations of velocity slip and temperature jump coefficients, gas flow through a tube due to pressure and temperature gradients, and gas flow through a thin orifice. Results on the two last problems are given over the whole range of gas rarefaction. A methodology for modelling the Holweck pump is described. An extensive list of publications on these topics is given.

1 Brief history of rarefied gas dynamics

Rarefied gas dynamics is based on the kinetic approach to gas flows. In 1859 Maxwell [1] abandoned the idea that all gaseous molecules move with the same speed and introduced the statistical approach to gaseous medium, namely, he introduced the velocity distribution function and obtained its expression in the equilibrium state. Thus Maxwell gave the origin to the kinetic theory of gases. Then, in 1872 Boltzmann [2] deduced the kinetic equation which determines the evolution of the distribution function for gaseous systems being out of equilibrium.

In 1909 Knudsen [3], measuring a flow rate through a tube, detected a deviation from the Poiseuille formula at a low pressure. Such a deviation was explained by the fact that at a certain pressure the gas is not a continuous medium and the Poiseuille formula is not valid anymore. A description of such a flow required the development of a new approach based on the kinetic theory of gases. This can be considered as the beginning of rarefied gas dynamics.

Later, advances were made by Hilbert [4], Enskog [5] and Chapman [6] to solve the Boltzmann equation analytically via an expansion of the distribution function with respect to the Knudsen number. The main result of this solution was a relation of the transport coefficients to the intermolecular interaction potential, but no numerical calculation of rarefied gas flows could be realized at that time.

In 1954 the so-called model equations [7,8] were proposed to reduce the computational efforts in calculations of rarefied gas flows. Using these models it was possible to obtain numerical results on rarefied gas flows in the transition regime. Thus in 1960 a numerical investigation of rarefied gas flows began in its systematic form. For a long time, it was possible to solve only the model equations. Practically, all classical problems of gas dynamics (Poiseuille flow, Couette flow, heat transfer between two plates, flow past a sphere, etc.) were solved over the whole range of gas rarefaction by applying the model equations. In 1989 first results based on the exact Boltzmann equation were reported, see, for example, Ref. [9]. However, even using the powerful computers available nowadays, a numerical calculation based on the Boltzmann equation itself is still a very hard task, which requires great computational efforts. Thus, the model equations continue to be a main tool in practical calculations.

Below, the main concepts of rarefied gas dynamics and some examples of its application will be given. In the last section, the main results of rarefied gas dynamics that could be applied to vacuum technology are listed.

2 Basic concepts of rarefied gas dynamics

2.1 Knudsen number and rarefaction parameter

The principal parameter of rarefied gas dynamics is the Knudsen number (Kn) which characterizes the gas rarefaction and is defined as the ratio

$$\mathrm{Kn} = \frac{\ell}{a} \ , \tag{1}$$

where ℓ is the equivalent molecular mean free path given as

$$\ell = \frac{\mu v_{\mathrm{m}}}{P} \ , \qquad v_{\mathrm{m}} = \left(\frac{2k_{\mathrm{B}}T}{m} \right)^{1/2} \ , \tag{2}$$

v_{m} is the most probable molecular velocity, $k_{\mathrm{B}} = 1.380662 \times 10^{-23}$ J/K is the Boltzmann constant, m is the molecular mass of the gas in kg, T is the temperature of gas in K, P is its pressure in Pa, and μ is the gaseous viscosity in Pa s.

Regarding the value of the Knudsen number, we may distinguish the following three regimes of gas flow. If the Knudsen number is small ($\mathrm{Kn} \ll 1$), the gas can be considered as a continuous medium and the hydrodynamic equations [10] can be applied. This regime is called *hydrodynamic*. If the Knudsen number is large ($\mathrm{Kn} \gg 1$), the intermolecular collisions can be neglected. Under this condition we may consider that every molecule moves independently from each other and, usually, the test particle Monte Carlo method [11,12] is employed. This regime is called *free-molecular*. When the Knudsen number has some intermediate value, we can neither consider the gas as a continuous medium nor discount the intermolecular collisions. In this case the kinetic equation should be solved [13–21], or the DSMC method [11,12] is used. This regime is called *transitional*.

Usually another quantity characterizing the gas rarefaction is used instead of the Knudsen number, *viz*. the rarefaction parameter defined as

$$\delta = \frac{a}{\ell} = \frac{1}{\mathrm{Kn}} \ . \tag{3}$$

Large values of δ correspond to the hydrodynamic regime and the small values of δ are appropriate to the free molecular regime. This parameter is more convenient because many solutions are given in terms of this parameter.

2.2 Velocity distribution function

The state of a monoatomic gas is described by the one-particle velocity distribution function $f(t,\mathbf{r},\mathbf{v})$, where t is the time, $\mathbf{r}$ is a vector of spatial coordinates, and $\mathbf{v}$ is a velocity of molecules. The distribution function is defined so that the quantity $f(t,\mathbf{r},\mathbf{v})\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{v}$ is the number of particles in the phase volume $\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{v}$ near the point $(\mathbf{r}, \mathbf{v})$ at the time t.

All macro-characteristics of gas flow can be calculated *via* the distribution function:

number density

$$n(t,\mathbf{r}) = \int f(t,\mathbf{r},\mathbf{v})\,\mathrm{d}\mathbf{v} \ , \tag{4}$$

hydrodynamic (bulk) velocity

$$\mathbf{u}(t,\mathbf{r}) = \frac{1}{n} \int \mathbf{v}\, f(t,\mathbf{r},\mathbf{v})\,\mathrm{d}\mathbf{v} \ , \tag{5}$$

pressure

$$P(t,\mathbf{r}) = \frac{m}{3}\int V^2 \; f(t,\mathbf{r},\mathbf{v})\,\mathrm{d}\mathbf{v}\,, \tag{6}$$

stress tensor

$$P_{ij}(t,\mathbf{r}) = m\int V_i V_j \; f(t,\mathbf{r},\mathbf{v})\,\mathrm{d}\mathbf{v}\;, \tag{7}$$

temperature

$$T(t,\mathbf{r}) = \frac{m}{3nk_{\mathrm{B}}}\int V^2 \; f(t,\mathbf{r},\mathbf{v})\,\mathrm{d}\mathbf{v}\,, \tag{8}$$

heat flow vector

$$\mathbf{q}(t,\mathbf{r}) = \frac{m}{2}\int V^2\mathbf{V} \; f(t,\mathbf{r},\mathbf{v})\,\mathrm{d}\mathbf{v}\,, \tag{9}$$

where $\mathbf{V}$ is the peculiar velocity

$$\mathbf{V} = \mathbf{v} - \mathbf{u}\,. \tag{10}$$

2.3 Boltzmann equation

The distribution function obeys the Boltzmann equation [11–21] which in the absence of external forces reads

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{r}} = Q(ff_*)\;, \tag{11}$$

where $Q(ff_*)$ is the collision integral, which has a complicated expression and is omitted here. Till now, a numerical solution of Eq.(11) with the exact expression of the collision integral is a very difficult task, that is why some simplified expressions of $Q(ff_*)$ are used. These expressions are called the model kinetic equations. They maintain the main properties of the exact collision integral, but they allow us to reduce significantly the computational efforts when the kinetic equation is solved numerically.

The most usual model equation was proposed by Bhatnagar, Gross and Krook (BGK) [7] and by Welander [8]. They presented the collision integral as

$$Q_{\mathrm{BGK}}(f\;f_*) = \nu\left[\,f^{\mathrm{M}} - f(t,\mathbf{r},\mathbf{v})\right]\;, \tag{12}$$

where f^{M} is the local Maxwellian

$$f^{\mathrm{M}} = n(t,\mathbf{r})\left[\frac{m}{2\pi k_{\mathrm{B}}T(t,\mathbf{r})}\right]^{3/2}\exp\left\{-\frac{m[\mathbf{v}-\mathbf{u}(t,\mathbf{r})]^2}{2k_{\mathrm{B}}T(t,\mathbf{r})}\right\}\;. \tag{13}$$

The local values of the number density $n(t,\mathbf{r})$, bulk velocity $\mathbf{u}(t,\mathbf{r})$, and temperature $T(t,\mathbf{r})$ are calculated *via* the distribution function $f(t,\mathbf{r},\mathbf{v})$ in accordance with the definitions (4), (5), and (8), respectively. The quantity ν is the collision frequency assumed to be independent of the molecular velocity and equal to $\nu = P/\mu$. However, this model does not provide the correct value of the Prandtl number.

The S-model proposed by Shakhov [22] is a modification of the BGK model giving the correct Prandtl number. The collision integral of this model is written down as

$$Q_{\mathrm{S}}(ff_*) = \frac{P}{\mu}\left\{ f^{\mathrm{M}}\left[1 + \frac{2m}{15n(k_{\mathrm{B}}T)^2}\mathbf{q}\cdot\mathbf{V}\left(\frac{mV^2}{2k_{\mathrm{B}}T} - \frac{5}{2}\right)\right] - f(t,\mathbf{r},\mathbf{v})\right\}. \tag{14}$$

This model has another shortcoming: the H-theorem can be proved only for the linearized S-model. In the non-linear form one can neither prove nor disprove the theorem. This property sometimes leads to non-physical results. However, the linearized S-model works very well for non-isothermal flows.

A critical analysis and comparison of results based on the exact Boltzmann equation, BGK model, and S-model are given in the review paper [20]. From this analysis we may conclude that the model equations significantly reduce the computational efforts. However, to obtain reliable results one should apply an appropriate model equation. If a gas flow is isothermal and the heat transfer is not important the BGK equation is the most suitable model equation. If a gas flow is non-isothermal it is better to apply the S-model.

3 Methods of computation in the transition regime

The discrete velocity method is the most used one. A set of values of the velocity $\mathbf{v}_i$ is chosen. The collision integral is expressed *via* the values $f_i(t,\mathbf{r}) = f(t,\mathbf{r},\mathbf{v}_i)$. Thus, the integro-differential Boltzmann equation is replaced by a system of differential equations for the functions $f_i(t,\mathbf{r})$. The differential equations can be solved numerically by a finite difference method. Then, the distribution function moments are calculated using some quadrature. Details of the method are given by Kogan [15] (Section 3.13) and elsewhere [20,23].

The Direct Simulation Monte Carlo (DSMC) method is also widely used. The region of the gas flow is divided into a network of cells. The dimensions of the cells must be such that the change in flow properties across each cell is small. The time is advanced in discrete steps of magnitude Δt, such that Δt is small compared with the mean time between two successive collisions. The molecular motion and intermolecular collision are uncoupled over the small time interval Δt by the repetition of the following procedures:

- The molecules are moved through the distance determined by their velocities and Δt. If the trajectory passes the boundary a simulation of the gas–surface interaction is performed in accordance with a given law. New molecules are generated at boundaries across which there is an inward flux.
- A representative number of collisions appropriate to Δt and the number of molecules in the cell is computed. The pre-collision velocities of the molecules involved in the collision are replaced by the post-collision values in accordance with a given law of the intermolecular interaction.

After a sufficient number of repetitions we may calculate any moment of the distribution function. Details of the method are given in the books by Bird [11,12].

4 Typical problems of rarefied gas dynamics

4.1 Velocity slip and temperature jump coefficients

A moderate gas rarefaction can be taken into account via the so-called velocity slip and temperature jump boundary conditions. It means that the bulk velocity is not equal to zero on the wall, but its tangential component u_y near the wall is proportional to its normal gradient

$$u_y = \sigma_{\mathrm{P}}\, \ell \frac{\partial u_y}{\partial x} \ , \quad \text{at} \quad x = 0 \ , \tag{15}$$

where σ_{P} is the viscous slip coefficient, x, y are a reference frame with the origin at the surface, x is the normal coordinate, while y is the tangential one.

A gas flow also can be induced by a tangential temperature gradient. In this case the tangential velocity is proportional to the temperature gradient

$$u_y = \sigma_{\mathrm{T}} \frac{\mu}{\varrho} \frac{\partial \ln T}{\partial y} \ , \quad \text{at} \quad x = 0 \ , \tag{16}$$

where σ_{T} is the thermal slip coefficient and ϱ is the local mass density.

The temperature jump condition on a solid surface reads as

$$T_{\mathrm{g}} = T_{\mathrm{w}} + \zeta_{\mathrm{T}}\, \ell \frac{\partial T_{\mathrm{g}}}{\partial x} \ , \quad \text{at} \quad x = 0 \ , \tag{17}$$

where T_{g} is the gas temperature near the surface and ζ_{T} is the temperature jump coefficient. Physically, Eq.(17) means that the temperature of the gas is not equal to that of the wall, but there is a discontinuity (jump) of the temperature.

The detailed technique of calculations of the slip and jump coefficients on the basis of the kinetic equation and their numerical values can be found in Refs. [20,24–27]. The values recommended in practical calculations are as follows

$$\sigma_{\mathrm{P}} = 1.018 \ , \qquad \sigma_{\mathrm{T}} = 1.175 \ , \qquad \zeta_{\mathrm{T}} = 1.954 \ . \tag{18}$$

These values were obtained under an assumption of the diffuse scattering of gaseous particles on the surface. In most practical situations this assumption is fulfilled. Only in some special situations can the deviation from the diffuse scattering be significant, see, for example, Ref. [28].

Numerical data of the velocity slip and temperature jump coefficients for gaseous mixtures can be found in Refs. [29–32].

4.2 Gas flow through long capillaries

Consider a monoatomic rarefied gas flowing through a long tube due to small longitudinal gradients of pressure P and temperature T denoted as

$$\xi_{\mathrm{P}} = \frac{a}{P} \frac{\mathrm{d}P}{\mathrm{d}x} \ , \qquad \xi_{\mathrm{T}} = \frac{a}{T} \frac{\mathrm{d}T}{\mathrm{d}x} \ , \tag{19}$$

respectively. Here, x is the longitudinal coordinate coinciding with the tube axis, a is the tube radius. We assume that the tube is so long that the end effects can be neglected and the flow is considered to be one-dimensional.

The mass flow rate $\dot{M}$ through the tube can be calculated as

$$\dot{M} = \frac{\pi a^2 P}{v_{\mathrm{m}}} \left(-G_{\mathrm{P}} \xi_{\mathrm{P}} + G_{\mathrm{T}} \xi_{\mathrm{T}} \right) \ , \tag{20}$$

where G_{P} and G_{T} are reduced flow rates calculated on the basis of the kinetic equation. They are determined by the rarefaction parameter δ. Note that a temperature gradient causes a gas flow of rarefied gas, i.e., the gas flows from a cold region to a hot one. This phenomenon is called *thermal*

creep. The details of calculations of the coefficients G_{P} and G_{T} and recommended data can be found in Ref. [20]. Here, just the main results corresponding to diffuse scattering will be given.

In the free molecular regime, i.e., at $\delta = 0$, the coefficients G_{P} and G_{T} can be calculated analytically

$$G_{\mathrm{P}} = \frac{8}{3\sqrt{\pi}} \ , \qquad G_{\mathrm{T}} = \frac{1}{2} G_{\mathrm{P}} \ . \tag{21}$$

Applying the Navier–Stokes equation with the slip boundary conditions (15) and (16) the coefficients G_{P} and G_{T} are obtained as

$$G_{\mathrm{P}} = \frac{\delta}{4} + \sigma_{\mathrm{P}} \ , \qquad G_{\mathrm{T}} = \frac{\sigma_{\mathrm{T}}}{\delta} \ . \tag{22}$$

These expressions are valid near the hydrodynamic regime, i.e. $\delta \gg 1$.

Table 1: Reduced flow rates G_{P} and G_{T} vs δ, Ref. [33]

δ	G_{P}	G_{T}	δ	G_{P}	G_{T}
0.001	1.5008	0.7486	1.0	1.4758	0.3959
0.01	1.4800	0.7243	1.4	1.5550	0.3514
0.02	1.4636	0.7042	2.0	1.6799	0.3016
0.05	1.4339	0.6637	5.0	2.3666	0.1752
0.1	1.4101	0.6210	10.0	3.5749	0.1014
0.2	1.3911	0.5675	20.0	6.0492	0.05426
0.5	1.4011	0.4779	50.0	13.495	0.02212

In the transition regime the kinetic equation (11) with the S-model (14) is applied. The coefficients G_{P} and G_{T} for some intermediate values of δ obtained in Ref. [33] are given in Table 1.

If the drops of pressure and temperature are large, then the technique elaborated in the works [33–35] must be used. The data on the coefficients G_{P} and G_{T} for a rectangular channel are given in Refs. [36,37]. The method of calculations of mass flow rate through a tube of variable diameter is given in Ref. [38]. A methodology of calculation of gaseous mixtures flowing through long tubes and channels is described in Refs. [39–43].

4.3 Orifice flow

Consider an orifice in an infinitely thin partition that separates two semi-infinite reservoirs. One of them contains a monoatomic gas at a pressure P_0. The other reservoir is maintained at a smaller pressure $P_1 < P_0$. The mass flow rate $\dot{M}$ through the orifice is determined by the rarefaction parameter δ with reference to the pressure P_0 and by the pressure ratio P_1/P_0. It is usual to use the reduced flow rate defined as

$$W = \frac{\dot{M}}{\dot{M}_0} \ , \qquad \dot{M}_0 = \frac{\sqrt{\pi} a^2}{v_m} P_0 \ , \tag{23}$$

where $\dot{M}_0$ corresponds to the mass flow rate into vacuum, i.e., at $P_1 = 0$, in the free molecular regime, i.e., at $\delta = 0$. If $P_1 \neq 0$ then there are two opposite flows through the orifice in the free molecular regime. Since they are independent the total flow rate is given as

$$W = 1 - \frac{P_0}{P_1} \ . \tag{24}$$

The papers [44,45] report numerical data on the flow rate W, where the problem was solved by the DSMC method. The number of particles fluctuates during the calculations, but it was maintained in the interval from 5×10^6 to 15×10^6. The number of samples was sufficient to guarantee the statistical scattering of the flow rate with 1%.

The dependences of the reduced flow rate W on the rarefaction parameter δ for $P_1/P_0 = 0;\ 0.1;\ 0.5$ and 0.9 are represented in Fig. 1. It can be seen that the numerical results are in a good agreement with the corresponding experimental data [46,47]. For $\delta > 100$ the variation of the flow rate is within the numerical accuracy for all pressure ratios considered here. Therefore the data presented here cover the whole range of the gas rarefaction δ. In the free-molecular regime ($\delta = 0$), the numerical value of W tends to its theoretical expression (24). In many practical applications it is necessary to know the correction to this limit value related to the intermolecular collisions. Near the free-molecular regime the flow rate W can be written as

$$W = \left(1 - \frac{P_1}{P_0}\right)(1 + A\delta) \quad \text{at} \quad \delta < 1 \ . \tag{25}$$

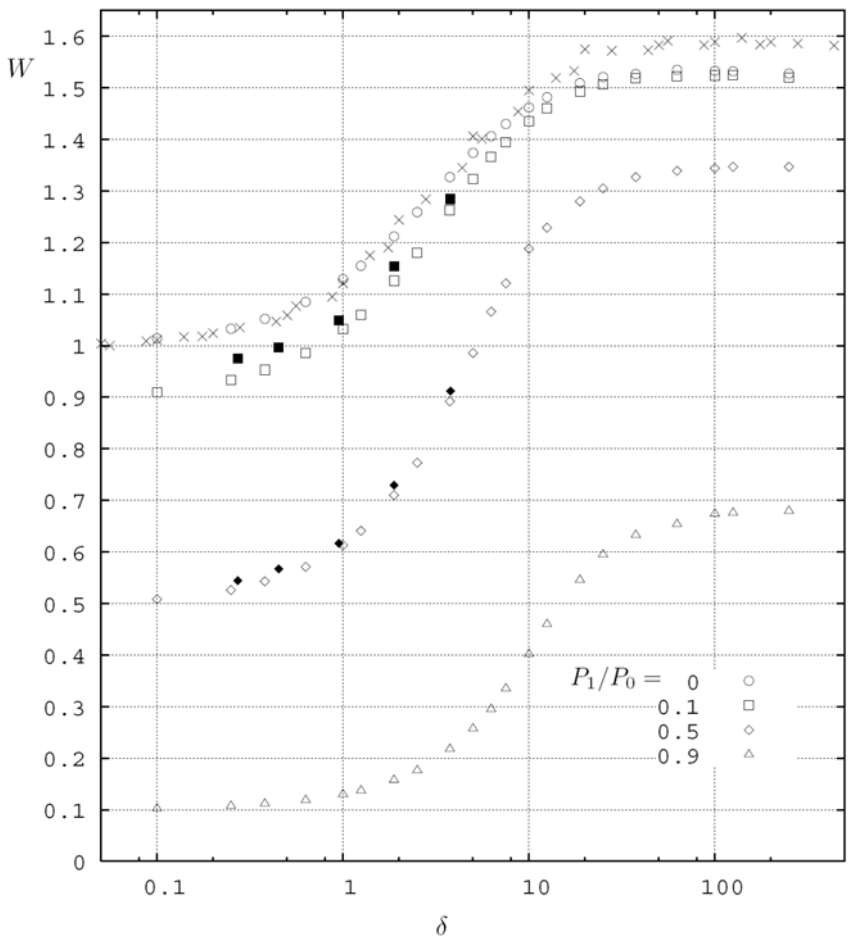


Fig. 1: Reduced flow rate W vs rarefaction parameter δ: open symbol—DSMC simulation [44,45]; filled symbols—experimental data from Ref. [46]; crosses—experimental data from Ref. [47]

The values of the constant A are 0.13; 0.15; 0.23; and 0.31 for the pressure ratio $P_1/P_0 = 0;\ 0.1;\ 0.5;$ and 0.9, respectively. These values were obtained by the least-squares method on the basis of the numerical results.

5 Modelling of Holweck pump

A Holweck pump is composed of two coaxial cylinders. One of them has grooves in a spiral form and the other is smooth. A rotation of the smooth cylinder causes the gas flow, i.e., the pumping effect. So, in the general case the gas flow through such a pump is three-dimensional. To reduce the computational efforts we consider a two-dimensional flow, i.e., we shall neglect the groove curvature and the end effects. More exactly, we consider a plane surface having several grooves regularly distributed over it as shown in Fig. 2(a). Another surface, which is smooth, moves in the direction ξ over the grooved surface and causes a gas flow in the direction η. The cross-section of one groove, i.e., AA, is presented in Fig. 2(b).

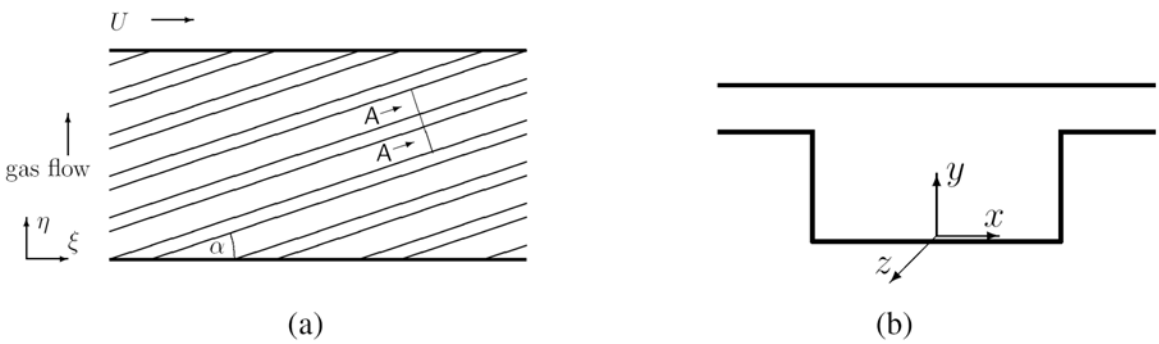


Fig. 2: Scheme of pump and cross-section of groove AA

The solution of the problems is divided into two stages. In the first stage we solve four independent problems over the whole range of the gas rarefaction:

i) longitudinal Couette flow, i.e., the gas flow due to a surface motion along the axis z. The coordinate system (x, y, z) is shown in Fig. 2(b);

ii) longitudinal Poiseuille flow, i.e., the gas flow caused by a pressure gradient along the axis z;

iii) transversal Couette flow, i.e., the gas flow due to a surface motion along the axis x;

iv) transversal Poiseuille flow, i.e., the gas flow caused by a pressure drop in the direction x through a pair of grooves and ridges. The solution of these four problems is determined by the groove and ridge sizes and by the rarefaction parameter δ.

In the second stage a linear superposition of the four solutions obtained previously is realized in accordance with the methodology described in Refs. [33–37]. This stage does not require much computational effort and allows us easily to change many parameters such as groove inclination, fore vacuum and high vacuum pressures, angular velocity of rotating cylinder, species of gas, temperature of the gas, etc.

Applying the present approach, the compression ratio and pumping speed were calculated. The results related to the compression ratio are shown in Fig. 3, from which it can be seen that the numerical results are in good agreement with the experimental data. The details of the numerical calculations and the measurements can be found in Ref. [48].

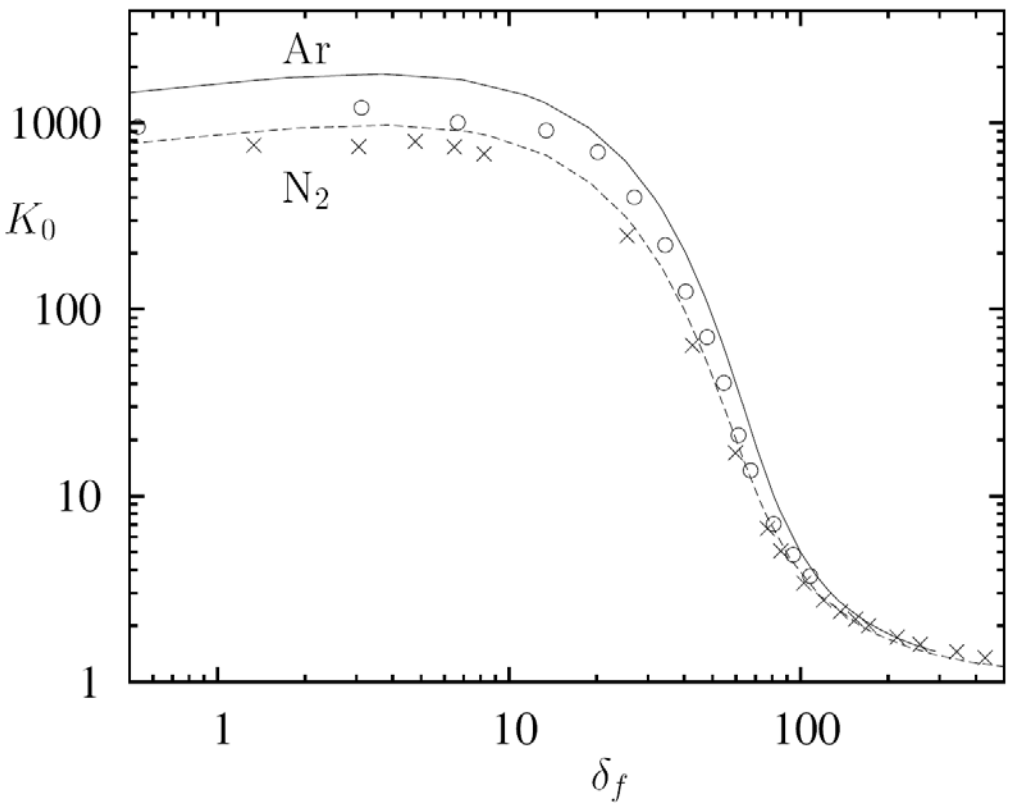


Fig. 3: Limit compression pressure ratio K_0 vs fore vacuum rarefaction δ_f: lines—theoretical results; symbols—experimental data

6 Summary of recent theoretical results

- Velocity slip coefficients for a single gas Refs. [20,27,49–51].
- Velocity slip coefficients for gaseous mixtures Refs. [29–31,52,53].
- Temperature jump coefficient for a single gas Refs. [27,54–56].
- Temperature jump coefficient for gaseous mixtures Refs. [32,57–59].
- Single gas flows through long tubes and channels over the whole range of the gas rarefaction Refs. [20,36–38,60–64].
- Mixture gas flows through long tubes and channels over the whole range of the gas rarefaction Refs. [39–43,65].
- Gas flow through orifices and slits Refs. [44,45,66–68].
- Couette flow of a single gas over the whole range of the gas rarefaction Refs. [69–74].
- Couette flow of mixtures over the whole range of the gas rarefaction Refs. [75–77].
- Modelling of vacuum pumps over the whole range of the gas rarefaction Refs. [48].

Some numerical calculations of the mass flow rate $\dot{M}$ through tubes, channels and orifices can be carried out on-line on the site http://fisica.ufpr.br/sharipov .

References

[1] *Scientific Papers of James Clerk Maxwell*, W. D. Niven, ed., (Dover Publications, Inc., New York, 1965).

[2] L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen, *Sitzung Berichte Kaiserl. Akad. der Wissenschaften* **66**, 275–370 (1872).

[3] M. Knudsen, Die Molekularstromung der Gase durch Offnungen und die Effusion, *Ann. Phys.* **28**, 999 (1909).

[4] D. Hilbert, *Grundüge einer allgemein Theorie der linearen Integralgleichungen* (Teubner, Leipzig, 1912).

[5] D. Enskog, Bermerkungen zu einer Fundamentalgleichung in der kinetischen Gastheorie, *Phys. Z.* **12**, 533–539 (1911).

[6] S. Chapman, On the law of distribution of molecular velocities, and on the theory of viscosity and thermal conduction, in a non-uniform simple monoatomic gas, *Phil. Trans. Roy. Soc.* London **216**, 279–341 (1916).

[7] P. L. Bhatnagar, E. P. Gross, and M. A. Krook, A model for collision processes in gases, *Phys. Rev.* **94**, 511–525 (1954).

[8] P. Welander, On the temperature jump in a rarefied gas, *Arkiv Fys.* **7**, 507–553 (1954).

[9] T. Ohwada, Y. Sone, and K. Aoki, Numerical analysis of the Poiseuille and thermal transpiration flows between two parallel plates on the basis of the Boltzmann equation for hard sphere molecules, *Phys. Fluids A* **1**, 2042–2049 (1989).

[10] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1989).

[11] G. A. Bird, in *Molecular Gas Dynamics*, L. C. Woods, W. H. Wittrick, and A. L. Cullen, eds., (Clarendon Press, Oxford, 1976).

[12] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Oxford University Press, Oxford, 1994).

[13] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938).

[14] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-uniform Gases* (University Press, Cambridge, 1952).

[15] M. N. Kogan, *Rarefied Gas Dynamics* (Plenum, New York, 1969).

[16] J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport Processes in Gases* (North-Holland Publishing Company, Amsterdam, 1972).

[17] C. Cercignani, *Theory and Application of the Boltzmann Equation* (Scottish Academic Press, Edinburgh, 1975).

[18] C. Cercignani, *The Boltzmann Equation and its Application* (Springer, New York, 1988).

[19] C. Cercignani, *Rarefied Gas Dynamics. From Basic Concepts to Actual Calculations* (Cambridge University Press, Cambridge, 2000).

[20] F. Sharipov and V. Seleznev, Data on internal rarefied gas flows, *J. Phys. Chem. Ref. Data* **27**, 657–706 (1998).

[21] Y. Sone, *Kinetic Theory and Fluid Mechanics* (Birkhäuser, Boston, 2002).

[22] E. M. Shakhov, Generalization of the Krook kinetic equation, *Fluid Dynamics* **3**, 142–145 (1968).

[23] F. M. Sharipov and E. A. Subbotin, On optimization of the discrete velocity method used in rarefied gas dynamics, *Z. Angew. Math. Phys. (ZAMP)* **44**, 572–577 (1993).

[24] S. Albertoni, C. Cercignani, and L. Gotusso, Numerical evaluation of the slip coefficient, *Phys. Fluids* **6**, 993–996 (1963).

[25] S. K. Loyalka, Temperature jump and thermal creep slip: Rigid sphere gas, *Phys. Fluids A* **1**, 403–408 (1989).

[26] T. Ohwada, Y. Sone, and K. Aoki, Numerical analysis of the shear and thermal creep flows of a rarefied gas over a plane wall an the basis of the linearized Boltzmann equation for hard-sphere molecules, *Phys. Fluids A* **1**, 1588–1599 (1989).

[27] F. Sharipov, Application of the Cercignani–Lampis scattering kernel to calculations of rarefied gas flows. II. Slip and jump coefficients, *Eur. J. Mech. - B/Fluids* **22**, 133–143 (2003).

[28] O. V. Sazhin, S. F. Borisov, and F. Sharipov, Accommodation coefficient of tangential momentum on atomically clean and contaminated surfaces, *J. Vac. Sci. Technol. A* **19**, 2499–2503 (2001), erratum: **20** (3), 957 (2002).

[29] F. Sharipov and D. Kalempa, Velocity slip and temperature jump coefficients for gaseous mixtures. I. Viscous slip coefficient, *Phys. Fluids* **15**, 1800–1806 (2003).

[30] F. Sharipov and D. Kalempa, Velocity slip and temperature jump coefficients for gaseous mixtures. II. Thermal slip coefficient, *Phys. Fluids* **16**, 759–764 (2004).

[31] F. Sharipov and D. Kalempa, Velocity slip and temperature jump coefficients for gaseous mixtures. III. Diffusion slip coefficient, *Phys. Fluids* **16**, 3779–3785 (2004).

[32] F. Sharipov and D. Kalempa, Velocity slip and temperature jump coefficients for gaseous mixtures. IV. Temperature jump coefficient, *Int. J. Heat Mass Transfer* **48**, 1076–1083 (2005).

[33] F. Sharipov, Rarefied gas flow through a long tube at any temperature difference, *J. Vac. Sci. Technol. A* **14**, 2627–2635 (1996).

[34] F. Sharipov and V. Seleznev, Rarefied gas flow through a long tube at any pressure ratio, *J. Vac. Sci. Technol. A* **12**, 2933–2935 (1994).

[35] F. Sharipov, Rarefied gas flow through a long tube at arbitrary pressure and temperature drops, *J. Vac. Sci. Technol. A* **15**, 2434–2436 (1997).

[36] F. Sharipov, Rarefied gas flow through a long rectangular channel, *J. Vac. Sci. Technol. A* **17**, 3062–3066 (1999).

[37] F. Sharipov, Non-isothermal gas flow through rectangular microchannels, *J. Micromech. Microeng.* **9**, 394–401 (1999).

[38] F. Sharipov and G. Bertoldo, Rarefied gas flow through a long tube of variable radius, *J. Vac. Sci. Technol. A* **23**, 531–533 (2005).

[39] F. Sharipov and D. Kalempa, Gaseous mixture flow through a long tube at arbitrary Knudsen number, *J. Vac. Sci. Technol. A* **20,** 814–822 (2002).

[40] S. Naris, D. Valougeorgis, D. Kalempa, and F. Sharipov, Gaseous mixture flow between two parallel plates in the whole range of the gas rarefaction, *Physica A* **336**, 294–318 (2004).

[41] S. Naris, D. Valougeorgis, F. Sharipov, and D. Kalempa, Discrete velocity modelling of gaseous mixture flows in MEMS, *Superlattices and Microstructures* **35**, 629–643 (2004).

[42] S. Naris, D. Valougeorgis, D. Kalempa, and F. Sharipov, Flow of gaseous mixtures through rectangular microchannels driven by pressure, temperature and concentration gradients, *Phys. Fluids* **17**, 100607 (2005).

[43] F. Sharipov and D. Kalempa, Separation phenomena for gaseous mixture flowing through a long tube into vacuum, *Phys. Fluids* **17**, 127102 (2005).

[44] F. Sharipov, Rarefied gas flow into vacuum through a thin orifice. Influence of the boundary conditions, *AIAA Journal* **40**, 2006–2008 (2002).

[45] F. Sharipov, Numerical simulation of rarefied gas flow through a thin orifice, *J. Fluid Mech.* **518**, 35–60 (2004).

[46] A. K. Sreekanth, Transition flow through short circular tubes, *Phys. Fluids* **8**, 1951–1956 (1965).

[47] W. Jitschin, M. Ronzheimer, and S. Khodabakhshi, Gas flow measurement by means of orifices and Venturi tubes, *Vacuum* **53**, 181–185 (1999).

[48] F. Sharipov, P. Fahrenbach, and A. Zipp, Numerical modeling of Holweck pump, *J. Vac. Sci. Technol. A* **23**, 1331–1329 (2005).

[49] C. E. Siewert and F. Sharipov, Model equations in rarefied gas dynamics: Viscous-slip and thermal-slip coefficients, *Phys. Fluids* **14**, 4123–4129 (2002).

[50] L. B. Barichello, M. Camargo, P. Rodrigues, and C. E. Siewert, Unified solution to classical flow problem based on the BGK model, *Z. Angew. Math. Phys. (ZAMP)* **52**, 517–534 (2001).

[51] C. E. Siewert and J. R. Thomas, Half-space problems in the kinetic theory of gases, *Phys. Fluids* **16**, 1557–1559 (1973).

[52] S. Takata, S. Yasuda, S. Kosuge, and K. Aoki, Numerical analysis of thermal-slip and diffusion-slip flows of a binary mixture of hard-sphere molecular gases, *Phys. Fluids* **15**, 3745–3766 (2003).

[53] S. Yasuda, S. Takata, and K. Aoki, Numerical analysis of the shear flow of a binary mixture of hard-sphere gases over a plane wall, *Phys. Fluids* **16**, 1989–2003 (2004).

[54] L. B. Barichello and C. E. Siewert, The temperature-jump problem in rarefied-gas dynamics, *Eur. J. Appl. Math.* **11**, 353–364 (2000).

[55] S. K. Loyalka, Temperature jump - rigid-sphere gas with arbitrary gas surface interaction, *Nucl. Sci. Eng* **108**, 69–73 (1979).

[56] Y. Sone, T. Ohwada, and K. Aoki, Temperature jump and Knudsen layer in a rarefied gas over a plane wall: Numerical analysis of the linearized Boltzmann equation for hard-sphere molecules, *Phys. Fluids A* **1**, 363–370 (1989).

[57] S. K. Loyalka, Temperature jump in a gas mixture, *Phys. Fluids* **17**, 897–899 (1974).

[58] C. E. Siewert, The McCormack model for gas mixtures: The temperature jump problem, *Z. Angew. Math. Phys. (ZAMP)* **56**, 273–292 (2005).

[59] C. E. Siewert and D. Valougeorgis, The temperature-jump problem for a mixture of two gases, *J. Quant. Spectrosc. Radiat. Transfer* **70**, 307–319 (2001).

[60] F. Sharipov, Application of the Cercignani–Lampis scattering kernel to calculations of rarefied gas flows. I. Plane flow between two parallel plates, *Eur. J. Mech. - B/Fluids* **21**, 113–123 (2002).

[61] F. Sharipov, Application of the Cercignani–Lampis scattering kernel to calculations of rarefied gas flows. III. Poiseuille flow and thermal creep through a long tube, *Eur. J. Mech. - B/Fluids* **22**, 145–154 (2003).

[62] L. B. Barichello and C. E. Siewert, A discrete-ordinates solution for Poiseuille flow in a plane channel, *Z. Angew. Math. Phys. (ZAMP)* **50**, 972–981 (1999).

[63] C. Cercignani, M. Lampis, and S. Lorenzani, Plane Poiseuille flow with symmetric and nonsymmetric gas–wall interactions, *Transp. Theory Stat. Phys.* **33**, 545–561 (2004).

[64] K. Aoki, Y. Sone, S. Takata, K. Takahashi, and G. A. Bird, One-way flow of a rarefied gas induced in a circular pipe with a periodic temperature distribution, Preprint, Kyoto University (2000) .

[65] V. G. Chernyak, V. V. Kalinin, and P. E. Suetin, The kinetic phenomena in nonisothermal motion of a binary gas mixture through a plane channel, *Int. J. Heat Mass Transfer* **27**, 1189–1196 (1984).

[66] C. Cercignani and F. Sharipov, Gaseous mixture slit flow at intermediate Knudsen numbers, *Phys. Fluids A* **4**, 1283–1289 (1992).

[67] F. Sharipov, Rarefied gas flow through a slit: Influence of the gas–surface interaction, *Phys. Fluids* **8**, 262–268 (1996).

[68] F. Sharipov, Non-isothermal rarefied gas flow through a slit, *Phys. Fluids* **9**, 1804–1810 (1997).

[69] C. Cercignani and F. Sernagiotto, Cylindrical Couette flow of a rarefied gas, *Phys. Fluids.* **10**, 1200–1204 (1967).

[70] K. Nanbu, Analysis of cylindrical Couette flows by use of the direction simulation method, *Phys. Fluids* **27**, 2632–2635 (1984).

[71] F. M. Sharipov and G. M. Kremer, Nonlinear Couette flow between two rotating cylinders, *Transp. Theory Stat. Phys.* **25**, 217–229 (1996).

[72] F. M. Sharipov and G. M. Kremer, Linear Couette flow between two rotating cylinders, *Eur. J. Mech. - B/Fluids* **15**, 493–505 (1996).

[73] W. Marques Jr, G. M. Kremer, and F. M. Sharipov, Couette flow with slip and jump boundary conditions, *Continuum Mech. Thermodyn.* **16**, 379–386 (2000).

[74] Y. Sone, S. Takata, and T. Ohwada, Numerical analysis of the plane Couette flow of a rarefied gas on the basis of the linearized Boltzmann equation for hard-sphere molecules, *Eur. J. Mech. - B/Fluids* **9**, 273–288 (1990).

[75] D. Valougeorgis, Couette flow of a binary gas mixture, *Phys. Fluids* **31**, 521–524 (1988).

[76] F. Sharipov, L. M. G. Cumin, and D. Kalempa, Plane Couette flow of binary gaseous mixture in the whole range of the Knudsen number, *Eur. J. Mech. - B/Fluids* **23**, 899–906 (2004).

[77] C. E. Siewert, Couette flow for a binary gas mixture, *J. Quant. Spectrosc. Radiat. Transfer* **70**, 321–332 (2001).

Introduction to cryogenics

Ph. Lebrun
CERN, Geneva, Switzerland

Abstract
This paper aims at introducing cryogenics to non-specialists. It is not a cryogenics course, for which there exist several excellent textbooks mentioned in the bibliography. Rather, it tries to convey in a synthetic form the essential features of cryogenic engineering and to raise awareness on key design and construction issues of cryogenic devices and systems. The presentation of basic processes, implementation techniques, and typical values for physical and engineering parameters is illustrated by applications to helium cryogenics.

1 Low temperatures in science and technology

Cryogenics is defined as *that branch of physics which deals with the production of very low temperatures and their effect on matter* [1], a formulation which addresses aspects both of attaining low temperatures which do not naturally occur on Earth, and of using them for the study of nature or in industry. In a more operational way [2], it is also defined as *the science and technology of temperatures below 120 K*. The reason for this latter definition can be understood by examining characteristic temperatures of cryogenic fluids (Table 1): the limit temperature of 120 K comprehensively includes the normal boiling points of the main atmospheric gases, as well as of methane which constitutes the principal component of natural gas. Today, liquid natural gas (LNG) constitutes one of the largest—and fastest-growing—industrial domains of application of cryogenics, together with the liquefaction and separation of air gases. The densification by condensation, and separation by distillation of gases was historically—and remains today—the main driving force for the cryogenic industry, exemplified not only by liquid oxygen and nitrogen used in chemical and metallurgical processes, but also by the cryogenic liquid propellants of rocket engines and the proposed use of hydrogen as a 'clean' energy vector in transportation.

Table 1: Characteristic temperatures of cryogenic fluids [K]

Cryogen	Triple point	Normal boiling point	Critical point
Methane	90.7	111.6	190.5
Oxygen	54.4	90.2	154.6
Argon	83.8	87.3	150.9
Nitrogen	63.1	77.3	126.2
Neon	24.6	27.1	44.4
Hydrogen	13.8	20.4	33.2
Helium	2.2*	4.2	5.2

* λ point.

The quest for low temperatures, however, finds its origin in early thermodynamics, with Amontons's gas pressure thermometer (1703) opening the way for the concept of absolute zero inferred a century later by Charles and Gay-Lussac, and eventually formulated by Kelvin. It is,

however, with the advent of Boltzmann's statistical thermodynamics in the late nineteenth century that temperature—until then a phenomenological quantity—could be explained in terms of microscopic structure and dynamics. Consider a thermodynamic system in a macrostate which can be obtained by a multiplicity W of microstates. The entropy S of the system was postulated by Boltzmann as

$$S = k_B \cdot \ln W \tag{1}$$

with $k_B \simeq 1.38\ 10^{-23}$ J/K. Adding heat dQ to the system produces a change of its entropy dS, with a proportionality factor T, temperature

$$T = \frac{\mathrm{d}Q}{\mathrm{d}S} \ . \tag{2}$$

Thus a low-temperature system can be defined as one to which a minute addition of heat produces a large change in entropy, i.e., a large change in its range of possible microscopic configurations.

Boltzmann also found that the average thermal energy of a particle in a system in equilibrium at temperature T is

$$E \sim k_B\, T \ . \tag{3}$$

Consequently, a temperature of 1 K is equivalent to a thermal energy of 10^{-4} eV or 10^{-23} J per particle. A temperature is therefore low for a given physical process when $k_B\, T$ is small compared to the characteristic energy of the process considered. Cryogenic temperatures thus reveal phenomena with low characteristic energy (Table 2), and enable their application when significantly lower than the characteristic energy of the phenomenon of interest. From Tables 1 and 2, it is clear that 'low-temperature' superconductivity requires helium cryogenics.

Table 2: Characteristic temperatures of low-energy phenomena

Phenomenon	**Temperature [K]**
Debye temperature of metals	few 100
High-temperature superconductors	~ 100
Low-temperature superconductors	~ 10
Intrinsic transport properties of metals	< 10
Cryopumping	few
Cosmic microwave background	2.7
Superfluid helium 4	2.2
Bolometers for cosmic radiation	< 1
Low-density atomic Bose–Einstein condensates	~ 10^{-6}

2 Cryogenic fluids

The simplest way of cooling equipment with a cryogenic fluid is to make use of its latent heat of vaporization, e.g., by immersion in a bath of boiling liquid. As a consequence, the useful temperature range of cryogenic fluids is that in which there exists latent heat of vaporization, i.e., between the triple point and the critical point, with a particular interest in the normal boiling point, i.e., the saturation temperature at atmospheric pressure. This data is given in Table 1. In the following, we shall concentrate on two cryogens: helium which is the only liquid at very low temperature, and nitrogen for its wide availability and ease of use for pre-cooling equipment and for thermal shielding.

To develop a feeling about properties of these cryogenic fluids, it is instructive to compare them with those of water (Table 3). In both cases, but particularly with helium, applications operate much closer to the critical point, i.e., in a domain where the difference between the liquid and vapour phases is much less marked: the ratio of liquid to vapour densities and the latent heat associated with the change of phase are much smaller. Owing to the low values of its critical pressure and temperature, helium can be used as a cryogenic coolant beyond the critical point, in the supercritical state. It is also interesting to note that, while liquid nitrogen resembles water as concerns density and viscosity, liquid helium is much lighter and less viscous. This latter property makes it a medium of choice for permeating small channels inside magnet windings and thus stabilizing the superconductor.

Table 3: Properties of helium and nitrogen compared to water

Property	Helium	Nitrogen	Water
Normal boiling point [K]	4.2	77	373
Critical temperature [K]	5.2	126	647
Critical pressure [bar]	2.3	34	221
Liquid density* [kg/m^3]	125	808	960
Liquid/vapour density ratio*	7.4	175	1600
Heat of vaporization* [kJ/kg]	20.4	199	2260
Liquid viscosity* [μPl]	3.3	152	278

* At normal boiling point.

The factor of ten in latent heat of vaporization between helium and nitrogen, combined with the lower density of the former, induces a large difference in vaporization rates under the same applied heat load (Table 4). This illustrates the need for implementing much better insulation techniques in liquid helium vessels to achieve comparable holding times.

Table 4: Vaporization of liquid helium and liquid nitrogen at normal boiling point under 1 W applied heat load

Cryogen	[mg/s]	[l/h liquid]	[l/min gas NTP]
Helium	48	1.38	16.4
Nitrogen	5	0.02	0.24

For both fluids, the sensible heat of the vapour over the temperature range from liquid saturation to ambient is comparable to or larger than the latent heat of vaporization. This provides a valuable cooling potential at intermediate temperature, which can be used for thermal shielding or for pre-cooling of equipment from room temperature. The heat balance equation for cooling a mass of, say iron m_{Fe} of specific heat $C_{\mathrm{Fe}}(T)$ at temperature T by vaporizing a mass dm of cryogenic liquid at saturation temperature T_{V}, latent heat of vaporization L_{V} and vapour specific heat C (taken as constant), assumes perfect heat exchange with the liquid and the vapour

$$m_{\mathrm{Fe}}\, C_{\mathrm{Fe}}(T)\, \mathrm{d}T = [L_{\mathrm{V}} + C\,(T - T_{\mathrm{V}})]\, \mathrm{d}m\,. \tag{4}$$

Hence the specific liquid cryogen requirement for cool-down from temperature T_0

$$\frac{m}{m_{\mathrm{Fe}}} = \int_{T_0}^{T} \frac{C_{\mathrm{Fe}}(T)\, \mathrm{d}T}{L_{\mathrm{v}} + C\,(T - T_{\mathrm{v}})}\,. \tag{5}$$

The term $C\,(T - T_v)$ adding to L_v in the denominator brings a strong attenuation to the specific liquid requirement, provided there is good heat exchange between the solid and the escaping vapour. Calculated values of specific liquid cryogen requirements for iron are given in Table 5, clearly demonstrating the interest of recovering the sensible heat of helium vapour, as well as that of pre-cooling equipment with liquid nitrogen.

Table 5: Volume [l] of liquid cryogens required to cool down 1 kg of iron

Using	Latent heat only	Latent heat and enthalpy of vapour
Liquid helium from 290 K to 4.2 K	29.5	0.75
Liquid helium from 77 K to 4.2 K	1.46	0.12
Liquid nitrogen from 290 K to 77 K	0.45	0.29

Typical operating domains with cryogenic helium are shown in Fig. 1, superimposed on the—peculiar—phase diagram of the substance: the solid phase only exists under pressure and the normal liquid He I undergoes below 2.2 K a transition to another liquid phase, He II, instead of solidifying. There is no latent heat associated with this phase transition, but a peak in the specific heat, the shape of which gave the name 'λ-line' to the phase boundary. He II exhibits superfluidity, a macroscopic quantum behaviour entailing very high thermal conductivity and very low viscosity. While operating in saturated He I provides fixed (saturation) temperature and high boiling heat transfer at moderate heat flux, it may develop instabilities in two-phase flow and is prone to boiling crisis above the peak nucleate boiling flux (about 1 W/cm^2). The use of mono-phase supercritical helium in forced-flow systems avoids the problems of two-phase flow. However, the strongly varying properties of the fluid in the vicinity of the critical point may create other issues, such as density wave oscillations. More fundamentally, supercritical helium exhibits no latent heat, so that applied heat loads result in temperature increases which must be contained by high flow-rate or periodic re-cooling in extended systems. At lower temperature, He II demonstrates excellent transport properties which make it a coolant of choice for advanced superconducting devices [3]. Besides the thermodynamic penalty of lower temperature, the use of He II imposes that at least part of the cryogenic circuits operate at sub-atmospheric pressure, thus requiring efficient compression of low-pressure vapour and creating risks of dielectric breakdown and contamination by air in-leaks.

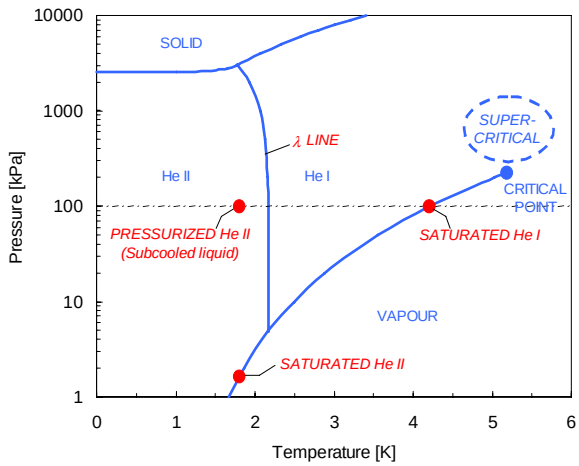


Fig. 1: Phase diagram of helium, showing typical operating domains

Thermo-physical properties of cryogenic fluids are available from tables, graphs and software running on personal computers, a selection of which are listed in the bibliography.

3 Heat transfer and thermal design

With the exception of superfluid helium, the heat transfer processes at work in cryogenics are basically the same as for any engineering temperature range. The strong variation of thermal properties of materials and fluids at low temperature, however, has two consequences: the relative and absolute magnitudes of the processes may be very different from those at room temperature, and the equations become non-linear, thus requiring numerical integration. Cryogenic thermal design is the art of using these processes adequately, either for achieving thermal insulation (cryostats, transfer lines) or for improving thermal coupling between equipment and coolant (cool-down and warm-up, thermal stabilization, thermometry) [4].

3.1 Solid conduction

Heat conduction in solids is represented by Fourier's law, expressing proportionality of heat flux with thermal gradient

$$Q = k(T)\, A \frac{\mathrm{d}T}{\mathrm{d}x}. \tag{6}$$

This equation also defines the thermal conductivity $k(T)$ of the material, which varies with temperature. Conduction along a solid rod of length L, cross-section A spanning a temperature range $[T_1, T_2]$, e.g., the support strut of a cryogenic vessel, is then given by the integral form

$$Q = \frac{A}{L} \int_{T_1}^{T_2} k(T)\ \mathrm{d}T\ . \tag{7}$$

Thermal conductivity integrals

$$\int_{T_1}^{T_2} k(T)\ \mathrm{d}T$$

of standard materials are tabulated in the literature. A few examples are given in Table 6, showing the large differences between good and bad thermal conducting materials, the strong decrease of conductivity at low temperatures, particularly for pure metals, and the interest of thermal interception to reduce conductive heat in-leak in supports. As an example, the thermal conductivity integral of austenitic stainless steel from 80 K to vanishingly low temperature is nine times smaller than from 290 K, hence the benefit of providing a liquid-nitrogen-cooled heat sink on the supports of a liquid helium vessel.

Table 6: Thermal conductivity integrals of selected materials [W/m]

From vanishingly low temperature up to	**20 K**	**80 K**	**290 K**
OFHC copper	11 000	60 600	152 000
DHP copper	395	5890	46 100
Aluminium 1100	2740	23 300	72 100
2024 aluminium alloy	160	2420	22 900
AISI 304 stainless steel	16.3	349	3060
G-10 glass-epoxy composite	2	18	153

3.2 Radiation

Black-body radiation strongly and only depends on the temperature of the emitting body, with the maximum of the power spectrum given by Wien's law

$$\lambda_{\max} T = 2898 \ [\mu\text{m K}] \tag{8}$$

and the total power radiated given by the Stefan–Boltzmann law

$$Q = \sigma A T^4 \tag{9}$$

with the Stefan–Boltzmann constant $\sigma \approx 5.67\ 10^{-8}$ W m^{-2} K^{-4}. The dependence of the radiative heat flux on the fourth power of temperature makes a strong plea for radiation shielding of low-temperature vessels with one or several shields cooled by liquid nitrogen or cold helium vapour. Conversely, it makes it very difficult to cool equipment down to low temperature by radiation only: in spite of the 2.7 K background temperature of outer space (and notwithstanding the Sun's radiation and the Earth's albedo which can be avoided by proper attitude control), satellites or interplanetary probes can make use of passive radiators to release heat only down to about 100 K, and embarked active refrigerators are required to reach lower temperatures.

Technical radiating surfaces are usually described as 'gray' bodies, characterized by an emissivity $\varepsilon < 1$

$$Q = \varepsilon \sigma A T^4 \ . \tag{10}$$

The emissivity ε strictly depends on the material, surface finish, radiation wavelength and angle of incidence. For materials of technical interest, measured average values are found in the literature [5], a subset of which is given in Table 7. As a general rule, emissivity decreases at low temperature, for good electrical conductors and for polished surfaces. As Table 7 shows, a simple way to obtain this combination of properties is to wrap cold equipment with aluminium foil. Conversely, radiative thermal coupling requires emissivity as close as possible to that of a blackbody, which can be achieved in practice by special paint or adequate surface treatment, e.g., anodizing of aluminium.

Table 7: Emissivity of some technical materials at low temperature

	Radiation from 290 K, surface at 77 K	**Radiation from 77 K, surface at 4.2 K**
Stainless steel, as found	0.34	0.12
Stainless steel, mech. polished	0.12	0.07
Stainless steel, electro-polished	0.10	0.07
Stainless steel + aluminium foil	0.05	0.01
Aluminium, black anodized	0.95	0.75
Aluminium, as found	0.12	0.07
Aluminium, mech. polished	0.10	0.06
Aluminium, electro-polished	0.08	0.04
Copper, as found	0.12	0.06
Copper, mech. polished	0.06	0.02

The net heat flux between two 'gray' surfaces at temperature T_1 and T_2 is similarly given by

$$Q = E\sigma A\left(T_1^4 - T_2^4\right) \tag{11}$$

with the emissivity factor E being a function of the emissivities ε_1 and ε_2 of the surfaces, of the geometrical configuration, and of the type of reflection (specular or diffuse) between the surfaces. Its precise determination can be quite tedious, apart from the few simple geometrical cases of flat plates, nested cylinders, and nested spheres.

If an uncooled shield with the same emissivity factor E is inserted between the two surfaces, it will 'float' at temperature T_S given by the energy balance equation

$$Q_S = E\sigma A\left(T_1^4 - T_S^4\right) = E\sigma A\left(T_S^4 - T_2^4\right) \quad . \tag{12}$$

Solving for T_S yields the value of $Q_S = Q/2$: the heat flux is halved in the presence of the floating shield. More generally, if n floating shields of equal emissivity factor are inserted between the two surfaces, the radiative heat flux is divided by $n + 1$.

3.3 Convection

The diversity and complexity of convection processes cannot be treated here. Fortunately, in the majority of cases, the correlations established for fluids at higher temperature are fully applicable to the cryogenic domain [6], and reference is made to the abundant technical literature on the subject. In the case of forced convection, one should keep in mind that the high density and low viscosity of cryogenic fluids often result in flows with high Reynolds number *Re* and hence strong convection. The Nüsselt number *Nü* which characterizes the efficiency of convective heat transfer relative to conduction in the fluid, is an increasing function of the Prandtl *Pr* and Reynolds numbers, respectively representing the ratio of mass to heat transport, and the ratio of inertial to viscous forces

$$Nü = f\,(Pr, Re) \quad . \tag{13}$$

The case of natural convection at low temperature, however, deserves particular mention, as this mechanism, usually weak at room temperature except on very large scales, becomes dominant in cryogenic equipment. In this case, the Nüsselt number is an increasing function of the Prandtl and Grashof *Gr* numbers, with the latter representing the ratio of buoyancy to viscous forces

$$Nü = f\,(Pr, Gr) \quad . \tag{14}$$

For gases, while *Pr* is about constant and independent of temperature, *Gr* is proportional to the heated volume, temperature difference, and to the coefficient of volume thermal expansion which scales as $1/T$ in the ideal case. As a consequence, there may exist in laboratory helium cryostats strong natural convection processes with Grashof numbers up to the 10^{12} range, i.e., higher than those encountered in the general circulation of the Earth's atmosphere. This has been used by hydrodynamics specialists to study turbulent convection in extreme conditions. The cryogenic engineer sees it as a powerful mechanism for cooling equipment and homogenizing its temperature.

3.4 Gas conduction

Since J. Dewar's invention (1898) of the cryogenic vessel which bears his name, evacuated envelopes provide the best insulation against heat transport in gaseous media. At low pressure, convection becomes negligible and only residual gas conduction is at work. This process operates in two distinct regimes, depending upon the value of the mean free path of gas molecules ℓ relative to the typical distance d between the cold and warm surfaces.

The mean free path of gas molecules, as predicted by kinetic theory, scales with the square root of temperature and inversely with pressure and the square root of molar mass. It therefore becomes large at low pressure, high temperature, and for light gas species.

When $\ell << d$ corresponding to higher residual pressure, the probability of interaction of a given molecule with others before it travels distance d is high (viscous regime), and heat diffuses as in any continuous medium

$$Q = k(T)\, A \frac{\mathrm{d}T}{\mathrm{d}x} \quad . \tag{15}$$

Note that the thermal conductivity $k(T)$ of the gas is independent of pressure.

When $\ell >> d$ at low residual pressure, the molecular regime prevails and the heat transfer between two surfaces at temperatures T_1 and T_2 is given by Kennard's law

$$Q = A\, \alpha(T)\, \Omega\, P\, (T_2 - T_1) \quad . \tag{16}$$

where Ω is a parameter depending upon the gas species, and α is the 'accommodation coefficient' representing the thermalization of molecules on the surfaces; its value depends on T_1, T_2, the gas species, and the geometry of the facing surfaces. Note that the conductive heat flux in molecular regime is proportional to pressure P and independent of the spacing between the surfaces (and therefore not amenable to the concept of thermal conductivity). Typical values of heat flux by gas conduction at cryogenic temperature are given in Table 8.

Table 8: Typical values of heat flux to vanishingly low temperature between flat plates [W/m^2]

Black-body radiation from 290 K	401
Black-body radiation from 80 K	2.3
Gas conduction (100 mPa helium) from 290 K	19
Gas conduction (1 mPa helium) from 290 K	0.19
Gas conduction (100 mPa helium) from 80 K	6.8
Gas conduction (1 mPa helium) from 80 K	0.07
MLI (30 layers) from 290 K, pressure < 1 mPa	1–1.5
MLI (10 layers) from 80 K, pressure < 1 mPa	0.05
MLI (10 layers) from 80 K, pressure 100 mPa	1–2

3.5 Multi-layer insulation

Multi-layer insulation (MLI) is based on multiple reflecting shields wrapped around the cryogenic piece of equipment to be insulated, with the aim of benefiting from the $n + 1$ reduction factor in radiative heat in-leak. In practice, this is implemented in the form of aluminium or aluminized polymer films, with low packing density achieved by crinkling or by insertion of a net-type spacer between layers. The wrapping can be made by winding the layers and spacer *in situ*, or by pre-fabricated blankets installed and held in place by insulating fasteners.

In all cases, MLI is a complex thermal system, involving the combination of radiation, solid-contact conduction, and residual-gas conduction between layers. As a result, increasing the number of layers, while beneficial for cutting radiation, usually results in increased packing with more contacts and trapped residual gas between layers, two effects which increase heat transfer. In view of the non-linearity of these elementary processes, thermal optimization requires layer-to-layer modelling and efficient control of the critical parameters. In practice, performance is measured on test samples and measured data is available from an abundant literature. Typical values for some practical MLI systems are given in Table 8.

Of particular interest is the case of operation in degraded vacuum, where the heat in-leak by molecular conduction is directly proportional to the residual pressure. The presence of a multilayer system which segments the insulation space into many cells thermally in series significantly contains the increase in heat in-leak to the low-temperature surface (Table 8). In this respect, the multilayer system is no longer used for its radiative properties, but for the reduction of molecular gas conduction. In the extreme case of complete loss of vacuum in a liquid helium vessel, MLI also efficiently limits the heat flux which would otherwise be very high due to condensation of air on the cold wall, thus alleviating the requirements for emergency discharge systems.

3.6 Vapour-cooling of necks and supports

The enthalpy of cryogen vapour escaping from a liquid bath can be used to continuously intercept conduction heat along solid supports and necks connecting the cryogenic bath with the room temperature environment. Assuming perfect heat exchange between the escaping vapour and the solid, the energy balance equation reads

$$k(T)\,A\,\frac{\mathrm{d}T}{\mathrm{d}x} = Q_{\mathrm{v}} + m'\,C\,\left(T - T_{\mathrm{v}}\right) \tag{17}$$

where Q_{V} is the heat reaching the liquid bath and m' is the vapour mass flow-rate. In the particular case of self-sustained vapour cooling, i.e., when the vapour mass flow-rate m' precisely equals the boil-off from the liquid bath,

$$Q_{\mathrm{V}} = L_{\mathrm{V}}\,m' \ . \tag{18}$$

Combining Eqs. (17) and (18) and integrating yields the value of Q_{V}

$$Q_{\mathrm{v}} = \frac{A}{L}\int_{T_{\mathrm{v}}}^{T_0} \frac{k\left(T\right)}{1 + \left(T - T_v\right)\dfrac{C}{L_v}}\mathrm{d}T \ . \tag{19}$$

The denominator of the integrand clearly acts as an attenuation term for the conduction integral. Numerical results for helium and a few materials of technical interest appear in Table 9. If properly used, the cooling power of the vapour brings an attenuation of one to two orders of magnitude in the conductive heat in-leak.

Table 9: Attenuation of heat conduction between 290 K and 4 K by self-sustained helium vapour cooling [W/cm]

Material	Purely conductive regime	Self-sustained vapour cooling
ETP copper	1620	128
OFHC copper	1520	110
Aluminium 1100	728	39.9
Nickel 99% pure	213	8.65
Constantan	51.6	1.94
AISI 300 stainless steel	30.6	0.92

Vapour cooling can also be used for continuous interception of other heat loads than solid conduction. In cryogenic storage and transport vessels with vapour-cooled shields, it lowers shield temperature and thus reduces radiative heat in-leak to the liquid bath. In vapour-cooled current leads,

a large fraction of the resistive power dissipation by Joule heating is taken by the vapour flow, in order to minimize the residual heat reaching the liquid bath [7].

A worked-out example of how these diverse thermal insulation techniques are implemented in a real design is given in Ref. [8].

4 Refrigeration and liquefaction

Refrigeration and liquefaction of gases are historically at the root of cryogenics, as they constitute the enabling technology which gave access to the low-temperature domain. They have developed over the years along several lines, to become a specialized subject which would deserve a thorough presentation. In the following, we shall briefly describe the basic thermodynamics, the cooling processes at work, and the corresponding equipment in the case of helium. For a more complete review, see Ref. [9].

4.1 Thermodynamics of refrigeration

A refrigerator is a machine raising heat Q_i from a low-temperature source T_i to a higher-temperature sink (usually room temperature) T_0, by absorbing mechanical work W_i; in doing so, it rejects heat Q_0 (see Fig. 2). These quantities are related through the application of the first (Joule) and second (Clausius) principles of thermodynamics

$$Q_0 = Q_i + W_i \tag{20}$$

$$\frac{Q_0}{T_0} \geq \frac{Q_i}{T_i} \,. \tag{21}$$

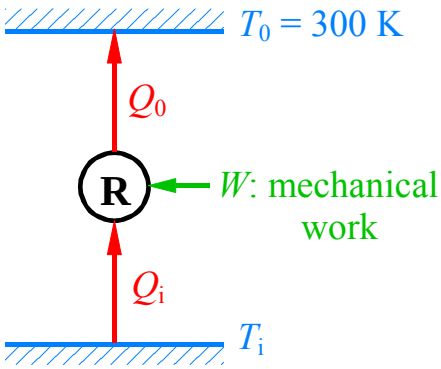


Fig. 2: Thermodynamic scheme of a refrigerator

In Eq. (21), the equality applies to the case of reversible process. From the above

$$W_i \geq T_0 \frac{Q_i}{T_i} - Q_i \,. \tag{22}$$

This expression can be written in three different ways. Introducing the reversible entropy variation $\Delta S_i = Q_i/T_i$,

$$W_i \geq T_0 \, \Delta S_i - Q_i \,. \tag{23a}$$

Another form isolates the group $(T_0/T_i - 1)$ as the proportionality factor between Q_i and W_i, i.e., the minimum specific refrigeration work

$$W_i \geq Q_i \left(\frac{T_0}{T_i} - 1 \right) \quad . \tag{23b}$$

As Carnot has shown in 1824, this minimum work can only be achieved through a cycle constituted of two isothermal and two adiabatic transforms (Carnot cycle). All other thermodynamic cycles entail higher refrigeration work for the same refrigeration duty.

A third form of Eq. (22) is

$$W_i \geq \Delta E_i \ . \tag{23c}$$

This introduces the variation of exergy $\Delta E_i = Q_i\ (T_0/T_i - 1)$, a thermodynamic function representing the maximum mechanical work content (Gouy's 'énergie utilisable') of a heat quantity Q_i at temperature T_i, given an environment at temperature T_0 .

Equation (23b) enables one to calculate the minimum mechanical power needed to extract 1 W at 4.5 K (saturated liquid helium temperature at 1.3 bar pressure, i.e., slightly above atmospheric) and reject it at 300 K (room temperature), yielding a value of 65.7 W. This is the power that would be absorbed by a refrigerator operating on a Carnot cycle between 4.5 K and 300 K. In practice, the best practical cryogenic helium refrigerators have an efficiency of about 30% with respect to a Carnot refrigerator, hence a specific refrigeration work of about 220.

Cryogenic refrigerators are often required to provide cooling duties at several temperatures or in several temperature ranges, e.g., for thermal shields or continuous heat interception (see Section 3.6 above). Equation (23b) can then be applied to the cooling duty at every temperature and every elementary mechanical power W_i summed or integrated in the case of continuous cooling. This also allows comparison of different cooling duties in terms of required mechanical work.

4.2 Helium refrigerators vs. liquefiers

A 4.5 K helium refrigerator absorbs heat isothermally at this temperature, usually by re-condensing cold helium vapour at saturation (saturation pressure 1.3 bar). By contrast, a liquefier also eventually condenses cold helium vapour at saturation, but starting from gaseous helium at 300 K which it must first pre-cool to 4.5 K. From Eq. (23a), the minimum mechanical power for helium liquefaction is

$$W_{liq} = W_{condens} + W_{precool} \tag{24}$$

$$W_{liq} = T_0\ \Delta S_{condens} - Q_{condens} + T_0\ \Delta S_{precool} - Q_{precool} \ . \tag{25}$$

The heat quantities $Q_{condens}$ and $Q_{precool}$ exchanged at constant pressure are given by the enthalpy variations $\Delta H_{condens}$ and $\Delta H_{precool}$. With $T_0 = 300$ K and the entropy and enthalpy differences taken from thermodynamic tables, one finds $W_{liq} = 6628$ W per g/s of helium liquefied. Given the minimum specific mechanical work of 65.7 at 4.5 K, this yields an approximate equivalence of about 100 W at 4.5 K for 1 g/s liquefaction. More precisely, a liquefier producing 1 g/s liquid helium at 4.5 K will absorb the same power (and thus have similar size) as a refrigerator extracting about 100 W at 4.5 K, provided they both have the same efficiency with respect to the Carnot cycle. For machines with mixed refrigeration and liquefaction duties, this equivalence can be approximately verified by trading some liquefaction against refrigeration around the design point and *vice versa*.

4.3 Real cycles and refrigeration equipment

So far we have addressed cryogenic refrigeration and liquefaction only through thermodynamics, i.e., through the exchanges of mass, heat, and work at the boundaries of machines seen as 'black boxes'. We shall now consider cycles, cooling methods, and equipment of real refrigerators.

In order to minimize the specific mechanical work requirement (and hence the size and power consumption), an efficient refrigerator should try and approximate the Carnot cycle, which is

represented by a rectangle on the temperature–entropy diagram: the two isotherms are horizontal lines, while the two isentropic transforms are vertical lines. To liquefy helium, the base of the rectangle should intercept the liquid-vapour dome (Fig. 3).

However, superimposing this cycle on the temperature-entropy diagram of helium shows that one should operate at a high pressure of about 613 kbar (!), with a first isentropic compression from 1.3 bar to 82 kbar (!), followed by an isothermal compression. This is clearly impractical, and real helium cycles are elongated along isobar (or isochoric) lines, thus involving transforms which require heat exchange between the high- and low-pressure streams. This heat exchange can be performed in recuperative or regenerative heat exchangers, for continuous or alternating flows, respectively. In the following, we focus on the continuous-flow cycles using recuperative heat exchangers which constitute the operating principles of large-capacity helium refrigerators and liquefiers.

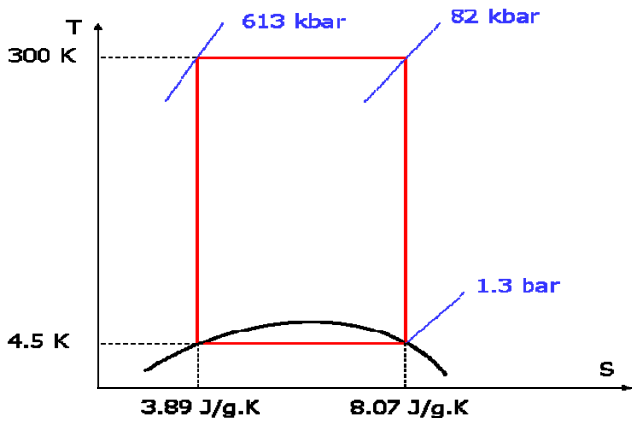


Fig. 3: A hypothetical Carnot cycle for helium liquefaction

Practical elementary cooling processes are shown on the temperature–entropy diagram in Fig. 4. Apart from the quasi-isobar cooling of the gas stream in a heat exchanger (segment AB_1), refrigeration can be produced by adiabatic (quasi-isentropic) expansion with extraction of mechanical work, usually in a gas turbine (segment AB'_2), and isenthalpic Joule–Thomson expansion in a valve or restriction (segment AB_3).

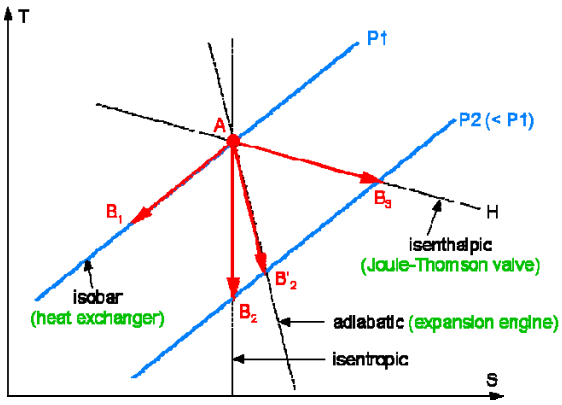


Fig. 4: Elementary cooling processes shown on temperature–entropy diagram

This latter process does not produce any cooling for ideal gases, the enthalpy of which is a sole function of temperature. For real gases, however, enthalpy depends both on temperature and pressure, so that isenthalpic expansion can produce warming or cooling, depending upon the slope of the isenthalps on the diagram. In order to cool the gas stream, Joule–Thomson expansion must start below a limit called the inversion temperature. The values of inversion temperature for cryogenic fluids (Table 10) show that while air can be cooled from room temperature by Joule–Thomson expansion

(the risk of freezing the pressure reducer on the air bottle is well known to scuba divers), helium must first be pre-cooled down to below its inversion temperature of 43 K. The moderate downward slope of isenthalps on the temperature–entropy diagram indicates that in any case, Joule–Thomson expansion generates substantial entropy. Its relative inefficiency with respect to adiabatic expansion is, however, accepted in view of the simplicity of its implementation, particularly when it results in partial condensation of the stream entailing two-phase flow conditions which would be difficult to handle in an expansion turbine.

Table 10: Maximum values of Joule–Thomson inversion temperature

Cryogen	Maximum inversion temperature [K]
Helium	43
Hydrogen	202
Neon	260
Air	603
Nitrogen	623
Oxygen	761

These elementary cooling processes are combined in practical cycles, a common example for helium refrigeration is provided by the Claude cycle and its refinements. A schematic two-pressure, two-stage Claude cycle is shown in Fig. 5: gaseous helium, compressed to HP in a lubricated screw compressor, is re-cooled to room temperature in water-coolers, dried and purified from oil aerosols down to the ppm level, before being sent to the HP side of the heat exchange line where it is refrigerated by heat exchange with the counter-flow of cold gas returning on the LP side. Part of the flow is tapped from the HP line and expanded in the turbines before escaping to the LP line. At the bottom of the heat exchange line, the remaining HP flow is expanded in a Joule–Thomson valve and partially liquefied.

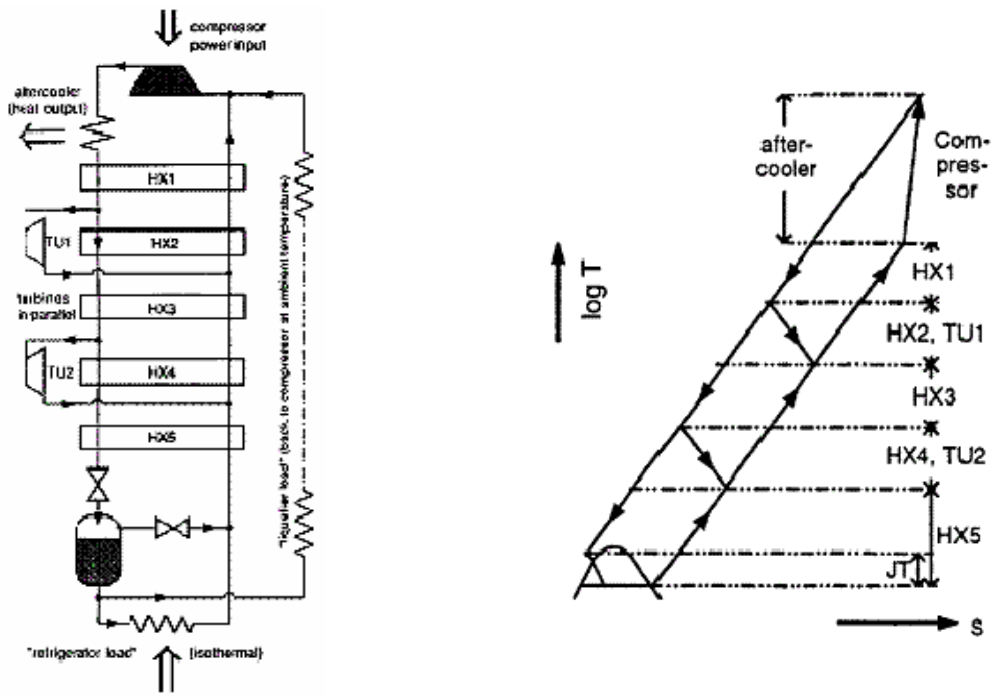


Fig. 5: Schematic example of two-pressure, two-stage Claude cycle: flow scheme (left) and T–S diagram (right)

Large-capacity helium refrigerators and liquefiers operate under this principle, however, with many refinements aiming at meeting specific cooling duties and improving efficiency and flexibility of operation, such as three- and sometimes four-pressure cycles, liquid nitrogen pre-cooling of the helium stream, numerous heat exchangers, many turbines in series or parallel arrangements, Joule–Thomson expansion replaced by adiabatic expansion in a 'wet' turbine, cold compressors to lower the refrigeration temperature below 4.5 K.

The capital cost of these complex machines is high, but scales less than linearly with refrigeration power, which favours large units. Operating costs are dominated by that of electrical energy, typically amounting to about ten per cent of the capital cost per year in case of quasi-continuous operation. For overall economy, it is therefore very important to seek high efficiency, which is also easier to achieve on large units. For a review of these aspects, see Ref. [10].

5 Conclusion

This brief paper has presented the basic ideas and principles of the most important aspects of cryogenics, i.e., cryogenic fluids, heat transfer, thermal design, and refrigeration. It has also provided the reader with typical numerical values of the relevant parameters, enabling him to perform order-of-magnitude estimates and apply his engineering judgement. There is of course much more to say on each of these topics, some of which have significantly developed over the years. Many other subjects not addressed here also pertain to cryogenic engineering, such as materials at low temperature, storage, handling and transfer of fluids, two-phase flow and discharge, vacuum and leak-tightness technology, instrumentation (in particular thermometry), process control, impurity control and safety. In all cases, the interested reader is referred to the selected bibliography for detailed information and to the proceedings of the cryogenic engineering conferences for recent developments.

Acknowledgements

The substance and the form of this course owe much to discussions with many colleagues who have provided useful inputs, particularly L. Tavian at CERN. While equations, tables and figures were carefully integrated in the text by E. Delucinge, any error which may remain in the document is evidently mine.

Selected bibliography

Introduction to low temperatures

K. Mendelssohn, *The Quest for Absolute Zero*, 2nd edition (Taylor & Francis, London, 1977).

J. Wilks and D.S. Betts, *An Introduction to Liquid Helium*, 2nd edition (Oxford Science Publications, Clarendon Press, Oxford, 1987).

R.G. Scurlock (ed.), *History and Origins of Cryogenics* (Clarendon Press, Oxford, 1992).

Ph. Lebrun, Cryogenics, key to advanced science and technology, *Bulletin of the IIR*, Paris, **LXXXIII** 2003-6 (2003) 4, http://www.iifiir.org/2enarticles_bull03_6.pdf .

Cryogenic engineering textbooks

R.B. Scott, *Cryogenic Engineering* (Van Nostrand, Princeton, 1959).

R.R. Conte, *Eléments de cryogéni*e (Masson, Paris, 1970).

G.G. Haselden, *Cryogenic Fundamentals* (Academic Press, London, 1971).

R.A. Barron, *Cryogenic Systems* (Oxford University Press, New York, 1985).

B.A. Hands, *Cryogenic Engineering* (Academic Press, London, 1986).

S.W. van Sciver, *Helium Cryogenics* (Plenum Press, New York, 1989).

K.D. Timmerhaus and T.M. Flynn, *Cryogenic Process Engineering* (Plenum Press, New York, 1989).

Fluid and material properties

R.D. Mc Carty, Thermodynamic properties of helium 4 from 2 to 1500 K at pressures to 10^8 Pa, *J. Chem. Phys. Ref. Data,* **2** (1973) 923.

W.C. Reynolds, *Thermodynamic Properties in SI: Graphs, Tables and Computational Equations for 40 Substances*, Department of Mechanical Engineering, Stanford University (1979).

Cryogenic Properties of Materials, NIST Cryogenic Technologies Group, www.cryogenics.nist.gov

V. Arp, *HEPAK, Thermophysical Properties of Helium*, www.cryodata.com

R.D. Mc Carty, *GASPAK, Thermophysical Properties of 36 fluids*, www.cryodata.com

Conference proceedings

Proceedings of the Cryogenic Engineering Conferences (CEC) and International Cryogenic Material Conferences (ICMC), *in* Advances in Cryogenic Engineering (multiple publishers).

Proceedings of the International Cryogenic Engineering Conferences (ICEC) (multiple publishers).

CERN Accelerator School proceedings

S. Turner (ed.), Proceedings of the CAS *Superconductivity in Particle Accelerators*, Hamburg, CERN 89-04, Geneva (1989).

S. Turner (ed.), Proceedings of the CAS *Superconductivity in particle accelerators*, Hamburg, CERN 96-03, Geneva (1996).

S. Russenschuck and G. Vandoni (eds.), Proceedings of the CAS *Superconductivity and Cryogenics for Accelerators and Detectors*, Erice, CERN-2004-008, Geneva (2004).

References

[1] *Oxford English Dictionary*, 2nd edition (Oxford University Press, 1989).

[2] *New International Dictionary of Refrigeration*, 3rd edition (IIR-IIF, Paris, 1975).

[3] Ph. Lebrun and L. Tavian, The technology of superfluid helium, CERN-2004-008, Geneva (2004) 375.

[4] G. Vandoni, Heat transfer, CERN-2004-008, Geneva (2004) 325.

[5] W. Obert *et al.*, Emissivity measurements of metallic surfaces used in cryogenic applications, *Adv. Cryo. Eng*. **27** (1982) 293.

[6] M.C. Jones and V. Arp, Review of hydrodynamics and heat transfer for large helium cooling systems, *in Advances in Refrigeration at the Lowest Temperatures,* IIR-IIF Commission A1-2, Zürich (1978) 41.

[7] Yu. L. Buyanov, Current leads for use in cryogenic devices, principle of design and formulae for design calculations, *Cryogenics* **25** (1985) 94.

[8] Ph. Lebrun, Design of a cryostat for superconducting accelerator magnets: the LHC main dipole case, CERN-2004-008, Geneva (2004) 348.

[9] U. Wagner, Refrigeration, CERN-2004-008, Geneva (2004) 295.

[10] S. Claudet *et al.*, Economics of large helium cryogenic systems: experience from recent projects at CERN, *Adv. Cryo. Eng*. **45B** (2000) 1301.

Design rules for vacuum chambers

C. Hauviller
CERN, Geneva, Switzerland

Abstract
Producing a vacuum chamber means defining the boundary conditions (inner and outer envelopes, operational constraints, etc.), choosing the material, designing the parts, manufacturing and assembling these parts together and putting them under vacuum. This paper gives the methodology, methods and hints for designing vacuum chambers.

1 Introduction

The first step in the mechanical design of a vacuum chamber is to clearly define what all the boundary conditions are and often this is not the easiest part of the job since most of the parameters are settled by the other systems making up the equipment, in the large sense, to be built. The vacuum chamber arrives late and should simply fit in what is left!

The material is a subject of long debate, still alive, even for 'usual' types of equipment.

The initial phase, the conceptual design, which usually does not need accurate studies, is followed by the detailed design, the latter strongly associated with the manufacturing, itself controlled by a quality assurance plan.

This paper on mechanical design is intended to go through all the steps mentioned above. It is difficult to be exhaustive and we shall not be, since what we deal with will be related not only to accelerator equipment on the beam lines but also to other vacuum vessels for services such as cryogenics. However, the vacuum vessels we consider are static, the ones which move being subject to specific rules not presented here.

Methodology, methods, and hints will be given, but consultation of some of the references is a must if you have to design such a system yourself. References on general mechanical design are not quoted here since there are many. Reference [1] from the 1999 CERN Accelerator School on Vacuum Technology could be very useful.

It is also important to remember that a classical rule, valid for any type of equipment, is that more than 70% of the final cost is already defined at the end of the design phase. Investing more during the initial phase is always rewarding!

2 Boundary conditions

2.1 Environment

To determine the environment is clearly the first step. The sub-sections are not necessarily exhaustive, but the list of physical phenomena influencing the design of vacuum chambers is probably complete.

2.1.1 Outside

If the external envelope is simply the volume where the chamber will sit, the situation should not be very difficult when it is located inside a large building, but fitting a vacuum chamber/beam pipe of a particle accelerator inside an optimized gap of magnets may be much more constraining. Adequate

space for supports will inevitably simplify the design of the chambers. In addition, space and access to the pumping ports and to the pumping and diagnostics equipment could also be very restricted leading to stringent consequences on the operation.

Make sure that you give your requirements before it is too late.

2.1.2 Inside

The beam envelope is the main parameter for the vacuum chamber of an accelerator, but the conductance is another one in case of lumped pumps: the pressure distribution between two pumping ports is parabolic and the maximum value of this parabola defines the operating pressure. Distributed pumping is a solution, at least partially. For a cryogenic system, the inner piping and its insulation layers define the inner envelope, but do not forget to include extra space to allow for the movements resulting from pressure and temperature variations.

2.1.3 Pressure and forces

The differential pressure on a vacuum chamber is obviously one bar during operation but it is recommended to check what all the intermediate steps are like, for example, an over-pressure test for qualification. In the presence of a closed end, the resulting forces could be large and the design should take into account the transmission of these forces to the fixing points through the vacuum chamber wall.

2.1.4 Temperature

Specifying operating temperature (usually room temperature), bake-out temperature (usually between 150ºC and 300ºC), and exceptional temperature in the case of an incident (like a cool-down due to cryogen losses) together with all the transitories is a must. The effects of temperature are dilatations, stresses and changes of material properties and they may have destructive effects if badly mastered. Slow temperature transients allow the resulting stresses to be minimized but it becomes an operational constraint that is not necessarily acceptable.

2.1.5 Specific cases

To be complete, one should go through cases of vacuum chambers submitted to specific constraints or dedicated to special applications.

Important local heat loads, like synchrotron radiation, should be evacuated through good thermal conducting materials. Electrical impedance is reduced with a good electrical conductivity but, on the other hand, non-conducting material is required for electric isolation.

Activation of material by the particles hampers the functioning of the equipment, delaying access to an accelerator after switching off and producing radioactive components that are delicate to handle. Elements with long half-lifetime must be avoided in the materials in order to reduce the activation level and minimize the decay time.

Should particles inside the vacuum chambers escape, either as a secondary beam going through a window to reach targets or generated by collisions and to be analysed by surrounding detectors, the parameters defining the transparency to particles, the radiation length (X_0), and the collision length must be considered.

2.2 Materials

Choosing the material for a vacuum chamber often leads to wide discussion of their principal properties. The material parameters design for the vacuum chamber in terms of mechanics could be

quite numerous, but there are only three main ones: the modulus of rigidity (Young's modulus: E), the elastic ($\sigma_{0.2}$) and rupture (σ_r) limits.

These parameters are usually easily available for any material. A simple analysis of traction/compression tests is a common practice. Figure 1 shows a typical shape of traction curves of metallic materials. It shows how to define the yield or elastic limit. The modulus of rigidity is obtained through the value of the slope of the quasilinear part of the curve.

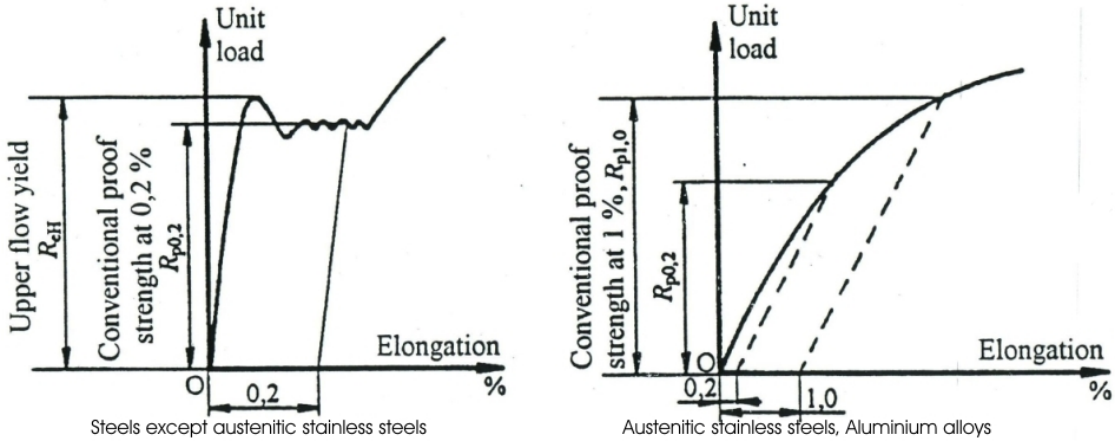


Fig. 1: Traction curves of metals

More sophisticated tests like biaxial ones are required only for very specific cases.

Other parameters like creep data, fatigue limit, fracture toughness could be useful for specific design but usually complicate the choice of material. These parameters decrease when the temperature increases, creep weakens the material and, as a rule of thumb, the upper temperature limit of usage of a metal is 70% of its melting temperature in degrees kelvin.

Another factor not to be forgotten is the vapour pressure which depends upon the temperature and the pressure. Materials like zinc, cadmium or magnesium, or standard resins which have significant vapour pressure at low temperature are not acceptable when a bake-out is foreseen.

Selection of materials for specific cases is a multi-parameters problem. It can be eased by the use of non-dimensional parameters [2]. These have been used for the specific type of vacuum chambers, which is the beam pipes for the experiments installed in colliders [3], LEP [4], the LHC [5]. As mentioned above, they should be transparent to particles. It has been shown that the selection can be based on the non-dimensional parameter ($X_0E^{1/3}$) that gives a figure of merit for the materials (Table 1) and, for this specific case of beam pipes, beryllium and carbon fibre composite outrun by a large factor aluminium, titanium, and steel.

Table 1: Figures of merit of materials in terms of transparency

	Be	**CFC**	**Al-Be**	**Al**	**Ti**	**Fe**
E (GPa)	290	200	193	70	110	210
X_0 (m)	0.353	0.271	0.253	0.089	0.036	0.018
$X_0E^{1/3}$	2.34	1.58	1.46	0.37	0.17	0.11

A good electrical or thermal conductivity could be required or banished. Copper and aluminium are the usual candidates for the former case and ceramics, like aluminium oxide, for the latter one.

Recent major advances in metallurgy and surface physics allow the creation of more optimized material which exhibits specific properties. The plain material is replaced by a combination of materials, each with a specific function with a better physical parameter. Two examples among many combinations are the adjunction of a layer of a conductive material (copper) on a structural material (stainless steel) and, as a barrier to gas diffusion, a thin layer of aluminium on a structural material (carbon fibre composite).

Besides the physical characteristics, specific technological properties should not be forgotten. Leak tightness of a vacuum chamber is a must but this may be difficult to obtain if the weldability of the material is poor and leads to the use of sophisticated and expensive techniques.

Cost is the final criterion. Besides the material quality, in particular the cleanliness in terms of inclusions and impurities required and that one cannot necessarily afford on a large scale, the availability is a predominant factor. It is strongly recommended to favour materials of general use in industrial application.

Raw material is not the single parameter for the cost criterion since a complex and expensive manufacturing process could totally hamper the final cost. But machining could be efficient and fast and therefore cheap, although a large amount of costly raw material is lost as chips. Moreover the precision required for a welded part could need expensive tooling not included in the initial estimate.

The most common materials are austenitic stainless steel (316LN, 316L, 304L), aluminium alloys (5000 and 6000 series) and copper (OFHC, Glidcop).

2.3 Legislation and codes

It is quite difficult to define a clear basis for the regulations to be applied to the design and construction of vacuum chambers. These vacuum chambers are generally special but they should behave properly and should be safe. It is important that guidelines at least should be provided. But this is not strictly the case since, to the knowledge of the author, no construction code dedicated to vacuum vessels exists but there is one for cryogenic vessels.

One's own regulations should therefore be applied and the common practice is to treat a vacuum chamber as a pressure vessel with some arrangements. This is in particular the case at CERN with the Safety Code D2 for pressure vessels and pressurized pipelines [6].

European legislation is now based upon directive 97/23/EC of the European Parliament and of the Council of 29 May 1997 on the approximation of the laws of the Member States concerning pressure equipment. "This directive applies to equipment subject to a maximum allowable pressure exceeding 0.5 bar" which "does not pose a significant hazard".

Based on this directive, standards, so-called norms or codes, are being prepared by qualified and specialized European and national bodies. They are intended to define common rules accepted by clients, manufacturers, technical centres, and inspection bodies.

These construction/safety codes have been developed for decades, before directive 97/23/EC, and are quite similar to each other and also to other international standards. These thick documents generally contain rules for materials, design, manufacturing, inspection and testing, certification of pressure vessels, including those submitted to external pressure.

The classical ones are:

- CODAP 2005, French code for construction of unfired pressure vessels issued by the Syndicat National de la Chaudronnerie, de la Tôlerie et de la Tuyauterie industrielle (SNCT) [7];
- EN 13445, Unfired pressure vessels issued by the European Committee for Standardization (CEN) [8];

– 2004 BPVC, Boiler and Pressure Vessels Code, Section VIII – Rules for Construction of Pressure Vessels, issued by the American Society of Mechanical Engineers (ASME).

There should also be a special mention for the European standard EN 13458-2 [9] which treats specifically the "static vacuum insulated vessels" of the "cryogenic vessels" in a way very similar to the codes mentioned above.

3 Design

3.1 Basics

Stresses generated by the loads enumerated before should usually remain in the elastic range. This means that the equivalent stress (according to the Von Mises or Tresca criterion for example) should not exceed the elastic limit ($\sigma_{0.2}$) anywhere in the structure. The directive 97/23/EC states in particular that "the permissible membrane stress for predominantly static loads…must not exceed…2/3 of the yield limit (stainless steel and aluminum alloys)".

But, vacuum usually corresponds to an external pressure and the resulting membrane stresses are compressive. Under these conditions, the membrane strain energy can be converted to bending strain energy, leading to an instability and a bifurcation point on the behavioural curve, a potential buckling (Fig. 2). Buckling is a non-linear phenomenon and it could be strongly influenced by the defects inherent to the manufacturing.

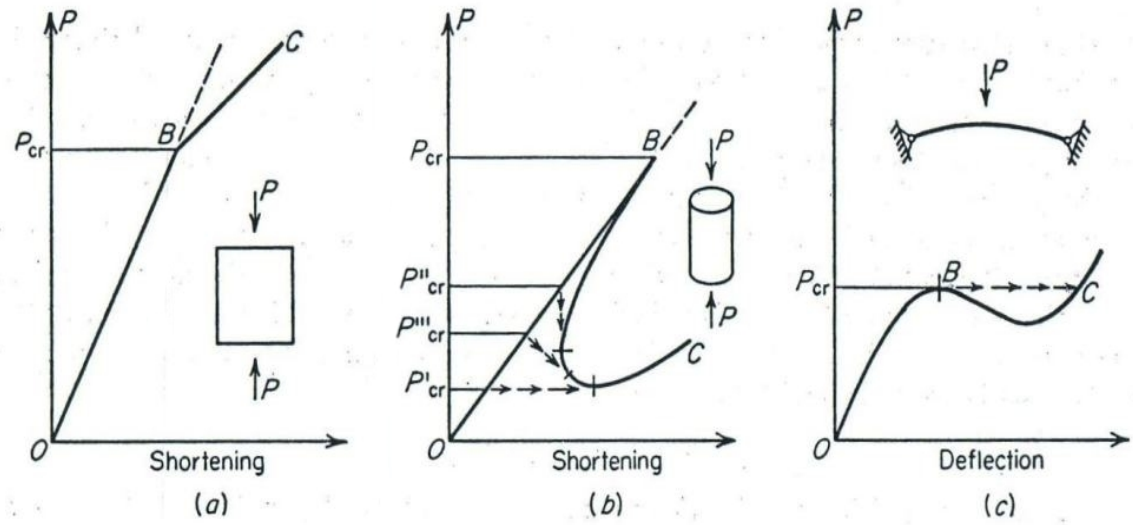


Fig. 2: Buckling behaviours

A safety factor must be applied to any computed buckling value: pressure vessel codes like CODAP use 3.0 but EN 13458-2 quotes only 2.0 for outer jackets.

In fact, in practice, it is quite easy to elaborate a design to stay in the elastic range of a material but buckling, a phenomenon factor, could be more difficult to deal with.

3.2 Methods and techniques

In the pre-design phases and when checking the final design, somewhat simple analytical methods could be more efficient than structural analysis programs. They permit a quick analysis of the effects of various geometrical parameters. Some of them will be quoted below.

However, the use of structural analysis packages, usually based on the Finite-Element Method (FEM) is a must for the detailed design of vacuum vessels. Two main families of elements ought to be used for these computations (Fig. 3): the solid elements for the 'thick' machined parts but the shell elements are more efficient for thin shell structure (t/R < 100). In the shell elements, the through-thickness stresses are assumed linear and integrated in membrane (constant term) and bending (linear term).

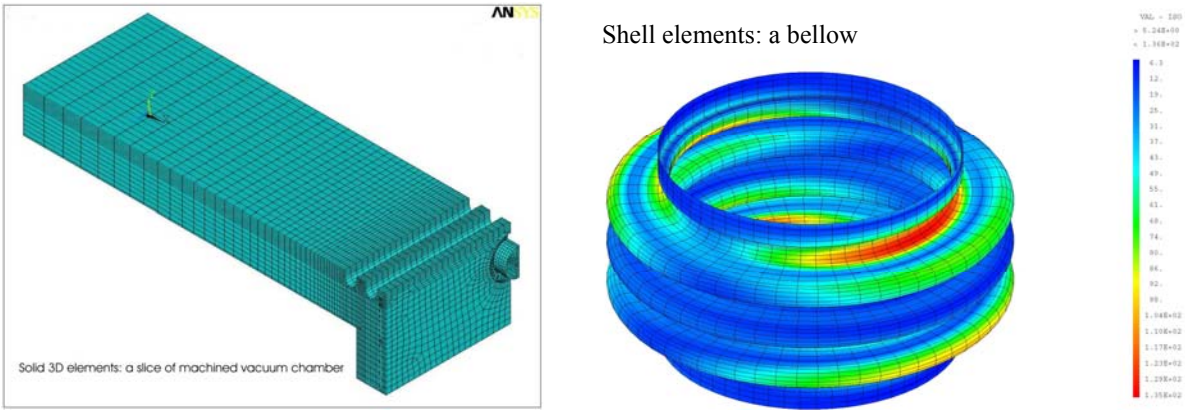


Fig. 3: Finite element models

The first step of the analysis is the linear elastic one which gives all the information on a structure: displacements, strains, and stresses. The second step is to search for buckling factors for the main modes (search for linear eigenvalues). A third one could be useful for thin shells where the non-linear pre-buckling analysis allows one to determine the shake-down factor on the linear buckling value; it could be as high as 10 for very thin shells.

3.3 Tubes

Tubes are obviously the most common shape for vacuum chambers.

The most stable ones are circular ones for which a quick analysis is quite straightforward.

The circumferential stress under external pressure p is simply:

$$\sigma_\theta = \frac{pR}{t} \ , \tag{1}$$

R and t being, respectively, the radius and the thickness of the circular tube.

If the tube is closed and subjected to an axial force F, the axial stress is

$$\sigma_z = \frac{pR}{2t} + \frac{F}{2\pi Rt} \tag{2}$$

and the Von Mises equivalent stress to be compared to the material maximum allowable stress is

$$\sigma_e = \frac{1}{2}\left[3\left(\frac{pR}{t}\right)^2 + \left(\frac{F}{\pi Rt}\right)^2\right]^{\frac{1}{2}} . \tag{3}$$

The buckling pressure can also be computed through analytical formulas, depending upon the geometrical parameters of the tube and the Young's modulus of the material. The most conservative one is for an infinite length of the tube:

$$p_{\mathrm{cr}} = \frac{0.25E}{1-\nu^2}\left(\frac{t}{R}\right)^3 , \tag{4}$$

ν being the Poisson ratio and $(1 - \nu^2)$ could be generally approximated by 0.9.

A rule of thumb is that the thickness of a stainless-steel circular tube should be at least one hundredth of its diameter (safety factor included).

A non-circular tube is more difficult to quickly estimate. However, quasi or approximately rectangular shapes can be approximated by plates or simply by beams (width of a unit length) either clamped or simply supported.

If approximated by a beam, the maximum deflection ($w_{\max}$) and stress ($\sigma_{\max}$) (upper and lower bounds) are

$$\frac{1}{32}\frac{pl^4}{Et^3} < w_{\max} < \frac{5}{32}\frac{pl^4}{Et^3} \tag{5}$$

$$\frac{1}{2}p\left(\frac{l}{t}\right)^2 < \sigma_{\max} < \frac{3}{4}p\left(\frac{l}{t}\right)^2 \qquad \text{(not at the same location)} , \tag{6}$$

l and t being, respectively, the span and the thickness of the tube.

Specific analytical programs have been developed to treat elliptical tubes.

3.4 Windows

Special care is required for these critical items. Designed to allow particle beams to escape the vacuum chambers without interaction, windows are thin and often manufactured with a special material.

The best shape for resisting external pressure is a spherical dome but the manufacture is difficult and the external ring must be quite rigid. If the window is oriented in such way that it can buckle, the classical buckling pressure is

$$p_{\mathrm{CL}} = \frac{2E}{\sqrt{3\left(1-\nu^2\right)}}\left(\frac{t}{R}\right)^2 , \tag{7}$$

R and t being, respectively, the radius of curvature and the thickness of the spherical dome.

A flat window is the other common option but, when put under pressure, the deflection ($w_{\max}$) is not negligible

$$w_{\max} = \frac{3}{16}\frac{p}{E\left(1-\nu^2\right)}\frac{R^4}{t^3} , \tag{8}$$

R and t being, respectively, the radius and the thickness of the circular window.

Non-circular windows exhibit high loads concentrated in the corners somewhat difficult to treat in case of a weld and a source of nightmares when clamped inside flanges.

3.5 Bellows

Bellows are another critical item which provide the capacity to minimize stresses due to displacements during commissioning and operation and to ease assembly of misaligned parts. Designed to bring flexibility, they are thin and inherently fragile. They are either mechanically formed or hydroformed from thin tubes, or assembled by welding a series of individual annular rings.

The design of bellows is treated in the general codes for pressure vessels but a specific code has been issued by the manufacturers, The Expansion Joint Manufacturers Association, Inc. (EJMA) [10], recognized as the authority on metallic bellows type expansion and a project norm pr-EN 14917 [11] was recently published by CEN.

Bellows are designed to withstand a given number of cycles of axial expansion/compression over the expected life of the system. They accept an angular movement but large offsets drastically decrease the fatigue limit. A double bellows 'à la cardan' (two bellows with a tube in between) is the recommended solution for the latter case if space exists. If the neighbouring tubes are not correctly supported, the bellows may buckle immediately or after cycling.

3.6 Special shapes

The imagination of the designer is without limits. Any shape can be designed and analysed with the available tools. But one should be careful with the analysis of the results, in particular with the interpretation of the local stresses (true or generated by the FEM mesh). Curved surfaces are preferable to flat panels wherever possible.

However, the inherent difficulties of manufacturing should not be forgotten. Major progress in extrusion technology now permits fancy shapes with multi-channels. Welded tubes could also be ring-stiffened to compensate for a thin wall or for a large width leading to large deflections.

4 Manufacturing and assembly

4.1 Machining and sheet metal work

Materials for manufacturing are available in various states: raw products like blanks or sheets, or semi-finished products like extruded elements (aluminium, copper), moulded, forged, or sintered (ceramics).

The choice is usually founded on cost but the final quality in terms of vacuum may be disturbed by the manufacturing techniques. Defects due to impurities internal to the material should not provoke leak-through and elaboration techniques properly handled help with that. Forging, extruding and laminating will squeeze the impurities but also enclose them. Machining will open them with a risk of leak, even with a good quality material. Laminated metal sheets will usually remain leak-tight except in the case of large deformations due to forming (e.g., deep drawing, 180° folding) or in the heat affected zones next to the welds.

4.2 Joining techniques

Joints are the main source of leaks in vacuum systems.

The first question to answer before choosing the assembly technique is: Is this assembly dismountable and, if yes, how often? This will help to choose knowledgeably between flanges and permanent sealing like welds. The cost in terms of money and space of flanges should be compared to that of a welding and subsequent cuttings and reweldings: parts, tooling, accessibility and operation costs.

4.2.1 Flanges

Flanges are industry standards available on the market. However, metal-sealed flanges can be developed for specific needs. The main parameters for their design are the strength of the material to withstand high forces for bolting or clamping, and the quality and hardness of the surface where the seal is positioned. The seal itself should keep its overall rigidity (elastic behaviour) while its surfaces in contact with the flanges should plastify to enforce leak-tightness.

4.2.2 Welding

Welds are a source of impurities and defects. Therefore, they should be carefully managed. Welding techniques will not be treated in detail here but hints in order to avoid potential problems are given.

The first point, obvious but sometimes forgotten because of the weight of other arguments, is the weldability of the material. To fulfil their role of mechanical resistance, the welds should be designed and executed according to the rules of the construction codes. Depending on the requirements, a reduction factor (0.85, 0.7) is applied in the calculation of the stress level and it is therefore recommended to avoid localizing a weld in a highly stressed zone.

But specific rules should be applied to high vacuum chambers to avoid contamination and virtual leaks generated by pockets. The welds should be performed on the vacuum side and trapped volumes between two welds are forbidden. Filler metal is not recommended. Grooves help to minimize heat propagation along the walls and can be useful in case of subsequent cutting and rewelding. Crossing welds lead to a remelting of the bath creating defects. Figures 4 and 5 present the typical weld configurations for vacuum components.

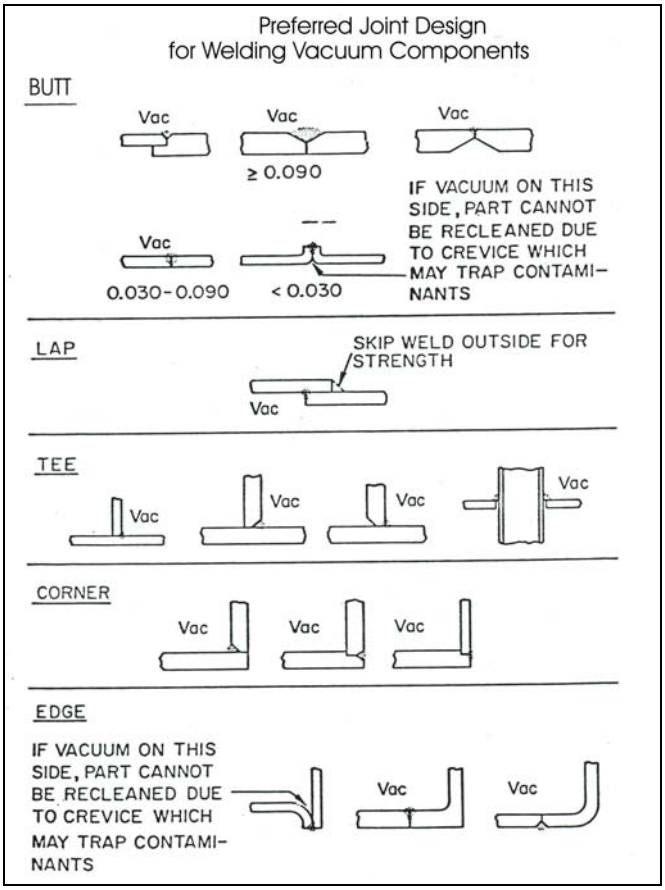


Fig. 4: Typical weld configurations for vacuum components

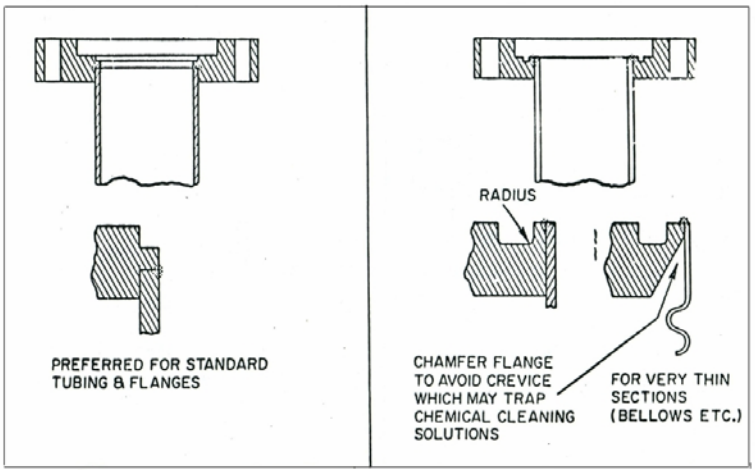


Fig. 5: Typical weld configurations for vacuum flanges

The preferred and most used technology is the Tungsten Inert Gas (TIG) welding. A very good inert gas (argon) protection is a prerequisite. Electron beam welding is another solution but it is not as flexible owing to the need for a vacuum envelope. And in case of thin shells, plasma welding is recommended.

Finally, one should remember that a good mechanical preparation of the parts to be welded is the key to success and that it is easier to weld in the workshop than *in situ* and the quality will be in accordance.

Whatever the techniques and the quality, no more than 99% of a series of welds will be leak-tight. Visual inspection must be systematic and X-rays are highly recommended if not compulsory.

4.2.3 Brazing

Brazing is an expensive solution which requires a furnace filled with a protective gas and jigs to keep the parts in position during the thermal cycle. It is often the only solution to join dissimilar metals or ceramics together. But the design of the interfaces to be brazed should be done according to strict rules taking onto account the expansion of the materials to allow the brazing to flow through the gap. The surfaces are etched before and a high-quality cleaning after brazing is important: one classical problem is the appearance of corrosion on stainless steel following a deficient cleaning of the chloride-based flux.

4.2.4 Gluing

Gluing is quoted here for completeness. Configuration of joints is similar to the ones used for brazing but the curing temperature is lower and even ambient. Degassing of glue joints is a problem, even for high-performance glues (epoxies) which are commonly used to repair leaks. There is still room for development of gluing of vacuum parts.

4.3 Cleanliness

Cleanliness of the delivered parts is essential to minimize outgassing. Cleaning can be performed through well-known processes which should be compatible with parts (materials, temperature, abrasion). But it is preferable to avoid polluting the parts during the fabrication, for example, to avoid machining with silicon oils difficult to remove. The last step of cleaning is to fully empty and dry the parts; the design should take that into account by suppressing zones where puddles will remain and avoiding closed volumes.

4.4 Ease the detection of leaks

An important and sometimes painful step in the commissioning of a vacuum system is leak detection. Simple actions will ease that work: provide access to the vacuum chamber, in particular to the joints (helium spraying, clam shell), avoid virtual leaks by suppressing closed volumes like bolts.

The time-of-flight technique will be used for large chambers: provide an easy way for helium to flow through radially in order to minimize the perturbation when measuring time of flight over long distances.

5 Quality assurance

A good design is the first step for good quality but it might be entirely spoiled by bad quality manufacturing. What is required should be clearly specified in the technical specification and in the drawings, using international standards. All the steps of the procurement, manufacturing, and assembly should be detailed beforehand and controlled according to agreed procedures. The construction codes contain specific chapters on this subject. But, in any case, the minimum requirement for choosing a manufacturer is that he be certified ISO 9001:2000, even if this is not foolproof.

To include regular inspections at the premises of the contractor either by your technicians and/or by an accredited inspection body could be considered a quite heavy workload but it is the only way to obtain the required quality. Look in particular at the material used versus the one specified, at the tolerances of the parts versus the manufacturing drawings, at the procedures and at the cleanliness.

6 Conclusion

Designing a vacuum chamber should not be very complex and a systematic approach will help to get what you expect with good quality at the right price.

Acknowledgements

The author thanks all his colleagues with whom he learned how to design vacuum chambers and all those who provided documents for this paper.

References

[1] L. Westerberg, Engineering, CAS, CERN Accelerator School: Vacuum Technology (1999).

[2] M.F. Ashby, *Materials Selection and Process in Mechanical Design* (Butterworth Heinemann, Oxford, 1999).

[3] C. Hauviller, Design of vacuum chambers for experimental regions of colliding beam machines, *IEEE 1993 Particle Accelerator Conference* (May 1993).

[4] O. Gröbner and C. Hauviller, LEP vacuum chambers for experimental regions: experience with the first generation, prospects for the second generation, *1990 European Particle Accelerator Conference* (June 1990).

[5] S. Karppinen *et al.*, Design of beampipes for LHC experiments, Vacuum **64** nos 2–4 (2002).

[6] CERN, Code D2 Rev. 2, Safety code for pressure vessels and pressurized pipelines (1998).

[7] Syndicat National de la Chaudronnerie, de la Tôlerie et de la Tuyauterie Industrielle (SNCT), CODAP, French code for construction of unfired pressure vessels (2005).

[8] European Committee for Standardization (CEN), Unfired pressure vessels, EN 13445 (2002).

[9] European Committee for Standardization (CEN), Cryogenic vessels, Static vacuum insulated vessels, EN 13458-2 (2002).

[10] Expansion Joint Manufacturers Association Inc. (EJMA), 2003 Eighth Edition, EJMA Standard and 2005 Addenda.

[11] European Committee for Standardization (CEN), Metal bellows expansion joints for pressure application, pr EN 14917 (2004).

Mechanical vacuum pumps*

A.D. Chew
BOC Edwards, Crawley, United Kingdom

Abstract
This presentation gives an overview of the technology of contemporary primary and secondary mechanical vacuum pumps. For reference a brief history of vacuum and a summary of important and basic vacuum concepts are first presented.

1 Introduction

Vacuum? It is just like turning on a tap. We have come to take for granted vacuum and its application, in all its degrees from vacuum forming, metallurgy, integrated circuitry fabrication to space simulation. The current provision, the engineering innovations, and global availability bear testament to the technical and commercial pioneers of the 19th and 20th centuries. The full history itself can be traced back to the ancient philosophers, however, here we touch briefly on the revolutionary work of the Natural Philosophers of the period of the Scientific Enlightenment.

There is some debate as to who achieved the first deliberate experimental vacuum (Berti or Torricelli), however, it is pretty certain that it was in Italy in the 1640s. The concept of a vacuum or void was a huge discussion amongst philosophical heavyweights across Europe in the 17th century. Indeed in reasoning against a vacuum (the Aristotolean 'horror vacuii'), the Torricellian void in an inverted mercury column was said to have been filled with elements from the liquid mercury itself. As well as creating a vacuum, they also recognised the concept of outgasssing—a concept familiar to many practitioners today.

Galileo and Descartes were party to the debate be it via the concept of the weight of air (measured as 2.2 g/l compared to a 'modern' value of 1.3 g/l). The work spread from Italy (Baliani) to Mersenne and Pascal in France. It is worth noting that these intellectual endeavours were patronized by the great realms of the time both for application to live issues (for example, suction of water in mines) and also for the reflected glory on the patrons themselves. Science as Natural Philosophy was truly revered in the 17th century.

Von Guericke raised significant issues with the Torricellian void and combined the glass tube with his own suction pumps (reported by the Roman engineer Vitruvius) to produced the famous Magdeburg hemispheres. His problems with the correction of the (degassing) effects of the mercury column also paved the way for the 'real' gas correction to the ideal gas laws. Along the way his correction to the mercury column also led to its use for weather prediction (barometry) and in 1660 he predicted the onset of a heavy thunderstorm.

Pascal had by then joined the debate (as a fully fledged vacuist) as to what exactly could constitute a void. He led against the Aristotelean view, a victory which was to lead to the development of the gas laws.

---

*Disclaimer. The author and his employer, BOC Edwards, disclaim any and all liability and any warranty whatsoever relating to the practice, safety and results of the information, procedures or their applications described herein.

1.1 The gas laws

"If I have seen further it is by standing on the shoulders of giants" a famous quotation from a letter written by Isaac Newton in 1676. This may have been a way of giving credit to his correspondent, rival, and contemporary Robert Hooke, but I would like to think that the giants Newton had in mind would include Robert Boyle who was, amongst many other things, a great vacuum pioneer.

Boyle had a young Hooke as his assistant and circa 1660 they jointly developed an air pump, arguably the first vacuum-producing apparatus suitable for experiments. These included the investigation and extinguishing of the ringing of a bell. This is a common vacuum demonstration now in classrooms but was used by Boyle as evidence for his attack on the plenists (and for theological applications). This vacuum pump allowed the accurate investigation of the (inverse) relationship between pressure and volume: $P_1V_1 = P_2V_2$—Boyle's (or Mariotte's) Law.

At this time no temperature scale existed and he could not determine the relationship between 'hotness' and volume. Later work (1702) by Amontons developed the air thermometer—it relied on the increase in volume of a gas with temperature rather than the increase in volume of a liquid and gave us Amontons' Law: $P_1T_2 = P_2T_1$.

Although Amontons had done the work before him, Charles is credited with the law $V_1T_2 = V_2T_1$ as measured on the Kelvin temperature scale. Gay-Lussac (also famed for his Law of Combined Volumes) in 1808 made definitive measurements and published results showing that every gas he tested obeyed this generalization. Dalton's work on the law of the partial pressure of gases (1801) and Avogadro's article (1811) stating that at the same temperature and pressure, equal volumes of different gases contain the same number of molecules would lead to the concept of the mole and Avogadro's number. By 1860 our modern day view of the Ideal Gas Law was in existence and the Maxwell–Boltzmann theory further explained the gas laws in terms kinetics of individual molecules. This incorporated Bernoulli's 1730s kinetic model of tiny gas molecules moving about in otherwise empty space (a foretelling of the van der Waals equation—perhaps the most commonly known of many Real Gas equations).

Boyle's Law was the first of the gas laws and was a truly remarkable display of experimental procedure with the application of leading-edge contemporary technology. A vacuum pump was the catalyst and facilitating element as it would prove to be in many other major scientific milestones.

1.2 First industrial applications of vacuum

The development of vacuum technology made many contributions in the 17th to 19th centuries ranging from evidence to resolve to the philosophical confrontation between plenists and vacuuists, to the development of the gas laws. By the mid 19th century there seems to have been a major application of vacuum awaiting discovery. Step forward the humble light bulb.

Davy moved the world on from gas/oil/candle lighting in 1809 by creating an incandescent light arc-lamp. De la Rue in 1820 followed with an evacuated glass bulb (thence filled with gas) to provide an incandescent light-bulb though it proved non-economic. From the 1850s work intensified in which vacuum was used to form the void in which an electric filament was heated to incandescence by an electric current. Edison is generally attributed with the design in 1879 of the modern-day practical light-bulb though others (Swan, Weston, Maxim etc.) made many contributions.

Althought they were intended to cause the gas to glow, certain glasses were seen to glow (e.g., Geissler's tube) at the end of the tube and/or around the filament. Many investigations of these cathode rays followed (Lenard and Crookes etc.) and it was a 'short' step for a third electrode to be used in the vacuum tube to see their effect on these cathode rays. Thomson developed from these studies his corpuscular theory and discovered the negatively charged electron (1897).

1.3 Vacuum-aided particle discoveries

The understanding of the strange phenomenon of cathode rays presented a great scientific challenge in the latter stages of the 19th century. In 1871 Crookes suggested that these rays were negatively charged and in 1879 invented the Crookes vacuum tube (of glass). This was an early form of the much known 20th century cathode-ray tube and eight years later he performed the famous Maltese cross experiment. Varley and Crookes in 1871 had suggested that cathode rays were particles and Perrin proved this in 1895. Positive (canal) rays, or ions, were observed by Goldstein in 1896.

Stoney in 1874 had estimated the charge on the cathode rays (calling them electrons in 1891) but it was not until 1897 that J.J. Thomson performed his famous series of experiments to demonstrate the particle or corpuscular, nature of cathode rays. He showed them to be a constituent of an atom—the first subatomic particle to be discovered. It is interesting to note that after Röntgen discovered X-rays in 1895 the first commercial X-ray machine was developed for medical applications in the following year (by E. Thomson): one of the shortest times to market of any vacuum application! Röntgen picked up his Nobel prize in physics in 1901 for his discovery of the X-ray.

A fundamental and at least enabling factor in these experiments was the development of the mean free path theory by Clausius in 1858 without the knowledge that a vacuum is needed to allow particles to travel unimpeded by collision with gas molecules (as a guide, the mean free path of nitrogen is 66 cm at 10^{-4} mbar). As with today, experimental necessity proved the catalyst for the development of vacuum generation mechanisms and vice versa new vacuum techniques facilitated advances in experimental developments. Various pumping techniques for glass vacuum systems had been developed in parallel to these significant discoveries. These included the Roots pump 1859, diffusion pumps (Geissler 1855, Topler 1862 and Sprengel 1865), Bunsen's 1870 water jet pump and Dewar's 1892 cryogenic (liquid air-cooled charcoal) pump. Fleuss' oil piston pump (1892) pump and Gaede's 1905 mercury sealed rotary vacuum and molecular drag pump of 1912 also played a role.

Subatomic particles were further isolated in 1918 with Rutherford's discovery of the first nucleon (alpha particle): the proton, and by Chadwick in 1932 of the neutron.

Since then vacuum has continued to be an instrumental tool in driving subatomic studies and more. Perhaps the most demanding and fundamental studies are still to reveal themselves: the use of UHV in interferometric experiments for the isolation, capture of gravitational waves. Gravitational waves were predicted from Einstein's General Theory of Relativity and are disturbances ('ripples') in the curvature of space–time caused by motions of matter. As these waves pass through matter, their strength weakens and the wave shrinks and stretches.

Mechanical vacuum pumps developed and used from ~ the mid 20th century to the present day will be discussed after some basic vacuum concepts.

2 Basic vacuum concepts

The commonly used and basic vacuum concepts utilize the gas quantities:

- number of molecules, N

  molar mass, M

  gas constant R_0

  temperature, T

  gas mass, w
- number of moles, n_m
- pressure (P) volume (V) units; gas quantity q

$$q = PV \ , \tag{1}$$

for example, mbar litre/s could be expressed in joules

$$q = PV = w\frac{R_0 T}{M} \ . \tag{2}$$

$$PV = n_{\mathrm{m}} R_0 T = \frac{w}{M} R_0 T \ . \tag{2a}$$

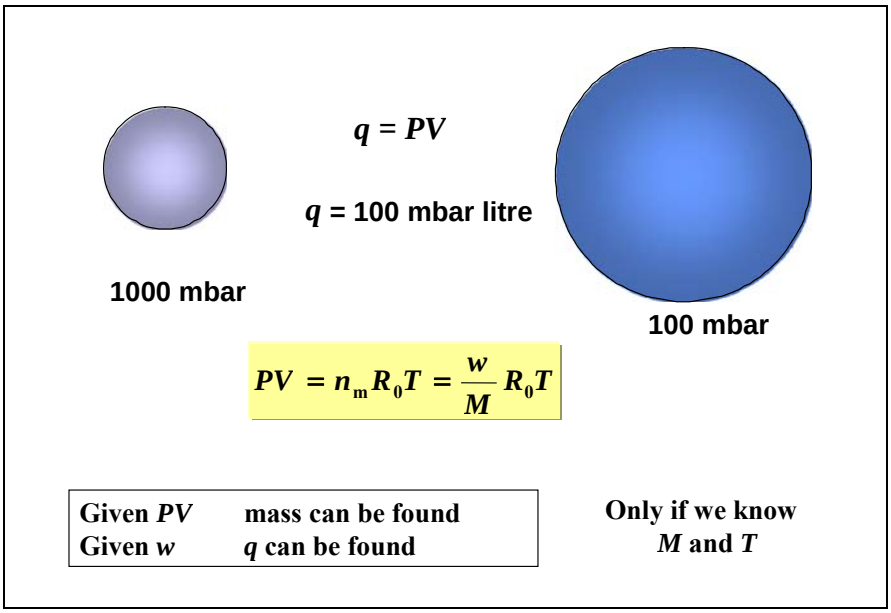


Fig. 1: Gas quantity

– Speed, $S \equiv$ volume rate

$$S \equiv \dot{V} \ . \tag{3}$$

– Throughput

Q = pressure × volume rate (where pressure is constant)

$$Q = P\dot{V} = PS \ . \tag{4}$$

$$Q = P\dot{V} = \dot{w}\frac{R_0 T}{M} \ . \tag{5}$$

Q tells us nothing about pressure and volume rate separately—only the product does that.

By speed, manufacturers generally mean the volume flow rate measured under standard conditions. Generally units are m^3/h, l/m or cfm for primary and l/s for secondary pumps but many other units are used (standards specify between 15ºC and 25ºC). Displacement is usually referred by manufacturers to the swept volume rate D, i.e., the trapped or isolated inlet volume/unit time. The maximum possible flow rate of the pump, or speed $S < D$.

2.1 Flow regimes

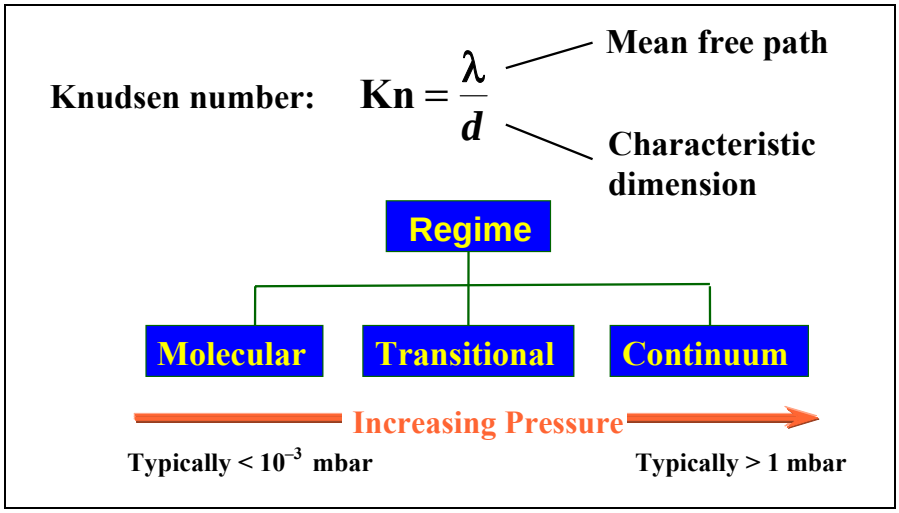


Fig. 2: Flow regimes

$$\mathrm{Kn} = \frac{\lambda}{d} \tag{6}$$

Knudsen number = mean free path/typical dimension.

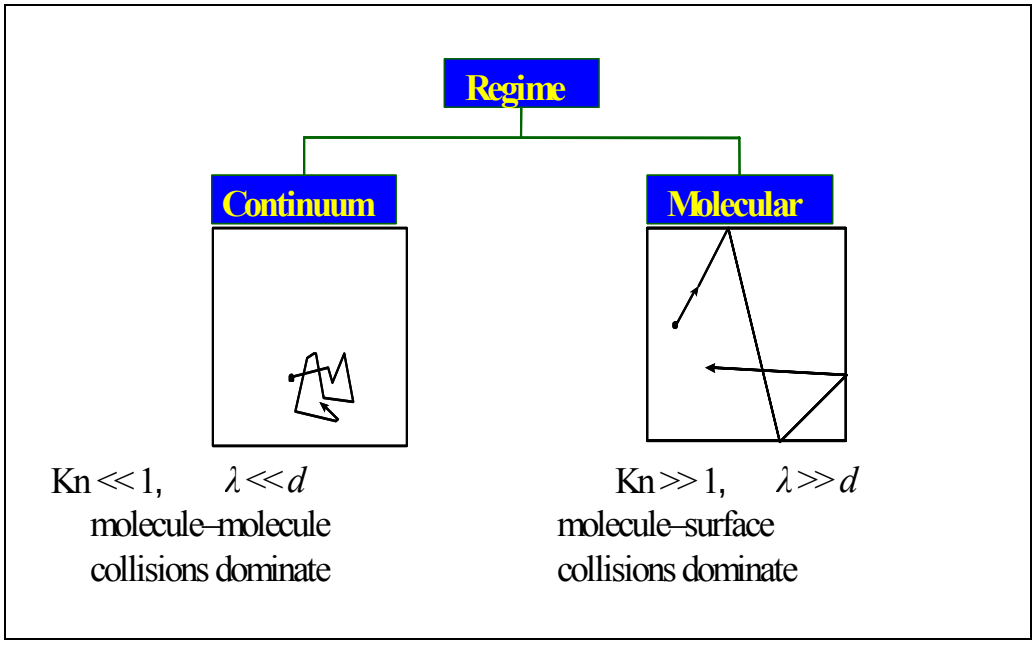


Fig. 3: Flow regime classification

Definitions:

- Kn < 0.01 continuum state
- 0.01 < Kn < 1 transitional state
- Kn > 1 molecular state.

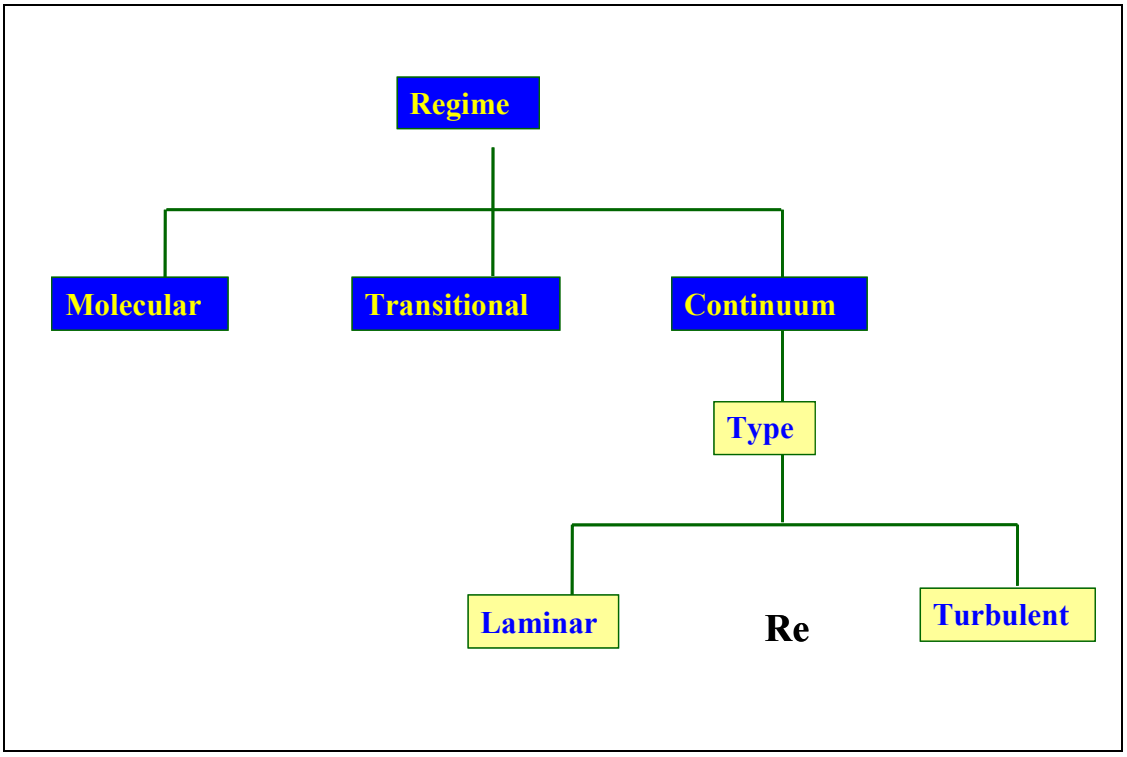


Fig. 4: Flow regimes and types

2.2 Conductance, *C*

Definition

$$C = \frac{Q}{P_u - P_d} \tag{7}$$

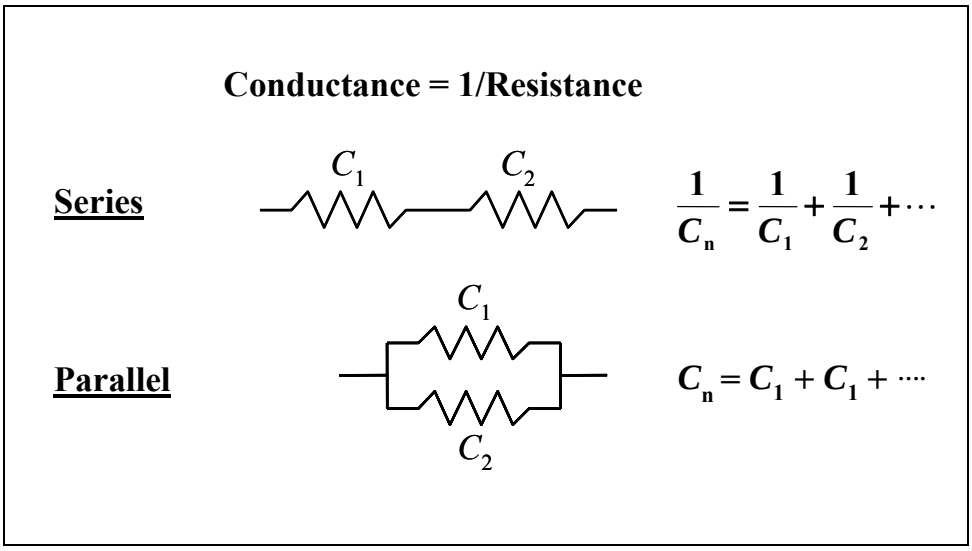


Fig 5: Conductance definition

Pumping speed can be combined with a conductance in the same way as conductances in series, see Fig. 6. N.B. in molecular flow we need to introduce the concept of transmission probability.

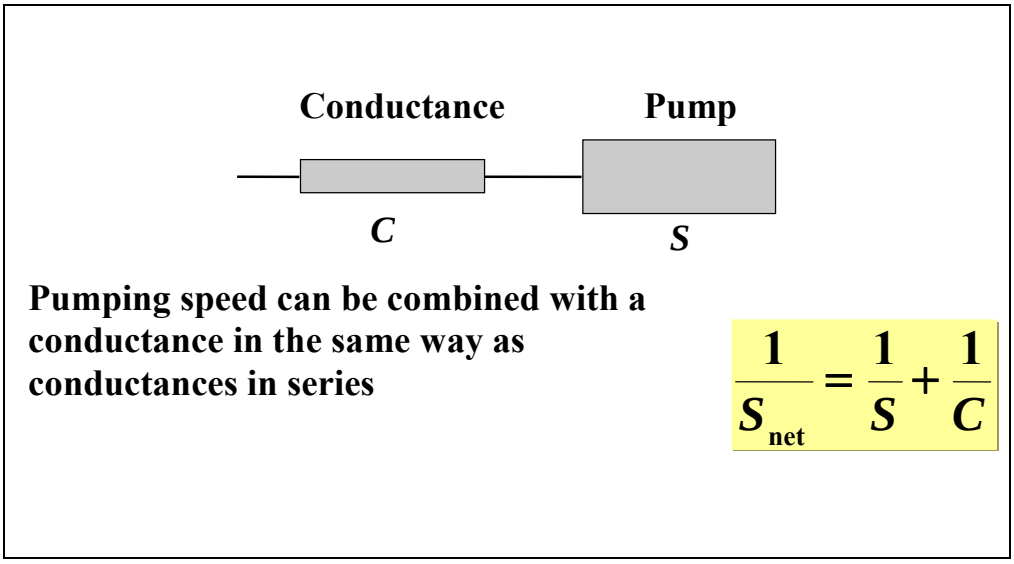


Fig. 6: Combination of speed and conductances

Speed, pressure ratio K and conductances are combined as in Fig. 7 so that the zero flow compression ratio K (i.e. ratio of the upstream to downstream pressures) is given as

$$Q = P_d S = (P_u - P_d)C$$
$$\therefore\ C = \frac{S}{K-1} \quad \text{or} \quad K = 1 + \frac{S}{C}\ . \tag{8}$$

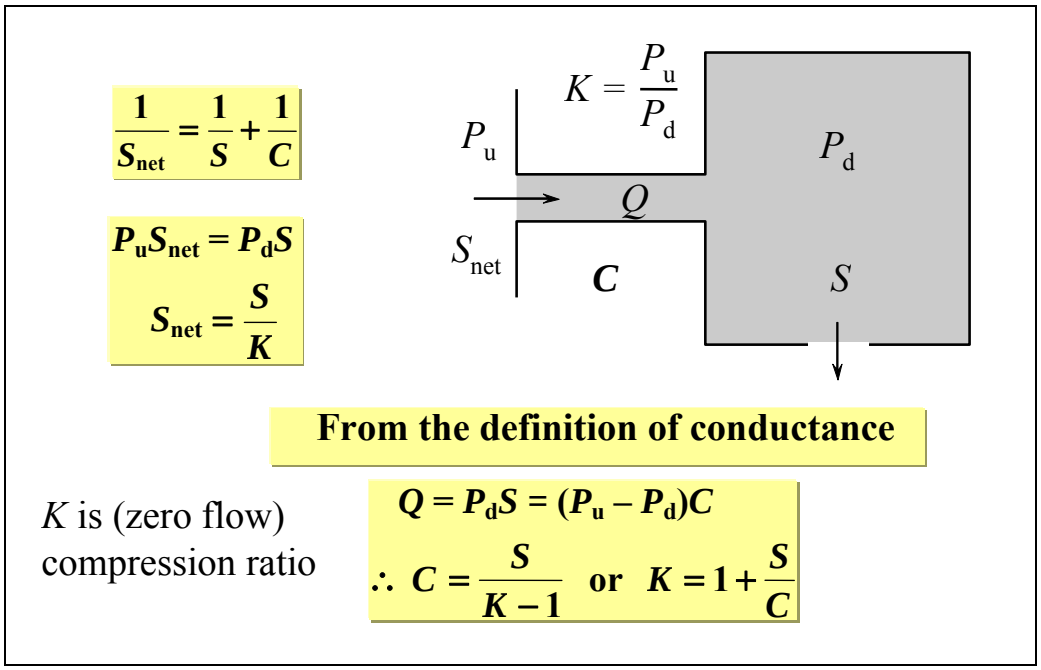


Fig. 7: Speed, pressure ratio and conductance relationship

The formula for the chamber exhaust: pressure at time t is given by

$$P_{(t)} = P_{(t=0)} e^{-\frac{S}{V}t}\ . \tag{9}$$

3 Mechanical pumps

A classification for mechanical pumps of speeds > 1 m^3/h is shown in Fig. 8.

Primary pumps – exhaust to atmosphere)

Secondary pump – exhaust to a backing (primary pump)

Max speeds shown exponent of maximum pump speed 10^n (m^3/h or l/s)

- Wet pumps

– Oil sealed rotary vane	Primary	n = 0 to 3 (m^3/h)
– Piston	Primary	n = 0 to 2 (m^3/h)
– Liquid ring	Primary	n = 1 to 3 (m^3/h)

- Dry Pumps

– Northey-claw	Primary	n = 1 to 3 (m^3/h)
– Roots	Primary or Secondary	n = 2 to 5 (m^3/h)
– Scroll	Primary	n = 1 to 3 (m^3/h)
– Screw	Primary	n = 1 to 3 (m^3/h)
– Regenerative	Primary	n = 1 to 3 (m^3/h)
– Piston	Primary	n = 1 to 2 (m^3/h)
– Diaphragm	Primary	n = 0 to 2 (m^3/h)
– Drag	Primary or Secondary	n = 0 to 2 (l/s)
– Turbomolecular	Secondary	n = 1 to 4 (l/s)

Fig. 8: Mechanical pump classification scheme

All theses pumps rely on the principle of positive displacement of gas (or vapour) except drag pumps, which exploit molecular drag and turbomolecular pumps (momentum transfer/capture technique). Table 1 indicates some of the range of choices involved between wet (oil-filled) and dry (oil-free) vacuum pumps.

Table 1: Wet and dry pump comparisons

	Wet pumps	**Dry pumps**
Capital cost	Low	High
Oil loss	Can be high at > 1 mbar	Very low
System contamination	Backstream at < 0.1 mbar	Very low
Add on costs	Oil return/filtration	Not necessary
Aggressive process	Not suitable	Resistant
Purge	Sometimes	Almost always

A consideration process involving more decision criteria is shown in Fig. 9.

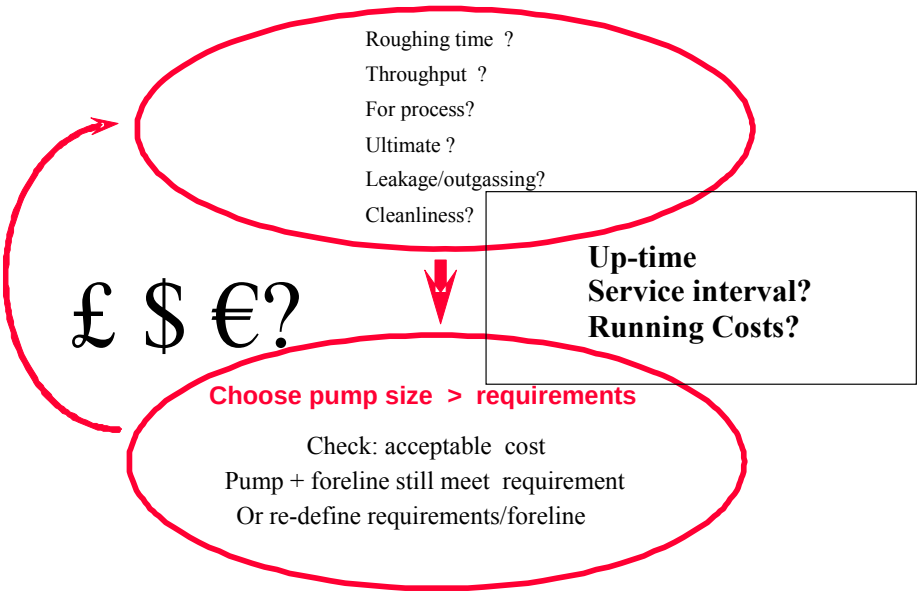


Fig. 9: Pump choice considerations

4 Oil-sealed rotary vane pumps

Oil-sealed rotary vane pumps (OSRVs shown in Fig. 10) were first developed in the early 1900s. Today, the two commonly used oil-sealed pumps are rotary vane and rotary piston pumps. Oil-sealed rotary vane pumps are often used for low inlet pressures and light gas loads. Oil-sealed rotary piston pumps are often large and are most often found in high-gas-load, high-inlet-pressure industrial applications.

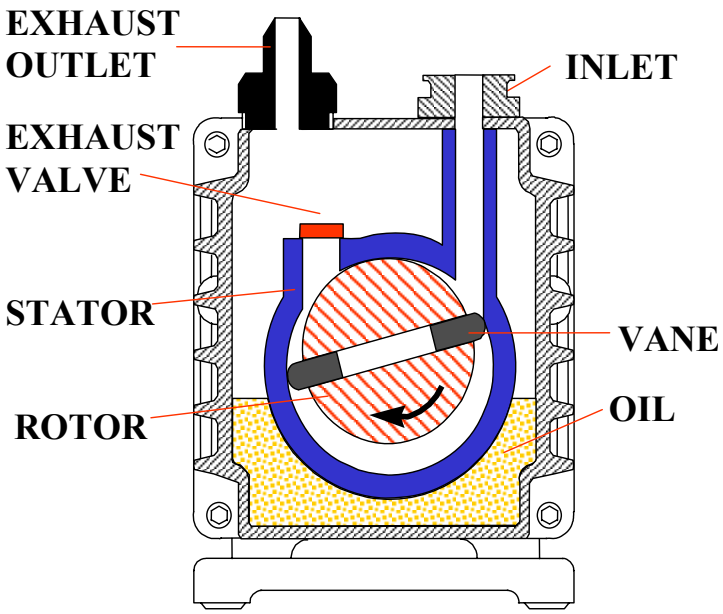


Fig. 10: Basic OSRV schematic

The OSRV pumping cycle is shown in Fig. 11.

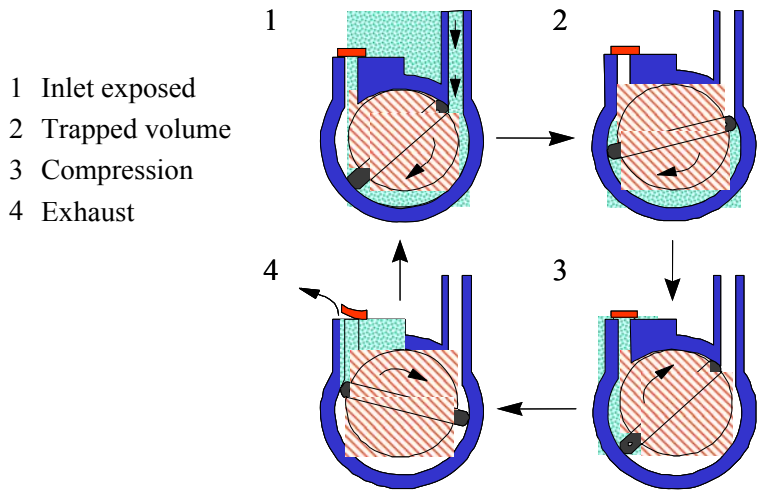


Fig. 11: OSRV pumping cycle

4.1 Functions of oil

The oil in OSRV and piston pumps has several functions:

- sealing: the oil surface tension seals the duo-seal and fills the gaps between the vanes, rotors, and stators;
- lubrication: of the bearing areas and blade contact surfaces;
- cooling: removes heat from rotors and stators;
- protects parts from rust and corrosion: coats and 'seals' surfaces to protect from aggressive gas.

4.2 Single- versus dual-stage pumps

A single-stage pump has one rotor and one set of vanes (ultimate pressure circa 10^{-2} mbar). They have a lower cost where low ultimate vacuum is not required and are used for higher inlet pressures or high gas loads due to lower compression. A dual-stage pump is most simply two single-stage pumps in series (ultimate pressure circa 10^{-3} mbar or lower). The effective higher compression ratio gives better ultimate vacuum. These are shown in Fig. 12 and a cutaway is shown in Fig. 13.

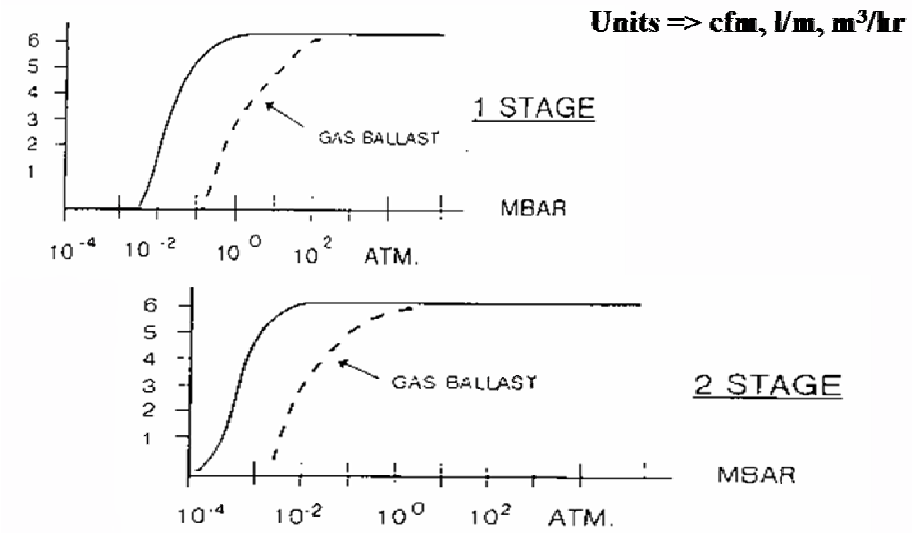


Fig. 12: Single- versus dual-stage pumps

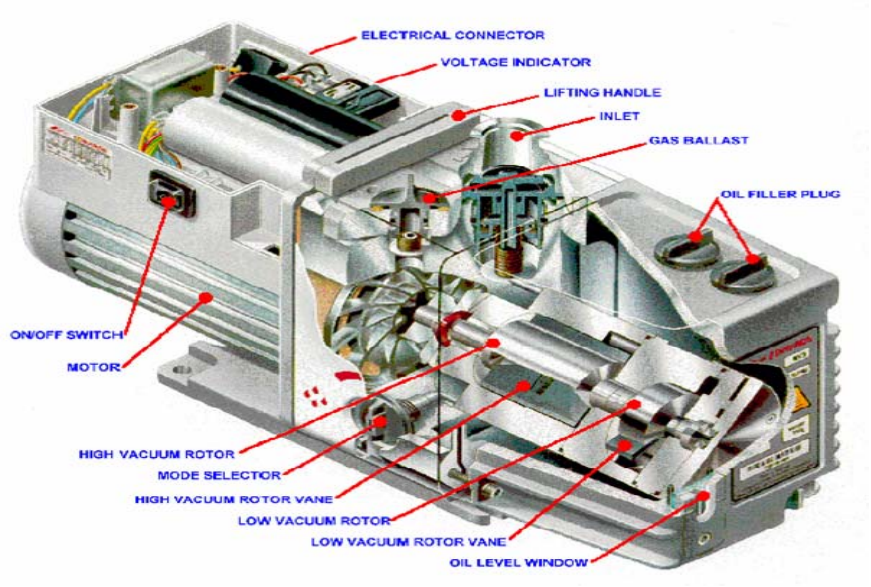


Fig. 13: OSRV pump cutaway view

4.3 OSRV gas loads

Pumped gas may contain both permanent gases and vapours, which can condense when compressed. Condensed vapours may include liquid water and solvents which can mix with pump oil to form an emulsion. Condensed vapours can limit the ultimate vacuum, cause corrosion, and possibly lead to pump seizure. Gas ballasting allows vapour pumping without condensation. The basic principle is that the ballast gas opens the exhaust before compression pressure allows vapours to condense.

5 Dry pumps – clearance mechanisms

Dry pumps do not have oil in the swept pumping volume and use a series of stages with small non-contacting clearances to displace gas/vapours and create compression. Ultimates can be 10^{-2} mbar and lower; higher frequencies giving lower ultimates as they provide reduced (ultimate-limitating) gas back-leakage. A typical design is shown in Fig.14.

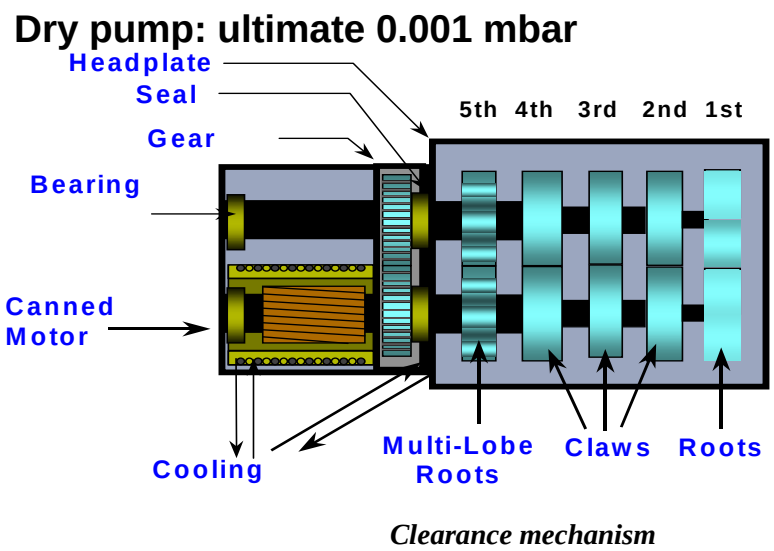


Fig. 14: Roots–claw dry pump schematic

5.1 Diaphragm pumps

Figure 15 shows the schematic of a diaphragm pump whose best ultimates are of the order 0.1 mbar.

The crankshaft rotates and the connecting rod pulls the diaphragm down, creating a vacuum in the chamber. This opens the inlet valve and closes the exhaust valve and the chamber fills with gas. As the crankshaft continues to rotate, the connecting rod forces the diaphragm to the top of the chamber. This compresses the gas, opens the outlet valve, and closes the inlet valve. The valves on the inlet and outlet to the chamber are flapper types, which are operated by pressure. Figure 16 shows a typical speed curve.

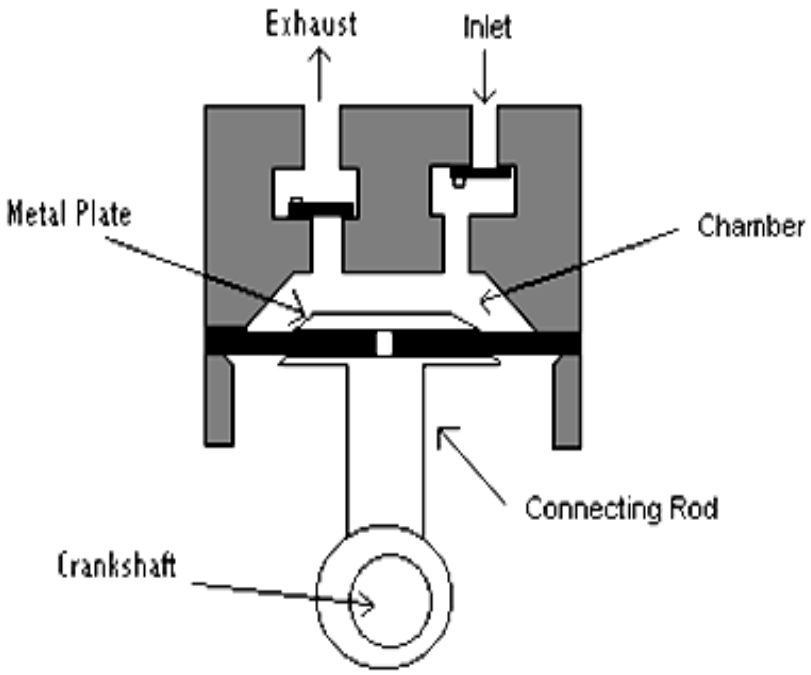


Fig. 15: Diaphragm pump operation schematic

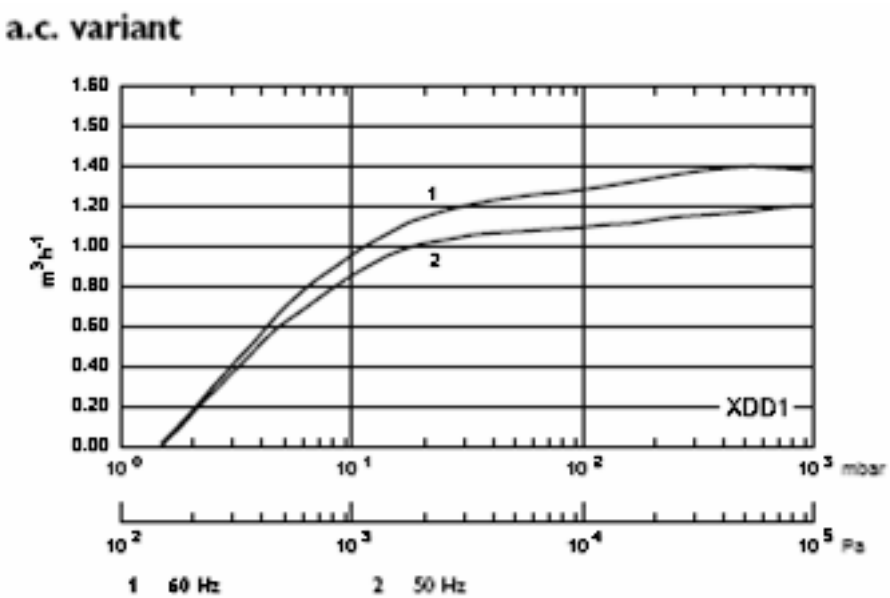


Fig. 16: Vacuubrand/BOC Edwards MD1/XDD1 diaphragm pump speed curve

5.2 Scroll pumps

Eccentrically mounted matching scroll forms move to isolate gas and transport it to an atmospheric pressure exhaust as in Fig. 17.

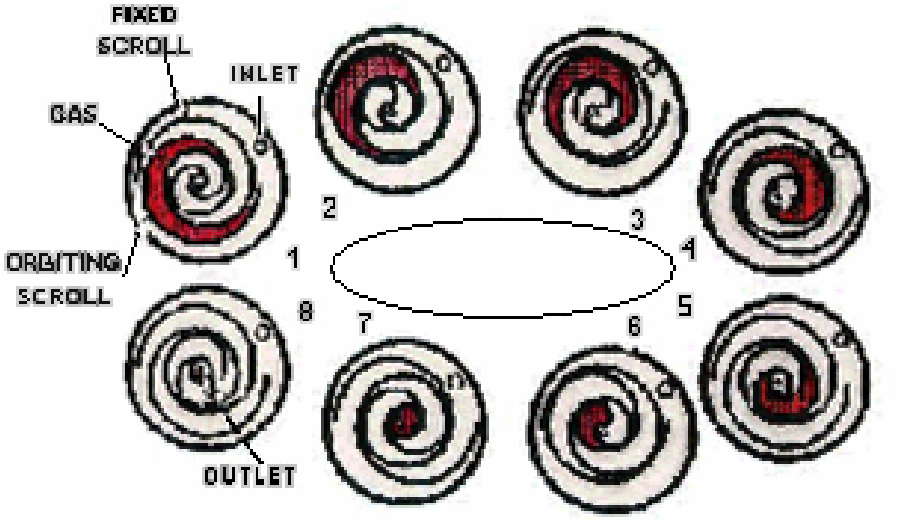


Fig. 17: Isolation of gas in scroll pump mechanism

A typical scroll pump speed curve is shown in Fig. 18.

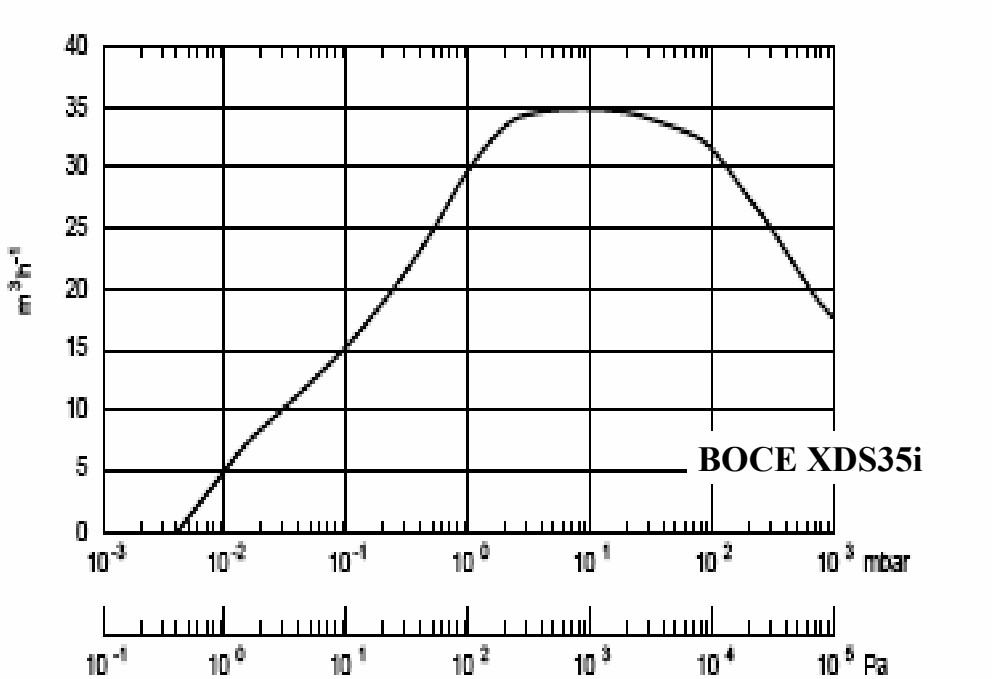


Fig. 18: BOC Edwards XDS35i speed curve

5.3 Dry piston pumps

The operational principle of dry piston pumps is the same as with oil-filled versions and is shown in Fig. 19. In the 'dry' version contact can be made with 'sacrificial' coatings.

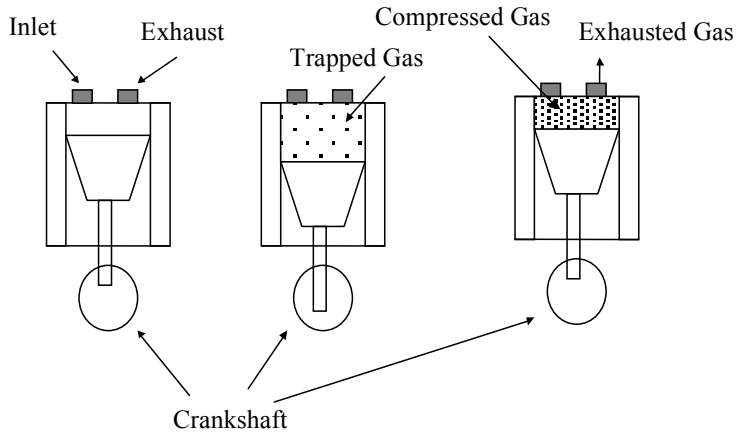


Fig. 19: Dry piston pump

As the crankshaft rotates, it moves the piston vertically through the cylinder, which traps, compresses, and then exhausts the gas from the pump. A coating around the outside of the piston creates the seal between the piston and the cylinder wall.

5.4 Roots/booster pumps

These are non-contacting secondary pumps and the sequence of operation is shown Fig. 20. Rotors trap a volume of air against the stator body and sweep it around, exhausting the gas at a point 180º from the inlet. Tight clearances between the rotors and the stator are critical to trap and move through the pump body.

Normally the booster displacement is greater than the backing pump speed therefore large outlet–inlet pressure differentials ΔP are generated at high inlet pressure. Unless this ΔP differential is limited, power demand increases rapidly. Limiting methods, shown in Figs. 21 and 22 include hydrokinetic drive, inverter motor control and pressure relief valves (outlet to inlet).

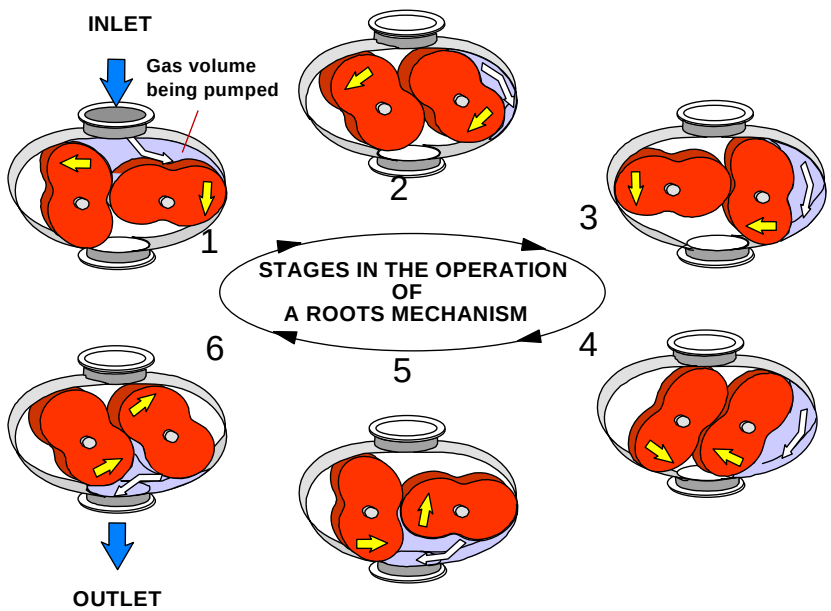


Fig. 20: Booster pump operation

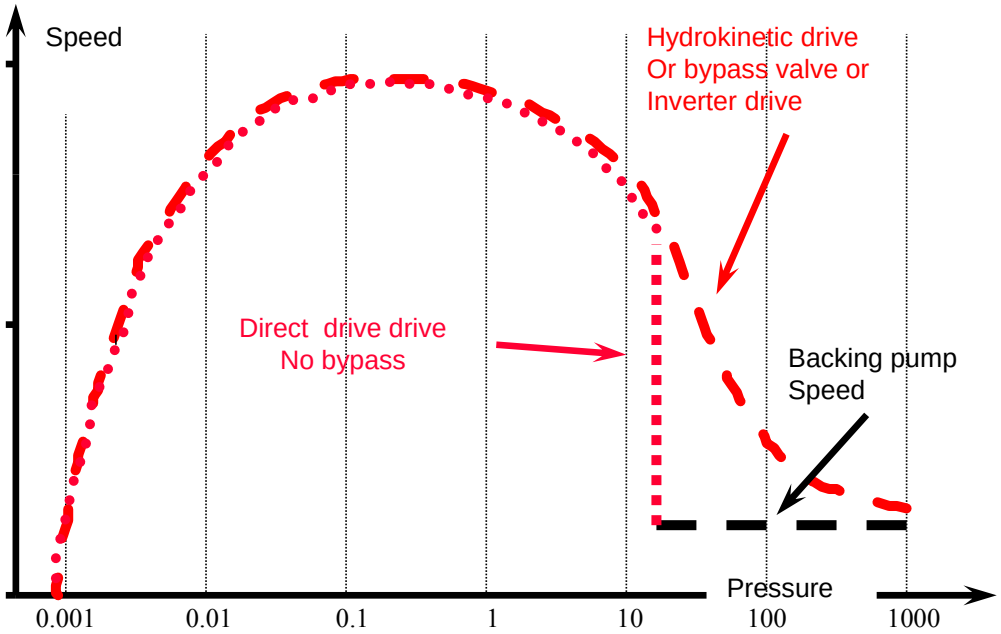


Fig. 21: Mechanical booster limiting methods

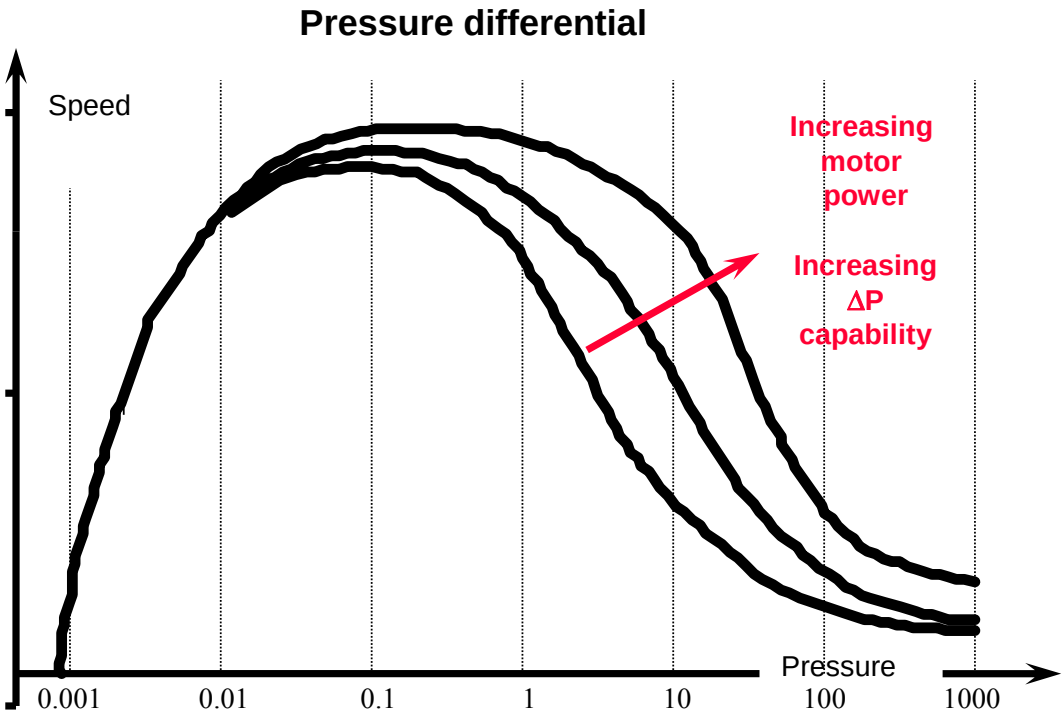


Fig. 22: Mechanical booster ΔP versus power

To achieve speed, ultimate or pump-down requirements, more than one booster may be used. With boosters in parallel, performance improvement is limited to approximately one decade improvement in ultimate with modest improvement in pumping speed at high pressures and with high pumping speed improvement at low pressures. With boosters in series, a lower ultimate is achieved (limited by outgassing and back-leakage). Also higher pumping speeds at higher pressures are achieved. This configuration limits the pumping speed at lower pressures. Schematically these are shown in Fig. 23.

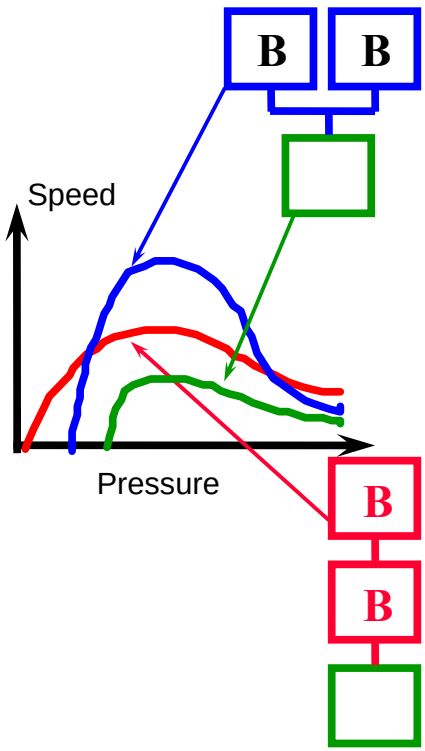


Fig. 23: Schematic of boosters in series and parallel

6 Turbomolecular pumps

The internal configuration of a bladed and compound/'drag' turbomolecular (secondary) pump is shown in Fig. 24. They can generally achieve UHV and exhaust to approximately 0.1 mbar. Different configurations and compound drag stages can increase throughputs and exhaust pressures significantly (> 20 mbar).

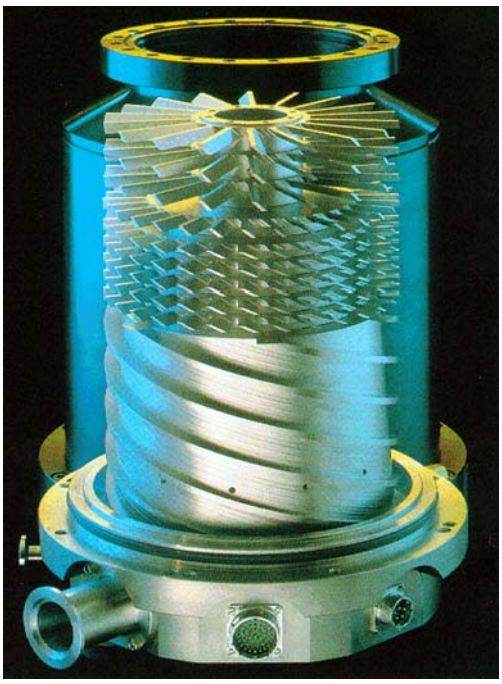
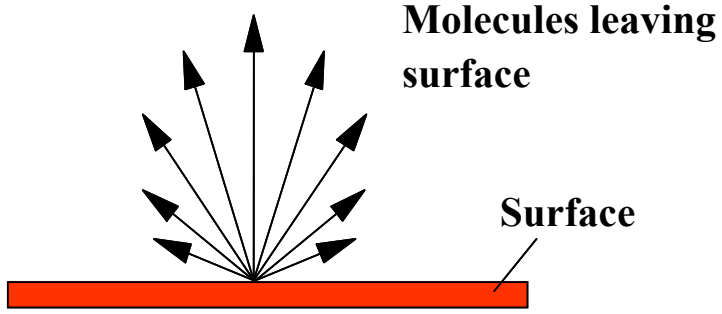


Fig. 24: Compound turbomolecular pump

6.1 Principle of operation

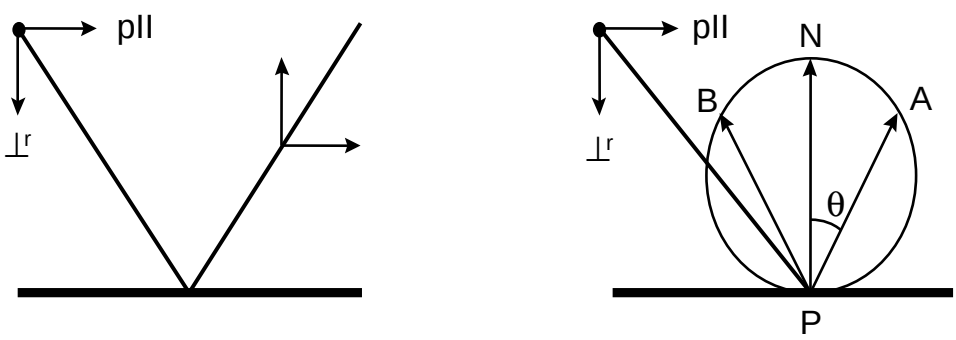
This is shown in Figs. 25 to 30 below.



1. The direction of the arrow indicates the direction of travel of the molecule.
2. The length of the arrow indicates the 'probability' that the molecule will depart in that direction - Knudsen

Fig. 25: Principle of operation

Consider (a) text book collision and (b) reality



(a) (b)

High rotational speeds (> 1000 Hz) tip velocity = molecular thermal velocities

Fig. 26: Molecular collisions with surfaces

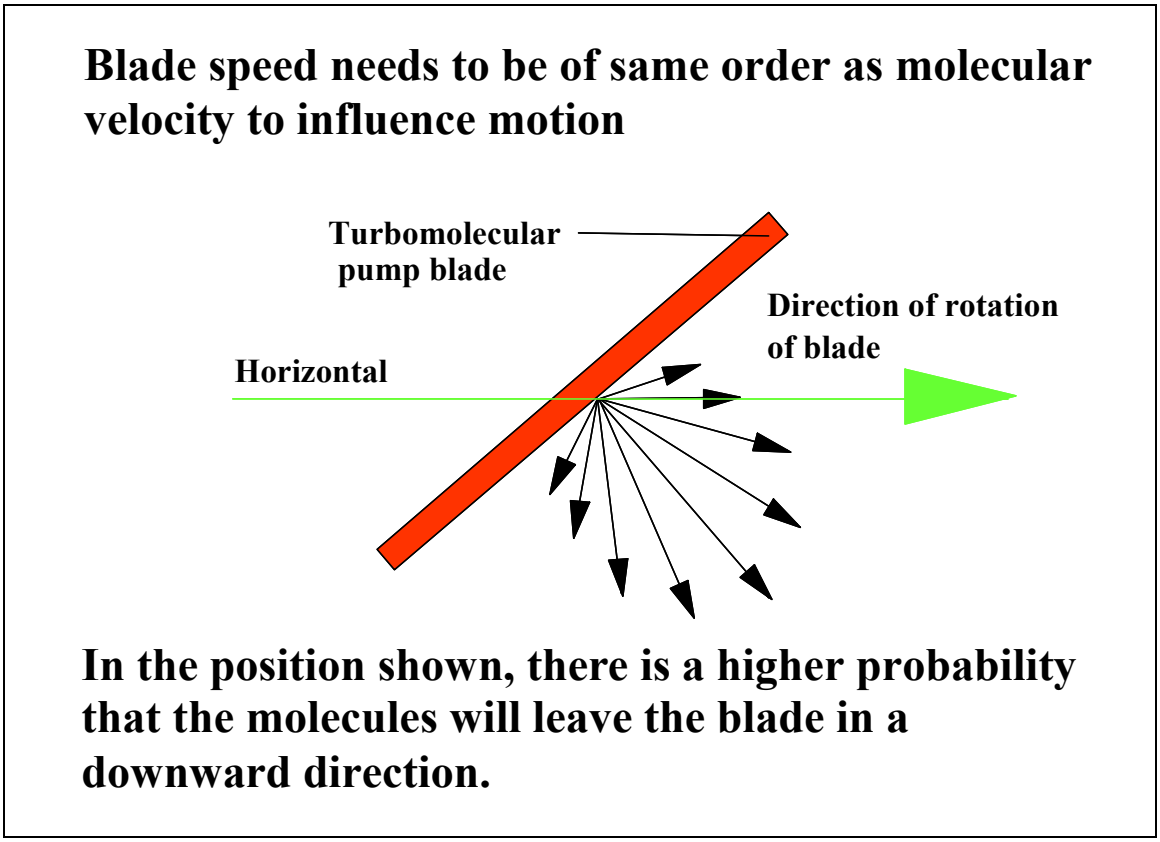


Fig. 27: Molecules leaving a surface

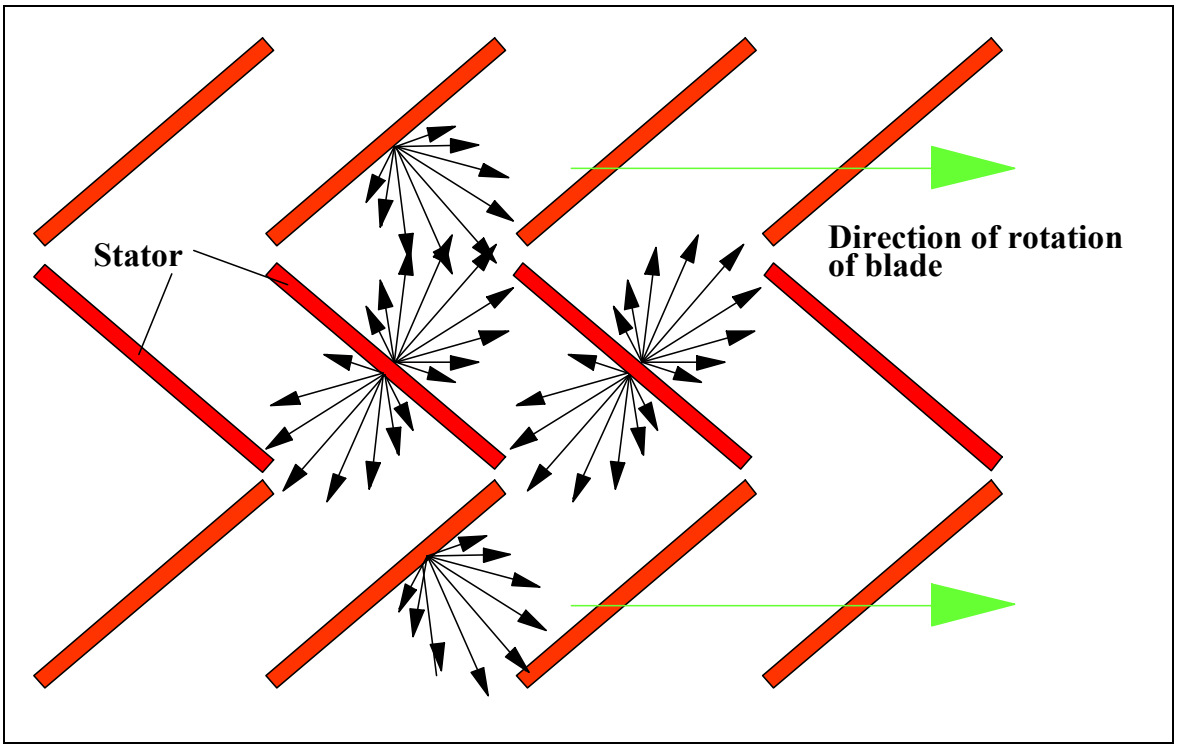


Fig. 28: Molecules leaving rotors/stators

- The blades at the top of the pump have an open blade structure.
- This gives a high pumping speed and low compression ratio.

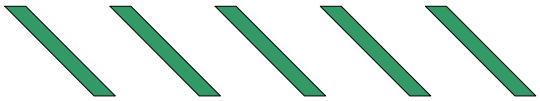


Fig. 29: Open blade structure

- The blades at the bottom of the pump have a closed blade structure.
- This gives a low pumping speed and high compression ratio.

Fig. 30: Closed blade structure

The Gaede and Holweck 'drag' stages are shown in Fig. 31. These allow operation to a higher exhaust pressure.

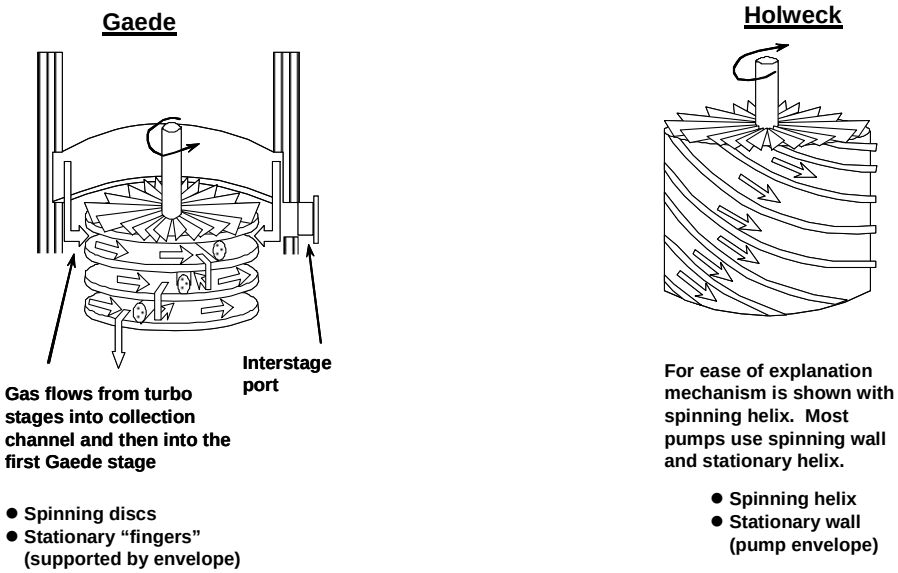


Fig. 31: Compound pumping technologies

'Conventional' bearing turbomolecular pumps have the rotor supported by a combination of greased/oil-bearings and permanent magnets, as shown in Fig. 32.

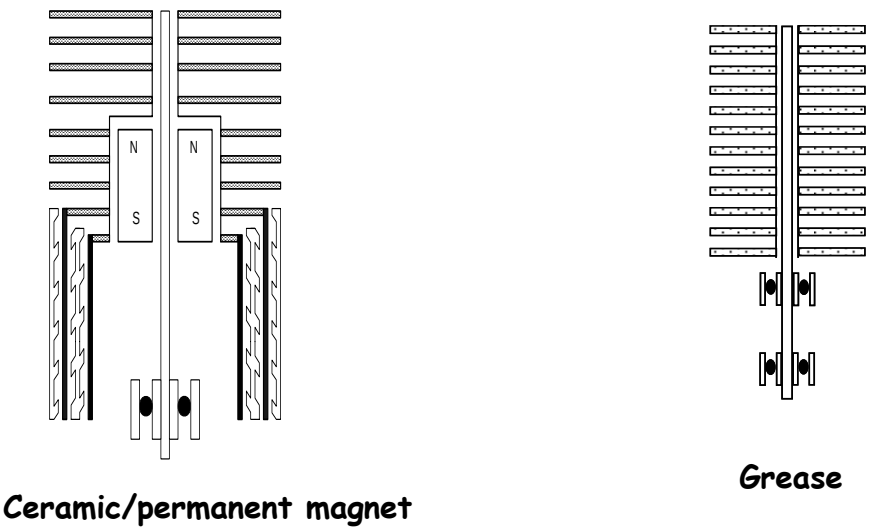


Fig. 32: Conventional turbomolecular-pump bearing technology

Magnetic levitation turbomolecular pumps shown in Fig. 33 are generally more expensive but are non-contacting with the following advantages:

- no bearings to wear and
- high reliability
- permanently low vibration characteristic with time
- no hydrocarbon lubricants present
- can be mounted in any orientation
- designed to work with (harsh) semiconductor process gases, radiation environments, etc.

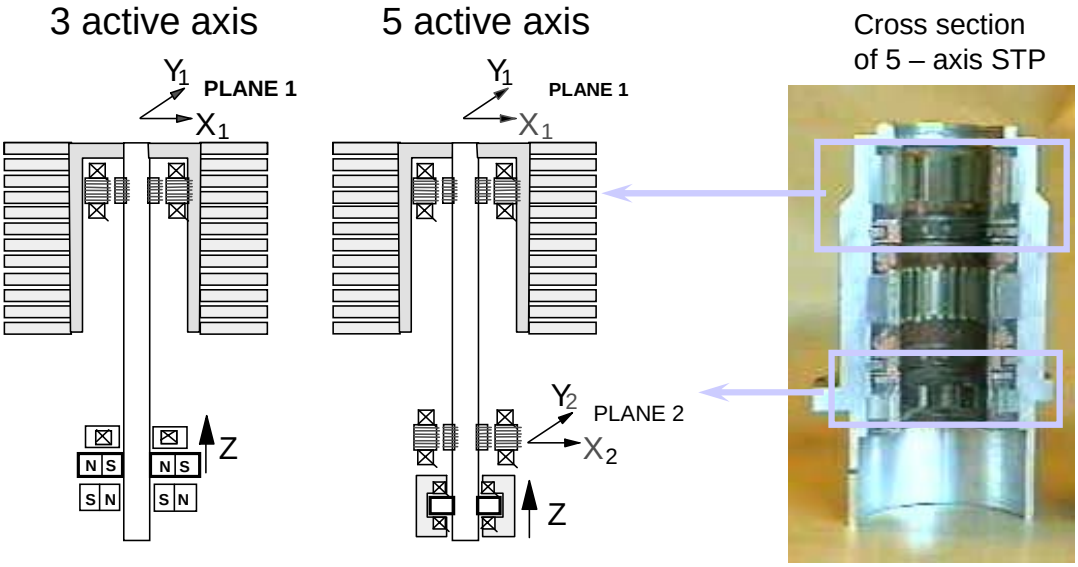


Fig. 33: Magnetic bearing turbomolecular pumps

A BOC Edwards STPA2203 (nominal speed 2200 l/s) is shown in Fig. 34 with respect to its throughput performance

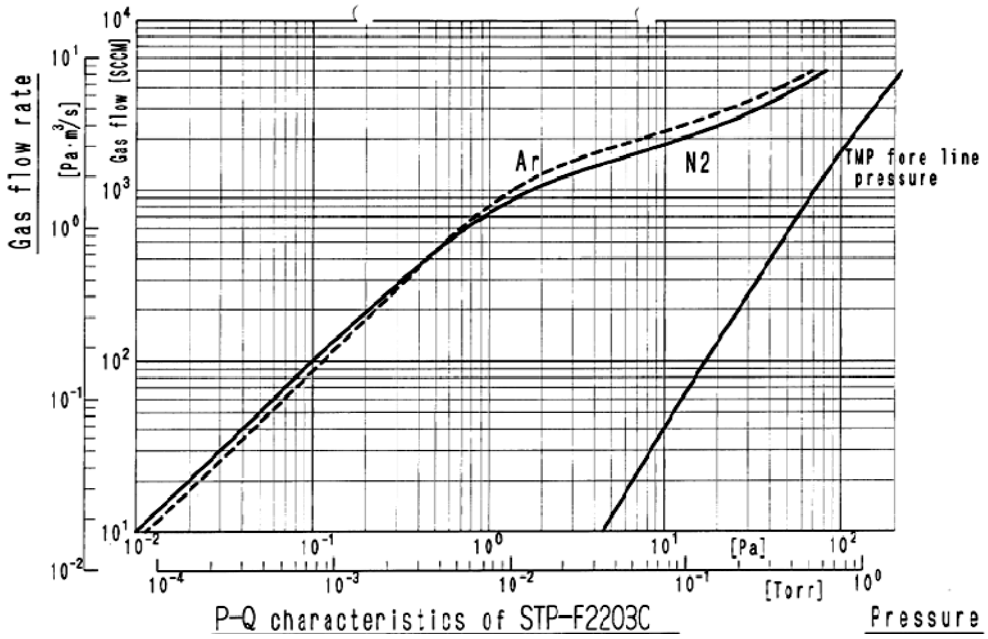


Fig. 34: BOC Edwards STPA2203 throughput performance

Other pumps designed for mass spectrometry: 'split flow'/multi-inlet pumps Ref. [1] and pure drag pumps are also extensively used, especially in machinery for scientific instrumentation.

7 Regenerative/drag pump

This pump is a single-axis mechanical clearance pump capable of achieving < 10^{-6} mbar and exhausting directly to atmospheric pressure. It is fully described in Ref. [2] and a cross-section shown in Fig. 35.

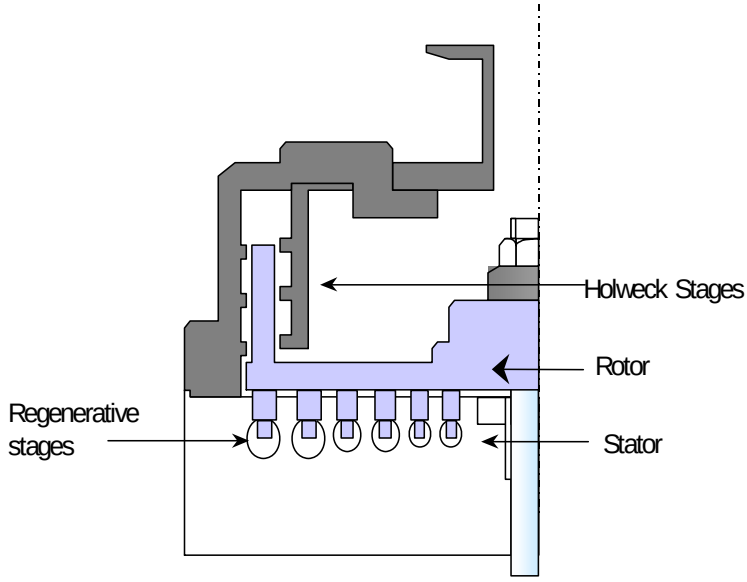


Fig. 35: Cross-section of the regenerative/drag (EPX) pump

The pump employs a combination of Holweck drag stages (which enable the achievement of the low ultimate and the molecular pumping speed—up to 140 l/s) and 'regenerative' stages which operate from a few mbar to atmospheric exhaust pressure. Figure 36 shows the pumping speed curves for a BOC Edwards EPX500 pump which show the typical molecular nature of the pump's performance.

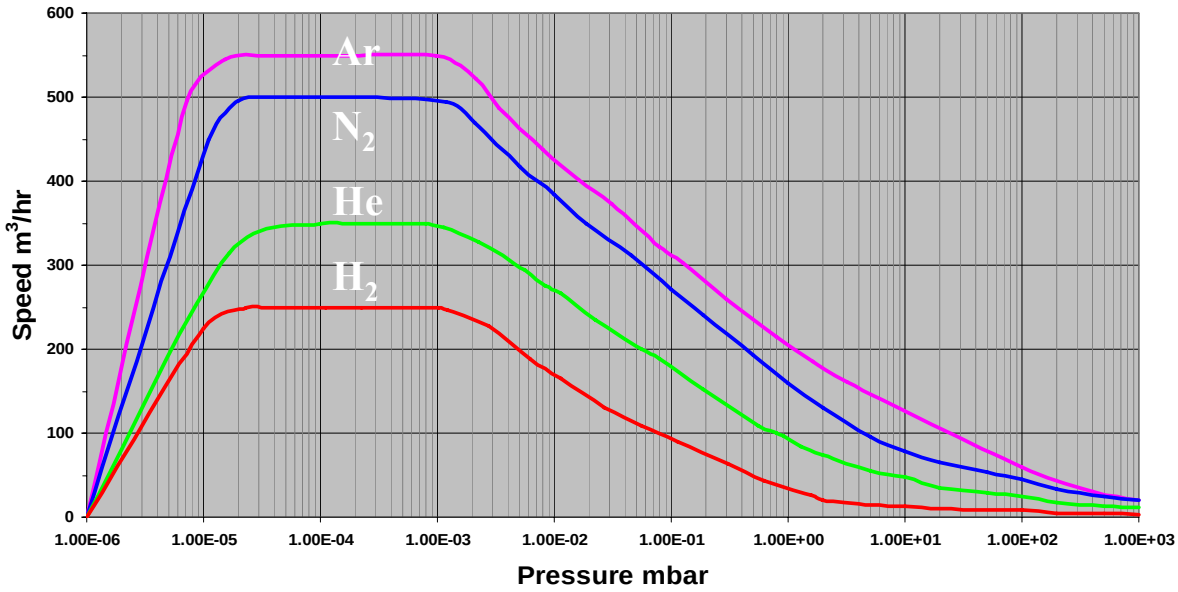


Fig. 36: EPX500 speed curves for different gases

References

[1] A. D. Chew, A. Cameron, D. Goodwin, J. Hamilton, T. Hawley-Jones, P. Meares, J. Pumfrey, J. Ramsden and D. Steele, Considerations for primary vacuum pumping in mass spectrometry systems, *Spectroscopy* **20,** January 2005.

[2] A. D. Chew, M. Galtry, R. G. Livesey and I. Stones, Towards the single pump solution: Recent development in high speed machines for dry vacuum pumping, *J. Vac. Sci. Technol. A* **23** (2005) 1314.

Gauges for fine and high vacuum

K. Jousten
Physikalisch-Technische Bundesanstalt, 10587 Berlin, Germany

Abstract
Vacuum gauges for use in accelerators have to cover about 17 decades of pressure, from 10^{-12} Pa to 10^{5} Pa. In this article we describe the history, measurement mode, design, accuracy and calibration of the gauges used down to 10^{-5} Pa. We focus on commercially available types of gauges, i.e., mechanical gauges, piezoresistive and capacitance diaphragm gauges, thermal conductivity gauges, and spinning rotor gauges.

1 Introduction

For about 350 years, from 1644 to about 1900, the Torricelli tube was the only instrument able to measure vacuum. It was based on the counterbalance of the gravitational force of a mercury column against a pressure difference in two volumes separated by the liquid mercury. If one of the volumes was under 'vacuum' conditions (i.e., the mercury vapour pressure), it was an absolute instrument and pressure was measured in 'mmHg' and later in 'Torr'.

Unfortunately these units are still being used today in some areas, even though the Torricelli tube is out of practical use and the Metre Convention (CGPM) was implemented almost 50 years ago in 1960 when the Système International (SI) of physical units replaced the Torr by the pascal. The pascal (Pa) is the force of 1 newton on 1 m^2 as pressure is defined by

$$p = F/A \ . \tag{1}$$

One can still use the bar for 10^5 Pa, hence 1 mbar is equal to 100 Pa. In this article we shall use mainly the unit Pa.

The measurement of vacuum pressure follows Eq. (1)

- by a direct measurement of the force per area (direct gauges)
- or indirectly by measurement of a quantity that is proportional to pressure, e.g., the molecular density, the impingement rate of molecules, the thermal conductivity, etc.

The direct measurement of pressure is limited to pressures larger than about 1 mPa. At this pressure, the force on 1 cm^2 is only 10^{-7} N which already needs an electrically amplified signal.

Figure 1 gives a classification scheme of vacuum gauges measuring pressure directly, Fig. 2 gives the most well-known types of vacuum gauges measuring pressure indirectly by another quantity.

The direct measuring vacuum gauges have the great advantage that their reading is independent of the gas species. They truly measure a total pressure of a gas mixture or a pure gas. The signal of indirect measuring vacuum gauges, on the other hand, depends for a given pressure on the gas species and for this reason it may not be possible to convert the signal into a correct pressure reading if the gas composition of a mixture is not known exactly.

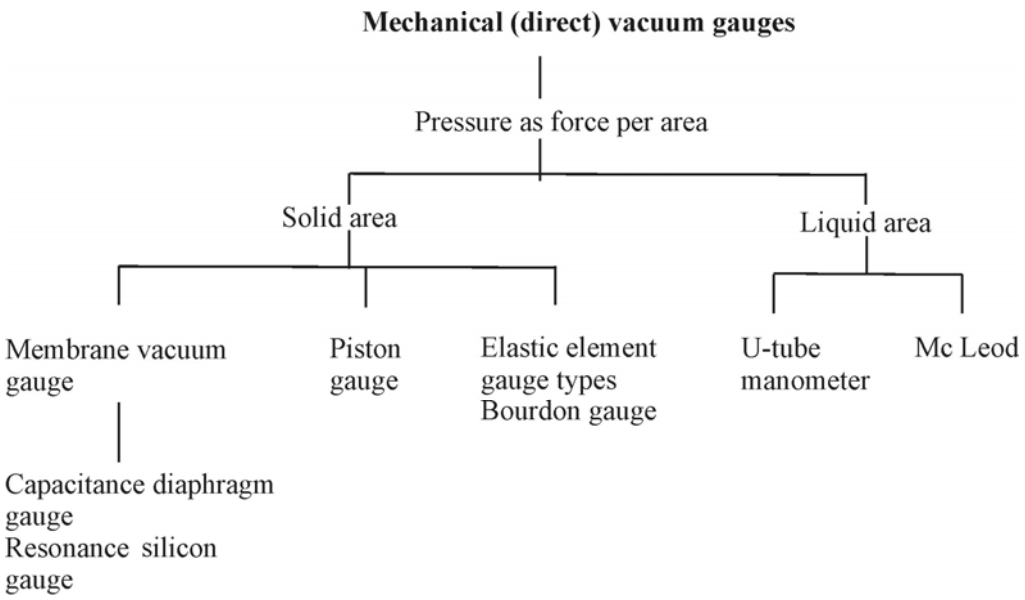


Fig. 1: Classification of directly measuring vacuum gauges

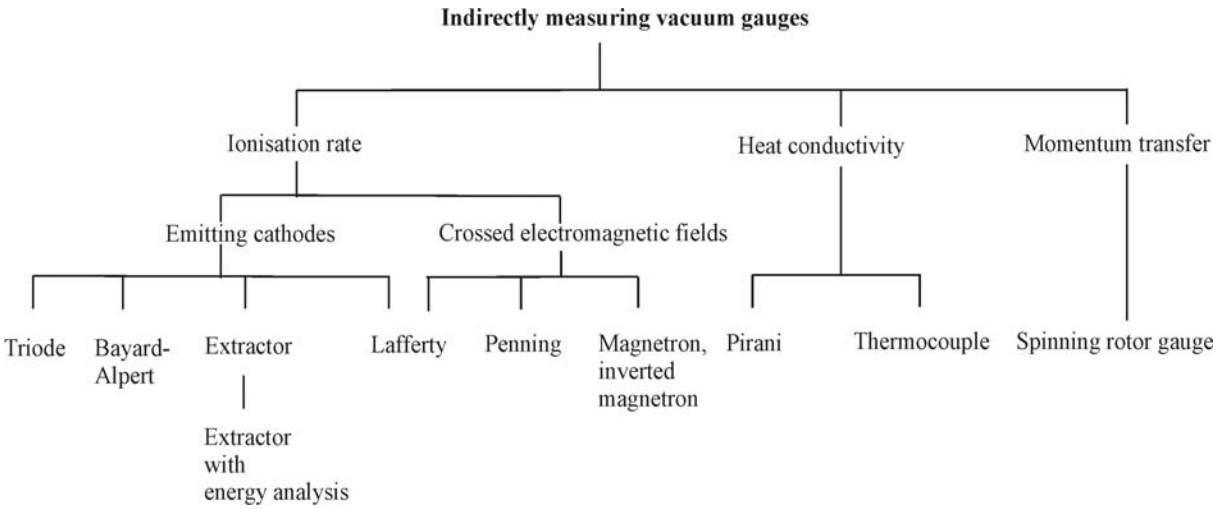


Fig. 2: Classification of indirectly measuring vacuum gauges according to their principle of measurement

Before we go into any details of vacuum measurement, let us say a word that is true for any measurement device. To measure something always means to compare a quantity with a unit for this quantity. The unit is internationally defined and realized in National Metrological Institutes and further disseminated to calibration laboratories and manufacturers and users of measuring equipment. It is hoped that each measuring instrument is somehow traceable to a so-called primary standard for this unit. This can be visualized by the pyramid of the calibration chain (Fig. 3). On each level some accuracy is lost. In the case of vacuum gauging there are several National Institutes that realize the pressure scale for vacuum pressures, which are published on the Web site of the BIPM [1], the Bureau International des Poids et Mesures.

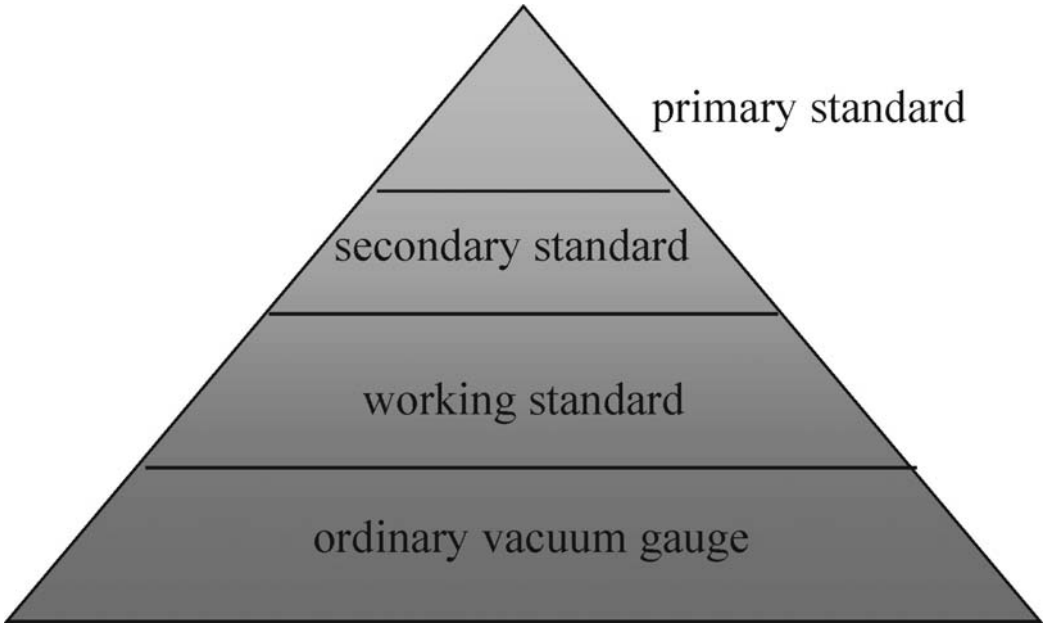


Fig. 3: The pyramid of the calibration chain. On each level downwards some information and accuracy are lost

One should also note the distinction between error and uncertainty of a measurement device. An error of the reading of a measurement device is the difference of the measured value from the true value of the quantity as defined by the SI system. After errors have been removed and all corrections known were made, still, each determined value of a physical quantity is an estimation. The uncertainty of a measured value gives the possible range by which the reading may not reflect the true value defined by the SI unit (Fig. 4). A high quality measurement will give a low uncertainty and vice versa. Without the concept of uncertainty, measured values cannot be compared, either among themselves or with reference values obtained somewhere else. Every measured value must be accompanied by its uncertainty in order to achieve comparability in a community.

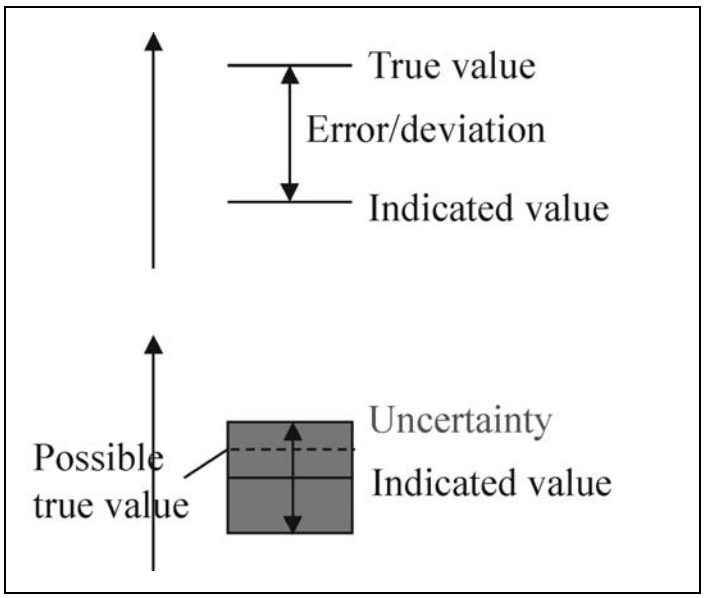


Fig. 4: The concept of error and uncertainty

2 Direct measuring vacuum gauges

2.1 Mercury manometer and the McLeod gauge

The mercury manometer (Fig. 5) is still the most accurate pressure gauge when operated carefully. Owing to the hazard of mercury it has almost disappeared from the commercial market, but it is used by the National Metrological Institutes where highest accuracy is required.

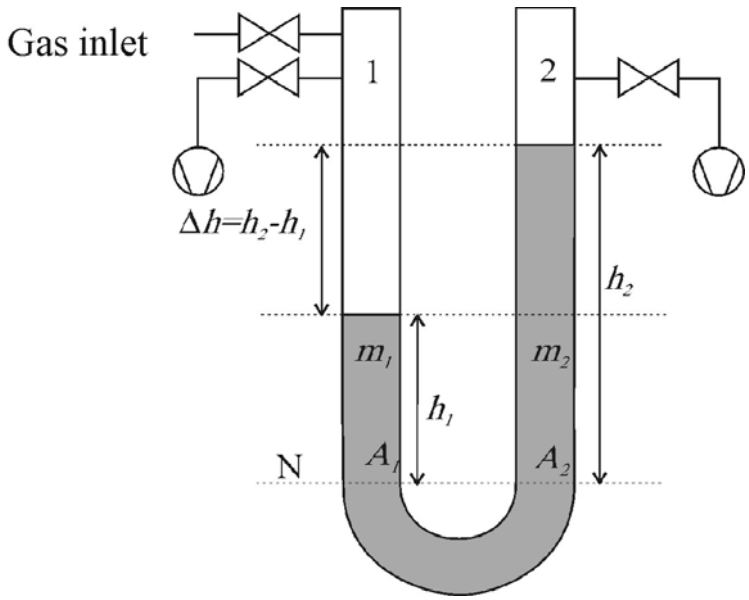


Fig. 5: The principal measurement scheme of the mercury manometer

The pressure difference between the two volumes 1 and 2 (Fig. 5) is given by

$$p_1 - p_2 = \rho g \Delta h \tag{2}$$

where ρ is the well-known density of mercury, g the local acceleration due to gravity. Volume 2 is usually pumped down to the fine vacuum and p_2 is approximately given by the vapour pressure of mercury. Since both ρ and g can be determined with high accuracy and Δh can also be determined with very low uncertainty by either optical interferometrical methods [2], [3] or phase-sensitive ultrasound detection [4], p_1 can be measured with the highest accuracy (relative uncertainties can be as low as $2 \cdot 10^{-6}$).

Mercury manometers are the primary standards for vacuum pressures from about 100 Pa to 10^5 Pa and therefore the traceability of pressure to the SI units of mass, time, and length.

Of historical importance is also the so-called McLeod gauge, which allowed the range of the mercury U-tube to be extended by several orders of magnitude by compressing the gas to be measured by moving the mercury in capillaries. It was invented in 1873 by H.G. McLeod and served as primary standard for pressures from 10^{-4} Pa up to the range of the mercury manometer until the 1960s.

2.2 Piston gauges

Figure 6 shows the principal measurement scheme of a rotating piston gauge. A cylindrical piston rotates in a closely fitted circular cylinder. The pressure at the base of the piston is defined as the ratio of the total downwards force on the piston to the effective area of the piston-cylinder assembly when floating at its operating level.

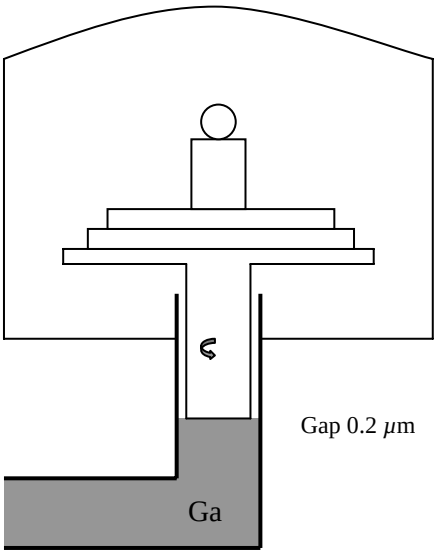


Fig. 6: The principal measurement scheme of a rotating piston gauge

The rotating piston gauge is a pressure generator, since the piston will float only when the pressure underneath exactly cancels the force from the top that may be varied by weights put on the piston. The pressure in the evacuated bell jar must be considered for absolute pressure measurement as volume 2 (Fig. 5) in the mercury manometer. The gap between piston and cylinder is typically a few tenths of a micron at a cross-section of 10 cm^2.

The rotation of the piston is needed to avoid any friction effects. Piston gauges can serve as primary standards for vacuum with somewhat higher uncertainty than mercury manometers from 10^3 Pa up to atmospheric pressure.

Non-rotating piston-cylinder assemblies are also used. In that case a balance is used to measure the force F, and the effective area A is determined by calibration with a mercury manometer (Fig. 7). These gauges have a resolution down to 1 mPa and have a good accuracy from 10 Pa up to about 10 kPa which is their upper pressure limit.

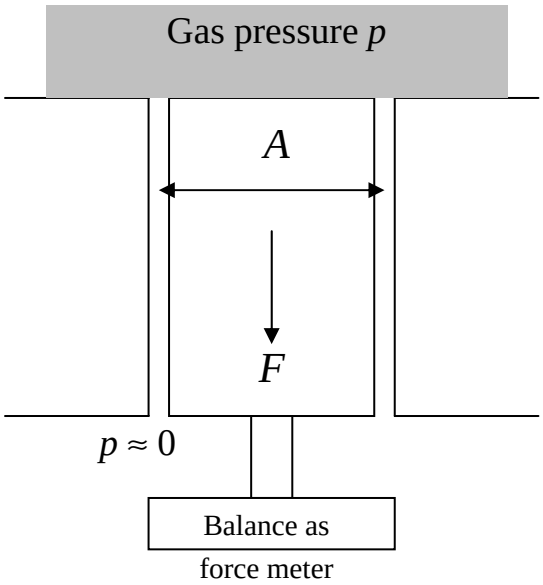


Fig. 7: The principal measurement scheme of a pressure balance

2.3 Mechanical gauges

Most mechanical gauges use a membrane to detect the force of the pressure (Fig. 8). On this membrane the force

$$F = (p_1 - p_2)A \tag{3}$$

is exerted, which will cause a deflection x of the membrane that can be used for measurement. In most cases x is converted into an angle φ that can be used for a needle indicator (see Fig. 8). When the reference pressure p_2 is negligible compared to p_1 the instrument shows the absolute pressure. Most common mechanical vacuum meters can be classified in three groups.

- The reference pressure is the prevailing atmospheric pressure. The measurement device is located on the reference side (a).
- The reference pressure is zero. The measurement device is located on the other (test) side (b).
- The reference pressure is zero. The measurement device is located on the same reference side (c).

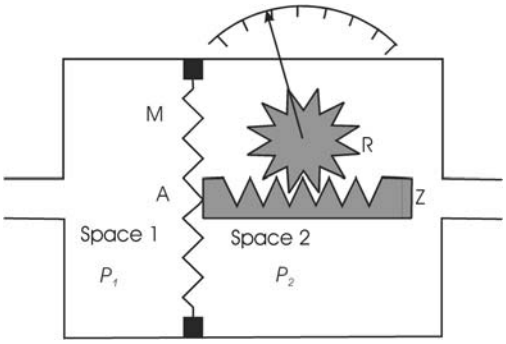


Fig. 8: The measurement scheme of a membrane gauge. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

An example of the latter type is shown in Fig. 9. This type has the advantage that the gas to be measured is not in contact with the sensitive instrumentation. The gauge is more resistant against corrosion. Only the membrane bellows normally made of a copper–beryllium alloy have to be protected against corrosive gases.

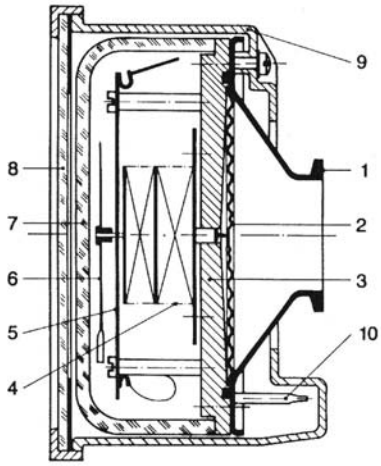


Fig. 9: Schematic of a membrane gauge with mechanical reading. 1 connecting flange; 2 membrane; 3 ground plate; 4 indicating system of membrane deformation; 5 indicating scale; 6 needle pointer; 7 glass vacuum enclosure; 8, 9 housing; 10 evacuation stem. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

A common example of type (a) is the so-called Bourdon gauge. Here a bent tube with a normally elliptical cross-section is closed at one end, and open to the gas to be measured at the other end. When the pressure inside the tube changes, the radius of curvature changes, since the force change on the outer side of the tube will be smaller due to the larger area than on the inner side of the bended tube. The bending is transformed by a lever to a needle pointer. The disadvantage of this type of cheap and robust gauge is that the reading depends on the atmospheric pressure. This can be corrected by turning the scale against the pointer.

2.4 Direct gauges with electrical output

The Bourdon gauge was greatly refined by the Quartz Bourdon Spiral (QBS) gauge (Fig. 10). A quartz helix is used as a tube. The bending of the helix is compensated by a force exerted by two electromagnetic coils. The balancing current through the coils is proportional to pressure. The null position is sensed optically via a mirror on the free end of the helix which reflects a beam light detected by two photocells. Since the helix remains fixed, hysteresis effects are negligible and the device can be linearized very well. The relative measurement uncertainty is in the low 10^{-4} range near full scales available from 7 kPa up to a few 100 kPa. The lowest reasonably measurable pressure is about 1 kPa.

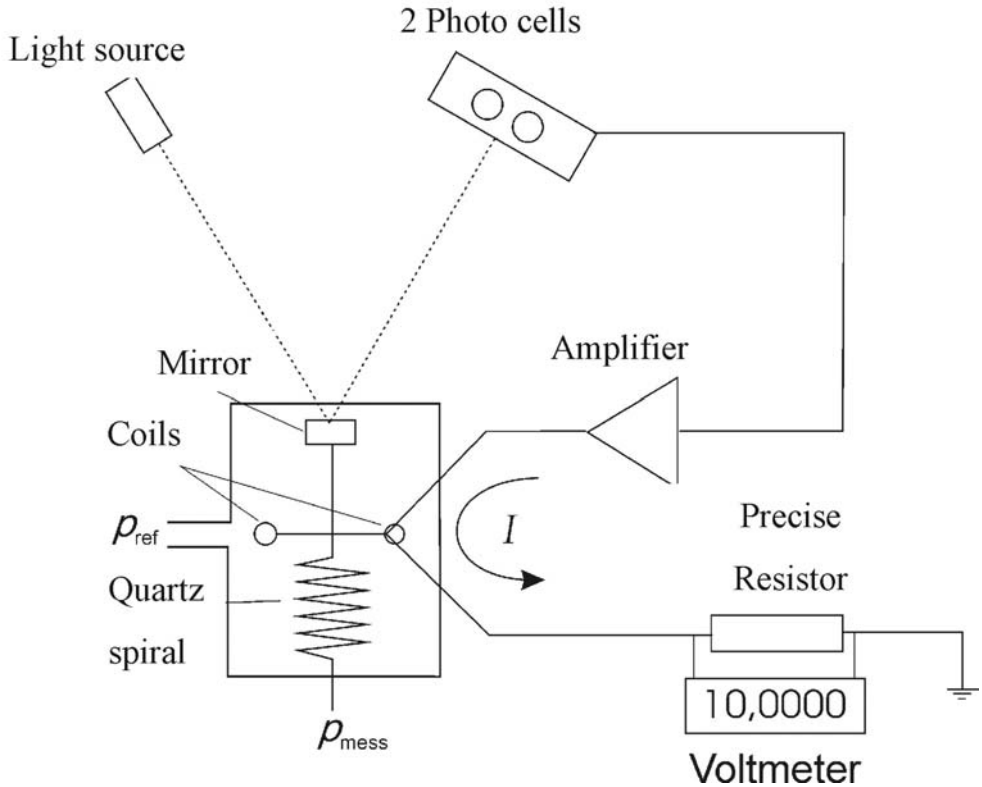


Fig. 10: Schematic of a Quartz Bourdon Spiral by Ruska, Houston, Texas. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

Another gauge of type (c) is the piezo-electrical vacuum meter (Fig. 11). A force on a quartz crystal generates a charge at the surface which may be measured by an electrometer. Since a pressure acting on all sides of a crystal will not do the job, a membrane is used that generates a force acting on one end of a quartz beam only. The generated charge is then guided to the electrical connector to the electrometer.

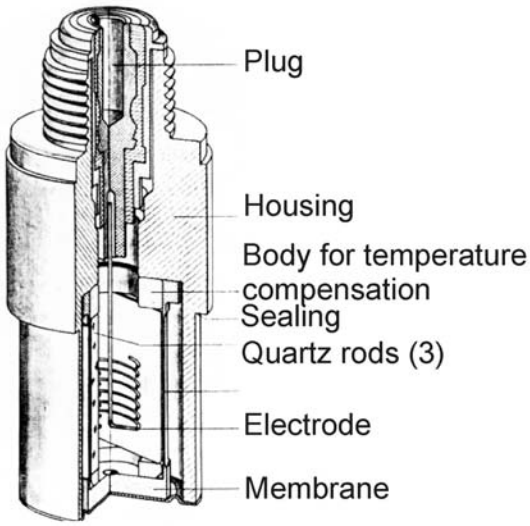


Fig. 11: Design of a piezo-electrical vacuum gauge

Other membrane gauges use the piezo-resistive effect to detect the force acting on the membrane (Fig. 12). In this effect the resistance is changed by a geometrical change of the material and for semiconductors, in addition, by a change of the specific resistance. Therefore mainly semiconductor materials are used for this type of gauge. A silicon crystal that has excellent elastic behaviour is used as the membrane. The low-ohmic resistive 'wires' are obtained by doting the silicon in two directions of the membrane. The resistors are part of a bridge circuit. A compensation for temperature effects has to be provided.

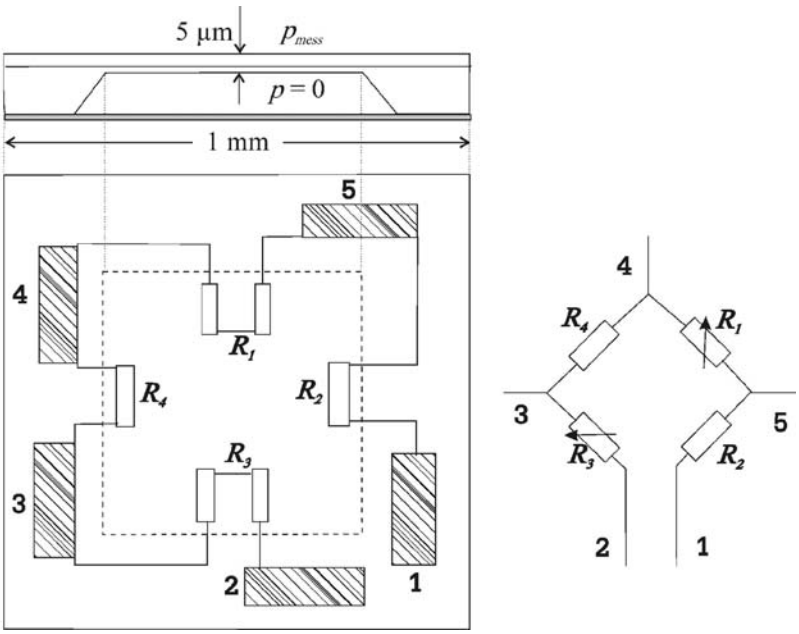


Fig. 12: Schematic diagram of a piezo-resistive vacuum gauge. The device is MEMS fabricated on silicon. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

In the capacitance diaphragm gauge (CDG) the membrane is used as one electrode of a capacitor (Fig. 13). When the membrane is deflected, the capacity will change and this change is converted into the measurement signal. Since capacity can be measured with very high accuracy, it is possible to measure deflections of the membrane of 0.4 nm. This resolution limit corresponds to about 1 mPa.

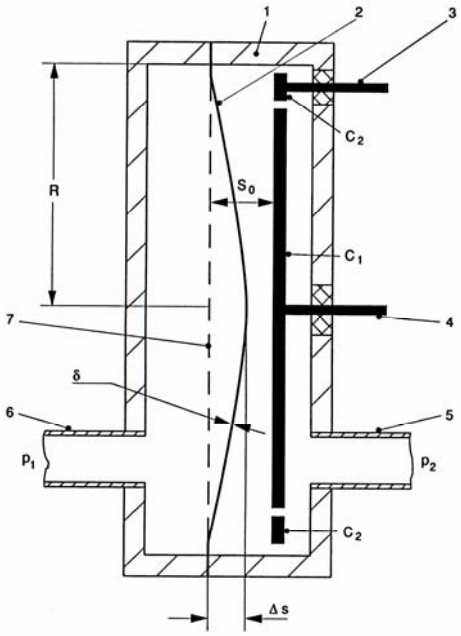


Fig. 13: Schematic diagram of the capacitance diaphragm gauge. 1 housing; 2 membrane; 3,4 electrical feedthroughs to capacitors; 5 gas inlet (reference side); 6 gas inlet (test side); 7 membrane in zero position. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

Such a good resolution and accuracy have been obtained by know-how that has been developed since 1949, mainly by the MKS company [5]. The circular membrane is manufactured from a material with very low expansion coefficient, namely the metal alloy INVAR, or ceramic. The membrane may be as thin as 25 μm. It is welded, soldered or glued to the housing. To improve zero stability, on the reference side there are two electrodes, a circular one in the centre, and a ring electrode around, forming two capacitors. The differences of the two capacities are used as measurand. Both capacitors are part of a bridge (Fig. 14). An oscillator sends a fixed signal of a frequency of about 10 kHz to the bridge which modifies it by amplitude and phase. The amplitude is proportional to pressure, the phase gives the direction of pressure. After demodulation with the oscillator signal a DC signal is produced which is linearized and amplified [5].

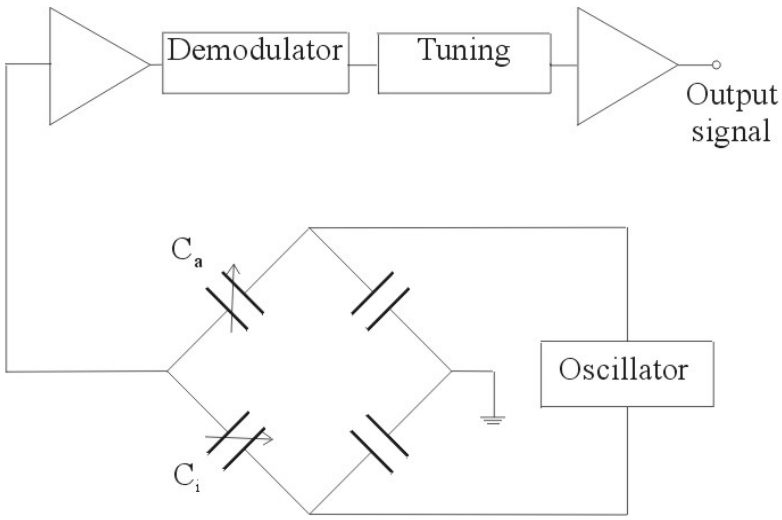


Fig. 14: The electrical circuit of a capacitance diaphragm gauge. C_i is the capacity between membrane and inner electrode, C_a between membrane and outer ring electrode (Fig. 13). From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

Since this instrument is so sensitive, temperature drifts may cause a large change of the null signal. For this reason CDGs for highest accuracy are stabilized to a temperature of about 45°C with a stability of a few millikelvin. This higher temperature causes at lower pressures (< 100 Pa) the so-called thermal transpiration effect which will be treated below.

CDGs are available with both sides of the membrane accessible from the outside (differential mode) or with the reference side enclosed and pumped with a getter material (absolute mode). Their full scales range from 0.1 Torr (13 Pa) to 1000 Torr (133 kPa), their resolution varies from 3½ digits to 5½ digits. The gauge head is normally separated from the controller, but also so-called 'active' gauges are available which have an analog voltage output to be measured with a DVM. They are very common in the semiconductor industry and in calibration laboratories or wherever high accuracy is required.

A MEMS fabricated version of a vacuum gauge with a membrane was presented in the late 1990s [6], the so-called resonance silicon gauge (Fig. 15). Here two 'H'-shaped resonators similar to tuning forks are fixed to a thin membrane of a Si crystal. When the membrane deflects, the eigenfrequency of the resonators is changed and a frequency signal is detuned. The resonator at the edge will reduce its frequency f_2, while the one in the centre will increase the frequency f_1. The difference ($f_1 - f_2$) is proportional to pressure, the sum ($f_1 + f_2$) reflects the common line pressure. Both 'H' resonators have to have their own vacuum enclosure, since otherwise the surrounding pressure would deteriorate the Q factor and also detune the frequency. With this casing only the deflection of the membrane causes the measurement signal. The measurement range of these gauges is at very low uncertainty (10^{-4} range) from 100 Pa to 100 kPa.

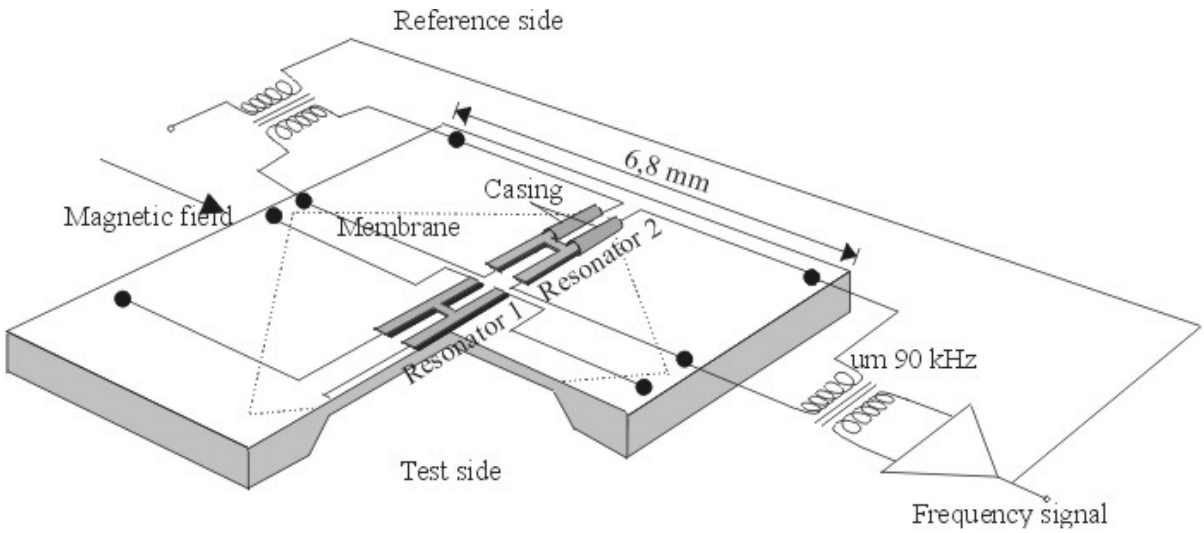


Fig. 15: Design of the resonance silicon gauge by Yokogawa Co, Japan. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

At the end of this section we would like to mention the thermal transpiration effect, since it is important for CDGs and in general for vacuum measurements. This effect develops when the pressure has dropped into the transition regime between viscous and molecular flow and when different temperatures on different locations of a vacuum system prevail (Fig. 16). In equilibrium of molecular flow, the same number of molecules must enter the left volume (Fig. 16) as leave it. Since the velocities of molecules is proportional to the square root of the temperature, this means that the molecular density must be lower when the temperature is higher. Because of the ideal gas law, $p = nkT$, the effect on pressure is reversed. The formulas are given on Fig. 16.

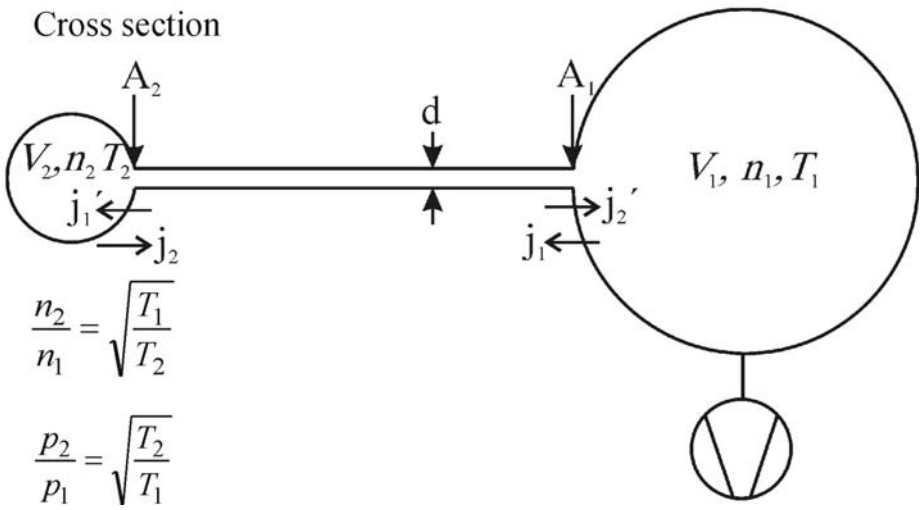


Fig. 16: Illustration of the effect of thermal transpiration. Two volumina of different temperature are connected together where the Knudsen number is > 0.5. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

Since the temperature in the CDG is about 45°C, the pressure in the gauge is increased by about

$$\sqrt{318/296} = 1.036 \ .$$

When CDGs are calibrated, this effect can clearly be seen. Both in the viscous flow regime and the molecular regime the error for the reading is constant with a transitional behaviour in between. Since the mean free path of helium extends to higher values, also the transition regime is shifted to higher pressures. The curve may be fitted accordingly [7], [8].

3 Indirect measuring vacuum gauges

3.1 Thermal conductivity gauges

In a certain pressure range gas conducts thermal energy proportionally to the number of molecules involved in the transport. This effect can be used to measure vacuum pressures: the power (heat) loss of a heated element, usually a wire, to an enclosure of stable temperature (‘room’ temperature) is measured. At higher pressures the gas density is so high and the mean free path of the molecules so short that the gas can be described as a continuum. In this case there is also a heat flow from an element at higher temperature to a wall of lower temperature, but this heat flow does not depend on the pressure. At lower pressures, however, where each molecule can freely travel from the heated element to the wall, each molecule transports some energy and the total amount of energy transport is proportional to the number of molecules. In the intermediate regime the proportionality becomes weaker.

Figure 17 shows the heat flows in a thermal conductivity gauge. The desired heat flow for measurement is $\dot{Q}_{\text{gas}}$, which is the heat transported by the gas molecules from the heated wire to the housing. A pressure-independent null signal gives $\dot{Q}_{\text{rad}}$, the loss due to radiation, and $\dot{Q}_{\text{end}}$, the thermal conduction from the wire to its support structure. There is also some convection $\dot{Q}_{\text{conv}}$ in the viscous flow regime. This is a disturbing effect since it is not stable and orientation dependent, but some gauges use this effect to extend the measurement range up to atmospheric pressure.

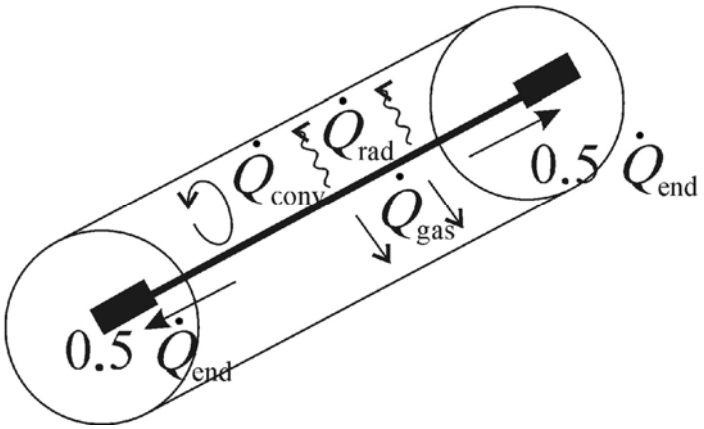


Fig. 17: The heat flow from a heated wire in a thermal conductivity gauge. Q_{gas} desired heat flow by the gas molecules. Q_{rad} heat flow by radiation. Q_{gas} heat flow by convection. Q_{end} thermal conduction in the wire to the holder.

Figure 18 shows the electrical power $\dot{Q}_{el}$ needed to keep the wire temperature constant in a thermal conductivity gauge dependent on pressure. At the very low end there is the pressure independent signal due to $\dot{Q}_{rad}$ and $\dot{Q}_{end}$, then $\dot{Q}_{gas}$ dominates, and finally only convection causes a pressure-dependent signal. Neglecting the convection, we can write

$$\dot{Q}_{el} = \dot{Q}_{gas} + \dot{Q}_{end} + \dot{Q}_{rad} \tag{4}$$

which can be approximated by the following formula:

$$\dot{Q}_{el} = \varepsilon\left(p_0 + \frac{p}{1+gp} \right) . \tag{5}$$

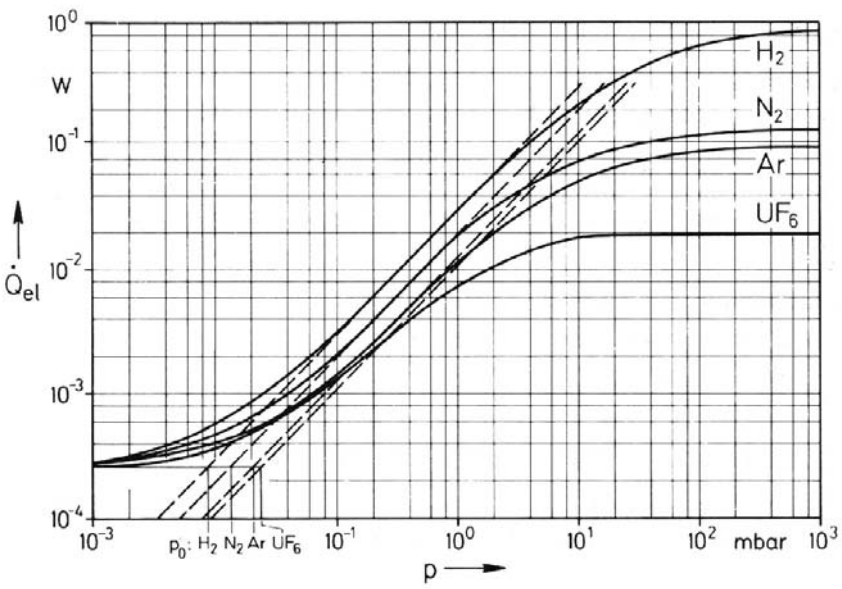


Fig. 18: The electrical power needed to keep the wire temperature constant in a Pirani gauge for different gas species. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

If one keeps the temperature of the wire constant and measures the electrical power Q_{el} needed for this, in some pressure regime $\dot{Q}_{el}$ will be proportional to p.

This type of gauge was invented in 1906 by Pirani [9]. He put the heated wire as part of a Wheatstone bridge which supplied the necessary electrical power. This is still done today and it is common to call a thermal conductivity gauge with a Wheatstone bridge a Pirani gauge. There are, however, several different operational modes: gauges where the temperature of the heated element is held constant are the most accurate ones with the largest measurement range, but also the most expensive. Alternatively, the heating voltage, current, or power can be kept constant (the latter was, what Penning did) and the temperature (resistance) of the wire is measured.

The electrical circuit of a Pirani gauge to keep the wire temperature constant is shown in Fig. 19. The wire has a diameter of 10μm or less. The resistances R_2, R_3, R_4, R_D, are of about the same value. R_T is used to compensate temperature changes of the surrounding enclosure.

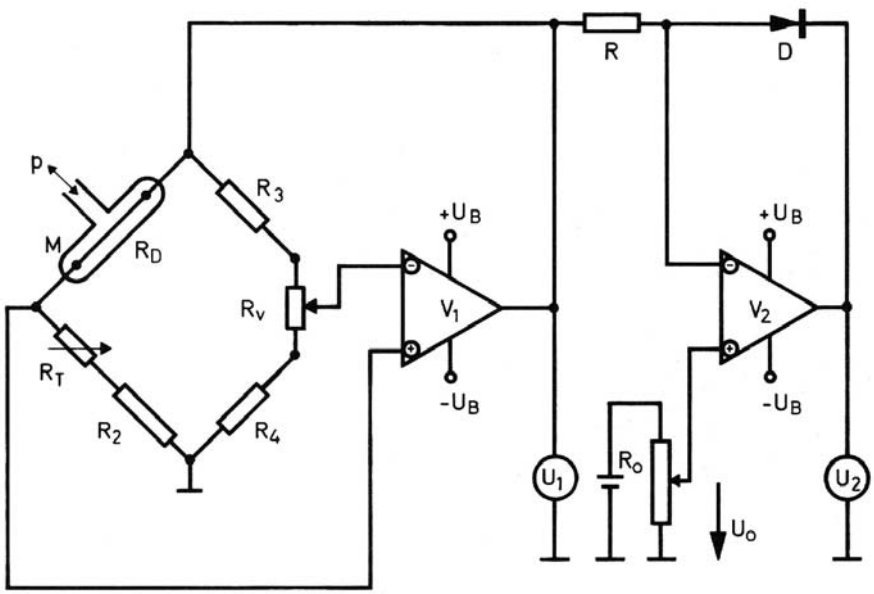


Fig. 19: Electrical circuit of a Pirani gauge. *M* measuring cell; R_D resistance of heated wire. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

A feature of Pirani gauges with constant temperature that may be important for accelerators is that they react faster with pressure changes, since no thermal relaxation times will cause delays which is the case in other configurations.

Most commercially available Pirani gauges need to be adjusted at atmospheric pressure and at a pressure below the resolution limit of about 10 mPa. The adjustment has to be repeated when the wire has changed (blackening, deposits etc.). Their measurement range is typically from 0.1 Pa to 10^4 Pa, sometimes up to atmospheric pressure.

Recently Pirani gauges have been produced by MEMS (Fig. 20). Instead of a heated wire, a heated sheet is used. The temperature was reduced to about 60°C, where macroscopic wires are normally at 120°C. A lower temperature will give a smaller null signal due to radiation and conduction, but increase the dependence on the environmental temperature changes. The small sizes by MEMS have the advantage that the Knudsen number is increased at the same pressure and therefore the molecular regime so that it can be expected that the reading will be more linear up to higher pressures. Another advantage of small structures is that no convection whirls can develop, so the gauge reading is independent of its orientation.

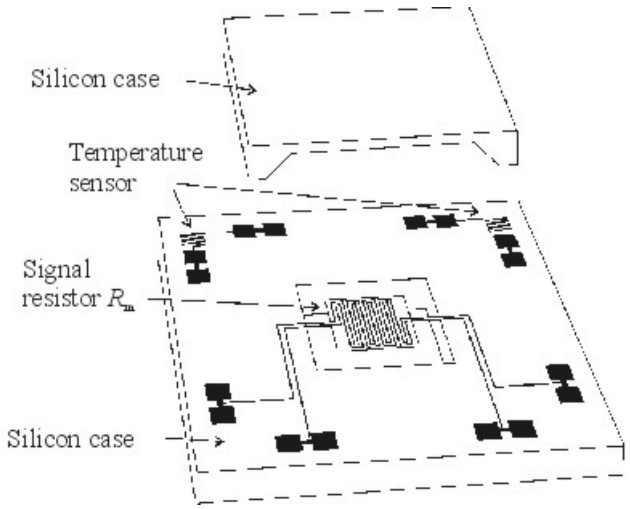


Fig. 20: MicroPirani of MKS produced with MEMS. Instead of a heated wire there is a heated slab. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

The reading of thermal conductivity gauges strongly depends on the gas species as could already be seen on Fig. 18. The physical parameters for this dependence are due to properties of the gas itself (like degrees of freedom, mean thermal velocity of the molecules), but also due to the accommodation of the gas molecules on the respective surfaces. These accommodation coefficients clearly depend on the surfaces and for this reason thermal conductivity gauges are very sensitive to any pollution.

General correction factors from text books for different gas species in thermal conductivity gauges should be handled with great care. Owing to the facts described above, the gas dependence is due to the specific surfaces involved and there is no such general value for a gas species in Pirani gauges. It can be seen from Fig. 21 that the values of different gauges scatter significantly and are also dependent on pressure.

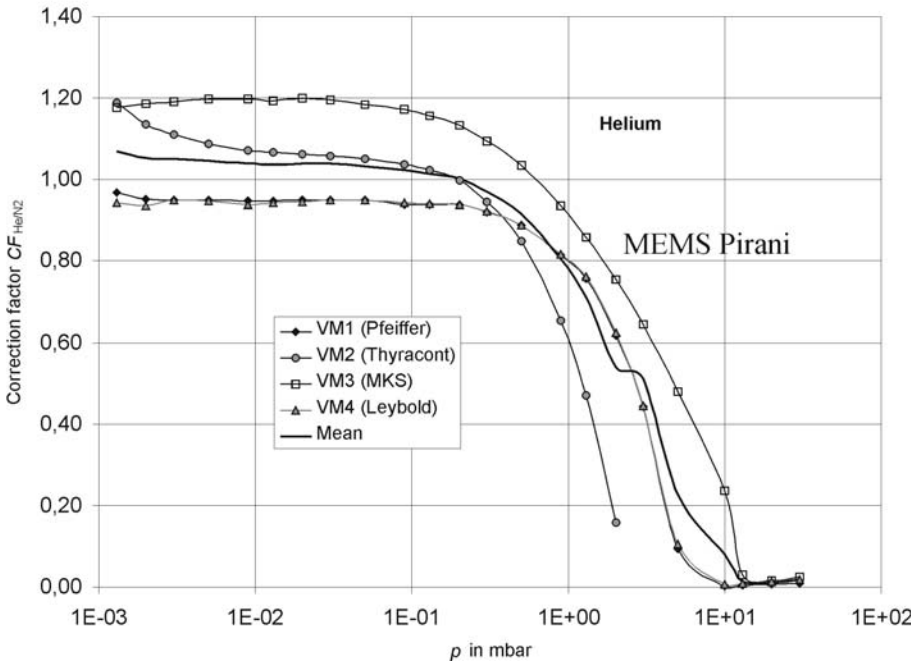


Fig. 21: Correction factor for helium compared to a reading for nitrogen for four different Pirani gauges vs. pressure

A very simple and cheap thermal conductivity gauge is the so-called thermocouple (Fig. 22). A constant current of typically 150 mA is sent through the heated wire and its temperature is measured by a thermocouple of chromel/alumel. The measurement range is rather limited (0.1 Pa–100 Pa) as well as the accuracy.

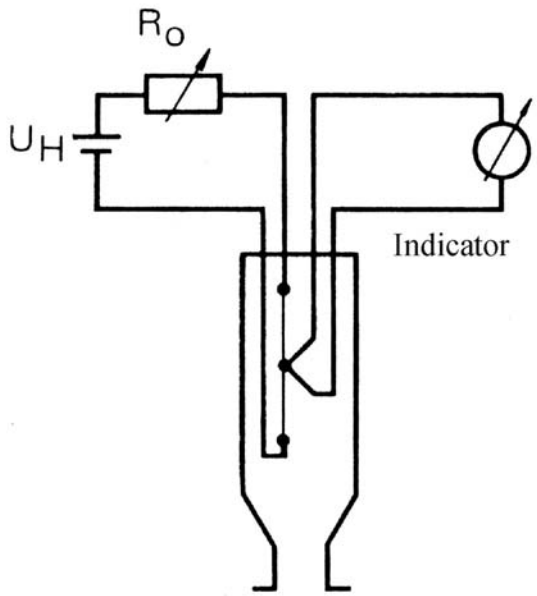


Fig. 22: Design of a thermocouple gauge. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

3.2 Spinning rotor gauge

Similar to the thermal conductivity gauge, the spinning rotor gauge has the best performance when the molecules can freely travel within the gauge and when the signal at the same time is significantly higher than a pressure-independent null signal. The spinning rotor gauge developed by Fremerey, [10]–[12], was the first commercially successful gauge based on the principle of gas friction. As for the thermal conductivity, the viscosity is present at all pressures, but only for pressures where the molecules freely travel within the gauge is it linearly proportional to pressure.

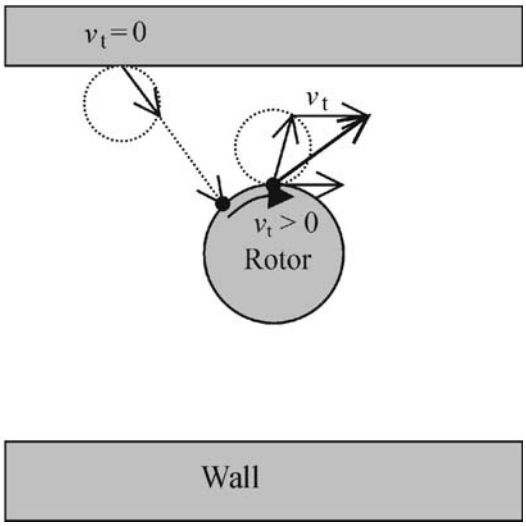


Fig. 23: Operating principle of the spinning rotor gauge. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

The operational principle of the spinning rotor gauge is shown in Fig. 23. A magnetically suspended ball of stainless steel rotates in a vacuum thimble. Molecules that hit the rotor from the wall will stick to it for a moment. If the ball did not rotate, the molecules would desorb again from it with a cosine distribution. Since the ball rotates, however, the molecules leaving the surface will have an additional velocity vector according to the tangential velocity of the rotor. The molecules will have the full tangential velocity of the rotor when they are completely accommodated to the surface. It turned out that this actually happens for technical surfaces and the accommodation factor of tangential momentum is very close to 1. This additional tangential momentum of the molecules is gained from the rotational energy of the rotor. This means that the rotor decelerates with each molecule hitting and leaving the rotor.

The whole effect can be nicely modelled and calculated and the result is that the relative decelaration rate $\dot{\omega}/\omega$ of the rotor is proportional to pressure. Considering also a pressure-independent residual drag *RD*, the pressure is obtained from the measurement signal $\dot{\omega}/\omega$ by

$$p = \sqrt{\frac{8kT}{\pi m}} \cdot \frac{\pi d \rho}{20\sigma} \left[\left(\frac{\dot{\omega}}{\omega} \right) - RD(\omega) \right] . \tag{6}$$

Here T is the temperature, m the molecular mass, d the diameter, ρ the density of the rotor, σ the accommodation factor, and ω the rotor frequency.

Figure 24 shows the schematic of a spinning rotor gauge. The weight of the rotor (1) is mainly balanced by a permanent magnet (3). Additional coils (4) on the top and the bottom suspend and stabilize the rotor vertically, four coils (8) stabilize the rotor in the two horizontal directions. The rotor is accelerated to about 415 Hz by another four coils (5) and the rotor frequency is detected by another two coils (6). They detect the rotating magnetic moment (normally a dipole) of the ball. By the generated current and also by eddy currents generated in the thimble and in the rotor itself by inhomogenities in the magnetic field, the rotor is slowed down independent of the gas molecules impinging on it. This causes the residual drag *RD*.

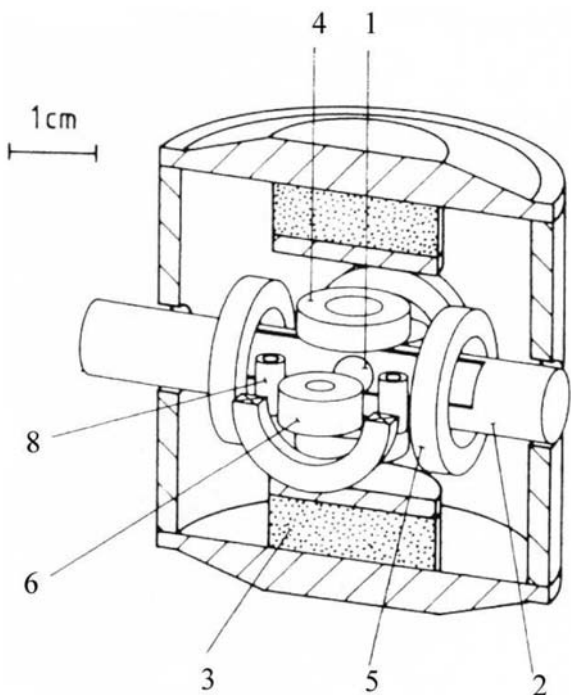


Fig. 24: Schematic of a spinning rotor gauge. 1 rotor; 2 vacuum tube; 3 permanent magnets; 4 two coils for vertical stabilization; 5 four drive coils; 6 two detection coils; 8 four coils for horizontal stabilization. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

Unfortunately this residual drag often depends on the frequency as can be seen in Fig. 25. Since the residual drag corresponds to a nitrogen pressure of 10^{-4} Pa to 10^{-3} Pa, pressure measurements in this range and lower need a careful evaluation of *RD* vs. frequency. Also a stable temperature is required since the pirouette effect by thermal expansion will slow down (or accelerate) the rotor.

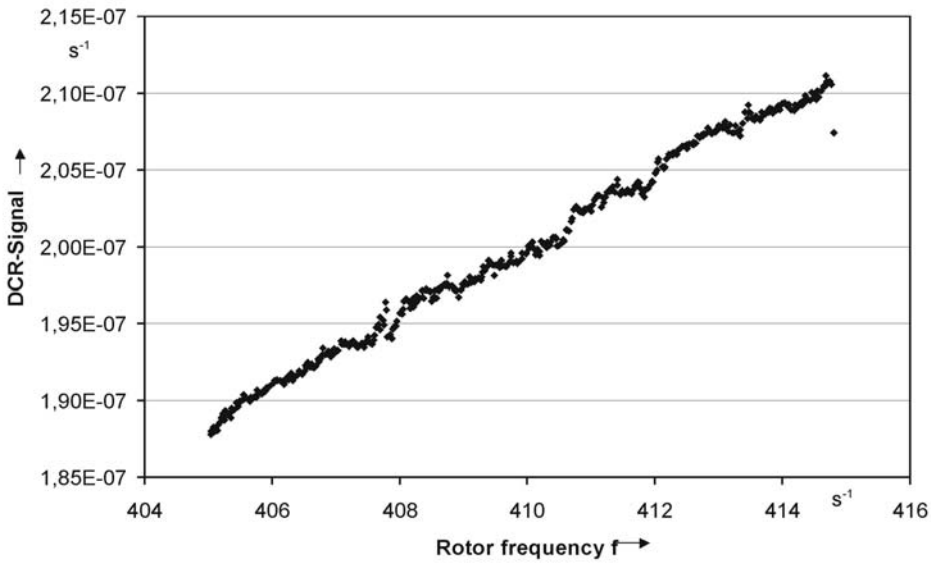


Fig. 25: Residual drag (DCR-signal) vs. rotor frequency for a given rotor. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

The measurement range of the spinning rotor gauge is from 10^{-4} Pa to about 1 Pa. Lower pressures are possible with vibration isolation; at higher pressures the rotor needs frequent re-accelerations, warms up, heats the surrounding gas and the accuracy is lost. At pressures of about 100 Pa the signal becomes independent of pressure, since the gas becomes a continuum.

The spinning rotor gauge is a completely inert vacuum gauge: it does not consume any gas (e.g., by ionization); it does not dissociate molecules (like a hot cathode); its outgassing rate is the same as the thimble wall material. It is an ideal instrument to measure outgassing rates by the pressure-rise method. In addition the spinning rotor gauge gives a very accurate signal (uncertainty as low as 0.3%) and the long-term stability is excellent. For this reason it is an ideal secondary and reference standard in high vacuum and can be used to calibrate ion gauges.

3.3 Ionization gauges for fine vacuum

Since ionization gauges are treated in another article on UHV gauges, we restrict ourselves here to ionization gauges for higher pressures (> 0,1 Pa). For the ionization principle we also refer to this article. Figure 26 gives the pressure reading obtained from the collector current for different types of ionization gauges. In all cases there is a turning point where a higher pressure causes a lower signal. There are several reasons for this.

- The mean free path of the electrons is reduced at higher pressures. With lower energy the ionization probability decreases.
- The mean free path of the ions is reduced. The collision with neutrals may reduce the probability that the ion reaches the collector.
- The secondary electron current from the ionized molecules becomes close to the primary electron current. Since the total current is kept constant, the ionization rate is reduced.

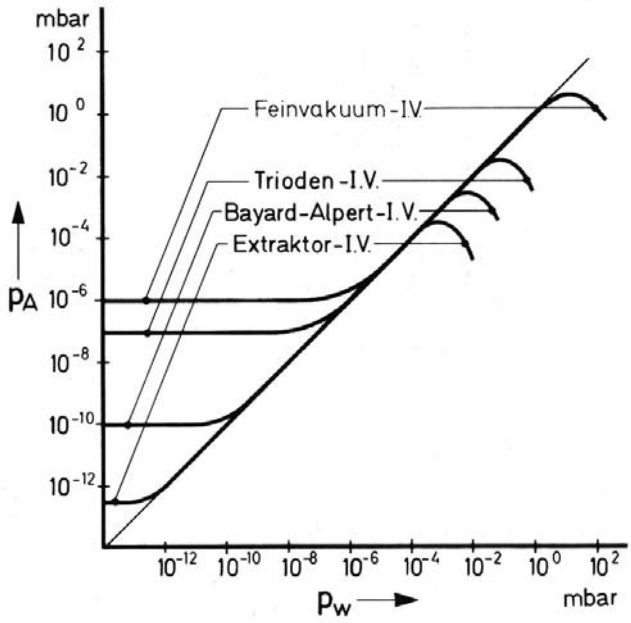


Fig. 26: Pressure reading obtained from the collector current vs. true pressure for different types of ionization gauges. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

To make an ionization gauge work at high pressures, the distances between the electrodes have to be reduced, the collector surface is enlarged, and the collector geometry must be such that the collection efficiency is optimized [13]. Figure 27 shows an example. Compared with the classical triode gauge, in the fine vacuum ionization gauge of Fig. 27 the anode and collector have changed position.

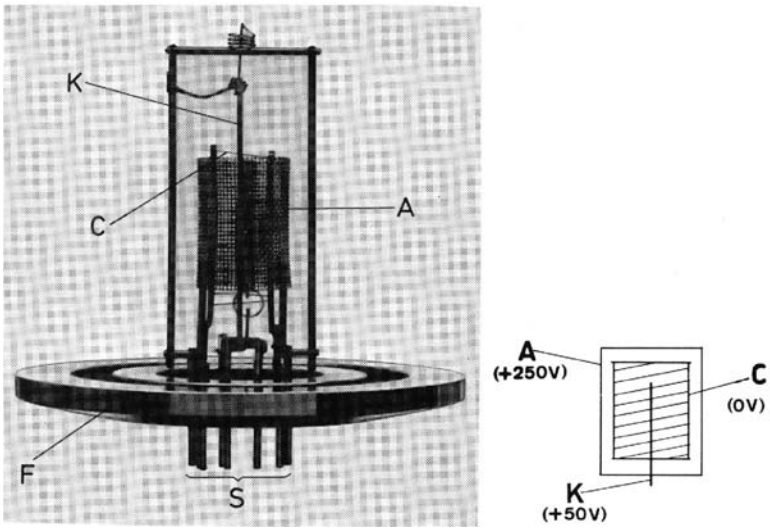


Fig. 27: Fine vacuum ionization gauge by Leybold Co. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

Last but not least, it is necessary to operate hot cathodes in ionization gauges for the fine vacuum with an oxide layer on an iridium cathode (ThO_2/Ir or Y_2O_3/Ir) to avoid rapid ageing due to corrosion by water vapour or air components. It is also necessary to operate the gauge at a low emission current (about 0.1 mA or less).

3.4 Optical methods

There were a number of classical optical methods attempted in the past, but the reader is referred to textbooks (e.g., Ref. [14]), since these attempts did not have a lasting effect. For accelerators and other long vacuum systems like gravitational wave detectors, it is worth noting, however, that there were a number of successful partial pressure measurements with tunable diode laser absorption spectroscopy ([15]–[17]). For example, with an infrared (IR) laser beam of a length of 100 m a resolution of 10^{-5} Pa for CO was described [15]. With multireflection cells a geometrical distance can easily be enlarged by a factor of 100 for the total path length. In long vacuum systems it seems feasible to reach resolutions in the UHV range for many IR absorbing gases like H_2O, CO, CO_2, CH_4. Clearly, all sensitive absorption measurements are gas-species selective and it is not possible to measure a total pressure. The advantage of this kind of partial pressure measurement is that it is an inert system (no dissociation, ionization, no chemical reactions, no outgassing etc.), you can average out gas density measurement over a long path (not a point measurement as with normal vacuum gauges), and areas are accessible that are otherwise not measurable.

4 Accuracy of vacuum gauges

For many users the accuracy of vacuum gauges is of minor importance, since either they need the vacuum pressure information only to switch between pumps, to open and close valves etc. and work otherwise at residual pressure or, when they start a process or experiment by admitting gas into their system, they ‘calibrate’ their vacuum gauges by the success or failure of the process or experiment. While the latter procedure certainly raises some questions, the accelerator community in the routine operation belongs mainly to the first group and they need the reading of vacuum gauges with perhaps an accuracy of a factor of 2. However, when the accelerator community exchanges data like photodesorption yields, outgassing rates, cross-sectional areas, pumping speeds etc., the values should be absolute in order to be comparable and in order to be usable at different places.

Considering the last point, this concluding section will make the reader pay attention to the uncertainty of measured values with vacuum gauges as already mentioned in the introduction. It should be made clear that in general any measured value is not of much use when there is no additional value about the estimated uncertainty of this value. Only this pair (value, uncertainty) makes it possible to have confidence in a measurement and to compare results.

Table 1 lists the sources of uncertainties that may contribute in general to the total uncertainty of a measured result of a vacuum gauge. Table 2 lists the sources that are directly related to the measurement with the vacuum gauge.

Table 1: Sources of uncertainties that contribute to the total uncertainty of a measured result of a vacuum gauge

General	Example
Uncertainties due to calibration chain	Has the vacuum gauge been ever calibrated? Against what standard?
Uncertainties due to installation	Pressure at gauge position may not reflect the pressure where the experiment takes place.
Uncertainties due to operation	Outgassing of an ion gauge may falsify an outgassing rate measurement.
Inaccuracies caused by the physical principle of measurement	Thermal conductivity or ion gauge is used, but gas mixture is not (accurately) known.
Uncertainties caused by the device itself	See Table 2.

Table 2: Sources of uncertainties due to the vacuum gauge itself

General	**Examples**
Offset measurement	Residual drag in SRG, zeroing of Pirani gauge, X-ray and ESD effect for ion gauges
Offset instability (drift)	Offset drifts with environmental temperature (pirouette effect in SRG), bridge is no more balanced with time
Resolution	Number of digits shown
Influence of environment (mainly temperature)	Enclosure temperature of Pirani changes, varies, thermal transpiration effect changes in CDG, amplifier changes amplification
Non-linearity	Ion gauge (sensitivity changes with pressure)
Integration time (scatter of data), repeatibility	Same signal at repeat measurements? Integration time in SRG, in picoammeter with ion gauge.
Reproducibility (stability of calibration constant)	Calibration constants change with time.
Hysteresis	Mechanical gauges (up, down measurement)
Prior usage, cleanliness	Surfaces change, accommodation coefficients change, secondary yield changes

A typical problem of uncertainty is the use of an indirectly measuring vacuum gauge whose reading is gas-species dependent but the composition of the measured gas is not known. In this case, most people use the trick to give the value ‘pressure as nitrogen equivalent’. This means that the gauge is calibrated or set to nitrogen and the reading is taken as nitrogen equivalent pressure. The true pressure of the real gas composition may be far different.

If the gas composition is approximately known (uncertainty!), very often a weighted mean may be taken to correct for a true pressure reading. For example, for the spinning rotor gauge an effective mass

$$m_{\mathrm{eff}} = \left(\sum_{\mathrm{i}=1}^{n} a_{\mathrm{i}} \sqrt{m_{\mathrm{i}}} \right)^2 \qquad \sum_{\mathrm{i}=1}^{n} a_{\mathrm{i}} = 1 \tag{7}$$

can be used, where a_{i} is the relative portion of molecule m_{i} of the total pressure with n components [18]. Similarly, the correction factors CF for gas species can be taken for thermal conductivity gauges or ion gauges:

$$CF_{\mathrm{eff}} = \sum_{\mathrm{i}=1}^{n} a_{\mathrm{i}} CF_{\mathrm{i}} \quad . \tag{8}$$

Table 3 lists the total measurement range, the typical relative uncertainty in this range, the optimum measurement range, and the relative measurement uncertainty in the latter for several commercial vacuum gauges including gauges for the UHV. ‘Relative’ means compared to the indicated value. Figure 28 illustrates this table graphically, but also the uncertainties of the vacuum pressure scale as realized by the three primary standards (mercury manometer, static expansion, and continuous expansion [19]) are given.

Table 3: Relative measurement uncertainty of commercially available vacuum gauges

Gauge type	Measurement range in Pa	Normal uncertainty	Optimum range in Pa	Lowest uncertainty
Piston gauge	$10–10^5$		$10^2–10^5$	$1 \cdot 10^{-4}–1 \cdot 10^{-5}$
Quartz Bourdon manometer	$10^3–10^5$		$10^3–10^5$	$3 \cdot 10^{-4}–2 \cdot 10^{-4}$
Resonance silicon gauge	$10–10^5$	$3 \cdot 10^{-3}–5 \cdot 10^{-4}$	$100–10^5$	$2 \cdot 10^{-4}–5 \cdot 10^{-5}$
Mechanical vacuum gauge	$10^2–10^5$	0.1–0.01		
Membrane vacuum gauge	$10^2–10^5$	0.1–0.01		
Piezo	$10^2–10^5$	1–0.01		
Thermocouple gauge	$10^{-1}–10^2$	1–0.3		
Pirani gauge	$10^{-1}–10^5$	1–0.15	1–100	0.1–0.05
Capacitance diaphragm gauge	$10^{-4}–10^5$	$10^{-1}–3 \cdot 10^{-3}$	$10^{-1}–10^5$	$4 \cdot 10^{-3}–1 \cdot 10^{-3}$
Spinning rotor gauge	$10^{-5}–10$	$0.1–7 \cdot 10^{-3}$	$10^{-3}–10^{-1}$	$4 \cdot 10^{-3}$
Penning gauge	$10^{-7}–1$	$5 \cdot 10^{-1}–2 \cdot 10^{-1}$	$10^{-5}–1$	$3 \cdot 10^{-1}–1 \cdot 10^{-1}$
Magnetron gauge	$10^{-8}–1$	$1–1 \cdot 10^{-1}$	$10^{-6}–1$	$1 \cdot 10^{-1}–2 \cdot 10^{-2}$
Ionization gauge (emission cathodes)	$10^{-10}–10^{-2}$	$1.5 \cdot 10^{-1}–5 \cdot 10^{-2}$	$10^{-8}–10^{-2}$	$5 \cdot 10^{-2}–2 \cdot 10^{-2}$

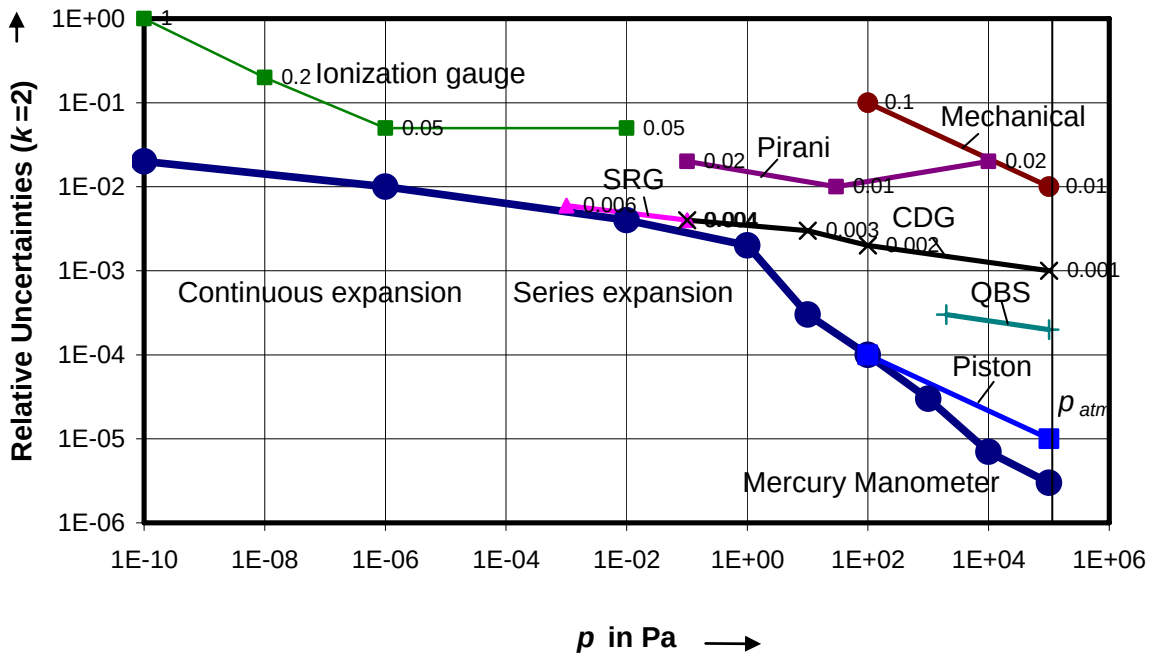


Fig. 28: Illustration of the relative uncertainties of the vacuum pressure scale by primary standards (bold line) and commercially available vacuum gauges

References

[1] http://www.bipm.org

[2] S. J. Bennet, P. B. Clapham, J. E. Dadson and D. I. Simpson, Laser interferometry applied to mercury surfaces, J. Phys. E: Sci. Instrum. **8** (1975) 5–7.

[3] J. Jäger, Use of a precision mercury manometer with capacitance sensing of menisci, Metrologia **30** (1993/94) 553–558.

[4] P. L. Heydemann, A fringe-counting pulsed ultrasonic interferometer, Rev. Sci. Instrum. **42** (1971) 983–986.

[5] J. J. Sullivan, Development of variable capacitance pressure transducers for vacuum applications, J. Vac. Sci. Technol. A **3** (1985) 1721–1730.

[6] K. Harada, Various applications of resonant pressure sensor chip based on 3-D micromachining, Sensors and Actuators A **73** (1999) 261–266.

[7] T. Takaishi and Y. Sensui, Trans. Faraday Soc. **59** (1963) 2503–2514.

[8] J. Setina, Metrologia **36** (1999) 623–626.

[9] M. Pirani, Deutsche Phys. Gesellschaft, Verh. **8** (1906) 686.

[10] J. K. Fremerey, Rev. Sci. Instrum. **44** (1973) 1396–1397.

[11] J. K. Fremerey, Vacuum **32** (1982) 685–690.

[12] J. K. Fremerey, J. Vac. Sci. Technol. A **3** (1985) 1715–1720.

[13] G. J. Schulz and A. V. Phelps, Rev. Sci. Instrum. **28** (1957) 1051.

[14] A. Berman, *Total Pressure Measurements in Vacuum Technology* (Academic Press, Orlando, FL, 1985), p.156.

[15] E. Lanzinger, K. Jousten and M. Kühne, Vacuum **51** (1998) 47–51.

[16] R. S. Inman and J. F. McAndrew, Anal. Chem **66** (1994) 247–2479J.

[17] G. J. Padilla Viquez, Dissertation, TU Berlin (2005).

[18] L. D. Hinkle and P. R. Jacobs, J. Vac. Sci. Technol. A **11** (1993) 261–263.

[19] J. 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, FL, 1985).

Saul Dushman, *Scientific Foundations of Vacuum Technique*, 2nd edition (John Wiley & Sons, New York, 1962).

Karl Jousten, *Wutz Handbuch Vakuumtechnik*, 9th edition (Vieweg, Wiesbaden, 2006), ISBN 3-8348-0133-X.

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, 1989).

P.A. Redhead, J.P. Hobson and 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.

Non-thermal outgassing

N. Hilleret
CERN, Geneva, Switzerland

Abstract
Non-thermal outgassing, i.e., the outgassing stimulated by the impact of energetic particles, is the predominant source of gas during the operation of accelerators and storage rings. This source of gas is not constant and changes with the integrated beam intensity or according to the circulating beam(s) characteristics. In this contribution, the various aspects of stimulated desorption will be reviewed in the case of ions, electrons, or photons impacting on technical materials commonly encountered in accelerators. Finally the consequences of these phenomena on the operation of accelerators will be described.

1 Introduction

The pressure achieved in the absence of beams in accelerators is often much smaller than the pressure required for reasons of beam life-time, particle stability, or beam-induced background or radiation in experimental areas. The typical operating pressure for an accelerator (with the exception of special cases like low-energy, heavy-ion storage rings) is in the 10^{-7} Pa range but pressures in the absence of beams could be as low as 10^{-11} Pa in some parts of the Intersecting Storage Rings (ISR) or of the Large Electron–Positron storage ring (LEP) at CERN. In the presence of beams, the situation changes drastically and the initially low pressure increases by orders of magnitude owing to the impact of energetic particles on the vacuum chamber walls. The large difference in pressure between the two situations demonstrates that the Non-Thermal Outgassing (NTO) phenomena create a gas load that is greater by orders of magnitude than the thermal outgassing. During the design of accelerator vacuum systems, it is hence of utmost importance to adequately estimate the gas load produced by NTO in order to adapt the cleaning treatments and the pumping speed to the pressure requirements. The aim of this lecture is to summarize the actual knowledge on NTO (often called ‘dynamic pressure effects’ in the field of accelerators).

After a brief historical reminder, the various forms of desorption, at room and cryogenic temperature, induced by energetic particle bombardment (ions, electrons, photons, heavy ions), will be treated in the context of accelerator construction and operation.

2 History

As is frequently the case in the history of vacuum technology, the vacuum tube industry was a major contributor to the study of Electron-Stimulated Desorption (ESD). The creation of ions by the electron bombardment of surfaces was first reported by Dempster [1] in 1918 during his study of ‘positive rays’. This positive ion formation under electron bombardment was a serious source of difficulties for the operation of electron tubes as it led to the destruction of the oxide-based electron emitters [2]. Empirically a threshold for this effect was found around 10 eV, a value remarkably close to what has been measured more recently [3]. First measurements of neutral molecule desorption were published in 1942 by Isagawa [4]. With the development of Ultra High Vacuum techniques in the 1960s, it became possible to produce

surfaces clean enough to study ESD more precisely. This resulted in the birth and the development of the Menzel–Gomer–Redhead (MGR) theory for ESD in 1964 [5–7]. This model was the main framework for the study of ESD until the publication of the Knotek and Feibelman model in 1978 [8]. Nowadays, ESD is regrouped with Photon Stimulated Desorption (PSD) to form the field of Desorption Induced by Electronic Transition (DIET). DIET is still a flourishing subject [9] having its own forum in the form of a series of biennial conferences since 1983.

All these studies were/are related mainly to the desorption of ions. The desorption of neutral particles is nevertheless at least 10 times more abundant and has many consequences for the vacuum specialist [10], especially in accelerators. For example, the study of spurious signals in Bayard–Alpert ionization gauges was P.A. Redhead's major interest when he started his studies on ESD [11]. In the field of accelerators many parasitic effects are due to NTO. Amazingly, during the year of the publication of the MGR model, i.e., 1964, a procedure was presented simultaneously by G.E. Fischer and R.A. Mack [12] and M. Bernardini and L. Malter [13] following an initial model published by E. Garwin [14] to predict the pressure increase in electron storage rings. They described a two-step process: photoemission followed by desorption stimulated by the photoelectrons and used neutral ESD to predict the pressure behaviour in electron storage rings. With the development of large and powerful electron storage rings, and later of multiTeV proton storage rings, numerous experiments of ESD and/or PSD have been carried out by the accelerator community to anticipate the performance of future vacuum systems [15–22].

Although sputtering was discovered by W.R. Grove in 1852 [23], Ion-Induced Desorption (IID) was for a long time not a subject for studies but an old recipe to clean surfaces [24]. A strong interest in the release of gases by ion bombardment came in the 1970s. Simultaneously in the accelerator and the fusion community, this release was identified as a main limitation in the performance of storage rings [25] and tokamaks [26]. In 1974, H.F. Winters and P. Sigmund [27] showed that the classical theory of sputtering [28] can partly explain the desorption of chemisorbed gases by low-energy ions. The validity of this approach was confirmed experimentally [29] with some restriction in the case of technical surfaces due to the absence of a well-defined binding energy for the desorbed molecules [30]. Surface physics investigations of technical materials were successfully launched at that time [31–34] in order to limit the detrimental consequences of IID, they contributed to defining a set of cleaning treatments [35–37], mainly based on ion bombardment, which resulted in a significant improvement for the performance of tokamaks [38] and storage rings (in the ISR the stored currents rose from 20 A to 60 A).

3 Mechanisms

In particle accelerators, energetic particles bombard the vacuum chamber walls, they can be (non-exhaustive list):

- ions created by the ionization of the rest gas by the circulating beam [25, 39],
- secondary electrons emitted by the vacuum chamber and accelerated across the chamber by bunched beams [40] ('electron clouds'),
- photons radiated by a relativistic bent beam,
- primary particles escaping from their trajectory and reaching the vacuum chamber (case of heavy-ion accelerators) [41].

The energy dissipated by these particles during their slowing down in the vacuum chamber leads to the desorption of molecules which could be present on the surface or synthesized before their liberation. The mechanisms involve either separately or simultaneously the excitation of the substrate electrons

and/or collisions with the substrate atoms. In the energy range of interest, the processes involved during the slowing down of particles are reviewed in one session of this CERN Accelerator School [42].

3.1 Ion-induced desorption

The mechanisms involved in the desorption by ions (usually eV to keV in accelerators) [43, 44] depend on the velocity of the incoming ion compared to the Bohr velocity of the substrate electrons. A comprehensive review of the various processes involved during the slowing down of this type of ion in matter can be found in Ref. [45]. The model described by H.F. Winters and P. Sigmund [27] is shown in Fig.1 and is relevant for low ion energies. Further development and experimental work based on this model can be found in Refs. [29, 30, 46–51]. It can be summarized in the following way: if the adsorbed molecule is isolated, three main contributions leading to desorption can be identified:

- direct knock-off of the adsorbate by the incoming ion (mechanism 1),
- desorption due to primary ions reflected from the substrate (mechanism 2),
- desorption due to the displaced (sputtered) substrate atoms which received enough energy to liberate the adsorbed molecule (mechanisms 3A and 3B).

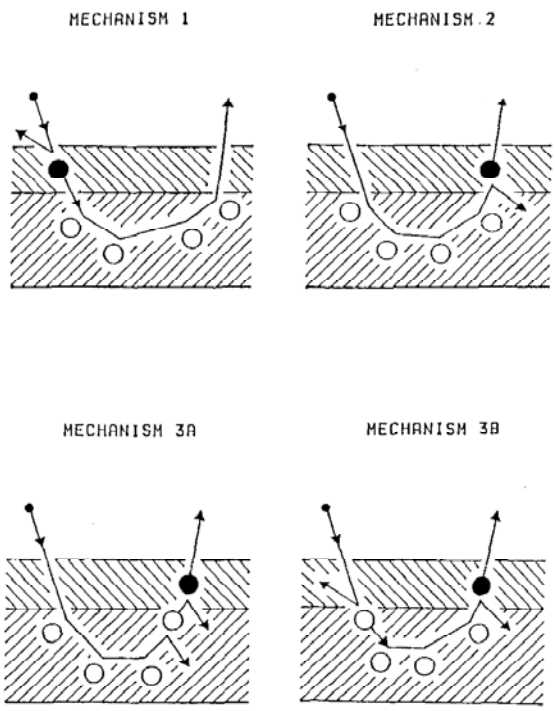


Fig. 1: Ion desorption mechanisms according to Winters and Sigmund [27] as represented by Yamamura and Kimura [46]

This mechanism is well adapted to the study of the desorption of chemisorbed gases from metals (i.e., binding energy in the eV range). Based on this model of binary collisions, a number of simulation programs have been developed which have been applied to the study of desorption [47, 52, 53]. In the case of physisorption (binding energy in the meV range), for example, the desorption of condensed gases, other desorption processes, e.g., electronic sputtering, can be predominant [54]. This stresses the importance of the adsorbate binding energy in the determination of the dominant process leading to desorption.

In the case of swift heavy ions (MeV or higher energy), often partially or totally stripped, the situation is totally different. The slowing down of the fast ions in matter is dominated by electronic stopping processes [42]. Models based on thermal spikes [55, 56] are used to predict sputtering and desorption from metallic surfaces or insulators [57–59]. Furthermore, for higher energies (~GeV), the picture is made more complex by the generation of showers releasing high-energy electrons and photons which can also desorb molecules along their trajectory.

3.2 Electron- and photon-induced desorption

In both cases desorption is the result of an electronic excitation as described in the MGR model [5–7] (Fig. 2). Schematically the process is the following: a binding electron of the adsorbate is excited in an antibounding state after excitation by an electron or photon. The ion formed has a high probability of being neutralized by tunnelling of an electron from the substrate resulting in the desorption of a neutral atom. In the absence of neutralization, the same process leads to the emission of a positive ion (a fraction between 10^{-2} to 10^{-4} of the total amount of desorbed species). A more detailed discussion on the mechanisms of deexcitation leading to desorption is presented in Refs. [60–62]. Special cases led to appropriate developments: for condensed gases (especially rare gas solids) excitons play a significant role in the desorption [54, 63, 64], at very low impact energy the formation of transient anions results in marked maxima in the variation of desorption with incident particle energy [65].

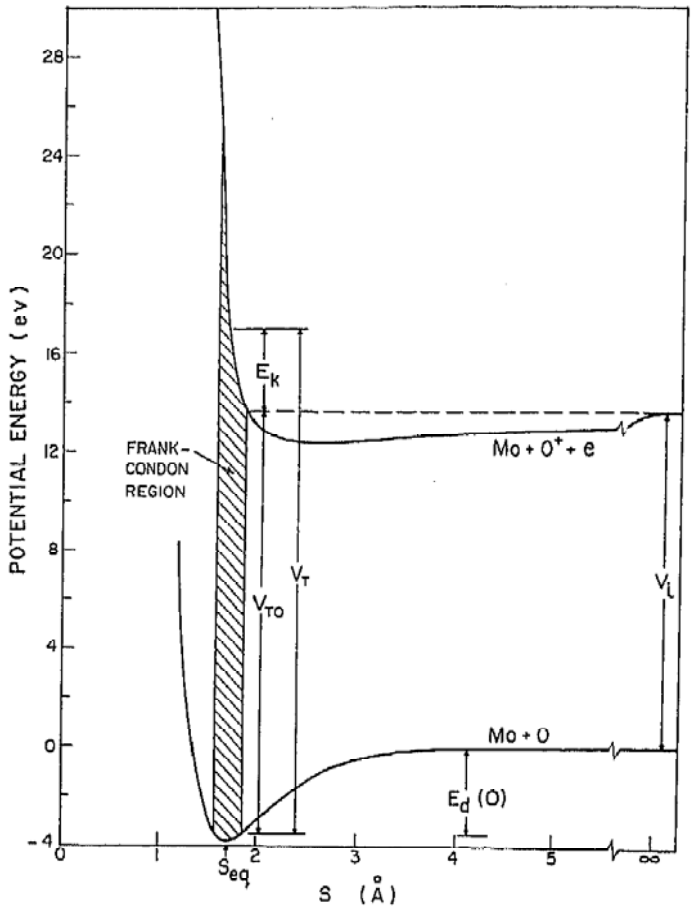


Fig. 2: Schematic potential energy diagram for desorption of oxygen from a molybdenum surface (copied from P.A. Redhead [7])

3.3 The case of technical surfaces

For technical surfaces, the situation is more complicated as the surface contains a lot of physical and chemical imperfections, which makes predictions of desorption cross-sections difficult. Furthermore other effects related to the energy deposition of the primary particle such as defect creation, stimulated diffusion and excitation of electrons can influence the release of impurities contained in the solid. This can justify the very intense argon glow discharge cleaning (10^{17} A^{+}ions/cm^2) required to obtain a low desorption yield for a stainless steel surface, releasing an equivalent of 77 monolayers of CO [66]. The surface oxide present in all technical materials plays a determinant role in the process of desorption as has been shown in the case of aluminium and copper for ESD and PSD [67–69]. The impurities contained in that layer and their migration are considered to be at the origin of the pressure evolution observed at the start of long photon irradiation as shown on Fig. 3 [67]. On this figure the C-containing species follow an evolution that significantly differs from the one observed for molecules containing oxygen. As mentioned above, diffusion is also considered responsible for the evolution of the desorption yields with the dose of photons or electrons received in the case of PSD [70, 71] and ESD [72]. Clearly the refinements of models used to study the desorption of pure gases from monocrystals are of little use in the case of technical surfaces.

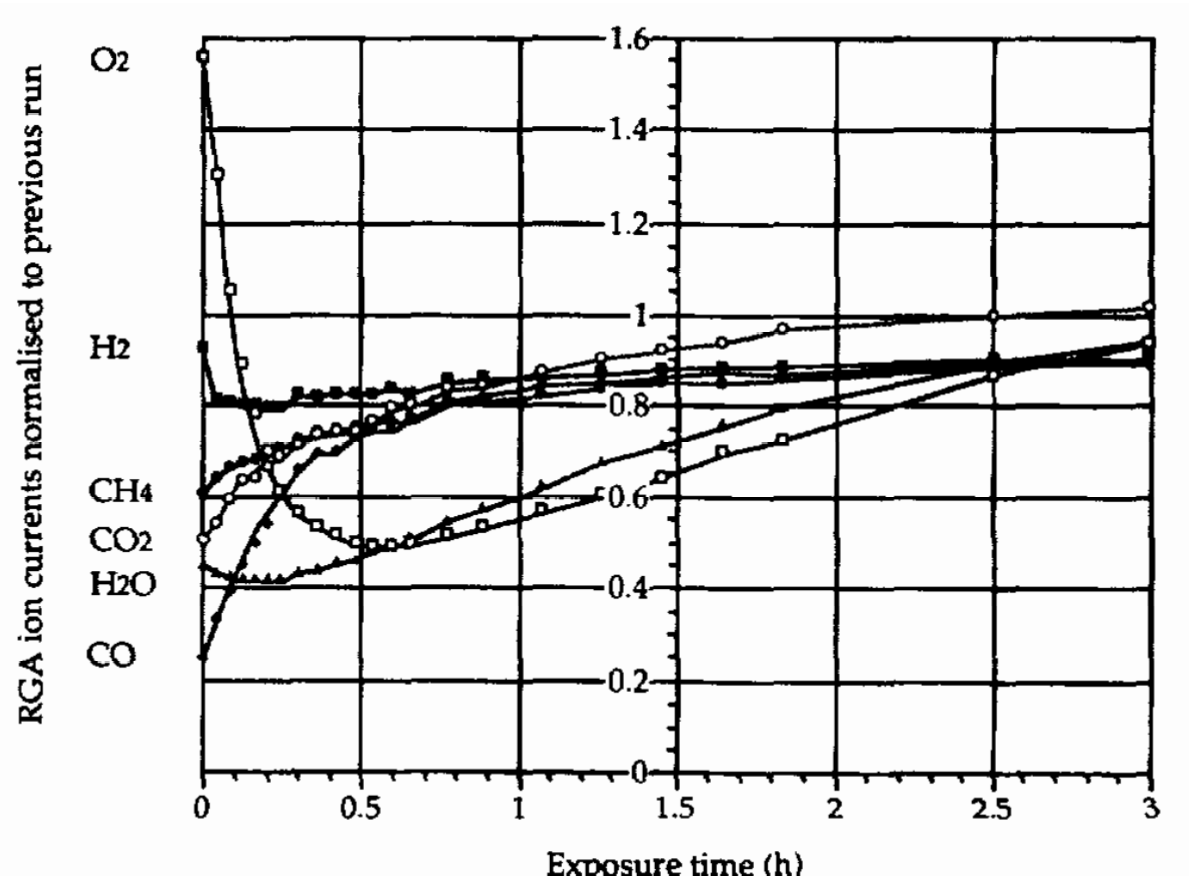


Fig. 3: Partial pressures evolution at the start of a long irradiation following an interruption of several hours [67]

4 Definitions and measurement procedures

4.1 Desorption yields

The desorption yield (η) is the number of molecules (N_m) of a given gas species desorbed per incident particle according to the formula:

$$\eta = \frac{N_m}{N_i} \ .$$

4.2 Desorption cross-section

A cross-section σ can be defined for the desorption processes as

$$\sigma = \frac{\eta}{N_d} \ ,$$

where N_d is the number of molecules on/in the solid, available for desorption per unit surface area and η the desorption yield. In the case of condensed gases, the coverage can be measured accurately and the evaluation of the formula is straightforward to obtain the cross-section.

In the case of room-temperature technical surfaces, the number of molecules per unit area is not known and the cross-section can be obtained from desorption measurements using the following relations:

$$Q_d(t) = N_0 - N_d(t) \ ,$$

$$N_d(t) = \frac{\eta(t)}{\sigma} \ ,$$

$$Q_d(t) = N_0 - \frac{\eta(t)}{\sigma} \ ,$$

$$\eta(t) = \sigma \times \left[N_0 - Q_d(t) \right] \ ,$$

where N_0 is the number of molecules on/in the solid available for desorption per unit surface area at time t_0 and $Q_d(t)$ the total number of desorbed molecules per unit surface area at time t.

Plotting η as a function of the number of desorbed molecules Q_d should yield a straight line whose slope is the desorption cross-section and the intercept with ordinates the product of the initial number of molecules available for desorption with the desorption cross-section.

4.3 Measurement procedure

As described in the preceding section, the measurement of the desorption yield consists in the measurement of the incident particle flux and of the corresponding amount of desorbed molecules. The main ways of obtaining these two quantities are described in the following paragraphs. Other methods of measurement use the integrated flux of incident and desorbed particles to calculate the yields.

4.4 Measurement of the impinging particle flux

For charged particles, the measurement of the particle flux consists of a current measurement. The main difficulties are related to the suppression of parasitic emission or reflection leading to an uncertainty on

the number of incident particles and to the desorption of molecules originating from surfaces outside the studied area. This parasitic desorption from undefined surfaces is of great importance when surfaces with very low desorption yield are studied since a small number of particles impinging on a high-yield area may represent a predominant part of the total desorbed flux. When conditions are favourable for reflections, e.g., grazing incidence, it is very important to assess the contribution of the reflected particles to the desorbed flux. For PSD from synchrotron light the number of radiated photons can be calculated by using the formulae given in Ref. [15]. Finally, in all cases, the secondary electron flux generated by the incoming energetic particles must be suppressed (positive bias on the sample) or corrected (if possible) to measure accurately the number of incident particles.

4.5 Measurement of the desorbed flux

Several methods can be used to measure the desorbed flux:

- In the case of high-energy ions, surface analysis methods can be used to estimate the total number of molecules released:

  Elastic Recoil Detection Analysis (ERDA) is used for the study of H_2 release from C samples by 200 MeV I^+ ions [73].

  Rutherford backscattering is also used to study the desorption of gas ices by MeV H or He ions [74, 75].

- Mass loss, measured by a quartz microbalance, can also be used in the case of condensed gases. Mass loss is inferred from the change of the oscillating frequency of the quartz [76, 77]. This method is especially accurate for thick gas coverages of heavy gases.
- Calibrated mass spectrometers are very widely used in the accelerator community for desorption measurements [37, 78]. This last method will be studied in more detail in the next paragraph.

4.5.1 More details on the mass spectrometer method

The measurement of the desorbed flux using a mass spectrometer requires

- the calibration of the residual gas analyser,
- the determination of cracking patterns for each desorbed gas,
- the correction for the contribution of several gases to the same peak (e.g., contribution of CO_2, C_2H_4, C_3H_8, C_2H_6 to the peak corresponding to mass 28).

The sensitivity of the mass spectrometer can be checked (or measured), provided the relative sensitivity of the mass spectrometer is known for the main desorbed gases, by comparing the sum of the relative partial pressure increase to the total pressure increase as measured by a total pressure gauge using the following relations:

$$I_{\mathrm{Tg}} = i_{\mathrm{eg}} \Sigma \left(P_{\mathrm{g}} s_{\mathrm{Ag}} s_{\mathrm{rgg}} \right) ,$$

$$P_{\mathrm{g}} = \frac{I_{\mathrm{Ag}}}{s_{\mathrm{Aa}} s_{\mathrm{rAg}} i_{\mathrm{eA}}} ,$$

$$s_{\mathrm{Aa}} = \frac{i_{\mathrm{eg}}}{i_{\mathrm{eA}}} \frac{s_{\mathrm{Ag}}}{I_{\mathrm{Tg}}} \Sigma \left(\frac{s_{\mathrm{rgg}}}{s_{\mathrm{rAg}}} I_{\mathrm{Ag}} \right) ,$$

where

- I_{Tg} is the total gauge current
- I_{Ag} is the gas analyser current for the peak corresponding to gas g
- i_{eg} is the gauge emission current
- i_{eA} is the gas analyser emission current
- s_{Ag} is the absolute sensitivity of the gauge for the reference gas
- s_{Aa} is the absolute sensitivity of the gas analyser for the reference gas
- s_{rgg} is the relative sensitivity of the gauge for gas g
- s_{rAg} is the relative sensitivity of the gas analyser for gas g.

The relative sensitivity for a gas g is the ratio of the absolute sensitivity for this gas to the absolute sensitivity for the reference gas

4.5.1.1 Parameters defining the sensitivity of the desorption measurement method

The sensitivity of the measurement method can be defined as the minimum η (η_{min}) which can be measured

$$\eta_{\mathrm{min}} = K \frac{\Delta P_{\mathrm{min}}}{i_{\mathrm{max}}} S ,$$

$$\Delta P_{\mathrm{min}} = \frac{P}{100} = \frac{Q_{\mathrm{stat}}}{100 S} ,$$

$$\eta_{\mathrm{min}} = K \frac{Q_{\mathrm{stat}}}{100\, i_{\mathrm{max}}} ,$$

where ΔP_{min} is the minimum measurable pressure variation (fixed arbitrarily to 1/100 of the base pressure which is a small value, difficult to achieve), S is the pumping speed of the measuring system, Q_{stat} the total outgassing of the measuring system, and i_{max} the maximum (electron/ion) current available for desorption.

Hence η_{min} is independent of the pumping speed and determined only by the available current and the static outgassing of the measuring equipment. However, there is a limit to the measurable current (I_{min}) given by the current amplifiers and hence to ΔP_{min} independent of the static pressure in the system. If we consider the minimum current I_{min} fixed by the current amplifier to be 10^{-12} A, then another limit is set to η_{min}:

$$\eta_{\mathrm{min}} \geq K \frac{S}{i_{\mathrm{max}}} \frac{I_{\mathrm{min}}}{s_{\mathrm{RGA}} i_{\mathrm{e}}} .$$

In this case, η_{min} is proportional to the pumping speed and inversely proportional to the residual gas analyser sensitivity (s_{RGA}). Hence increasing the pumping speed is a limitation for the measurement of a small desorption yield. For room-temperature measurement, it is an unnecessary limitation, as most of the gas available for desorption is tightly bound in the oxide layer. Figure 4 summarizes the various limitations to η_{min}: decrease of η_{min} with the system outgassing rate down to a limitation only depending on the pumping speed (s_{RGA} is set to 10^{-2} Pa^{-1}).

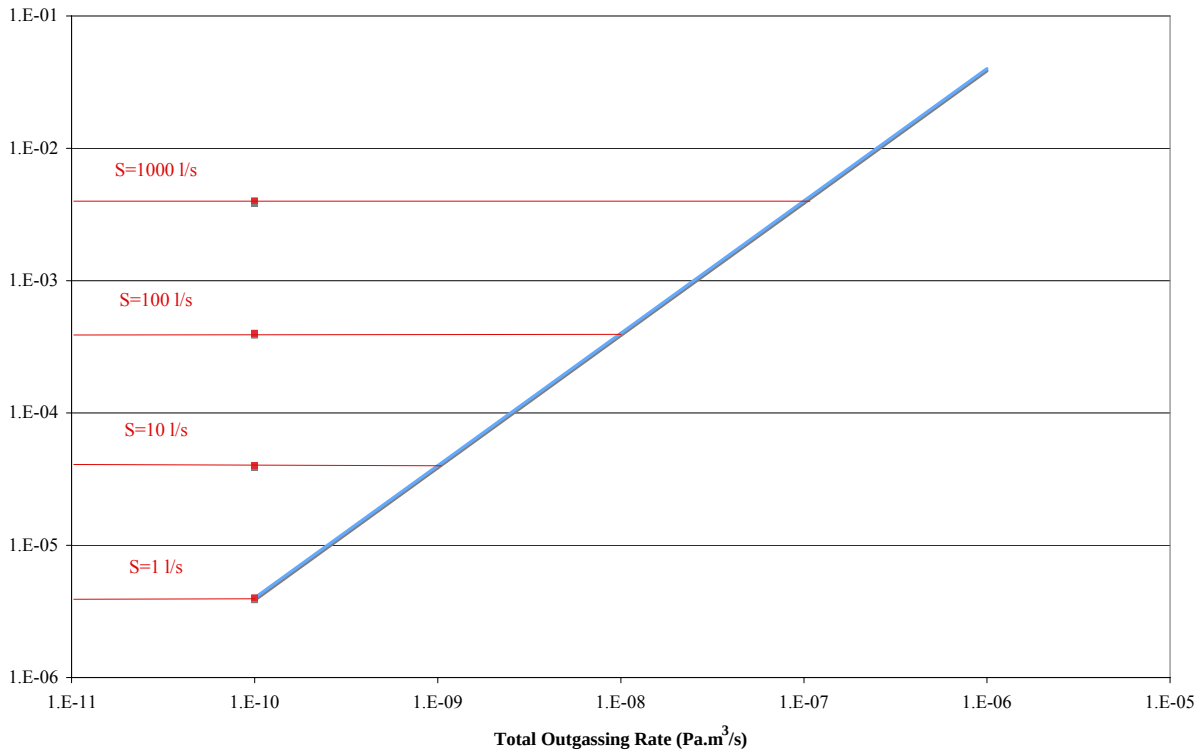


Fig. 4: The minimum measurable η as a function of total outgassing and pumping speed

4.5.2 Alternate method

In the case of an accelerator vacuum system study, using a uniform particle flux over a finite pipe length, it is possible to estimate the desorbed flux by measuring the pressure difference between the centre of a uniformly irradiated tube and its extremities where lumped pumps are installed [79]. The same system can also be used with a distributed pumping system (e.g., getter coated tubes or liquid helium cooled tubes [80]). This method has the advantage of allowing the measurement of the desorption yield of large pumping surfaces under conditions close to those encountered in real accelerators. Nevertheless, the system must be long enough to avoid end effects related to reflected particles which could introduce systematic errors.

5 Ion-induced desorption

This section will only consider the desorption yield of technical surfaces, prepared as accelerator vacuum chambers: for example, degreasing followed by alkaline etching and rinsing in demineralized water.

5.1 Variation of desorption yield as a function of ion energy

The desorption yield increases smoothly with ion energy (see Fig. 5), like the sputtering yield. The maximum is reached in the keV region (Fig. 6) and threshold energies close to 1 eV have been measured [83].

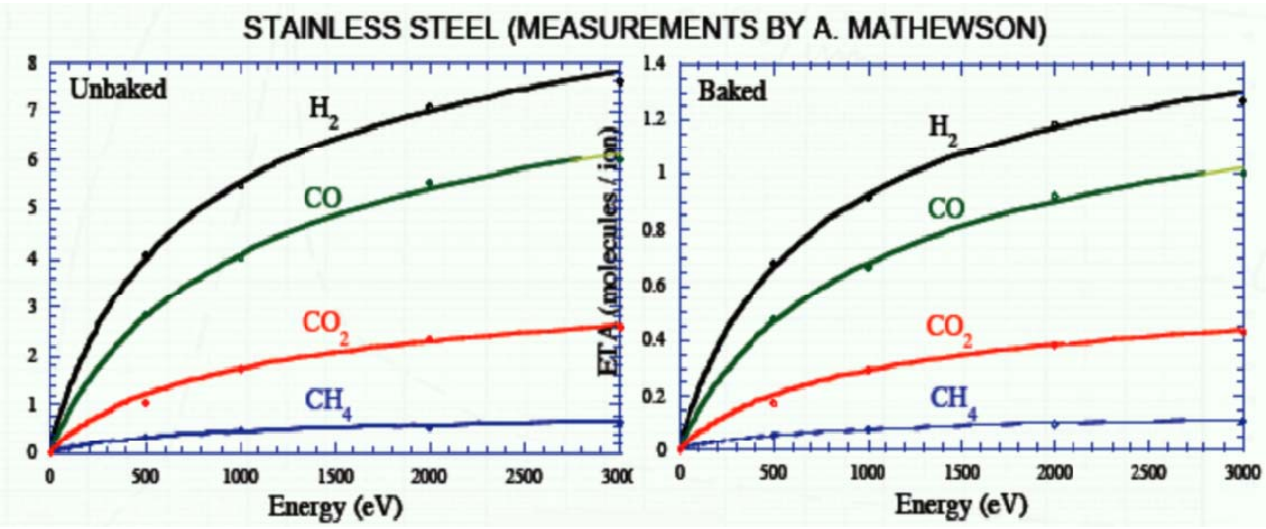


Fig. 5: Variation of ion-induced desorption with incident $^{15}N_2^+$ ion energy [81]

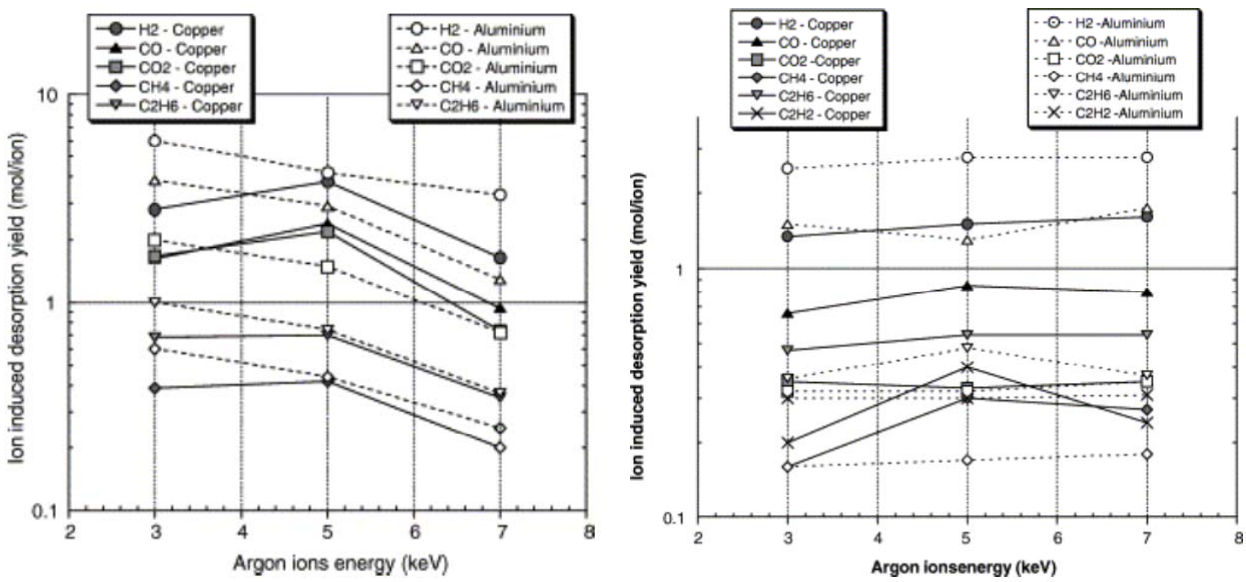


Fig. 6: Variation of ion-induced desorption with incident A^+ ion energy for baked and unbaked copper and aluminium [82]

5.2 Variation of desorption yield with surface treatment

Figure 7 [81] shows the desorption yield of stainless steel and aluminium as a function of the sample preparation. As can be seen, the desorption yield is more dependent on the surface treatment than on the base material for technical surfaces.

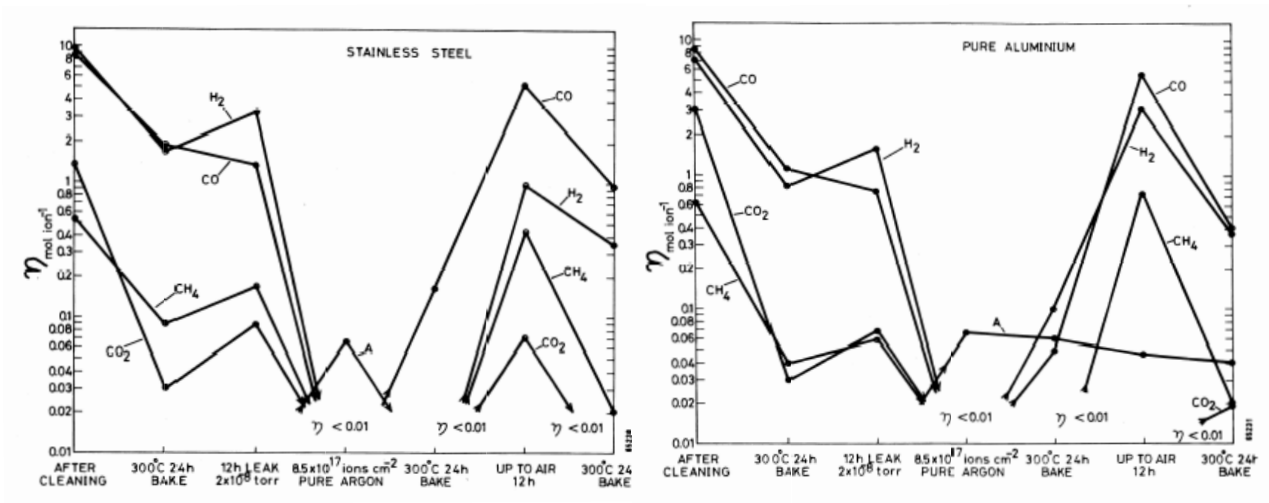


Fig. 7: Ion-induced desorption yields as a function of surface preparation for stainless steel and aluminium

5.3 Variation of the desorption yield with the nature of the incident ions

The desorption efficiency (D.E.) presented in the Fig. 8 is the ratio of the desorption yield measured with the ion i^+ to the desorption yield of the same molecule measured with the ion H_2^+. The pronounced increase of the desorption yield with the ion mass could be partly related to the enrichment of the surface with elements introduced in the substrate by the ion itself (e.g., C or O implanted by ions containing these elements as CO^+ or CO_2^+).

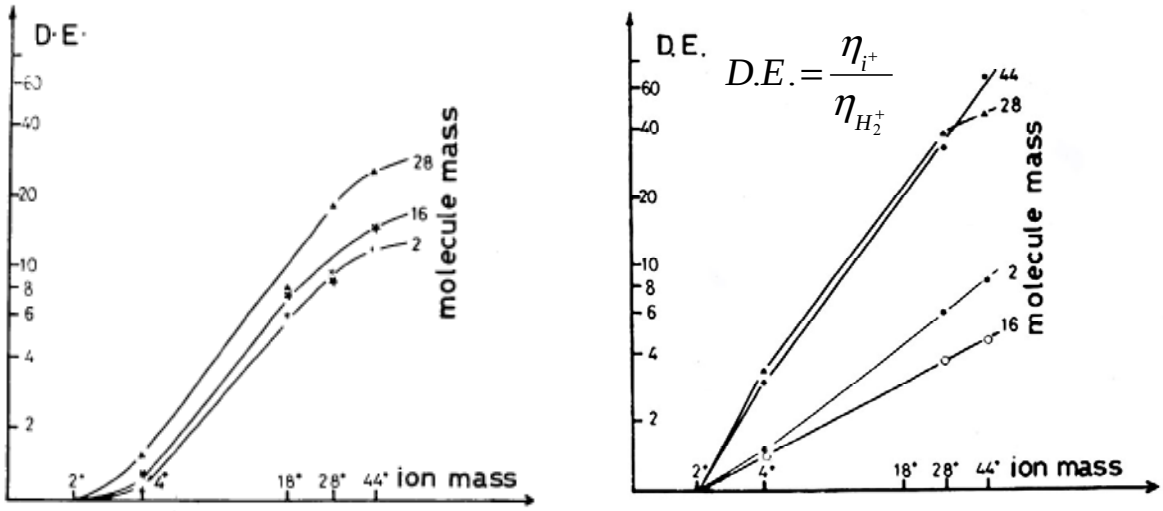


Fig. 8: Desorption efficiency of various incident ions [84] incident on a stainless steel surface

6 Swift heavy-ion-induced desorption

In heavy-ion storage rings or accelerators, beam losses linked to charge transfer processes could, via desorption, generate pressure increase degrading the accelerator performance [41, 85–87]. Very high desorption yields have been measured [73, 88–90]. Figure 9 illustrates the desorption yield of several metallic surfaces as a function of the incident beam dose for 52 MeV/u Pb^{53+}. In this energy range the incident ions lose energy mainly by electronic losses [57, 59] which determine the subsequent desorption. The desorption scales like $(dE/dx)^{1/n}$ with n between 1 and 3 and with the ion charge (q) as q^m with m close to 3 in the case studied in Ref. [57]. It is striking to see that a 2 μm thick Ag layer (i.e., much smaller than the range of the incident ion) can significantly alter the desorption yield.

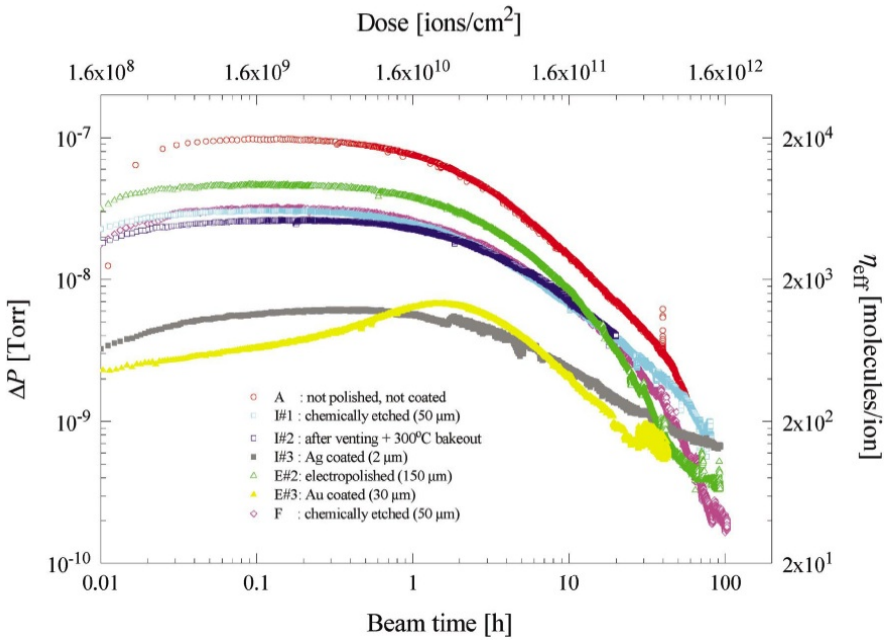


Fig. 9: Effective desorption yield of various materials bombarded by 52 MeV/u Pb^{53+}

7 Electron-induced desorption

Electron-induced desorption has a great historical importance for the study of desorption mechanisms and for the simulation of photon-induced desorption in electron storage rings. It is also an important source of desorption in regions of accelerators where electron multiplication can occur, e.g., Radio Frequency (RF) cavities, RF couplers and more recently over the entire accelerator when the electron cloud effect [40] takes place.

7.1 Variation of desorption yield with electron energy

In the case of as-received copper, the variation of the desorption yield with the electron energy $\eta(E)$ is presented in Fig. 10 [91] and can be fitted with the following equation:

$$\eta(E) = \eta_{300} \times \left(\frac{E - E_t}{300 - E_t} \right)^{0.85},$$

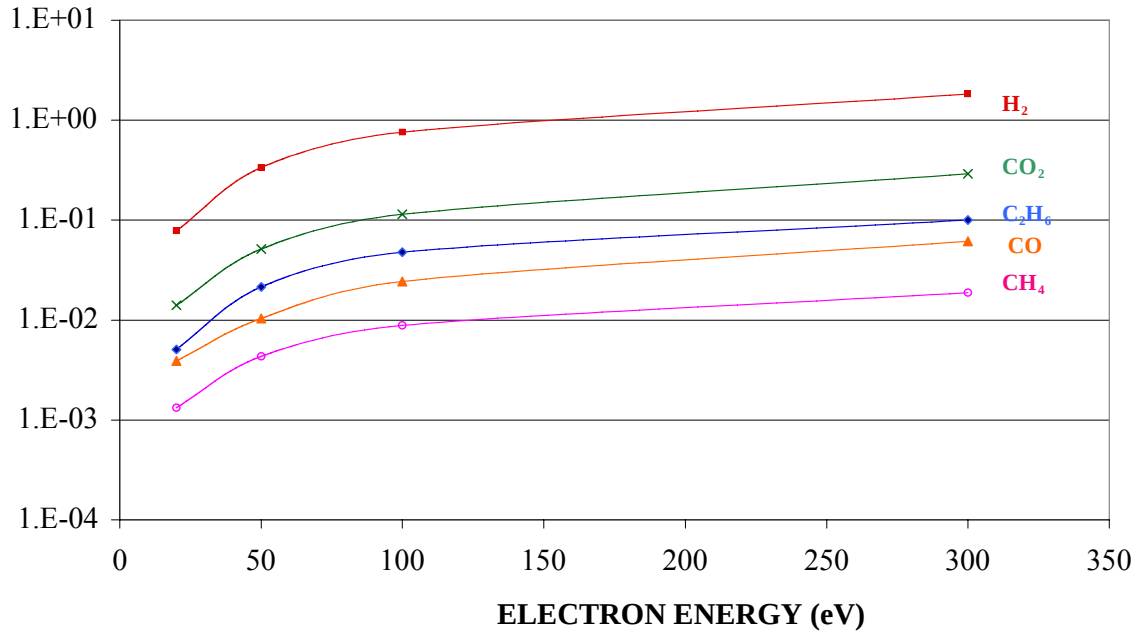


Fig. 10: Variation of desorption yield with electron energy

where η_{300} is the desorption yield at 300 V, and E_t is the threshold energy. The corresponding E_t are listed in Table 1.

Table 1: Threshold energies for various gases desorbed from as-received copper

Gas	Threshold energy (eV)
H_2	12.7
CH_4	7.5
CO	7.2
C_2H_6	11.4
CO_2	9.1

After a fast increase of the yield with energy, η saturates above 300 eV. The threshold for desorption appears around 10 eV [91–93].

7.2 Variation of desorption yield with electron dose

The impact of the electrons on a surface results in a decrease of the desorption yield. This effect is similar to what is observed under photon impact and is of great importance for the design and operation of electron storage rings. Figure 11 shows the evolution of the desorption yields for baked copper, stainless steel, and aluminium [72]. The decrease of desorption yield with electron dose D can be fitted by [72]:

$$\eta = cD^n \ .$$

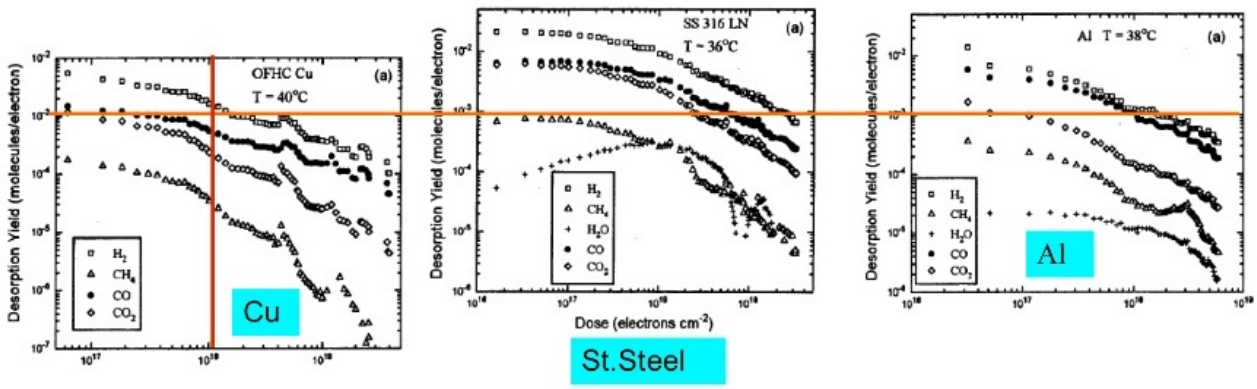


Fig. 11: Variation of desorption yield with electron dose [72]

Usually two distinct regions can be distinguished as can be seen in Fig. 12 illustrating the case of unbaked copper [94]. The fit parameters *c* and *n* in this case are given in Table 2 where it can be seen that *n* is of the order of –0.5. In the case of Fig. 12, the integration of the curves giving the total number of molecules released during the desorption shows that the equivalent of approximately 100 monolayers, in the case of H_2 and CO, have to be removed in order to obtain a decrease of the desorption yields by a factor of 100.

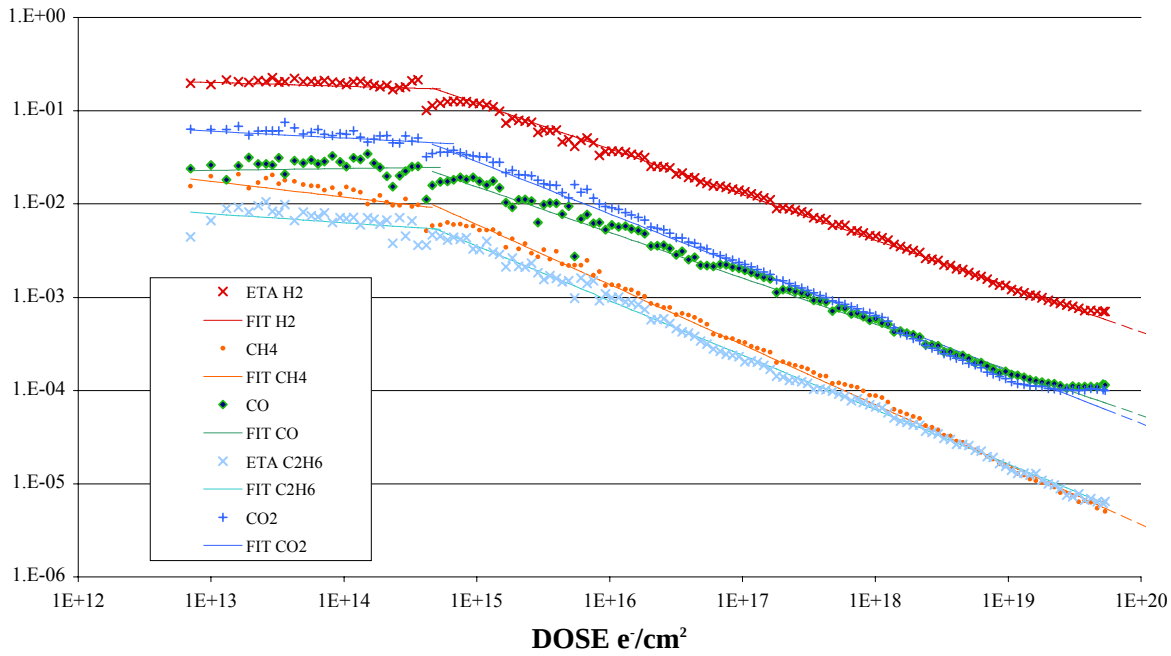


Fig. 12: Variation of desorption yield with incident electron dose in the case of unbaked copper

The case of water is not shown here as the desorption yield of this molecule increases during the initial part of the bombardment. The case of such transients observed also during photon bombardments has been discussed in Refs. [67, 69, 95].

Table 2: Fit parameters in the case of unbaked copper

	Gas				
Fit parameters	**H_2**	**CH_4**	**CO**	**C_2H_6**	**CO_2**
n (> 10^{15})	–0.489	–0.643	–0.489	–0.584	–0.561
c (> 10^{15})	2.68×10^6	2.67×10^7	3.36×10^5	2.02×10^6	7.36×10^6

8 Photon-induced desorption

Photon induced desorption has been studied in great detail because of its importance for the design and operation of synchrotron light sources and electron–positron storage rings. It has many features similar to electron-induced desorption and often electron bombardment is used to simulate photon bombardment. The required formulae to calculate photon fluxes in electron storage rings have been derived and discussed in the preceding CAS [95].

8.1 Variation of desorption yield with photon energy

In Fig. 13 the desorption of aluminium is shown as a function of the critical energy as measured for several electron storage rings. The desorption yield reaches a plateau from 2 keV up to 60 keV. Above that critical energy, the Compton effect plays a predominant role in the absorption of the photons which increases the total desorption rate because of the numerous energetic recoil electrons and scattered photons. For other materials (copper, stainless steel) the variation of η with the critical energy is similar, as shown in Ref. [96].

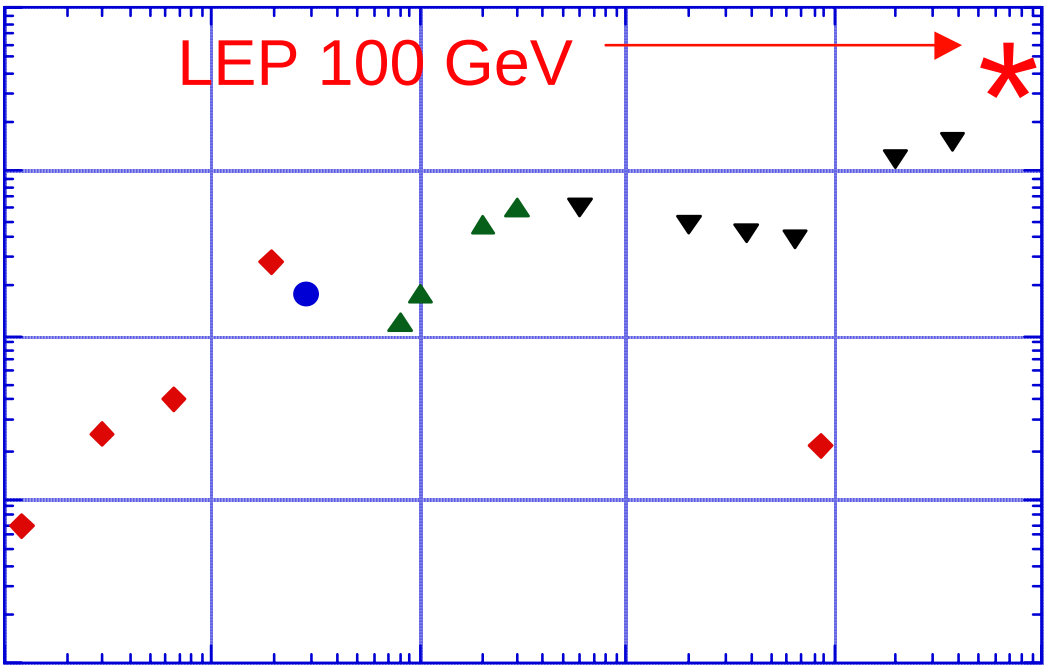


Fig. 13: Variation of desorption yield with photon critical energy [95]

8.2 Variation of desorption yield with photon dose

The knowledge of this variation is of utmost importance for the design of electron storage rings as it allows one to predict the evolution of the gas load during the exploitation of the accelerator and to adequately adapt the size and type of the pumping system. As a consequence, numerous measurements exist in the literature showing either η as a function of the photon dose or the specific pressure increase (the pressure variation normalized to the circulating current versus the integrated circulating current: mA.h). Many measurements have been published for the various technical materials: Al (6000 serie), Cu OFE, stainless steel [15, 18, 19, 21, 69, 96–100]. An illustration of typical desorption yields for technical materials is given in Figs. 14 and 15.

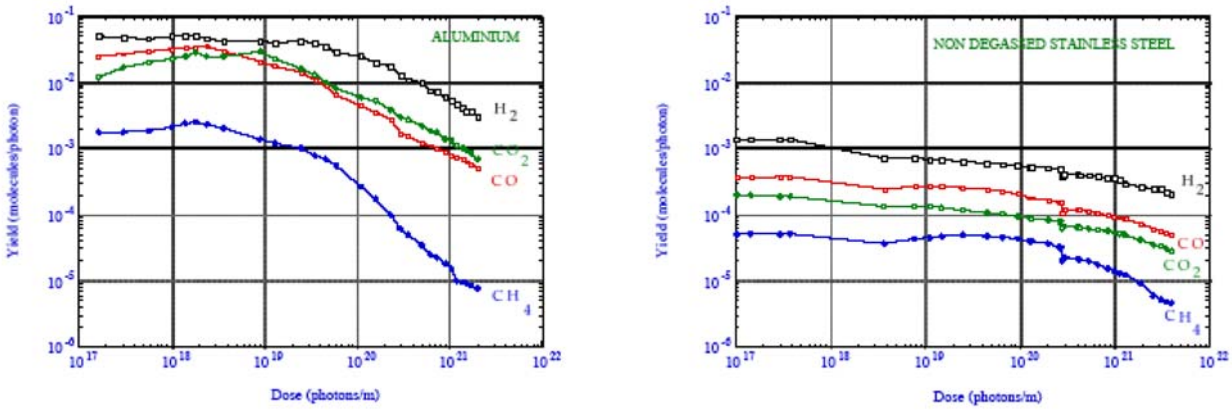


Fig. 14: The variation of desorption yield with photon dose for unbaked aluminium (6060) and stainless steel (304L) [95]

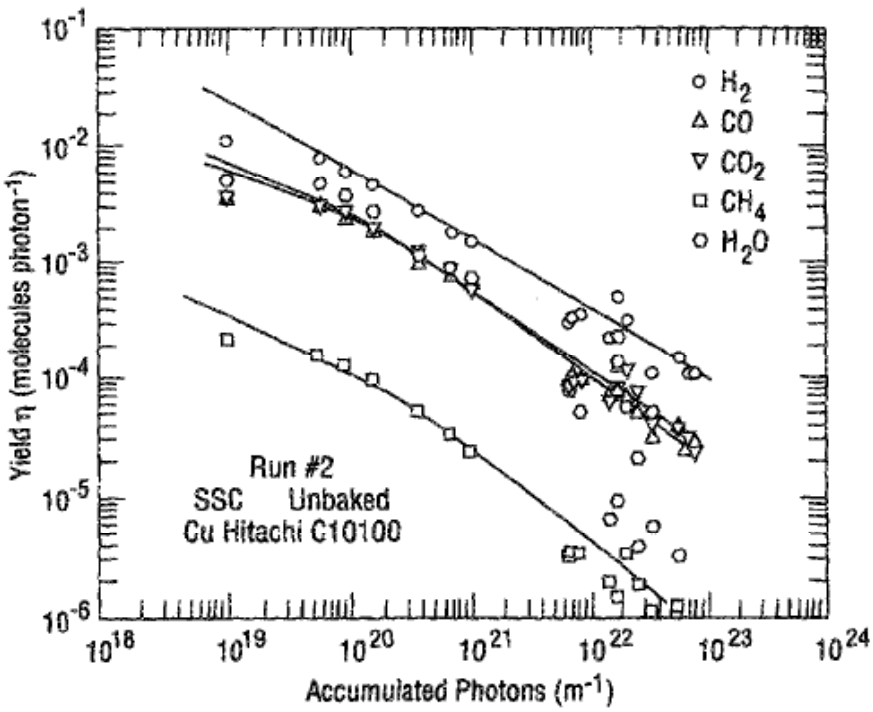


Fig. 15: Variation of desorption yield with photon dose for unbaked copper [98]

During these irradiations, time-dependent phenomena (see Fig. 3) depending on the type of desorbed molecules have been found, more pronounced in the case of oxygen-containing species like H_2O or O_2 [67, 69, 95]. In parallel to this effect a marked and persistent pumping has been measured, after the end of a long irradiation, for CO, CO_2 and O_2 [67]. Nothing similar has been measured for H_2 and CH_4. The pumping obviously tends to decrease the apparent desorption yield and strengthens the reduction of η with the beam dose. These effects illustrate the complexity of the mechanisms involved during photon desorption and the difficulty in measuring the desorption yield reliably under conditions close to those present in a real accelerator.

9 Desorption of pumping surfaces

With the development of new technologies and the introduction of strongly conductance-limited beam pipes, an increasing part of accelerator vacuum systems consists of surfaces with high sticking coefficients. This tendency is illustrated in Fig. 16 where the (approximate) fraction of the total vacuum system area capturing molecules is shown for three storage rings: the ISR (designed in 1960), LEP (designed in 1970–1980), and the LHC (designed in 1990). Although these surfaces are often protected against the direct impact of energetic particles, there has recently been great interest in the measurement of their desorption yields. Such a measurement is made difficult by the simultaneous pumping of these surfaces and, in many publications, this results in a poorly defined 'effective' desorption yield which combines the desorption and the pumping of the surfaces. This 'effective' desorption yield may be of technical interest for the design of a particular machine, but is very difficult to transpose for use in other machines with different geometries. The difficulties and limitations of this type of measurement are well described in Ref. [101].

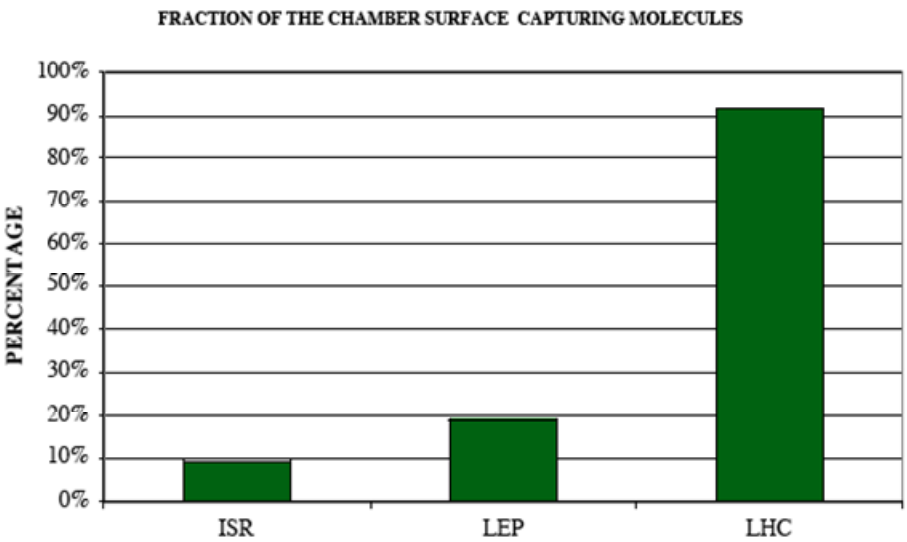


Fig. 16: Fraction of the area of the total accelerator vacuum system covered with gettering surfaces

9.1 Desorption of Non-Evaporable Getter (NEG) surfaces

NEG surfaces as thin-film coating on accelerator vacuum chambers provide a very efficient distributed pumping [102–104]. As mentioned above, many of the published results are 'effective' desorption yields and are only of comparative interest. Concerning desorption yields, Fig. 17 shows a comparison between two sets of results obtained in totally different systems for the same type of TiZrV NEG coating which

show reasonable coherence: a first set of data [80] was obtained using a three gauge method and 4.5 keV critical energy photons at 10 mrad incidence. The other measurements [105] were obtained for a fixed sample at normal incidence using 194 eV critical energy photons. In both cases the pumping speed or the sticking coefficients were measured *in situ* to calculate the desorption yields. These results give good orders of magnitude for the photon-induced desorption of TiZrV getter layers. During the experiment made in Novosibirsk [80], an amazing CO pumping effect stimulated by photon irradiation was recorded on a getter layer saturated with CO. For electron-induced desorption, useful data can be found in Ref. [106].

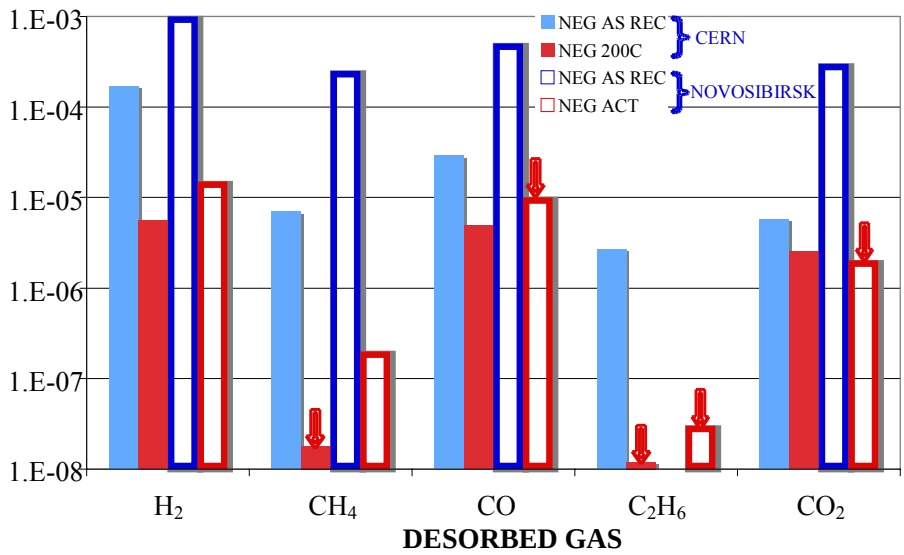


Fig. 17: Photon-induced desorption yield of TiZrV coatings [80, 105]. Measurements indicated by an arrow show upper limits.

9.2 Desorption of condensing surfaces

The design and construction of high-energy superconducting storage rings created a lot of interest in the accelerator community for the desorption of cryogenic surfaces. This field had previously been a subject of interest for the fusion community [107] and for the study of interstellar objects and of their modification under energetic particle impact [108]. For thick coverages of gases ('ice films') many measurements were obtained [54, 74, 75, 109–112]. The case of thin films (around the monolayer coverage) is less abundantly treated [107, 113–116]. Photon-induced desorption of condensed layer has been extensively studied for the design of new superconducting machines [20, 22, 117–122]. As in the case of getter surfaces, the parasitic pumping between the irradiated area and the measurement equipment must be carefully evaluated. For example, any additional pumping speed introduced by parasitic condensing surfaces could be interpreted as a reduction of the desorption yield when the temperature decreases. As the binding energy of condensed gases is very low (9 meV per molecule for solid hydrogen, 88 meV per molecule for solid CO), the yields are very high, as can be seen on Fig. 18 where the electron- and ion-induced desorption yield of H_2 is shown as a function of the condensed layer thickness. This dependence on the binding energy is clearly shown on Fig. 19 where, in the case of rare gases, the desorption yield under 300 eV electron impact is plotted as a function of the sublimation energy. In the case of photon impact the values published in Ref. [120] (Fig. 20) reach a maximum for a thickness of ~10^{19} molecules/cm^2. The maximum yields are approximately equal for CO_2, CH_4, H_2 between 0.4 and 0.6 molecule/photon. It is less for CO: ~0.04.

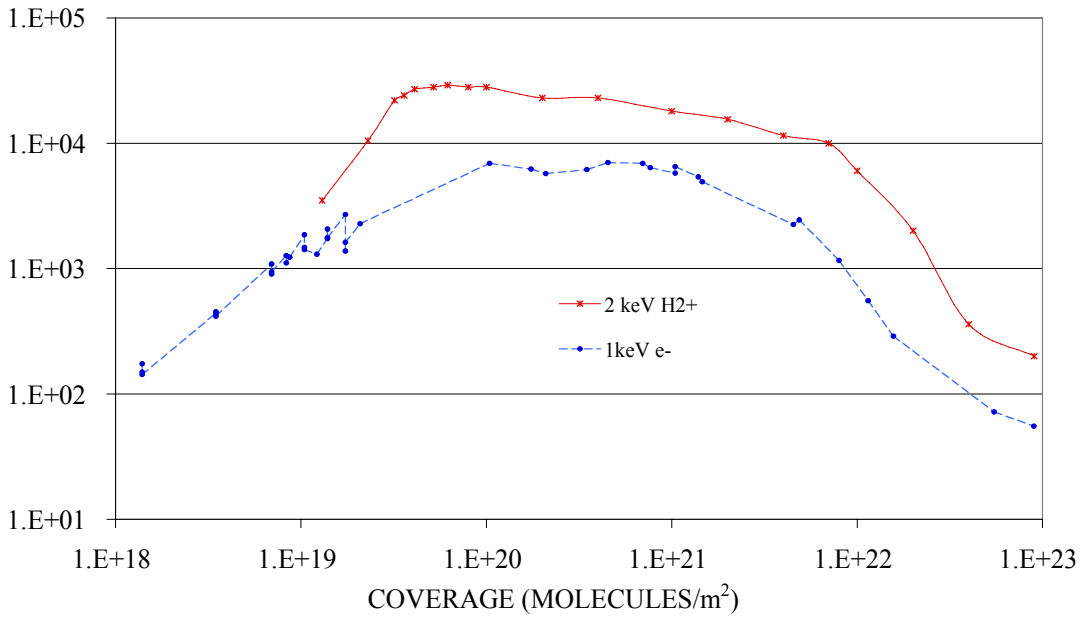


Fig. 18: Variation of the desorption yield of condensed hydrogen as a function of the layer thickness. Data for H_2^+: Hilleret unpublished; data for electrons [116].

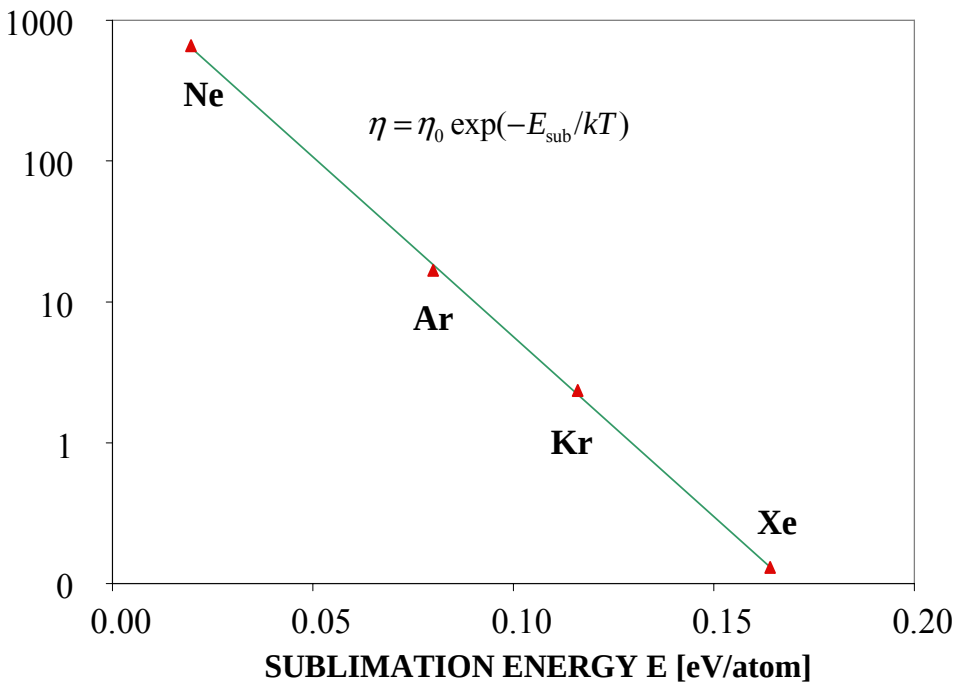


Fig. 19: Variation of the maximum desorption yield with the sublimation energy of condensed rare gases [116]

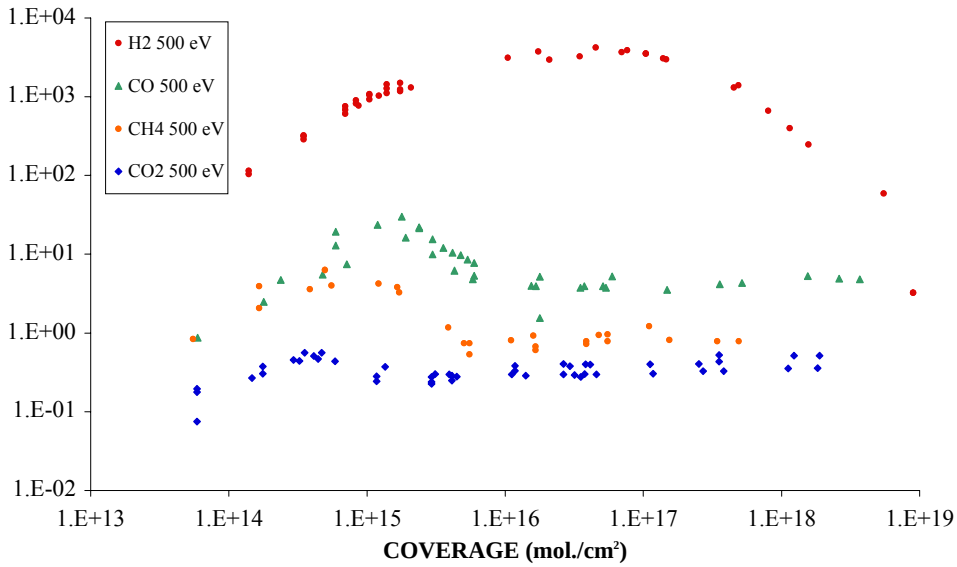


Fig. 20: Desorption yield of condensed gases under 300 eV electron impact as a function of gas coverage [123]

10 The consequences of non-thermal outgassing on accelerator operation

The preceding paragraphs have shown that the outgassing rates are increased by the impact of energetic particles. In accelerators, the circulation of high-energy particles produces photons and charged particles as well as the necessary electrical field to accelerate them. It follows that the gas loads are increased in the presence of beams which results in pressure variations related to the presence and/or intensity of the circulating beam(s). In this section, various processes leading to a variation of the pressure in the presence of beams (so-called 'dynamic pressure') will be described.

10.1 Ion-induced instability

This effect was first observed during the running-in of the Intersecting Storage Rings (ISR) at CERN: the attempt to store a 1 A proton beam at 26 GeV resulted in a large pressure increase (a factor 50) and an immediate loss of most of the stored beam. This effect, shown on Fig. 21, was attributed to a runaway process [25]: the ions created by the ionization of the rest gas are accelerated by the space charge of the beam and hit the vacuum chamber (energy between several 100 eV to some keV) where they release gas through ion-induced desorption. This leads to a pressure increase which further enhances the ion bombardment resulting in a pressure runaway and a beam loss. This process is described, in the case of a lumped pumping system, by the following equations where P_0 is the pressure in the absence of beam, $P(I)$ is the pressure of this gas in presence of a beam with intensity I, σ is the ionization cross-section, η is the desorption yield, S the pumping speed for that gas, and I_c the critical current at which the pressure becomes infinite. The complete calculation of the critical current can be found in Ref. [95]. To increase the critical current, it is possible either to increase the pumping speed or to decrease the desorption yield. This decrease cannot be obtained as easily for proton machines as lepton machines since the pressure is not stable when the current is increased: in the conditions where an intense ion bombardment can clean the vacuum chamber *in situ*, the pressure is not stable and the accelerator cannot be run under such conditions for a long period.

$$P(I) = \frac{P_0}{1 - I * \frac{\sigma}{e} * \frac{\eta}{S}} \ ,$$

$$I_c = \frac{e * S}{\sigma * \eta} \ .$$

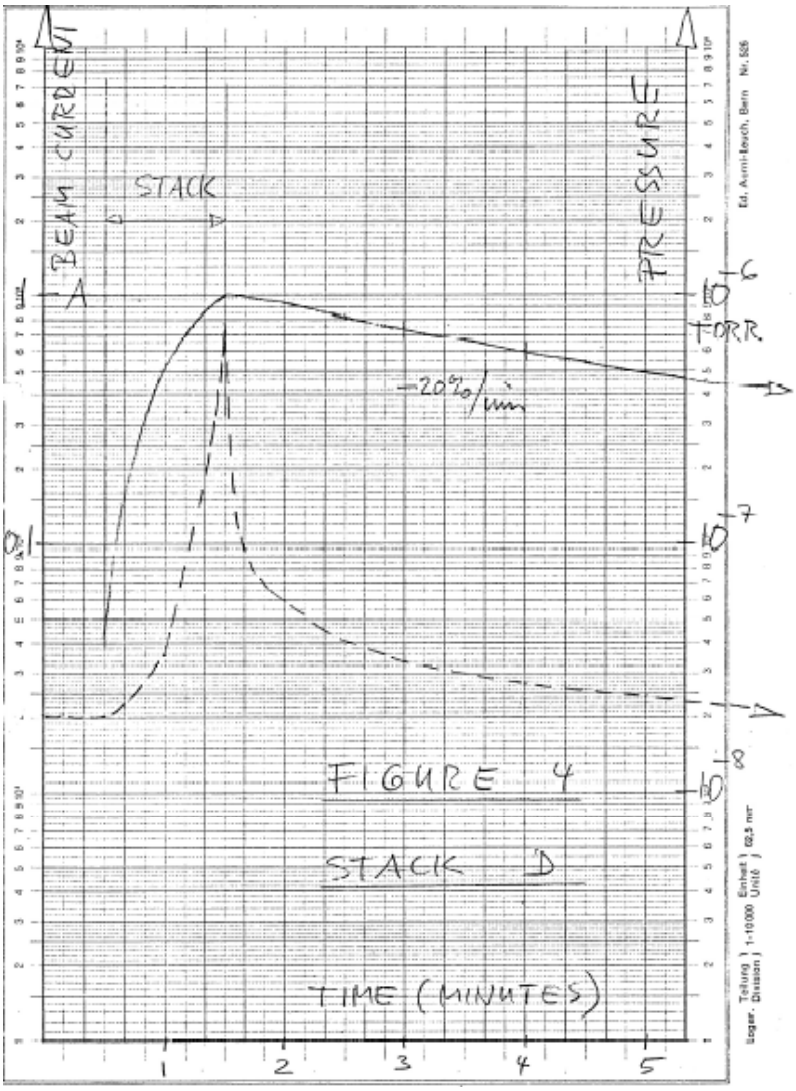


Fig. 21: Pressure bump in the ISR. ISR running-in note: E. Fischer, O. Gröbner, E. Jones, 18 November 1970.

10.1.1 The long-term pressure drift

On Fig. 22 the evolution of pressure with time shows a long-term drift (9 hours) for a stable stored current (41.6 A).

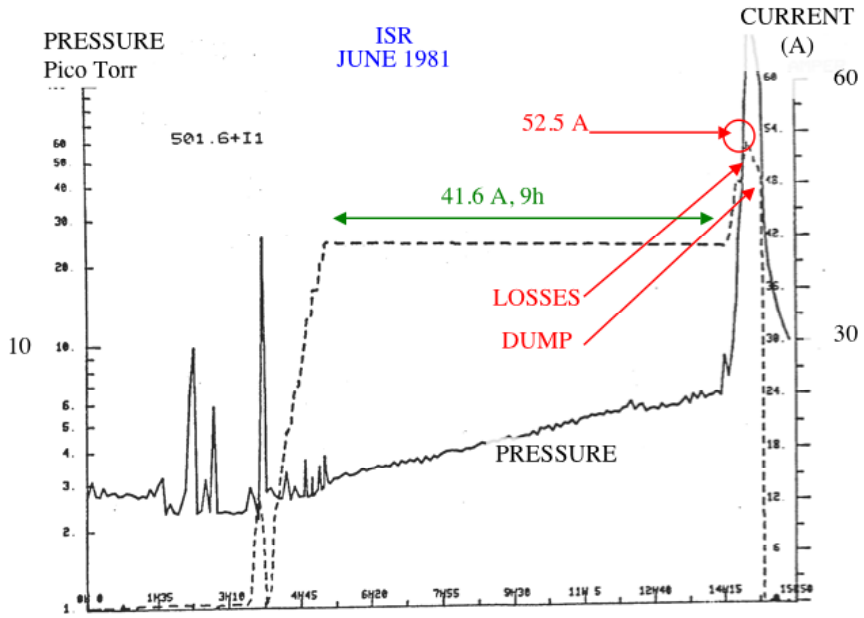


Fig. 22: Long-term pressure drift in the ISR

At the end of the stable period, stacking was resumed causing a quick runaway and the dumping of the beam. A first lesson can be drawn from that example: before the runaway, that is at least at 80% of the critical current, the pressure was $8 \cdot 10^{-10}$ Pa and hence physics experiments could have proceeded without pressure-induced background. The slow increase of the pressure in the first part of the run was due to a gradual modification of the gas composition. In the preceding equations, three terms depend on the nature of the gas: the pumping speed, the ionization cross-section, and the desorption yield (see Section 5.3). The cross-sections for various gases are given in Table 3. Initially the residual gas was mainly H_2, a gas with small ionization cross-section, and low desorption efficiency as ion. The residual gas was gradually enriched with heavier molecules desorbed by the H^+ ions. These heavier gases, which have a higher ionization cross-section and a higher desorption efficiency for heavy masses, become more and more abundant in the residual gas and are at the origin of this slow pressure drift. Experiments were made in an ISR test section [124], the corresponding results have been used to draw Fig. 23. During this experiment, the initial gas composition contained approximately 80% H_2; after 100 min at constant current, CO represented more than 50% of the residual gas.

Table 3: Ionization cross-section for various gases (unit 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	1.8
Ar	1.1	2.0
CO_2	1.6	2.8

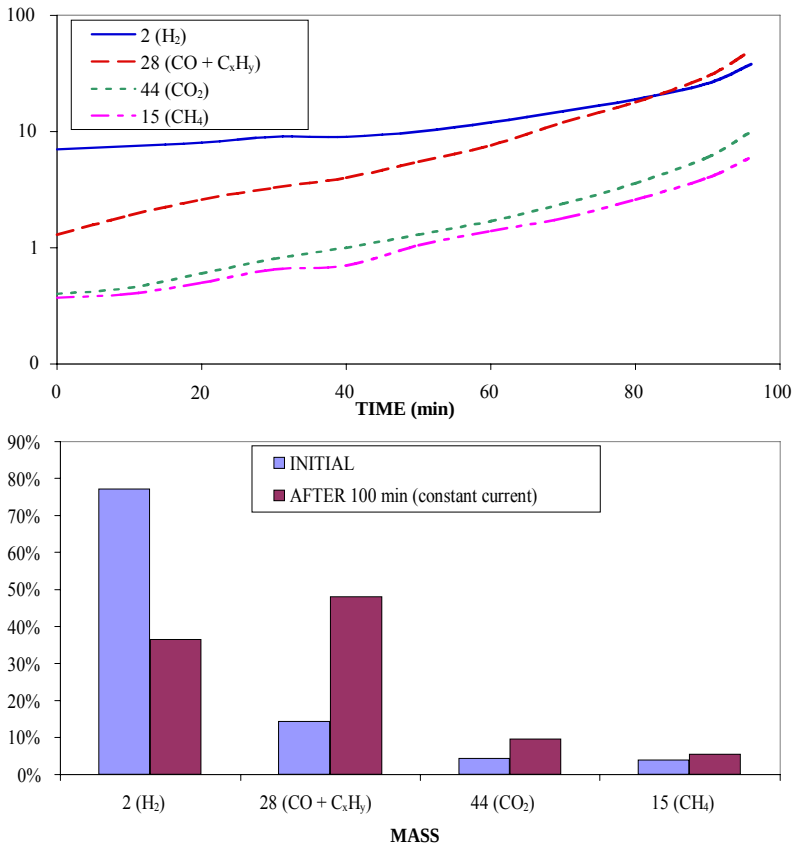


Fig. 23: Variation of the gas composition with time during a slow pressure bump [124]

10.1.2 Beam pumping

Here η is defined as the number of molecules liberated by one incident ion in the case of ion-induced desorption. In an accelerator, the incident ion stems from the residual gas, hence the total number of molecules, liberated by one ion is

$$\eta_{\text{mach}} = \eta_{\text{lab}} - \kappa \ ,$$

where η_{lab} is the desorption yield measured in the laboratory and κ is the trapping probability for the incident ion. For a clean surface, η_{lab} can be smaller than κ and hence η_{mach} is negative. In that case 'beam pumping' appears and the pressure decreases in the presence of the beam as shown in Fig. 24. The importance of this phenomenon depends on the installed pumping speed and on the ionization cross-section of the pumped gas: it is more visible on the total pressure when the main gas is heavy and poorly pumped. The trapping probability is maximum (= 1) for a clean surface in which no gas has been pumped. Then it decreases when the quantity of implanted gas increases. As an example [125] the initial κ is 0.65 for 4 keV argon impinging on copper, it decreases to 0 after a dose of $4 \cdot 10^{16}$ Ar ions/cm^2. The beam pumping speed (l/s/m) is given by the equation:

$$S_{\text{beam}} = \frac{I}{e}\sigma\kappa \quad \left(\text{m}^2\ \text{s}^{-1}\right) \ .$$

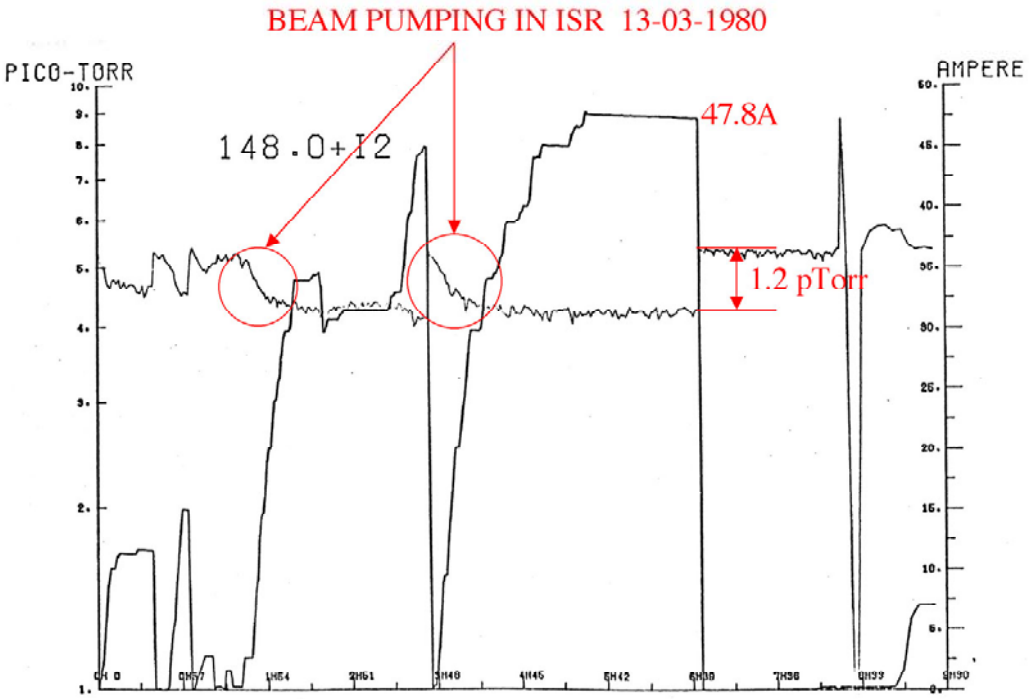


Fig. 24: Example of beam pumping in the ISR

The consequences of this effect can be significant if during ionization a molecule which is not pumped is fragmented in pieces which could be pumped: case of hydrocarbons in a getter-pumped system.

Finally, in lepton machines, beam pumping has also been observed on two occasions and seems to be related to modifications of the vacuum chamber surface due to photon bombardment:

- for CO, CO_2 and O_2 molecules [67] on a copper chamber after exposition to 10^{23}photons/m (no pumping observed for H_2 and CH_4);
- for CO on a TiZrV coated stainless steel chamber activated and saturated with CO [80].

10.2 Beam cleaning due to synchrotron light

In the sections on electron- and photon-induced desorption, it has been shown that the desorption yields decrease considerably with the dose of incident particles. This is of major importance during the design and the operation of lepton storage rings and of very high-energy proton storage rings like the LHC. The magnitude of this effect is shown on Fig. 25 representing the pressure rise per mA of stored beam along the lifetime of the electron–positron storage ring LEP at CERN. The pressure decreases, for a constant particle energy, like the inverse of the circulating beam dose (mA.h). This was due to a decrease of 4 orders of magnitude of the gas load during the life of the LEP accelerator. This effect must be taken into account to choose adequately the characteristics of the pumps and to avoid over-dimensioning of the pumping system. As a consequence, the beam dose at which the nominal performance of the accelerator vacuum system is reached must also be clearly evaluated. The cleaning rate changes with the gases, hydrocarbons cleaning at a faster rate than H_2 and CO (see Fig. 26) which was, at the end of the running period, the limiting gas in terms of lifetime considerations.

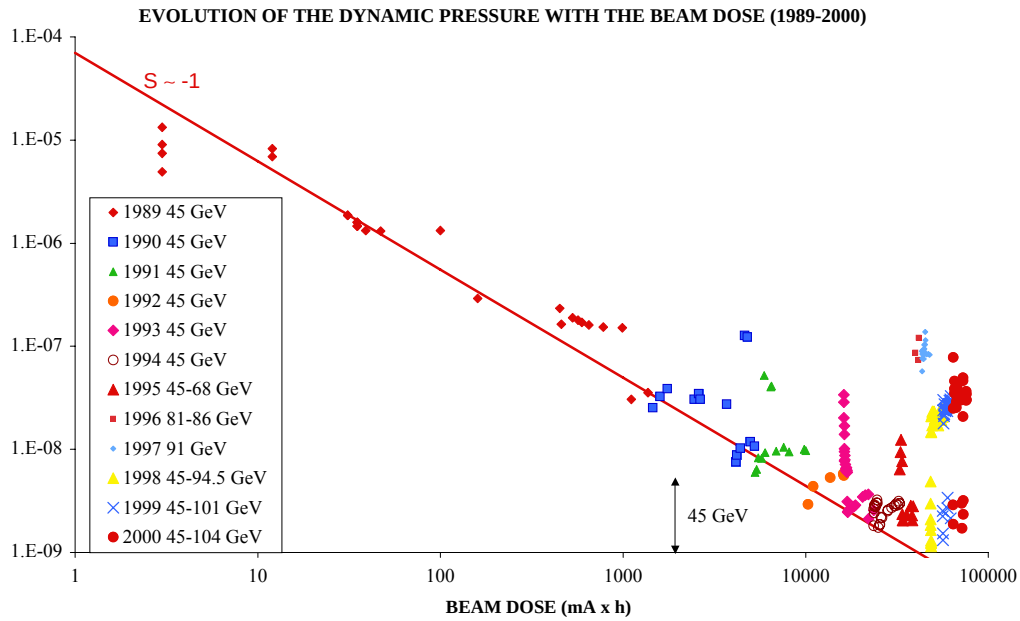


Fig. 25: Variation of the dynamic pressure in LEP arcs along the ten-year life of this machine (the slope –1 corresponds to points measured at constant beam energy: 45 GeV)

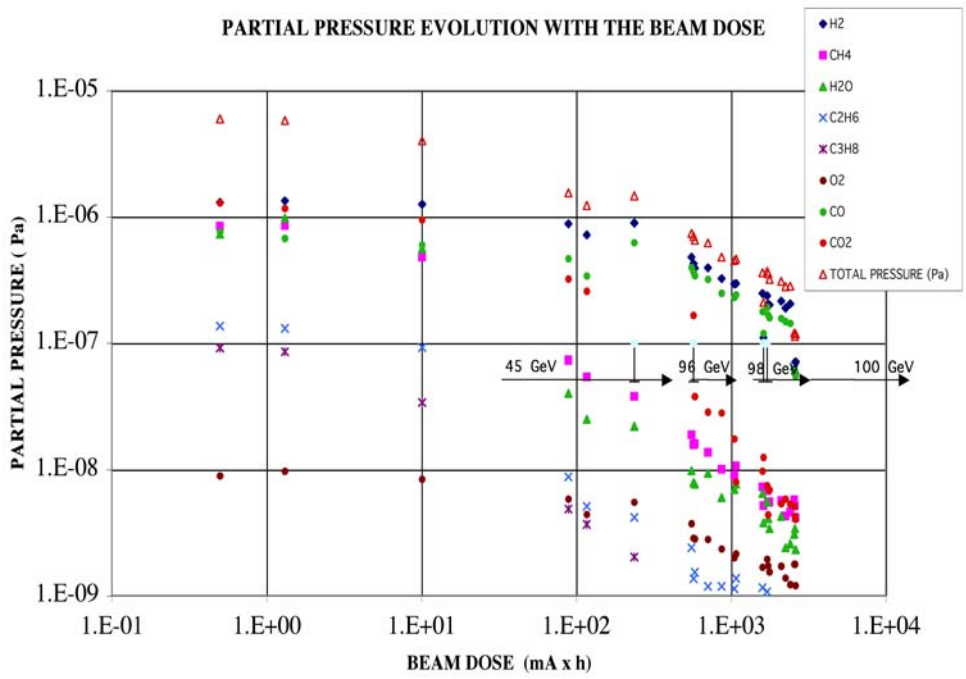


Fig. 26: Variation of the partial pressure in a freshly installed vacuum chamber during one year of LEP operation

The beam intensity decrease in an accelerator due to beam–gas interactions is given by the relation:

$$I = I_0 \, e^{-\frac{t}{\tau}}$$

$$\tau = \frac{1}{v \sigma n}$$

where τ is the total lifetime, v is the speed of the stored particles, σ is the cross-section of the considered process, and n the gas density. For a residual gas consisting in a mixture of gases, τ can be calculated from the lifetime determined by each individual gas (τ_g) by:

$$\frac{1}{\tau} = \sum \frac{1}{\tau_g}$$

$$F = \frac{\tau}{\tau_g} \ ,$$

where F represents the fraction of the total lifetime due to gas g. Figure 27 represents this contribution F measured on a new vacuum chamber during its first year in LEP. At the start of the running period, the three main contributors to lifetime were: CO_2 51%; CO 19%; and H_2O 12% (although the system was initially baked). At the end of the running period, they had changed to: CO 79%; CO_2 9%; H_2 3%, the heavier gases being a much stronger limitation for the lifetime. Note that hydrogen was never the limiting gas in terms of lifetime in that accelerator. Because of its small cross-sections, of its high conductance, and of the availability of very efficient pumps, hydrogen is not a limitation for the lifetime in most accelerators.

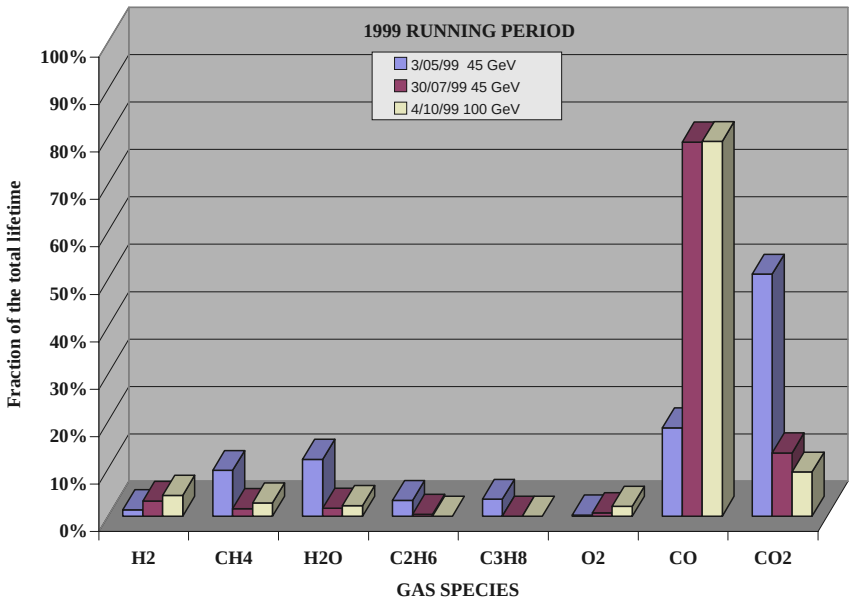


Fig. 27: Evolution of the fraction of the total lifetime for the main residual gases during one year of LEP operation (new vacuum chamber)

11 Conclusions

Non-thermal outgassing is the main gas load during the operation of accelerators. It must be carefully considered during the design phase of the machine and many data are available to estimate the ion, electron, and photon-induced desorptions. During the operation of accelerators, these various processes are responsible for pressure variations in the presence of beams which can be a strong limitation on the accelerator performance such as ion-induced instabilities or pressure increase due to electron cloud. On the other hand, electron- and photon-induced desorption leads to an important cleaning of the vacuum chambers and considerably reduces the gas load to the pumping system. This decrease is also an important input for the choice of the pumping system and of the expected duration to reach the nominal performance in a lepton accelerator. In cold systems, the condensation of a non-negligible amount of gases with low binding energy gives rise to very large desorption yields (often greater than 100). Hence in such accelerators, when operating at temperatures where condensation can occur, the condensing surfaces must be protected against the direct impact of energetic particles.

References:

[1] A.J. Dempster, *Phys. Rev.* **11** (1918) 316.

[2] E.R. Wagner, *Trans. Electrochem. Soc.* **59** (1931) 223.

[3] T.E. Madey and J.T. Yates, Jr., *J. Vac. Sci. Technol.* **8** (1971) 525.

[4] Y. Isagawa, *Rev. Phys. Chem. Jpn.* **16** (1942) 117.

[5] D. Menzel and R. Gomer, *J. Chem. Phys.* **40** (1964) 1164.

[6] D. Menzel and R. Gomer, *J. Chem. Phys.* **41** (1964) 3311.

[7] P.A. Redhead, *Can. J. Phys.* **42** (1964) 886.

[8] M.L. Knotek and P.J. Feibelman, *Phys. Rev. Lett.* **40** (1978) 964.

[9] T.E. Madey, *Surface Sci.* **299–30** (1994) 824.

[10] J.L. de Segovia, *Vacuum* **47** (1996) 333.

[11] P.A. Redhead, *Vacuum* **48** (1997) 585.

[12] G.E. Fischer and R.A. Mack, *J. Vac. Sci. Technol.* **2** (1965) 123.

[13] M. Bernardini and L. Malter, *J. Vac. Sci. Technol.* **2** (1965) 130.

[14] E.L. Garwin, Memorandum dated 14 August 1963 (1963).

[15] O. Groebner, A.G. Mathewson, H. Stoeri *et al.*, *Vacuum* **33** (1983) 397.

[16] G. Horikoshi, *J. Vac. Sci. Technol., A* **5** (1987) 2501.

[17] A.G. Mathewson, J.P. Bacher, K. Booth *et al.*, *J. Vac. Sci. Technol., A* **7** (1989) 77.

[18] C.L. Foerster, H. Halama and C. Lanni, *J. Vac. Sci. Technol. A*, **8** (1990) 2856.

[19] S. Ueda, M. Matsumoto, T. Kobari *et al.*, *Vacuum* **41** (1990) 1928.

[20] V.V. Anashin, G. Derevyankin, V.G. Dudnikov *et al.*, in *The 40th National Symposium of the American Vacuum Society* (AVS, Orlando, FL, USA, 1994), Vol. 12, p. 1663.

[21] C.L. Foerster, C. Lanni, C. Perkins *et al.*, *J. Vac. Sci. Technol., A* **13** (1995) 581.

[22] R. Calder, O. Gröbner, A.G. Mathewson *et al.*, *J. Vac. Sci. Technol., A* **14** (1996) 2618.

[23] G. Horikoshi and M. Kobayashi, *J. Vac. Sci. Technol.* **18** (1981) 1009.

[24] K.H. Kingdon and I. Langmuir, *Phys. Rev.* **22** (1923) 148.

[25] R.S. Calder, *Vacuum* **24** (1974) 437.

[26] G.M. McCracken and P.E. Stott, *Nucl. Fusion* **19** (1979) 889.
[27] H.F. Winters and P. Sigmund, *J. Appl. Phys.* **45** (1974) 4760.
[28] P. Sigmund, *Phys. Rev.* **184** (1969) 383.
[29] E. Taglauer, G. Marin, W. Heiland *et al.*, *Surface Sci.* **63** (1977) 507.
[30] E. Taglauer, W. Heiland and U. Beitat, *Surface Sci.* **89** (1979) 710.
[31] A.G. Mathewson, *Vacuum* **24** (1974) 505.
[32] M.-H. Achard, R. Calder and A. Mathewson, *Vacuum* **29** (1979) 53.
[33] H.F. Dylla, *J. Nucl. Mater.* **93–94** (1980) 61.
[34] F. Waelbroeck, J. Winters and P. Wienhold, *J. Vac. Sci. Technol., A* **2** (1984) 1521.
[35] R.P. Govier and G.M. McCracken, *J. Vac. Sci. Technol.* **7** (1970) 552.
[36] R.M. Lambert and C.M. Comrie, *J. Vac. Sci. Technol.* **11** (1974) 530.
[37] A.G. Mathewson, in *International Symposium on Plasma Wall Interactions,* Jülich, 1976 (Pergamon Press, Oxford, 1977), p. 517.
[38] M. Ulrickson, H.F. Dylla, P.H. LaMarche *et al.*, *J. Vac. Sci. Technol., A* **6** (1988) 2001.
[39] E. Fischer, *J. Vac. Sci. Technol.* **9** (1972) 1203.
[40] F. Ruggiero, in *Sixth European Particle Accelerator Conference, EPAC98*, Stockholm, 1998.
[41] J. Bosser, C. Carli, M. Chanel *et al.*, CERN-PS-99-033-DI (CERN, Geneva, 1999).
[42] J. Schou, in *Vacuum in Accelerators* (CERN Accelerator School, Platja d'Aro, Spain, 2006).
[43] W.C. Turner, *J. Vac. Sci. Technol., A* **14** (1996) 2026.
[44] O.B. Malyshev, CERN - Vacuum Technical Note 99-17 (1999).
[45] H. Niehus, W. Heiland and E. Taglauer, *Surface Sci. Rep.* **17** (1993) 213.
[46] Y. Yamamura and H. Kimura, *Surface Sci.* **185** (1987) L475.
[47] E. Taglauer, U. Beitat, G. Marin *et al.*, *J. Nucl. Mater.* **63** (1976) 193.
[48] E. Taglauer and W. Heiland, *J. Nucl. Mater.* **76–77** (1978) 328.
[49] E. Taglauer, *Nucl. Fusion* **Special Issue** (1984) 43.
[50] Y. Yamamura, T. Takiguchi and H. Kimura, *Nucl. Instrum. Methods Phys. Res., B* **78** (1993) 337.
[51] E. Taglauer, *Surface Sci.* **299–300** (1994) 64.
[52] M.T. Robinson and I.M. Torrens, *Phys Rev., B* **9** (1974) 5008.
[53] J.P. Biersack and L.G. Haggmark, *Nucl. Instrum. Methods* **174** (1980) 257.
[54] R.E. Johnson and J. Schou, *Mat. Fys. Medd. K. Dan. Vidensk. Selsk.* **43** (1993) 403.
[55] F. Seitz and J.S. Koehler, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic Press, New York, 1956), Vol. 2, p. 305.
[56] P. Sigmund, *Appl. Phys. Lett.* **25** (1974) 169.
[57] M. Caron, F. Haranger, H. Rothard *et al.*, *Nucl. Instrum. Methods Phys. Res., B* **179** (2001) 167.
[58] H.D. Mieskes, W. Assmann, F. Grüner *et al.*, *Phys. Rev., B* **67** (2003) 155414.
[59] M. Toulemonde, W. Assmann, C. Trautmann *et al.*, *Nucl. Instrum. Methods Phys. Res., B* **212** (2003) 346.
[60] P. Avouris, F. Bozso and R.E. Walkup, *Nucl. Instrum. Methods Phys. Res., B* **27** (1987) 136.
[61] P. Avouris and R. E. Walkup, *Annu. Rev. Phys. Chem.* **40** (1989) 173.
[62] J.W. Gadzuk and C.W. Clark, *J. Chem. Phys.* **91** (1989) 3174.

[63] T.M. Orlando and G.A. Kimmel, *Surface Sci.* **390** (1997) 79.

[64] T. Kloiber and G. Zimmerer, *Rad. Effects and Defects in Solids* **109** (1989) 219.

[65] L. Sanche, *Surface Sci.* **451** (2000) 82.

[66] R. Calder, A. Grillot, F. Le Normand *et al.*, in *7th International Vacuum Congress*, Vienna, Austria, 1977, edited by R. Dobrozemsky, F. Ruedenauer, F. P. Viehboeck and A. Breth (Techische Universität Wien, Vienna, 1977), Vol. 1, p. 227.

[67] O. Gröbner, A.G. Mathewson and P.C. Marin, *J. Vac. Sci. Technol., A* **12** (1994) 846.

[68] N. Ota, K. Kanazawa, M. Kobayashi *et al.*, *J. Vac. Sci. Technol., A* **14** (1996) 2641.

[69] G.Y. Hsiung, K.Y. Young, Y.J. Hsu *et al.*, *J. Vac. Sci. Technol., A* **19** (2001) 1657.

[70] M. Andritschky, O. Gröbner, A.G. Mathewson *et al.*, *Vacuum* **38** (1988) 933.

[71] G.F. Dugan, SSCL-610 (SSC, Dallas, 1993).

[72] J. Gómez-Goñi and A.G. Mathewson, *J. Vac. Sci. Technol., A* **15** (1997) 3093.

[73] R. Behrisch, V.M. Prozesky, H. Huber *et al.*, *Nucl. Instrum. Methods Phys. Res., B* **118** (1996) 262.

[74] W.L. Brown, W.M. Augustyniak, E. Brody *et al.*, *Nucl. Instrum. Methods* **170** (1980) 321.

[75] W.L. Brown, L.J. Lanzerotti, K.J. Marcantonio *et al.*, *Nucl. Instrum. Methods Phys. Res., B* **14** (1986) 392.

[76] J. Schou, H. Sorensen and P. Borgesen, *Nucl. Instrum. Methods Phys. Res., B* **5** (1984) 44.

[77] N.J. Sack and R.A. Baragiola, *Phys. Rev., B* **48** (1993) 9973.

[78] E.L. Garwin, E.W. Hoyt, M. Rabinowitz *et al.*, in *Proceedings of the Fourth International Vacuum Congress,* Manchester, 1968 (Institute of Physics and the Physical Society, London, 1969?), p. 131.

[79] V.V. Anashin, A.V. Evstigneev, A.P. Lysenko *et al.*, *Nucl. Instrum. Methods Phys. Res., A* **359** (1995) 110.

[80] V.V. Anashin, I.R. Collins, R.V. Dostovalov *et al.*, *Vacuum* **75** (2004) 155.

[81] A.G. Mathewson, CERN-ISR-VA/76-5 (CERN, Geneva, 1976).

[82] M.P. Lozano, *Vacuum* **67** (2002) 339.

[83] S.M. Liu, W.E. Rodgers and E.L. Knuth, *J. Chem. Phys.* **61** (1974) 902.

[84] N. Hilleret, in *Fourth International Conference on Solid Surfaces* (Societé Française du Vide, Cannes, 1980), Vol. 2, p. 1221.

[85] W. Fischer, M. Bai, J.M. Brennan *et al.*, in *EPAC 2002*, Paris, 2002, p. 1485.

[86] E. Mustafin, O. Boine-Frankenheim, I. Hofmann *et al.*, *Nucl. Instrum. Methods Phys. Res., A* **510** (2003) 199.

[87] A.W. Molvik, H. Kollmus, E. Mahner *et al.*, *Phys. Rev. Lett.* **98** (2007) 064801.

[88] N. Madsen, PS/DI/Note 99-21 (CERN, Geneva, 1999).

[89] E. Mahner, J. Hansen, J.M. Laurent *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **6** (2003) 013201.

[90] E. Mahner, J. Hansen, D. Kuchler *et al.*, *Phys. Rev. Spec. Top. Accel. Beams* **8** (2005) 053201.

[91] F. Billard, N. Hilleret and G. Vorlaufer, Vacuum Technical Note 00-32 (CERN, Geneva, 2000).

[92] T. Zhang, B. You, A. Hu *et al.*, *Colloid J. Russ. Acad. Sci. Kolloid. Zh.* **65** (2003) 21.

[93] J.C. Lin and R. Gomer, *Surface Sci.* **218** (1989) 406.

[94] G. Vorlaufer, CERN Thesis (CERN/Technische Universität Wien, Wien, 2002), p. 200.

[95] O. Gröbner, in *Cern Accelerator School: Vacuum Technology,* Snekersten, Denmark, 1999, CERN 99-05, p. 127.

[96] J. Gómez-Goñi, O. Gröbner and A.G. Mathewson, *J. Vac. Sci. Technol., A* **12** (1994) 1714.
[97] O. Gröbner, A.G. Mathewson, P. Strubin *et al.*, *J. Vac. Sci. Technol., A* **7** (1989) 223.
[98] C.L. Foerster, C. Lanni, I. Maslennikov *et al.*, (AVS, 1994), Vol. 12, p. 1673.
[99] C.L. Foerster, C. Lanni, J.R. Noonan *et al.*, (AVS, 1996), Vol. 14, p. 1273.
[100] C.L. Foerster, C. Lanni and K. Kanazawa, (AVS, 2001), Vol. 19, p. 1652.
[101] P. Chiggiato and R. Kersevan, *Vacuum* **60** (2001) 67.
[102] C. Benvenuti, J.M. Cazeneuve, P. Chiggiato *et al.*, *Vacuum* **53** (1999) 219.
[103] C. Benvenuti, P. Chiggiato, P. Costa Pinto *et al.*, *Vacuum* **60** (2001) 57.
[104] P. Chiggiato and P. Costa Pinto, *Thin Solid Films* **515** (2006) 382.
[105] V. Baglin, C. Benvenuti, P. Chiggiato *et al.*, in *International Vacuum Congress*, San Francisco, USA, 2001.
[106] F. Le Pimpec, O. Gröbner and J.M. Laurent, *Nucl. Instrum. Methods Phys. Res., B* **194** (2002) 434.
[107] S.K. Erents and G.M. McCracken, *J. Appl. Phys.* **44** (1973) 3139.
[108] R.E. Johnson, L.J. Lanzerotti and W.L. Brown, *Nucl. Instrum. Methods Phys. Res.* **198** (1982) 147.
[109] E.R. Moog, J. Unguris and M.B. Webb, *Surface Sci.* **134** (1983) 849.
[110] J. Schou, *Nucl. Instrum. Methods Phys. Res., B* **27** (1987) 188.
[111] J. Schou, Risoe-R-591 (Risoe National Laboratory, Roskilde, Denmark, 1991).
[112] R.A. Baragiola, R.A. Vidal, W. Svendsena *et al.*, *Nucl. Instrum. Methods Phys. Res., B* **209** (2003) 294.
[113] G.M. McCracken, *Vacuum* **24** (1974) 463.
[114] N. Hilleret and R.S. Calder, in *7th International Vacuum Congress*, Vienna, 1977), p. 227.
[115] J.C. Barnard, I. Bojko and N. Hilleret, *Vacuum* **47** (1996) 347.
[116] H. Tratnik, N. Hilleret and H. Störi, *Vacuum* **81** (2007) 731.
[117] A.G. Mathewson, *Proceedings of the IEEE Particle Accelerator Conference* **5** (1993) 3828.
[118] V.V. Anashin, O.B. Malyshev, R. Calder *et al.*, *Vacuum* **48** (1997) 785.
[119] V.V. Anashin, O.B. Malyshev, R. Calder *et al.*, *Nucl. Instrum. Methods Phys. Res., A* **405** (1998) 258.
[120] V.V. Anashin, O.B. Malyshev, R. Calder *et al.*, *Vacuum* **53** (1999) 269.
[121] V. Baglin, I.R. Collins, O. Gröbner *et al.*, in *European Particle Accelerator Conference*, Paris, 2002.
[122] D. Bintinger, P. Limon and R.A. Rosenberg, *J. Vac. Sci. Technol., A* **7** (1989) 59.
[123] H. Tratnik, CERN Thesis (CERN/Technische Universität Wien, Wien, 2005), p. 120.
[124] A.G. Mathewson and P. Strubin, ISR Performance Report ISR/VA/AM/sm 17-08-81 (CERN, Geneva, 1981).
[125] G. Hulla, private communication (2007).

Materials for high vacuum technology: an overview

S. Sgobba
CERN, Geneva, Switzerland

Abstract
In modern accelerators stringent requirements are placed on materials for vacuum systems. Their physical and mechanical properties, machinability, weldability, or brazeability are key parameters. Adequate strength, ductility, magnetic properties at room temperature as well as at low temperatures are important factors for vacuum systems of accelerators working at cryogenic temperatures, such as the Large Hadron Collider (LHC) under construction at CERN. In addition, baking or activation of Non-Evaporable Getters (NEGs) at high temperatures imposes specific choices of material grades of suitable tensile and creep properties in a large temperature range. Today, stainless steels are the dominant materials of vacuum constructions. Their metallurgy is extensively treated here. The reasons for specific requirements in terms of metallurgical processes are detailed, in view of obtaining adequate purity, inclusion cleanliness, and fineness of the microstructure. In many cases these requirements are crucial to guarantee the final leak tightness of the vacuum components.

1 An historical introduction

As far back as 1936, Springer in Berlin published a treatise on *Werkstoffkunde der Hochvakuumtechnik* by Espe and Knoll [1] which saw a five-fold increase in contents by the time of its first English edition of 1966 [2]. The Espe handbook is the first systematic and comprehensive presentation of high vacuum materials. Espe pointed out that the selection and handling of vacuum materials is based on a different viewpoint than ordinary industrial construction, mentioning the ease of degassing, an adequate strength at high as well as low temperature, matching values of thermal expansion coefficients, and purity of materials as principal factors to be considered in material selection for vacuum applications. Ease of fabrication and cost of vacuum materials were considered as being often 'of secondary importance'. Very constant properties of the raw materials, to be specially prepared, exact knowledge of the material properties, critical selection and careful controls were mentioned as aspects of paramount importance. Among the materials treated, glass is of a primary importance, followed by ceramics, precious metals (mainly Pt used in glass–metal sealing), iron and steels (little attention is devoted in the handbook to stainless steels), refractory metals and alloys, polymers, Al, Cu, Ni, and their alloys. At that time, electrode systems and vacuum tubes were the main applications and the first background of this voluminous treatise of several thousand pages.

A very comprehensive handbook on materials is the *Handbook of Materials and Techniques for Vacuum Devices* by Kohl [3], a revision and expansion of its predecessor *Materials and Techniques for Electron Tubes* [4]. Glass (called at the time 'the miracle maker' from an old expression of Phillips [5]) and ceramics are the main materials of interest, together with metallic materials. Additional handbooks containing more or less comprehensive reviews of materials for high vacuum were published in the 1960s [6–8]. In a recent book by O'Hanlon [9], containing a chapter on materials, stainless steels receive the largest attention compared to other metallic materials. Today austenitic stainless steels represent the reference material for many vacuum devices, owing to their corrosion resistance, strength and ductility retained at ambient and service temperatures, suitability for intended

cleaning procedures, stability of properties in service, toughness, magnetic properties, sharpness (retention of cutting edges in applications to vacuum seals), rigidity, and dimensional stability [10].

In addition to stainless steels, other families of materials relevant for vacuum applications in accelerators that are discussed in the present overview are Al and alloys, Cu and alloys, including oxide dispersion strengthened alloys. Special factors will be discussed that are relevant to manufacturing techniques for vacuum purposes such as welding (including laser and electron beam welding) or brazing (in particular metal to ceramic junctions). Finally, innovative technologies or processes will be presented, which are applied in the frame of the construction of the Large Hadron Collider (LHC) such as near net shaping techniques based on Powder Metallurgy (PM) and Hot Isostatic Pressing (HIP), or are foreseen for future developments such as HIP diffusion bonding. For each material family and process, examples of applications are given. Failure analyses, including corrosion issues are discussed.

2 Stainless steels

Stainless steel plays a crucial role in the construction of modern accelerators. The example of the dipole and quadrupole magnets of the LHC is representative in this respect. A special austenitic stainless steel has been developed [11] for the beam screen and the cooling capillaries of the machine vacuum system, retaining high strength, ductility, and low magnetic susceptibility at the working temperature between 10 K and 20 K. Several tens of kilometres of components have been produced in this special grade. The magnet cold bore is manufactured as a seamless 316 LN tube. 316 LN is also the reference grade for the shrinking cylinder of the dipole magnets (2500 bent plates of 15.35 m in length, 10 mm thick for a total weight of approximately 3000 t). The plates are welded longitudinally by a special Surface Tension Transfer (STT) technique combined with traditional pulsed Metal Inert Gas (MIG) welding. More than 2800 magnet end covers of complex shape, including several nozzles, have been fabricated starting from HIPed 316 LN powders and near net shaped into geometry close to the final form. Between the magnets, some 1600 interconnections consist of several thousand leak tight components to be integrated, mainly working at cryogenic temperature (1.9 K). Interconnection components are also mainly based on austenitic stainless steels. For the convolutions of the several thousands of bellows involved in the machine and working under cyclic load at 1.9 K, a special remelted 316 L grade has been selected showing an extremely low inclusion content and improved austenite stability at the working temperature. The grade is highly formable at Room Temperature (RT) [12].

2.1 Families of stainless steels

Stainless steel can be defined as a ferrous alloy containing a minimum of 12% Cr [13], with or without other elements [10,14]. Chromium, as well as additional alloying elements, imparts corrosion and oxidation resistance to steel. Figure 1 is an iron–chromium diagram, which is the foundation of stainless steels [13]. On the 100% Fe axis of the diagram, one recognizes the stability domains of the various phases of iron as a function of temperature: α- and δ-iron, corresponding to the ferromagnetic ferritic phase of body centred cubic (bcc) structure, which are present up to 912°C and in the ranges between 1394°C and 1538°C, respectively, and γ-iron, corresponding to the austenitic phase of face centred cubic structure (fcc), in the range between 912°C and 1394°C. This phase is non-ferromagnetic.

In the diagram of Fig. 1, the ferritic phase is extensive while the γphase is limited to a loop. The same diagram is the basis of identification of two first families of stainless steels: ferritic types, with Cr contents between 14.5% and 27% Cr, and martensitic types that are iron–chromium steels with small additions of C and other alloying elements, usually containing no more than 14% Cr (except some types with 16% to 18% Cr) and sufficient C to promote hardening. From the diagram it can be

seen that ferritic grades do not transform to any other phase up to the melting point, and hence can only be strengthened by cold working. Martensitic grades, owing to the addition of C to the Fe–Cr system enlarging the domain of stability of the γ-phase, can be hardened by a rapid cooling in air or a liquid medium from above a critical temperature. This results in grades that have excellent strength.

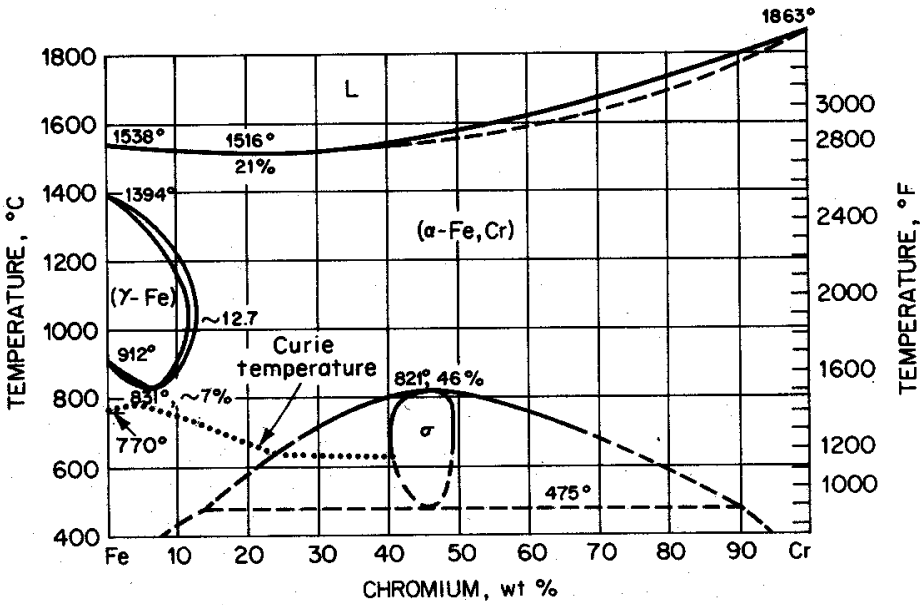


Fig. 1: The Fe–Cr phase diagram (from Ref. [15])

Ferritic and martensitic grades are generally inferior in terms of corrosion resistance than austenitic grades, are ferromagnetic, and undergo ductile-to-brittle transition at cryogenic temperatures. Moreover, at temperatures in the range of 1000°C or more (solution annealing), ferritic grades undergo grain growth, while martensitic grades are subject to a loss of the martensitic hardening phase due to a reversion of martensite to austenite in the annealing conditions associated to a high-temperature treatment. On the other hand, austenite is non-magnetic, does not undergo any ductile-to-brittle transition below RT and is less subject to grain growth during vacuum firing. For these reasons, the austenitic grades are a first choice for vacuum applications. Austenitic grades, because of their high Cr and Ni content, are the most corrosion resistant of the stainless steel group[1].

2.2 Austenitic stainless steels

They are formed by the addition of elements (Ni, Mn, N, etc.) to the Fe–Cr system of Fig. 1, expanding the domain and enhancing the stability of the γ phase. For sufficient additions of a balanced amount of adequate alloying elements, the formation of ferrite can be suppressed and the tendency to form martensite on cooling or during work hardening can be partially or completely suppressed. As an example, the addition of 8% to 10% Ni to a low C FeCr steel can already allow a relatively stable austenitic structure at RT to be obtained. The '304-type' stainless steel (18Cr8Ni, on an Fe basis) is a typical example of a very common iron–chromium–nickel austenitic stainless steel. This grade is generally applied in its low (L) carbon version 304L ($C \leq 0.030\%$), showing enhanced corrosion resistance and ductility especially in welded structures.

Figure 2 is a pseudobinary section of the FeCrNi ternary diagram for increasing Cr + Ni contents [16]. In this diagram, for a total Fe of 70%, it is straightforward to visualize the basic 304 type of austenitic stainless steel. Continuous lines in the diagram correspond to real transformations, dashed lines to transformations not observed in practical conditions. For a 20% Cr content and 10% of

[1] Because of their limited application in vacuum devices, precipitation-hardening grades are not mentioned here.

Ni (composition corresponding to the vertical line identified by an arrow), the diagram shows that cooling from the usual solution annealing temperature[2] of 1050°C can result in some residual ferrite. Indeed, the 304L grade retains, when cooled from solution annealing (or precipitates during a welding operation), a given amount of untransformed delta ferrite. Increasing Ni and reducing Cr (vertical lines to the left) enhances stability of the γ-phase against precipitation of ferrite and, for sufficient Ni, can totally suppress it.

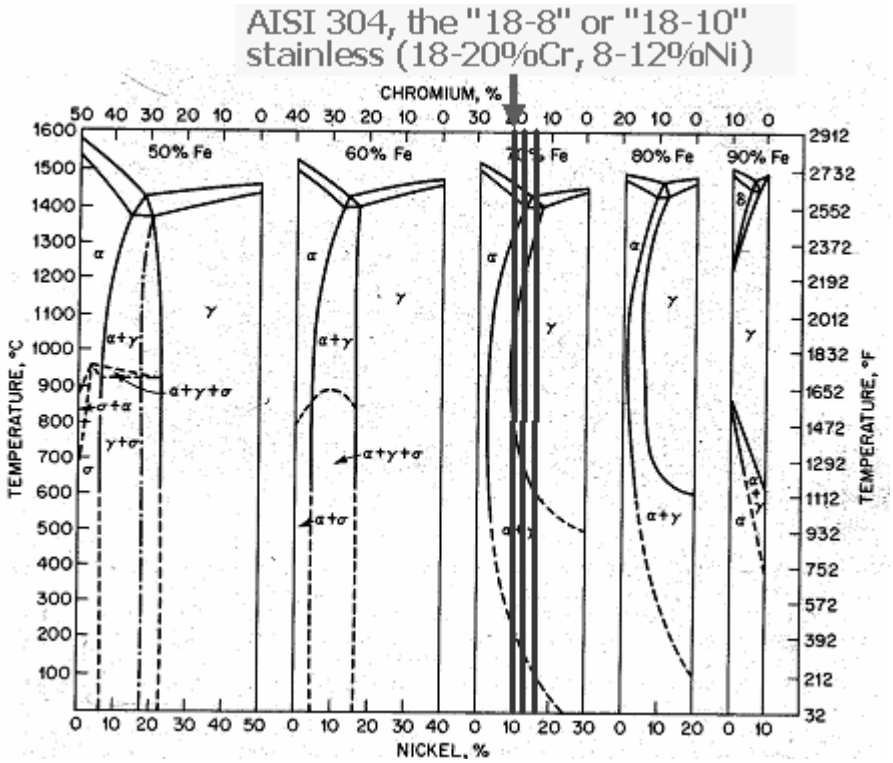


Fig. 2: Cross-section of the Fe–Cr–Ni ternary phase diagram [16]

From the diagram in Fig. 2, it appears clear that:

1) Austenitic stainless steels may contain residual amounts of delta ferrite, which can be critical on account of its reduced toughness and of its ferromagnetic nature for specific vacuum applications, particularly for components that have to be applied at cryogenic temperature and/or in critical magnetic environments.

2) Elements like Ni extend the domain of austenite, while Cr reduces it. Elements such as C, N, Mn (to some extent), or combinations of them play a role similar to Ni ('Ni-equivalents', Ni_{eq}), while elements like Mo, Si, Nb, or their combination act as 'Cr-equivalents' (Cr_{eq}). Schaeffler, DeLong or Hull diagrams allow the total ferrite formation effects to be predicted in a multielement system, such as an industrial stainless steel grade and its welds (Section 2.5.1).

2.3 Grades of practical interest for vacuum applications

The first three rows in Table 1 show the composition of the grades of stainless steels (304L, 316L, 316LN) of main interest for vacuum applications and covered by CERN specifications. The mentioned composition ranges are based on CERN specifications [17–22], with reference to the relevant standard grade designations.

[2] Solution annealing is the usual final step in the fabrication of several austenitic stainless steel products. For the different grades, ranges of solution annealing temperature are prescribed by standards.

Table 1: Grades of practical interest for vacuum applications. Composition ranges according to CERN specifications, wherever applicable, or EN 10088 [23]. Single values are maximum admitted values.

Grade (AISI, or 'commercial designation')	Grade (EN, symbolic and numeric)	C	Cr	Ni	Mo	Si	Mn	N	Others
304L [17–18]	X2CrNi19-11 1.4306	0.030	17.00–20.00	10.00–12.50		1.00	2.00		P ≤ 0.045, S ≤ 0.030 Co ≤ 0.2
316L [19]	X2CrNiMo17-12-2 1.4404	0.030	16.00–18.50	11.00–14.00	2.00–2.50	1.00	2.00	0.050	P ≤ 0.030, S ≤ 0.010 Co ≤ 0.2
316LN [20–22]	X2CrNiMoN17-13-3 1.4429	0.030	16.00–18.50	12.00–14.00	2.00–3.00	1.00	2.00	0.14–0.20	P ≤ 0.045, S ≤ 0.015 Co ≤ 0.2
316Ti [23]	X6CrNiMoTi17-12-2 1.4571	0.08	16.50–18.50	10.50–13.50	2.00–2.50	1.00	2.00		P ≤ 0.045, S ≤ 0.015, 5 × C ≤ Ti ≤ 0.70
'P506' [11]	–	0.030	19.00–19.50	10.70–11.30	0.80–1.00	0.50	11.80–12.40	0.30–0.35	P ≤ 0.020, S ≤ 0.002, B ≤ 0.002, Cu ≤ 0.15, Co ≤ 0.2

AISI 304L is a general-purpose grade. For vacuum applications, the grade should be purchased through a careful specification, aimed to achieve a substantially austenitic microstructure and a controlled maximum level of inclusions (Section 2.4). Owing to the limited amount of alloying elements, its magnetic susceptibility can be subject to increase by martensitic transformation upon cooling to RT or to cryogenic temperatures, and following work hardening at RT or lower [24]. Its selection should be carefully considered in applications where an increase of magnetic susceptibility might be of concern. The price of this grade is approximately 5 CHF/kg (at spring 2006 rates) for a general-purpose version, and up to 8.50 CHF/kg for a vacuum specified wrought product.

AISI 316L is a Mo-containing grade. Molybdenum enhances corrosion resistance and austenitic stability versus martensitic transformation. The ferrite-promoting characteristics of Mo (Section 2.2 and Section 2.5.1.1) have to be compensated by adjustments in Cr and Ni to achieve an almost fully austenitic microstructure. Owing to its formability, ductility and increased austenitic stability compared to 304L, this grade is covered by a special CERN specification for application to bellows convolutions of the LHC interconnections. In the form of thin sheet for these convolutions, the price rises up to 40–80 CHF/kg. Wrought products for vacuum applications have prices in between 304L and 316LN for equivalent metallurgy and metalworking processes and controls.

AISI 316LN is a nitrogen containing stainless steel. Nitrogen increases austenite stability against martensitic transformations and is a powerful austenite former with respect to ferrite. Nitrogen substantially increases strength, while allowing ductility to be maintained down to cryogenic temperatures [25]. Owing to limited softening compared to 304L and 316L, 316LN is the grade selected when vacuum firing is required. Prices in the order of 16–18 CHF/kg are common for wrought products such as sheets and tubes. Multidirectionally forged products issued from ElectroSlag Remelted (ESR) ingots (Section 2.4) have prices of the order of 50 (but up to 135) CHF/kg.

AISI 316Ti is an example of a Ti 'stabilized' grade. Stabilized grades contain higher C than low C grades (limited to 0.030% C max). They are alloyed with Nb, Ta or Ti to prevent carbide precipitation. AISI 316Ti, or similar stabilized grades (AISI 321, AISI 347), are offered by steel suppliers or component producers as an alternative to low-carbon grades. Stabilized grades should generally be avoided for demanding vacuum applications, since the addition of the stabilizer elements results in precipitate carbides, reducing the cleanliness of the steel (Section 2.4) and toughness in specific low temperature applications.

P506 is a stainless steel specially developed by CERN [11], belonging to the family of high Mn, high N stainless steels and successfully produced by Böhler /AT and Aubert et Duval /FR. This special composition allows low relative magnetic permeability (< 1.005) to be maintained down to cryogenic temperatures in the base material and in its welds. A full stability of the austenite versus delta ferrite precipitation in welds, and/or versus martensitic transformations when deformed at very low temperatures is guaranteed by the alloy composition [25]. Its price (40–60 CHF/kg) is comparable to that of 316L sheets or multidirectionally forged products in 316LN grade.

2.4 Special requirements on the microstructure of stainless steels for vacuum applications

Figure 3 shows an end-fitting welded to a thin bellows convolution, intended for application in the vacuum system of the CMS experiment at CERN. The bellow ends are machined from stainless steel forged bars, showing strings of non-metallic inclusions aligned parallel to the axis of the bar (the primary forming direction). These inclusions result in a leak through the material thickness of some 10^{-4} mbar · l · s^{-1} [26].

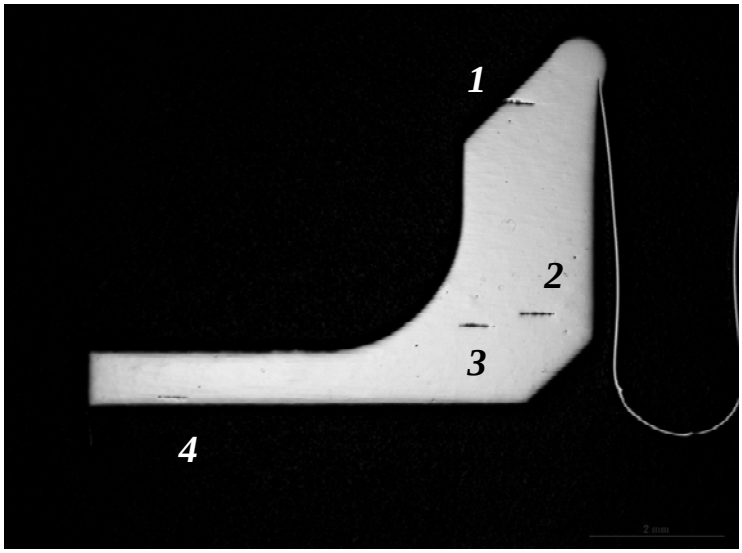


Fig. 3: Bellow end-fitting machined from an AISI 316LN round bar, showing alignment of oversized (1, 2, 3) and thick (4) B type inclusions up to class 2, classified according to standard ASTM E45-97ε2 (2002) [27]. The CERN specification [20] imposes a class of inclusions at most 1 for type B inclusions and a half-class above this limit in up to 2% of the fields counted.

Several leaks through the base material thickness of flanges of the plug in modules of LHC magnet interconnections have been reported and investigated [28]. In this case, leaks are due to exogenous macroinclusions (Fig. 4). Macroinclusions are understood in terms of the presence of entrapped slag or refractories added to protect the steel melt during the melting phases, segregated during the solidification and not fully removed because of an insufficient top discard of the steel ingot. Slag trapped in the ingot results in elongated macroinclusions during the hot transformation of the steel (forging of the ingot, rolling of the bar).

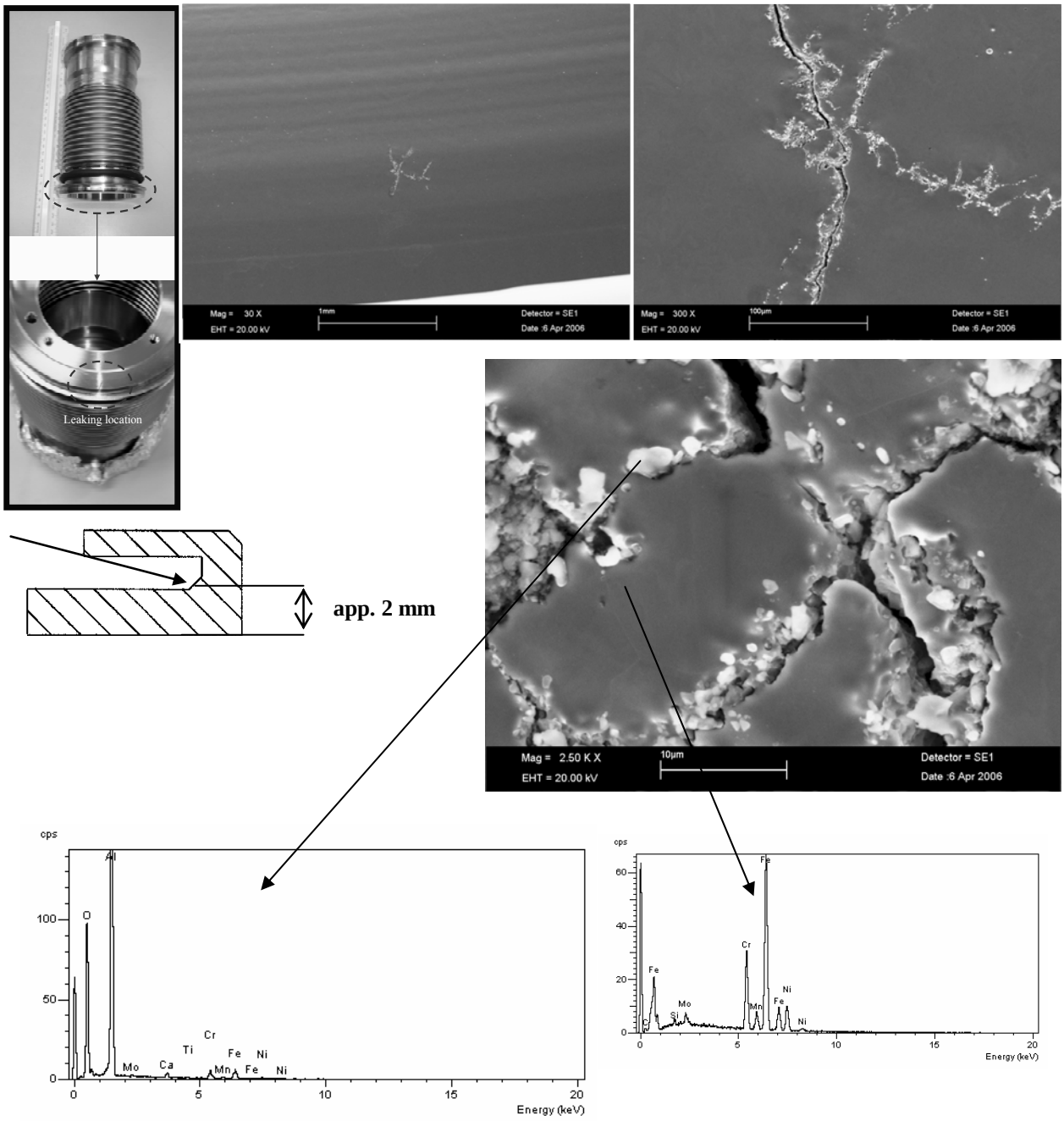


Fig. 4: Several plug-in modules of the LHC interconnections have been found leaking at the flange location. Flaws associated with the presence of Al, O and Ca were detected on the two opposite surfaces of the throats of the cooling exit tube. Al and Ca are typical elements contained in slag or refractories. Entrapped residues of slag resulted in macroinclusions. Elsewhere, cross-sectional micro-optical observations of a non-leaking area only showed B type inclusion up to class 1 (worst field) [27].

A complete specification of a stainless steel for vacuum applications must consider the risk of leaks due to the presence of non-metallic micro- and macroinclusions embedded in the microstructure of the final product. The risk of leaks can be reduced by one or more of the following actions:

1) In the technical specification, impose a maximum allowed content of non-metallic inclusions. CERN specifications impose maximum limits according to the international standards in force.

2) Since microscopical test methods for determining the inclusion content of steel are not intended for assessing exogenous macroinclusions, impose a steelmaking process. Stainless steels are primarily melted in an electric furnace and decarburized through Argon-Oxygen Decarburization (AOD) or Decarburization in Vacuum (VOD). These processes, consisting of a single melting step, might not guarantee on their own a sufficient homogeneity of the ingot. Imposing a remelting process such as Vacuum Arc Remelting (VAR) or ESR ensures the homogeneity of the material will be effectively influenced. In addition, remelting reduces the impurity and the microinclusion content of the final products, allows high density and a lack of macrosegregations with shrinkage cavities to be achieved [29].

3) Specify three-dimensionally forged products [21]. Upsetting steps allow the alignment of the inclusions which are elongated by two-dimensional steps of previous forging or rolling to be broken, thus reducing the risk of leak through the thickness.

4) Specify products with a fine and homogenous grain size.

5) Avoid breaking the fibre of the product by machining walls perpendicular to the direction of the material flow, such as in the example of Fig. 3.

6) Introduce Non-Destructive Testing (NDT) such as Ultrasonic Testing (UT) not only on the final product, but also on semifinished products resulting from intermediate steps of steel transformation. NDT procedures and acceptance criteria should be the object of an agreement between supplier and customer.

In conclusion, for vacuum applications for accelerators, stainless steels should never be selected on the basis of a mere material designation or availability of general-purpose stock. Critical parameters should be explicitly specified and stringently controlled, in some cases not only on the final product but through the definition and application of a quality control plan with tests associated with the different phases of material production.

2.5 Aspects related to joining processes of stainless steels in the frame of vacuum applications

2.5.1 Precautions during welding

Austenitic stainless steels are readily welded by conventional arc (TIG, MIG) or beam (laser, electron beam) techniques. Properly specified and qualified welds according to standards in force[3] can be fully sound. Tensile strengths equal to or higher than the minimum specified for the base metal can be obtained. For cryogenic applications, austenitic stainless steels can be welded whilst guaranteeing a minimum value of impact energy. Nevertheless, some additional aspects are relevant for welding austenitic stainless steels in the frame of a vacuum application.

[3] Welds specified according to level B of standards ISO 15614-1 [30] for arc welding, ISO 15614-11 [31] for electron and laser beam welding, with reference to standard ISO 6520-1 [32] and ISO 5817 [33] or ISO 13919-1 [34] for classification and quality levels for imperfections, are generally specified at CERN for vacuum applications.

2.5.1.1 Possible presence of delta ferrite in austenitic stainless steel welds

As mentioned in Section 2.2, austenitic stainless steels can contain residual amounts of delta ferrite, which can be critical in applications to accelerators owing to its reduced toughness and ferromagnetic nature. Constitution diagrams for stainless steel weld metals such as the one presented in Fig. 5 allow the ferrite content to be estimated in the as-deposited weld on the basis of the composition of the base and, when applicable, of the weld filler metal. Control of ferrite to minimum or zero level might be required in the context of specific accelerator applications (cryogenics, components close to the beam). Since fully austenitic stainless steel welds are more sensitive to microfissuring, specific precautions have to be taken such as reducing the weld heat input, minimizing restraint, designing for low constraint and keeping impurities in the stainless steel composition to minimum levels [14].

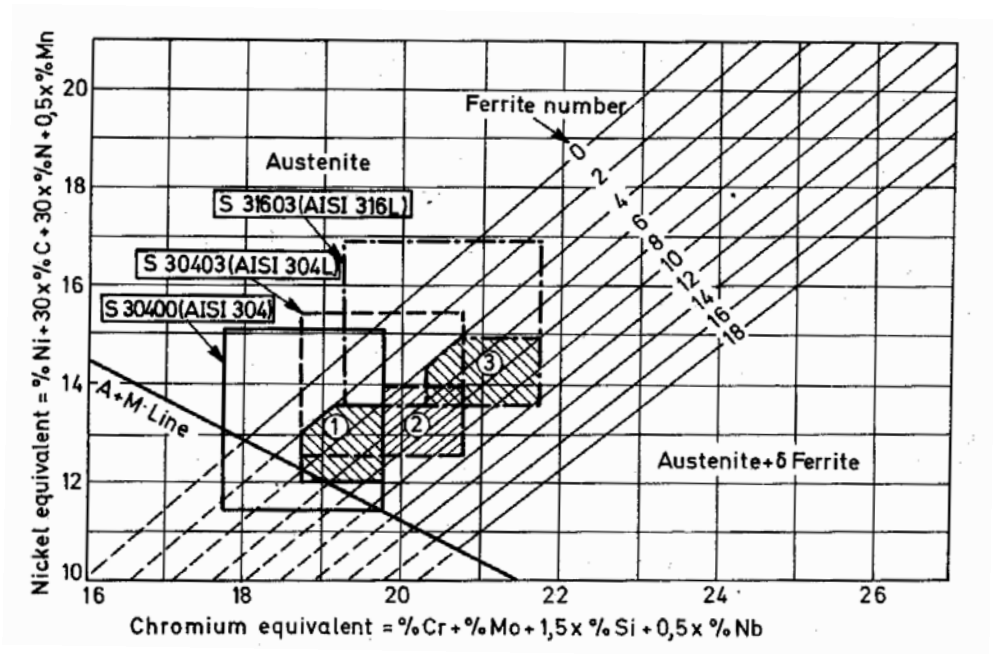


Fig. 5: Diagram for the determination of the ferrite content in austenitic stainless steel weld metal according to DeLong [35,36]

2.5.1.2 Minimizing the risk of microfissuring in fully austenitic stainless steel welds

Figure 6 shows a so-called 'Suutala diagram' allowing the risk of hot cracking to be predicted, as a function of the solidification mode (primary ferrite or austenite) and the impurity content of the steel (P, S). The primary mode of solidification, be it austenite or ferrite, is important for predicting the integrity of the weld. The ratio of Cr_{eq} to Ni_{eq} identifies the two modes in the diagram, primary ferrite for $Cr_{eq}/Ni_e > 1.5$ approx., primary austenite for $Cr_{eq}/Ni_e < 1.5$. In the domain of solidification to primary austenite, only the grades, fillers or combination of them corresponding to a very limited total residual element content can be considered safe in terms of hot cracking. Particular care should be taken when mixing base materials of very different origin and quality. Austenitic stainless steels like 316LN properly specified [20,21] generally solidify in the fully austenitic range and show limited impurity content. If mixed with a general-purpose stainless steel such as 304L or 316L, or so-called 'free machining' grades with added S to improve machining rates, a risk of cracking cannot be excluded depending on the dilution. Austenitic stainless steels with high impurity contents are generally designed to solidify in the primary delta ferrite range, in order to avoid risks of hot cracking. Their dilution with a fully austenitic grade might locally result in an austenitic weld with equivalent unacceptable level of impurities.

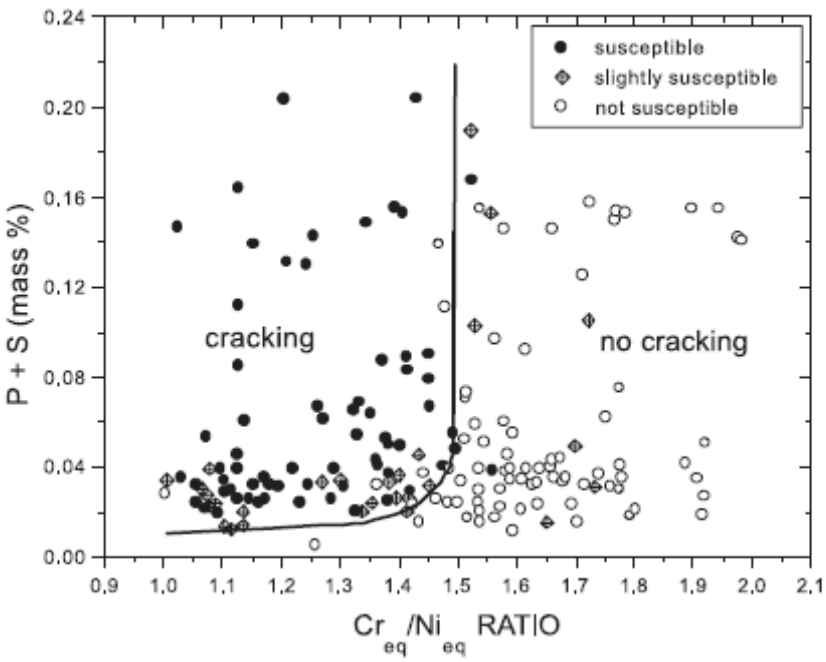


Fig. 6: Cracking susceptibility during arc welding of austenitic stainless steels ('Suutala' type diagram [37–40]). Equivalents based on Schaeffler equivalent formulae for Cr_{eq} and Ni_{eq}, $Cr_{eq} = Cr + 1.5Si + 1.37Mo$, $Ni_{eq} = Ni + 0.31Mn + 22C + 14.2N$. Variants of the diagram exist to also take into account the effect of B.

Because of this risk, the use of free-machining grades, where additions of 150 ppm to 300 ppm S are allowed [23], should be strictly avoided. Components to be welded together in the vacuum system of an accelerator can be of very different origin and supplies. Their impurity content and solidification mode should be carefully checked to avoid leaks in the welds due to microfissuring.

2.5.2 Vacuum brazing stainless steel to ceramic

When vacuum brazing a metal to a ceramic material, the compatibility of the thermal expansion properties of the two materials to be brazed has first to be checked. Figure 7 compares the thermal expansion curves of selected metals, alloys, and ceramics. Brazing a ductile metal of significantly different thermal expansion coefficient to a ceramic is made possible by the ability of the metal to deform whilst cooling down from the brazing temperature (the ductility of the 'unmatched metal' is required). An example is Al_2O_3 to Cu brazing. On the other hand, direct brazing of Al_2O_3 to stainless steel is seldom performed. In order to join the two materials, an intermediate component is used, typically made of Kovar, matching the thermal expansion coefficient of the ceramic. The metal-to-ceramic joint obtained in this way can be readily welded (e.g., by electron beam welding) to the stainless steel component.

The compatibility of the two materials to be assembled can be evaluated through the Thermomechanical Compatibility Factor (TCF) [41]. $TCF = (\sigma_Y)^{-1}$ for $\varepsilon_T \geq \varepsilon_Y$, while $TCF = (\sigma_Y\, \varepsilon_T/\varepsilon_Y)^{-1}$ for $\varepsilon_T < \varepsilon_Y$, where ε_T is the difference of the total elongation between metal and ceramics at the brazing temperature, ε_Y and σ_Y the yield strain and strength, respectively. High values of TCF correspond to high compatibility. If the difference in total strain exceeds the elastic limit of the metal ($\varepsilon_T \geq \varepsilon_Y$,), the compatibility will be the best for a ductile material with low yield strength. On the other hand, if the difference in total strain is under the yield strain, the compatibility will be the best for a couple with low difference in thermal expansion and, *ceteris paribus*, for a metal with low elastic modulus ($\approx \sigma_Y/\varepsilon_Y$) allowing the internal stresses resulting from brazing to be minimized.

As a general rule, owing to the strong anisotropy of the mechanical properties of ceramics, design of the joint to be brazed should result in a compressive state of stress of the ceramic. The compression strength of a 96% Al_2O_3 equals 2070 MPa at RT, while the tensile strength is limited to 180 MPa, further decreasing to 90 MPa at 1100°C [42].

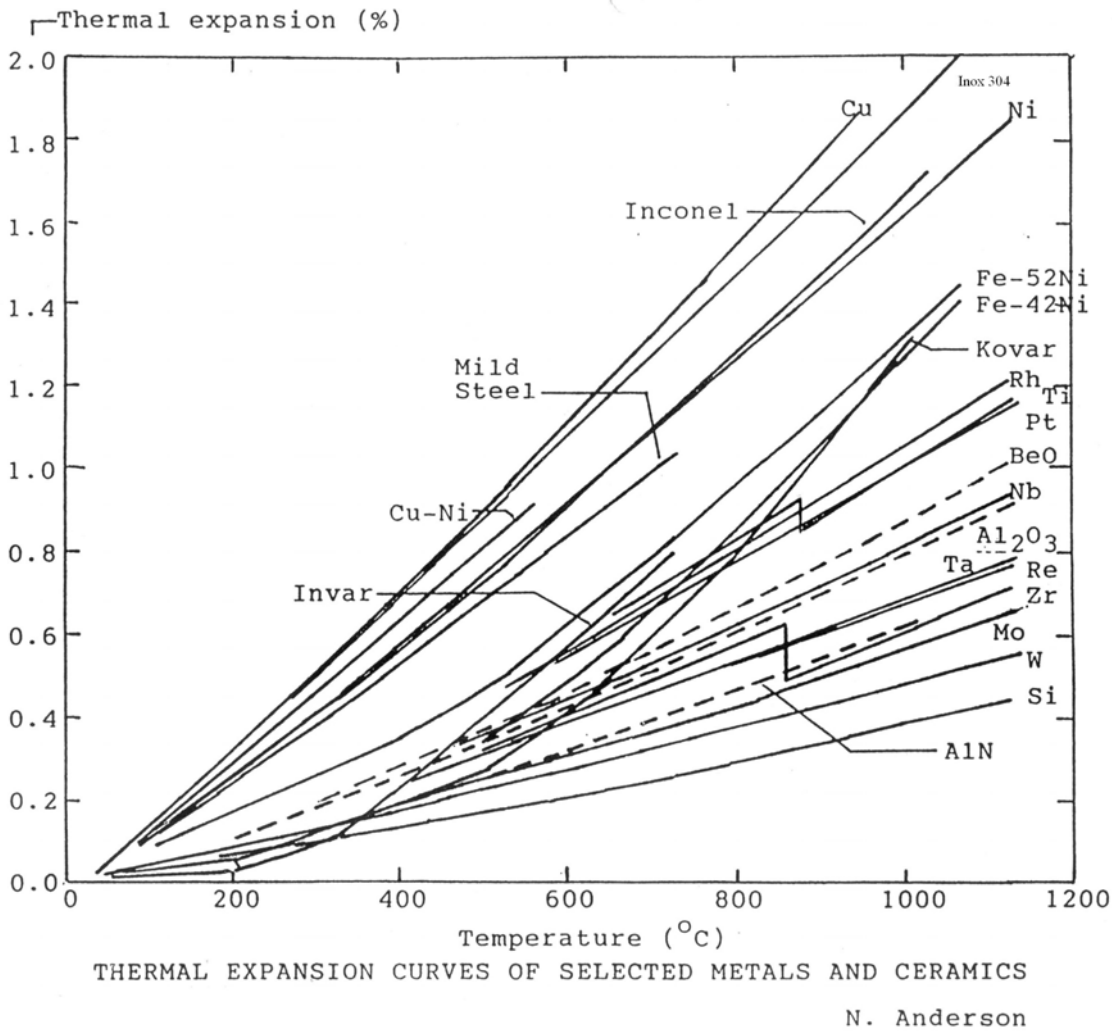


Fig. 7: Thermal expansion curves of selected metal and ceramics [41]

Metal-to-ceramic brazing requires an adherence of the brazing fillers on both metal and ceramic surfaces. In order to achieve this goal, two main possibilities are offered, the use of active brazing or metallization of the ceramic surface. A typical metallization applied among others to alumina-based ceramics is the 'Moly-Manganese' process [43].

3 Aluminium alloys

Wrought aluminium alloys are very attractive materials for ultrahigh vacuum systems for accelerators. Aluminium alloys allow the residual radioactivity after machine shutdown to be strongly reduced compared to stainless steels, show high transparency to radiation, can be shaped into complicated profiles by extrusion, are completely non-magnetic and, for given components, might be competitive in terms of cost. High electrical and thermal conductivity are an asset. High thermal conductivity and low thermal emissivity allow aluminium alloy components to tolerate high heat fluxes in spite of their

relatively low melting point [44]. An all-aluminium alloy vacuum system has been adopted in the frame of the TRISTAN electron–positron collider constructed at the National Laboratory for High-Energy Physics in Japan, following developments of components such as bakable aluminium vacuum chambers and bellows [44,45].

3.1 Aluminium alloy grades of interest for vacuum construction and their weldability

Aluminium alloys are classified in eight groups or series (Table 2), according to their main alloy element. The first digit identifies the family, the last two digits the type in the family (with the exception of the 1xxx series where subsequent digits identify the purity), the second digit a variation in the same family.

Table 2: Wrought aluminium alloy designation and grades of practical interest for vacuum applications. Heat treatable families are shown in bold characters. Some alloys of the 8xxx series are heat treatable as well.

Alloy group	AA designation	Readily weldable alloys of the family	Examples of relevant alloys for vacuum applications
Pure aluminium	1xxx series	EN AW-1060, –1100, –1199, –1350	
Al-Cu	**2xxx series**	**EN AW-2219, -2090 (AlCuLi)**	**EN AW-2219**
Al-Mn	3xxx series	EN AW-3003, –3004, –3105	EN AW-3003
Al-Si	4xxx series	weld fillers	weld fillers
Al-Mg	5xxx series	EN AW-5005, –5050, –5052, –5083, –5086, –5154, etc.	EN AW-5083
Al-Mg-Si	**6xxx series**	**EN AW-6061, –6063, –6070, –6082**, etc.	**EN AW-6082** **(6061)**
Al-Zn	**7xxx series**	**EN AW-7005, –7020**	
Al+other element (e.g., Li)	8xxx series	**EN AW-8090 (AlLiCu)**	

Aluminium alloys can be hardened by cold or warm working (strain hardened alloys) and/or by heat treatment (age hardened alloys), depending on the type of alloying elements. For strain hardened alloys, the level of mechanical properties that can be attained depends on the alloying elements and their content: 5xxx alloys have a potential level of mechanical strength achievable by work hardening that is superior to 1xxx and 3xxx series. Age hardening alloys, identified in bold in Table 2, are generally strengthened by a three-stage treatment (solution annealing, quenching or rapid cooling, holding at RT or higher temperature). A cold work step can be added between solution heat treating and ageing for certain alloys.

Extrudable (EN AW-6061, -6082, -5083, 7020, etc.) and weldable alloys (see Table 2) are of paramount importance for vacuum constructions for accelerators. Several alloys within the different families are considered as readily weldable. The weldability of aluminium alloys by electron beam techniques is rated by the Merkblatt DVS 3204 [46].

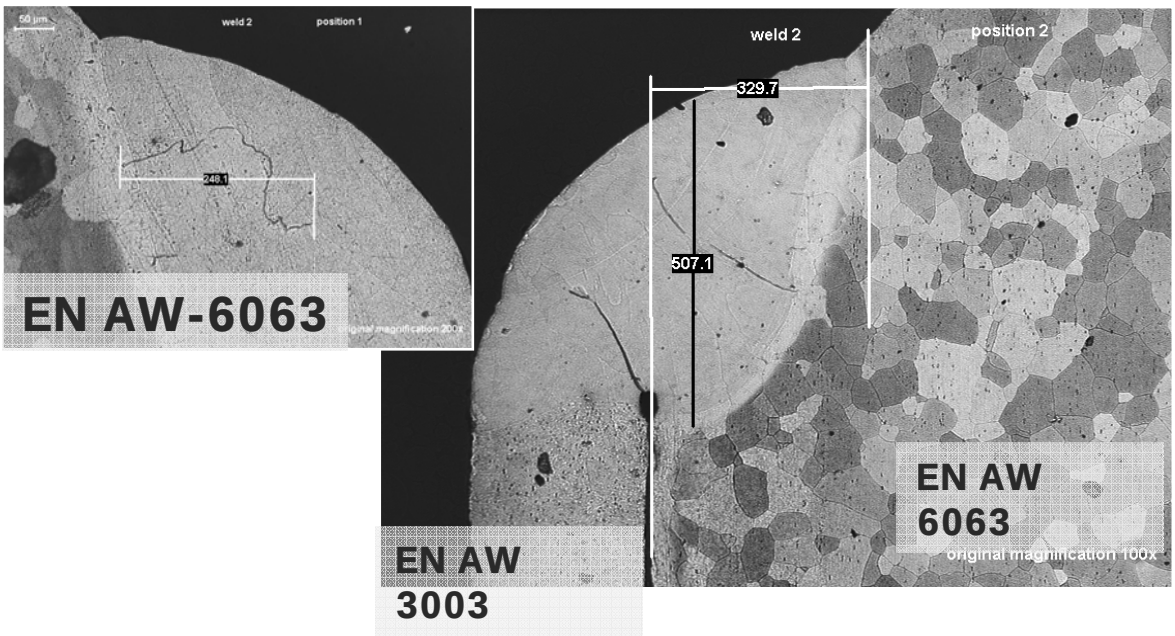


Fig. 8: Solidification cracks in autogeneous laser weld of tubes and fittings of the cooling circuit of the ATLAS Pixel experiment involving a 6xxx series alloy: a) laser weld developing a liquation crack; b) in mixed welds (EN AW-6063 diluted with EN AW-3003), where no filler is applicable (laser weld of thin walled components), cracks can develop due to an unfavourable dilution of the two alloys. Locally in the weld bead a critical Mg-Si balance can be present, resulting in solidification cracks [47]. Measured distances in μm.

Nevertheless, some families of weldable alloys are more sensitive than others to microcracking (solidification cracking, Fig. 8). In general, the Si and Mg contents of extrudable alloys of the 6xxx series (Al-Mg-Si) are optimized for extrusion purposes. Si and Mg build the intermetallic phase Mg_2Si that is soluble to 1.85% at the eutectic temperature. Solidification cracking appears in the solid-liquid coexistence zone whose temperature is in the range immediately above the solidus and occurs when solidifying weld metal undergoes tensile stresses during its solidification [48].

Whenever applicable (typically in arc welding of products above few mm thicknesses) the use of a filler metal can reduce the risk of weld cracking. Detailed tables are available suggesting the most suitable fillers for different alloys and their combinations. These tables rate the fillers as a function of ease of welding, strength of the joint, ductility, corrosion resistance, service at moderate or high temperatures, etc.

In addition to the risk of cracking, porosity is a common problem when welding aluminium alloys. Porosity arises from gas entrapment due to poor shielding, air, moisture, unclean wire or metal surface (hydrated oxides, oil, hydrocarbon contaminants), high cooling rates, etc. Hydrogen contamination is the cause of virtually all weld porosity in aluminium alloys, due to its high solubility in the molten pool and limited in solid aluminium.

For the specific case of pure Al, the role of minor elements, especially Na and Ca has been identified as cause of welding issues (cracks, voids and excessive porosity). The weld response of different pure Al heats containing different levels of minor elements [49] was compared. The role of Na in the crack sensitivity of Al alloys had already been identified by Ransley and Talbot [50]. Sodium is sometimes added to outgas pure aluminium in the melting phase [51]; Ca has a known interaction with hydrogen [52].

3.2 Compatibility of aluminium alloys with high-temperature service (baking, activation of non-evaporable getters)

Non heat-treatable alloys, such as EN AW-5083, one of the most common general-purpose Al alloys, can be supplied in a work hardened state, where they can reach significant strength. The same alloys in a fully annealed or as fabricated temper (designated as O and H111, respectively), show moderate to low tensile properties which might be of limited interest in several applications.

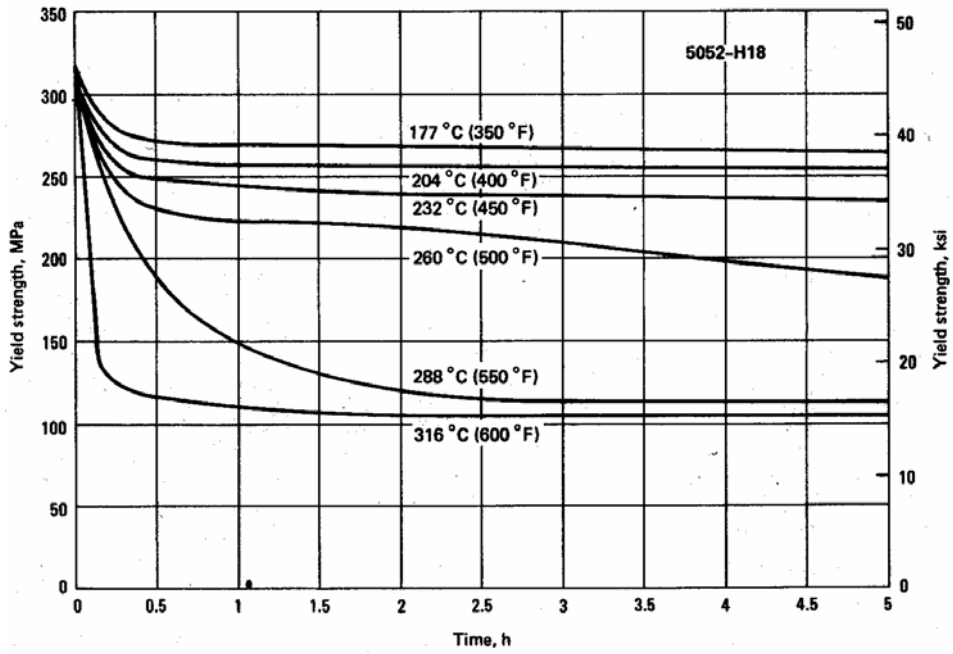


Fig. 9: Annealing curves of work hardened EN AW-5052 [10]. At a given temperature (curves for 177°C ≤ T ≤ 316°C are reported), increasing time at temperature results in a significant loss of the initial yield strength. The reported annealing temperature of this alloy is 343°C, corresponding to the temperature allowing for full annealing with no time at temperature required [53].

In the example of Fig. 9, the annealing curves of a typical wiredrawing alloy such as EN AW-5052 are presented. In a fully hard state designated as H18 temper [54] and achieved by about 75% cold reduction and temperature control during reduction (not to exceed 50°C), EN AW-5052 can reach yield strength above 300 MPa. The alloy is softened by the exposure at temperature for times of few hours.

If a time–temperature cycle has to be applied to an aluminium vacuum chamber for baking purposes or for activation of Non-Evaporable Getters (NEGs), the possible loss of properties induced by the cycle should be critically considered. Let us take as an example a cumulative effect of 30 activations during 24 h at 260°C. One-day cycles of activation at this temperature per year are applicable to TiZrV NEG coatings deposited on the internal walls of vacuum chambers for the intersection areas of the LHC. In the example shown in Fig. 9, a sudden loss of properties would already occur for EN AW-5052 H18 after the first hours of application of the heat cycle. The final properties will tend to the ones of a fully annealed state.

In addition to the properties at RT after exposure at high temperature, the properties at the temperature of activation or baking are relevant for the design of a component to be exposed to high temperature cycles. The total time of exposure at temperature should be taken into account. Creep considerations should not be forgotten in design, especially when dimensioning a chamber to resist buckling [55].

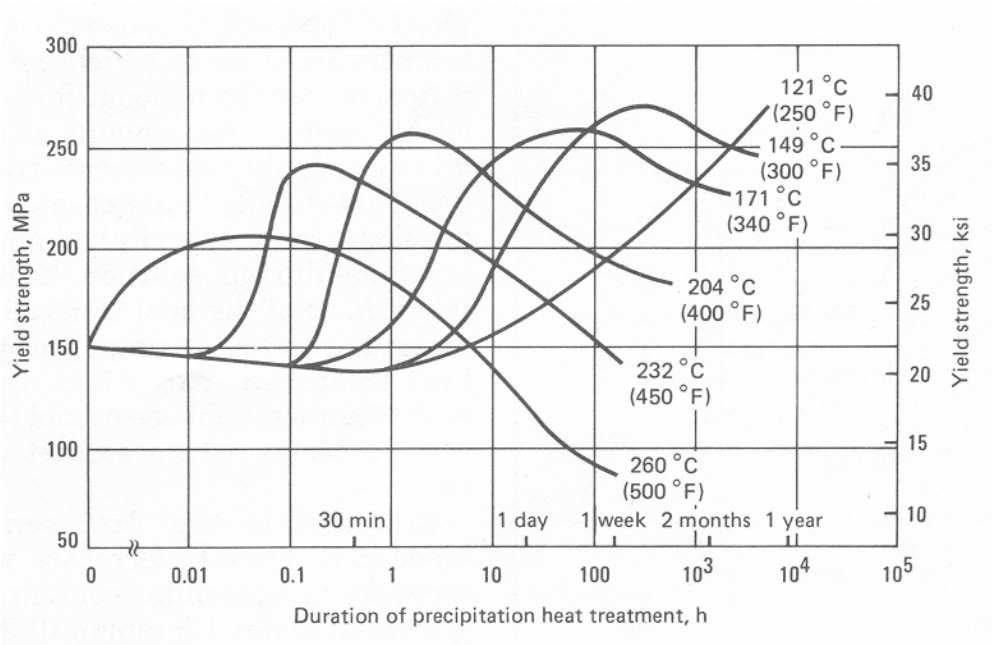


Fig. 10: Effect of time and temperature on the yield strength of EN AW-6061 [53]

Heat treatable alloys show a different response to a thermal cycle. In the example of Fig. 10, the heat treatable alloy EN AW-6061 is initially in a solution annealed and naturally (at RT) aged temper. Temperature cycles result in an initial increase of strength up to a maximum followed by a further decrease (overageing). At $T = 260°C$, a few days of exposure are sufficient to reduce strength to very poor values. The yield strength expected after 1000 h at 260°C is only 70 MPa [56], far under the initial 275 MPa of the product delivered in a peak-aged temper.

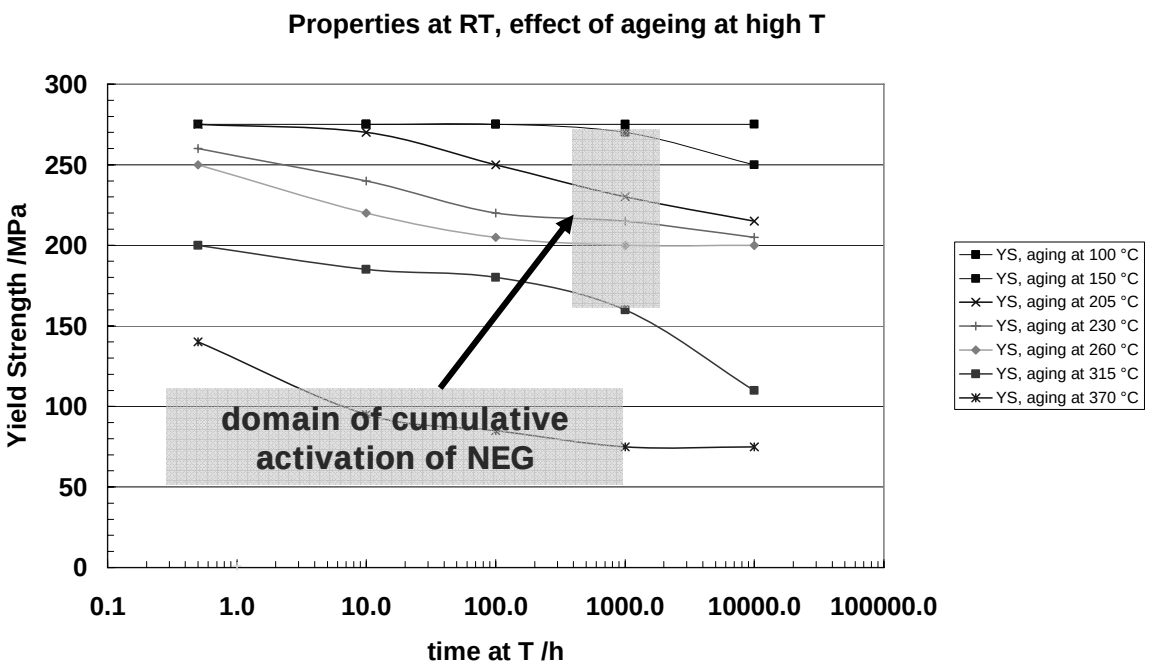


Fig. 11: Effect of time and temperature on the yield strength of EN AW-2219 (from data of Kaufman [56]). Activation of TiZrV NEG coatings typically occurs in a temperature range between 200°C and 300°C.

In summary, considerations of compatibility with heat treatment cycles severely restrict the selection of Al alloys for vacuum applications. The effect should be critically discussed in the phase of alloy selection. A weldable high strength alloy, EN AW-2219, developed for high temperature application by Alcoa in the 1960s, is a solution for applications where the loss of mechanical properties at high temperature is a concern. As can be seen from the plot of Fig. 11, starting from a proper temper state, this alloy maintains yield strength of approximately 200 MPa after 1000 h at 260°C.

3.3 Thin walled components: a failure analysis

EN AW-2219 bellows were machined from a forged rod. These bellows, to be integrated in one welded assembly (two flanges + two bellows + one tube), are intended for application to LHCb, one of the four major experiments at the LHC. One bellow revealed a leak, detected on a convolution and issued from the presence of corrosion pits on the inner surface of the convolution. SEM observations and EDX analyses were performed at the pitting corrosion location. They revealed the presence among other elements of Cl [57]. Cross-sectional observations (Fig. 12) confirmed the intragranular nature of the pits and showed a rough grain microstructure (Fig. 13), consisting of one or few grains in the wall thickness of the convolutions. The presence of Cl residues issued from the machining operations, performed with the help of Cerrobend contaminated by halogens from previous machining residues, was sufficient to provoke through thickness leaks [58].

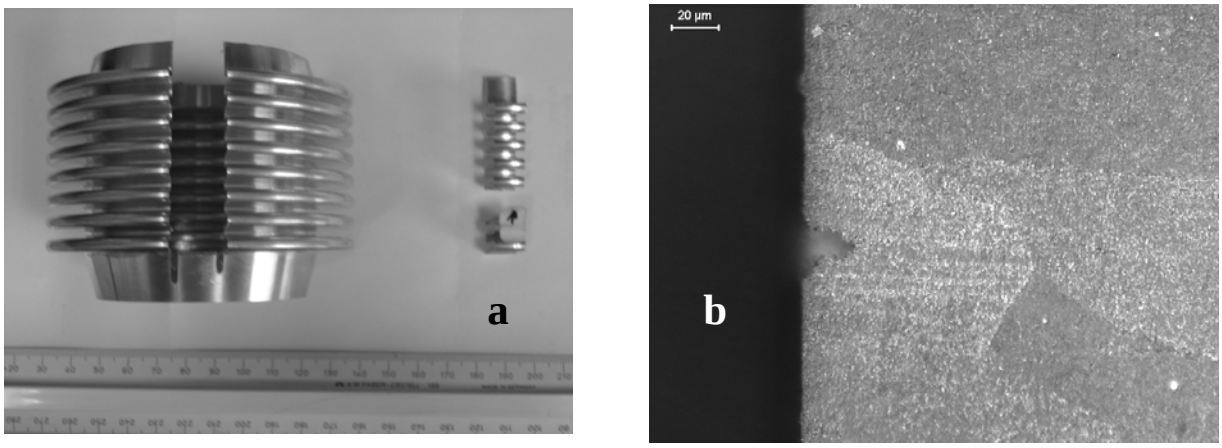


Fig. 12: Thin-walled (0.4 mm) bellow machined from a generic-purpose forged rod: a) a sample is removed from the bellow at the location of the leak; b) cross sectional observations

Hydraulic forming of seamless tubes to produce Al-alloy bellows, 0.3 mm thick, is also reported in the literature, as well as fabrication of bellows by welding and brazing. In the case of thin-walled products such as bellows and windows to be machined from bulk material, special three-dimensional forgings should be ordered with a very fine grain size. As an example, the starting material for the Velo Windows of the same LHCb experiment, including a thin-walled window and bellow, consisted of grains of a few micrometres in size. This microstructure could be achieved through free forging of heavy blocks of EN AW-6061 at controlled temperature. Non-destructive testing of the semifinished products at different stages of transformation was also applied to detect possible discontinuities at an early stage. Immersion or mechanized contact UT techniques allow submillimetre defects to be identified in due time.

The price of the special forgings for the Velo Windows was 18 CHF/kg, to be compared with the price of 5 CHF/kg to 8 CHF/kg of standard plates in 5xxx or 6xxx alloys.

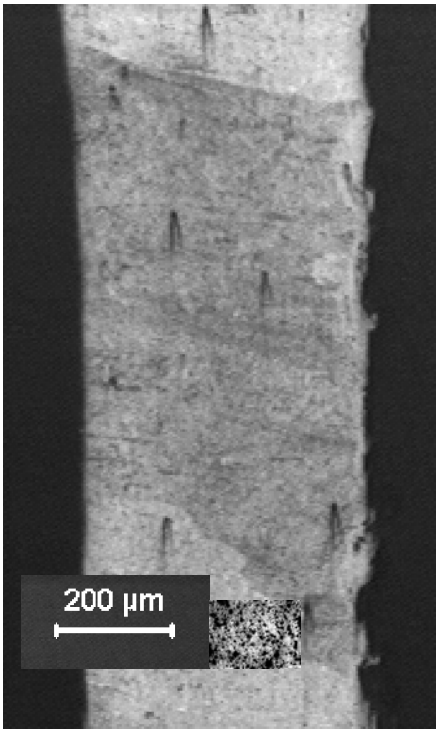


Fig. 13: Comparison at the same scale of the microstructure of the leaking bellow, showing one or few grains through the wall thickness of the convolution, and the fine-grained microstructure of a specially developed Al-alloy forging for the Velo Window of the same LHCb experiment

4 Copper and copper alloys

A relatively high corrosion resistance, machinability and formability into different shapes, ease of weld for several grades make copper and copper alloys very attractive for vacuum components, in particular for those applications where high electrical and/or thermal conductivity are relevant properties. Table 3 lists the distinct families of wrought copper and copper alloys according to their composition. Standard designation systems for coppers and copper alloys include the Unified Numbering System (UNS) [59], and a temper designation system [60].

Table 3: Classification of wrought coppers and copper alloys (from Ref. [61])

Generic name	UNS No.	Composition
Wrought alloys		
Coppers(a)	C10100–C15815	>99% Cu
High-copper alloys(b)	C16200–C19900	>96% Cu
Brasses	C20100–C28000	Cu-Zn
Leaded brasses	C31200–C38500	Cu-Zn-Pb
Tin brasses	C40400–C48600	Cu-Zn-Sn-Pb
Phosphor bronzes	C50100–C52480	Cu-Sn-P
Leaded phosphor bronzes	C53400–C54400	Cu-Sn-Pb-P
Copper-phosphorus and copper-silver-phosphorus alloys(c)	C55180–C55284	Cu-P-Ag
Aluminum bronzes	C60800–C64210	Cu-Al-Ni-Fe-Si-Sn
Silicon bronzes	C64700–C66100	Cu-Si-Sn
Other copper-zinc alloys	C66300–C69710	Cu-Zn-Mn-Fe-Sn-Al-Si-Co
Copper nickels	C70100–C72950	Cu-Ni-Fe
Nickel silvers	C73500–C79830	Cu-Ni-Zn

Relevant for vacuum applications are pure coppers and the so-called 'electrical coppers', belonging to the first group in Table 3 (coppers, C10100-C15815), high-strength copper alloys such as high and low Be Cu-Be alloys (C17200 and C17410, respectively), Cu-Ni alloys such as the non-magnetic '70–30 cupronickel' (C71500). On the other hand, alloys containing Zn, Pb, Cd, Se, S, etc. might result in unsuitable vapour pressure for vacuum applications [9].

4.1 Pure coppers

Oxygen-free copper (Cu OF, C10200, 99.95% min Cu) and oxygen-free electronic copper (Cu OFE, C10100, 99.99% min Cu) are high conductivity electrolytic coppers (in the annealed state, conductivity > 101% IACS at 20°C) with limits for oxygen and other impurities. For Cu OFE, there are specific limits for 17 elements [62]. These grades are typically used for the fabrication of RF cavities, bus bars, waveguides, vacuum seals, klystrons, sputtering targets, evaporation materials for thin film applications, etc. OFE grade is to be preferred for applications involving vacuum brazing or electron beam welding, and for cryogenic applications where a high Residual Resistivity Ratio (RRR) is required. Fine-grained products (CERN specifies a maximum grain size of 90 μm) are mandatory for vacuum applications, especially when thin-walled components are foreseen. The possibility to achieve a specified grain size depends not only on the sequence of the applied thermomechanical transformations, but also on the grain structure and texture of the material supplied for further forging or drawing (see Fig. 14). The supplied forged bars should be 100% ultrasonically inspected to detect possible continuity faults. The agreement between the supplier and the customer of an UT procedure is essential to define a control of the material quality in the frame of critical vacuum applications. As an example, CERN specifies a UT control at a frequency of 4 MHz, depending on the thickness. The homogeneity and the grain fineness of the bars are imposed through a criterion of 20% maximum allowed ultrasonic attenuation at any point in the product [63].

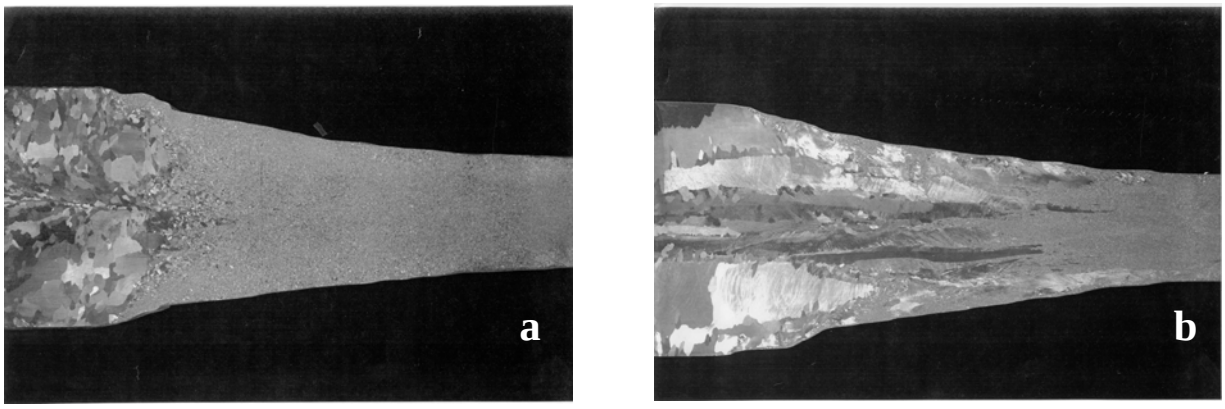


Fig. 14: Partially forged pure copper ingots from two different suppliers. Diameter of the incoming ingot, approx. 200 mm. The final grain size achieved by the same forging sequence applied to the two ingots strongly depends on the initial texture: a) fine-grained solidification structure, b) rough grain structure (courtesy of Stainless).

For ease of machining, half-hard tempers are preferred to fully soft conditions. OFE Cu at mid-2006 prices is supplied at a cost of 23–30 CHF/kg in drawn products, and 38 CHF/kg (3D forged). The base price of OF Cu is 10 CHF/kg. For comparison, the price of a general-purpose Electrolytic Tough Pitch (ETP) copper is approx. 8 CHF/kg (basis). ETP copper (99.90% min Cu, Ag is counted as Cu), because of the oxygen content, is subject to embrittlement when heated at 370°C or above in a reducing atmosphere, and might be critical for applications involving annealing, brazing, or welding.

4.2 Copper-silver and copper-zirconium alloys

As with many metals and alloys, electrical coppers, and in particular pure coppers, undergo thermal softening, a degradation of strength and hardness after exposure to elevated temperatures. Oxygen-free copper is available as silver-bearing copper having specific minimum Ag content. Silver increases resistance to softening, in particular in cold worked states, without substantially decreasing thermal and electrical conductivity at RT. Annealed oxygen-free silver coppers have an electric conductivity of 100% IACS at 20°C. Figure 15 shows the annealing curves of cold worked OF copper, and of different OF coppers containing increasing amounts of Ag [61]. For the same amount of initial cold work and annealing time (1/2h for the curves of Fig. 15), increasing amounts of Cu displace the annealing temperature toward high temperatures.

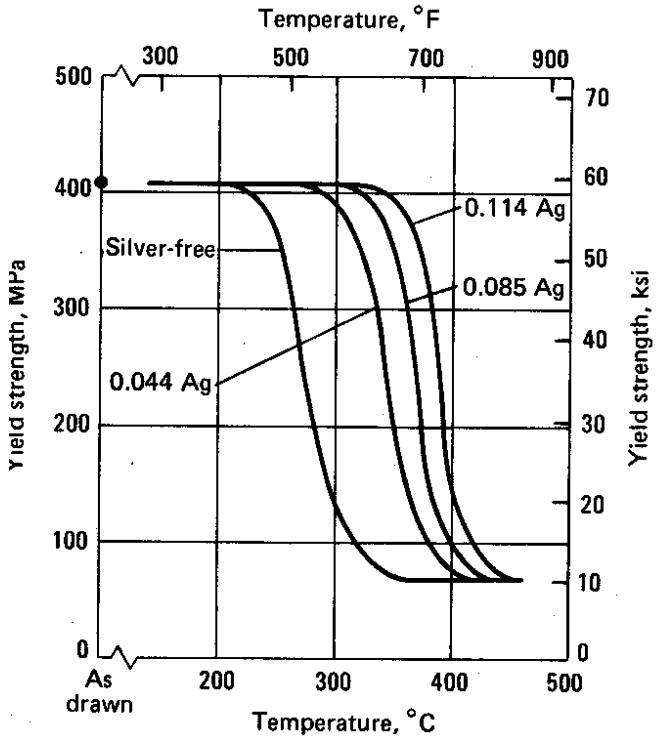


Fig. 15: Softening characteristics of OF Cu (cold worked 90%) containing various amount of Ag. Annealing time is 1/2h [61].

In addition, Ag improves the creep strength of pure copper. NEG coatings, deposited as thin magnetron sputtered TiZrV films on the internal wall of the several hundred drift chamber of the vacuum system of the Long Straight Sections (LSS) of the LHC, will be activated in a temperature range between 200°C and 300°C (see Section 3.2). In view of the repeated activations, alloy C10700 (99.95% min Cu + Ag, 0.085% min Ag, oxygen max: 0.0010%) has been selected for the chambers. A hard drawn state with specified maximum mechanical property limits has been preferred in order to maintain substantial strength after exposure to thermal cycles. Since increasing initial work hardening has the effect of decreasing, *ceteris paribus*, the annealing temperature of the alloy, maximum allowed initial hardness was limited to 90 HRF.

An alloying element also capable of increasing the annealing temperature of pure Cu is Zr. The resistivity increase of Zr per 1% addition is 80 nΩ·m, much higher than Ag (3.6 nΩ·m). A typical value of conductivity of CuZr alloys ($0.10\% \leq Zr \leq 0.20\%$) is 93–95 IACS. Zirconium coppers are heat treatable. The alloys show improved fatigue strength and retain much of their room-temperature strength up to 450°C. In a proper temper state, the alloy C15000 (99.85Cu – 0.15Zr) is a candidate for the fabrication of the matrix of the cavities of the CLIC accelerator (see Section 5.2).

4.3 Oxide Dispersion Strengthened (ODS) coppers

Pure copper can be strengthened by a small amount of finely dispersed oxides such as alumina in the matrix. The dispersoids are stable at elevated temperatures, up to the melting point of the matrix, and prevent recrystallisation and softening of the material when exposed to high temperatures [64]. ODS coppers, commercially known as GlidCop® (trademark of SCM Metal Products, Inc., manufacturer of metal powders and pastes for PM), are based on a Cu-Al_2O_3 system.

Since the oxides are immiscible in liquid Cu, PM techniques are applied followed by a consolidation passing through conventional thermomechanical steps or HIP. Several techniques are applied to achieve a fine and uniform dispersion of oxides, such as selective internal oxidation [65], mechanical mixing [66], coprecipitation from salt solutions [67]. Starting from a copper-aluminium solid solution, aluminium is converted into alumina during the process.

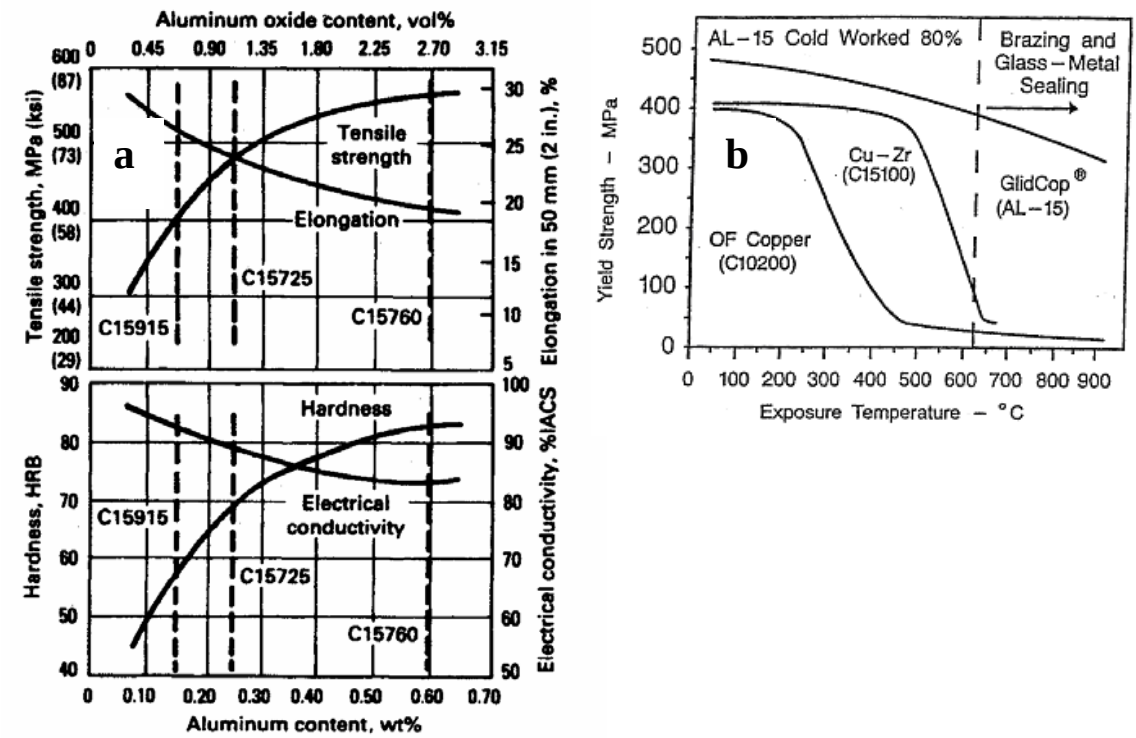


Fig. 16: a) Properties of ODS coppers (rod stock in hot extruded condition) [68]; b) Softening resistance of cold worked Glidcop® AL-15 compared to OF copper and copper-zirconium. Glidcop® AL-15 is reinforced with 0.15% of Al as Al_2O_3 and roughly corresponds to grade C 15725 of Fig. 16(a). Annealing time is 1/2h [69].

Figure 16(a) [68] shows the effect of increasing aluminium oxide contents on strength, hardness, ductility and conductivity of ODS coppers. Figure 16 (b) [69] shows the significant improvement of the resistance against high-temperature softening conferred by addition of 0.25% of Al under the form of Al_2O_3 to the alloy C15725. Compared to alloys based on additions of soluble elements, the improved resistance to recrystallization and softening is particularly relevant above 600°C (vacuum brazing temperature range). This is due to the inertness of the dispersoids compared to the easily dissolved precipitates in conventional precipitation hardened alloys such as CuZr or CuBe, based on diffusion controlled systems which are thermally unstable [64].

Alumina reinforcement particles are of a typical size ranging from few nm to few tens of nm [Fig. 17(a)]. Though of different shape, their size is comparable to that of the strengthening phases of the highest performing copper alloys such as CuBe [Fig. 17(b)].

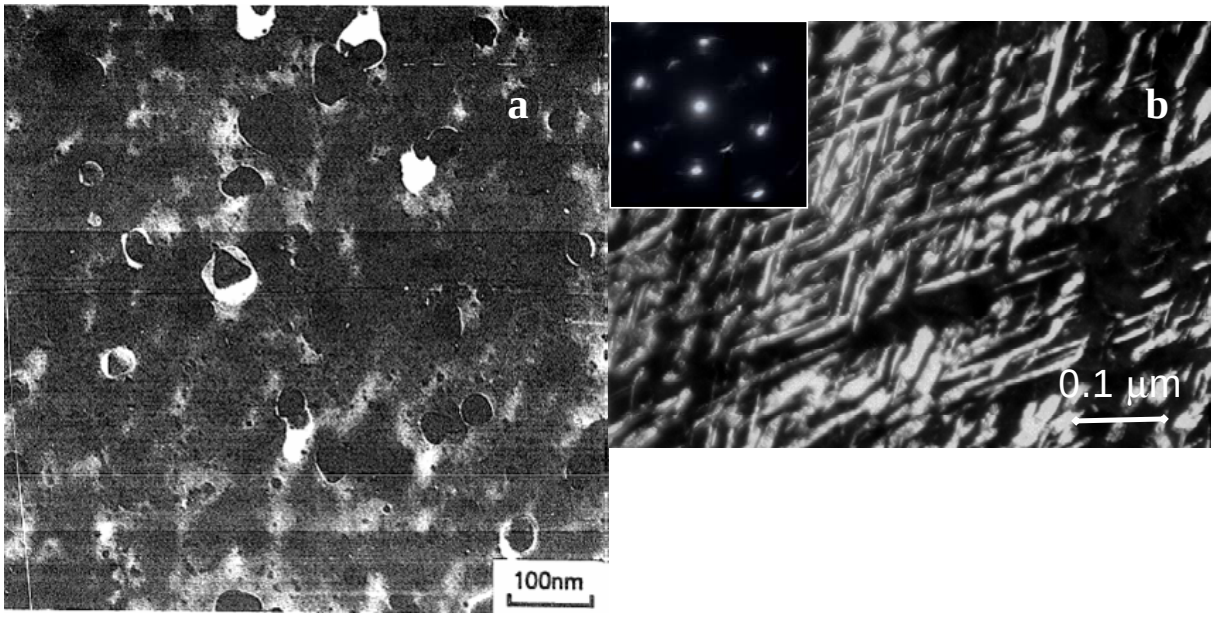


Fig. 17: a) Alumina dispersed precipitates [70]. b) At approximately the same scale, image of the reinforcing phase of a CuBe alloy C17670; dark field Transmission Electron Microscopy (TEM) view of γ' precipitates for a solution-annealed and artificially-aged (360°C – 6h) temper [71].

GlidCop® can be brazed with silver braze alloys, but requires a previous electroplating with Ni or Cu of the surfaces to be brazed. Gold-copper braze alloys of unplated GlidCop® have been successfully used and are reported in the literature [64]. EB welding tests have also been performed at CERN with successful results [72]. GlidCop® is largely used in modern accelerator components such as the crotches of the ESRF [73], exposed to intense bending magnet radiation, and in the collimators of the LHC. A typical current price for GlidCop® is $ 40/kg.

5 Special technologies for present and future accelerators

5.1 Near-net shaping of the covers of the LHC dipole magnets [74]

The 1232 superconducting dipole magnets of the LHC will operate at 1.9 K. The cold mass of the dipole magnets is enclosed by a shrinking cylinder and two end covers at each extremity of the cylinder. The end covers are domed and equipped with a number of protruding nozzles for the passage of the different cryogenic lines (Fig. 18). The covers are structural components that must retain high strength and toughness at cryogenic temperature. They are MIG welded onto the magnet shrinking cylinders. The protruding nozzles of the covers are welded to the interconnection pipes by an automatic orbital autogeneous TIG technique. Several thousand welds are foreseen. AISI 316LN has been selected because of its mechanical properties, ductility, stability of the austenitic phase against low-temperature spontaneous martensitic transformation. Owing to the complex geometry of the end covers, a PM + HIP technique has been selected for the fabrication of the covers. PM is an attractive near-net-shaping technique, allowing the final shape to be approached and the machining to be reduced to a minimum. The covers have been produced from atomized powders of the relevant steel grade, blended, homogenized and filled into capsules with geometry approaching the cover shape. After evacuation and sealing, a HIPing cycle has been performed. HIPing consists of a time–temperature–pressure cycle performed at 1180°C during 3 h under a stress of 100 MPa, allowing a fully dense structure to be achieved. The capsulated covers have been solution annealed, water quenched, pickled to remove the capsules, ground and machined to the final dimensions. Finally,

100% dye penetrant, visual and UT inspection (to measure the wall thickness and detect possible defects) have been performed on the finished covers, showing no relevant defects and full soundness and compactness of the components. Closed or open die forging would require significantly more machining, a welded product would need extensive inspections and stress relieving, whilst a cast solution would have poorer mechanical properties. PM was demonstrated as a technique fully adapted to the fabrication of complex-shaped components such as LHC end covers, working in a severe cryogenic environment and which have to be leak tight to superfluid helium. This is demonstrated by the excellent behaviour of dipole magnets equipped with PM covers, which have performed satisfactorily for several years in the so-called String 2 experiment. This near-net shaping technique, finally retained for the series production, was also selected on the basis of its price competitiveness.

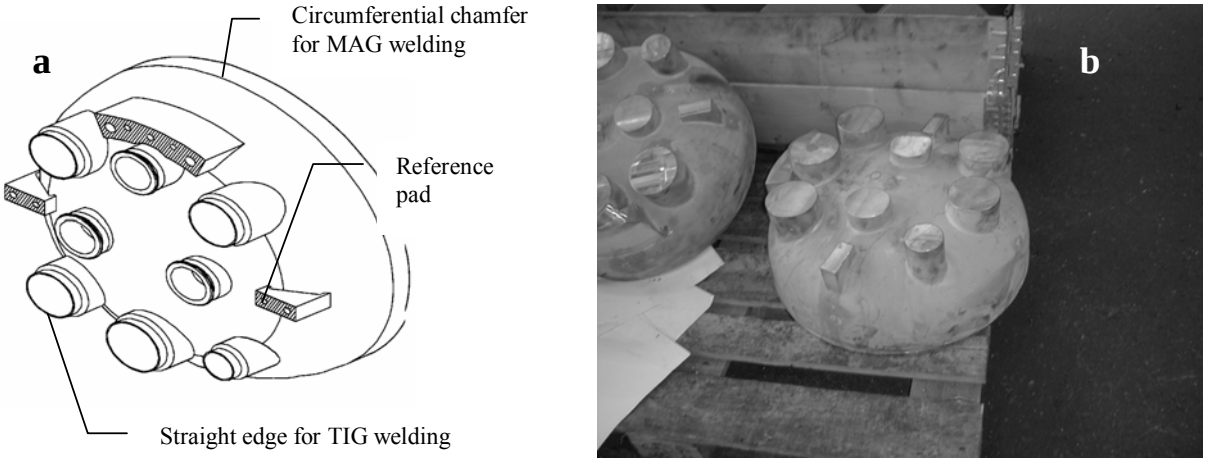


Fig. 18: a) 3-D view of an end cover. The diameter of the cover is 570 mm, the height 224 mm, and the overall weight 100 kg. The extremities of the nozzle to be welded on the interconnection pipes have an external (internal) diameter of 84 (80) mm. The main circumference to be welded to the shrinking cylinder has a thickness of 10 mm. The most severe dimensional tolerances are of the order of 0.05 mm. b) Near-net shaped covers, after capsule removal by pickling and heat treatment, before final machining.

5.2 Fabrication of bimetallic structures for CLIC by HIP-assisted diffusion bonding [75]

The accelerating structures of CLIC (acronym for a Compact LInear Collider) will withstand 10^{11} RF impulsions at 30 GHz and surface electrical fields up to 300 MV/m. Classical materials for the RF cavities such as pure Cu have shown their limits, and special technologies had to be developed for their fabrication. The structures have to be produced within tolerances of 1 μm, with materials that have to withstand extremely high surface fields, sparks, and thermal fatigue. Bimetallic structures consisting of a Mo insert and a CuZr matrix have been produced starting from a bimetal assembled through HIP-assisted diffusion bonding aimed at the production of prototype accelerating structures. Mo is foreseen for the irises, withstanding high fields and sparks; CuZr (C15000) is an electrical copper alloy (Section 4.2) showing, in this family, higher fatigue properties than pure Cu in a proper temper state (Fig. 19). HIP-assisted diffusion bonding might be a valid alternative to the assembly by brazing of premachined Mo inserts. Vacuum brazing, on account of the low cooling rates from the brazing temperature, is unfavourable to achieve a solution-annealed state of the heat treatable alloy CuZr. This alloy shows higher fatigue properties in a proper solution-annealed and artificially aged state, possibly with added cold working steps. Alternative techniques for the fabrication of the bimetal are explosion bonding (compatible with the couple CuZr-Mo) or coextrusion. On the other hand, casting a CuZr alloy around a Mo matrix might result in a porous and inhomogeneous copper alloy structure. First HIPed and explosion bonded bimetals have been tested in terms of mechanical

properties, adherence of the interface, and aptitude to machining. In Fig. 19, a prototype component machined from the HIPed bimetal is shown.

High purity (99.97%) Mo has a price of 1800 CHF/kg. Bimetallic prototypes have been produced by HIP diffusion bonding at a cost of 40 000 CHF/m for an external diameter of the bar of 85 mm.

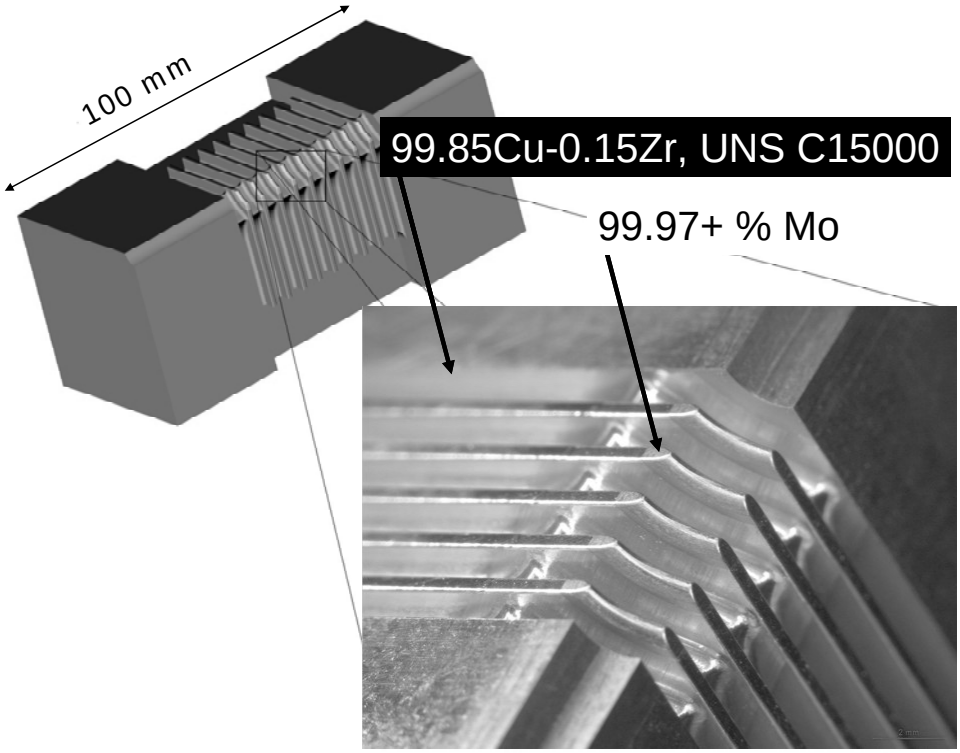


Fig. 19: 3-D view of a quadrant of a Hybrid Damped Structure (HDS) of CLIC for the main beam acceleration. Four quadrants will be assembled to produce a HDS. The picture of a HDS quadrant machined from the HIP diffusion bonded bimetallic rod is shown on the right.

6 Conclusions

The description of several applications of conventional and special metals and alloys for high vacuum purposes shows that a material consists not only of a 'chemical composition' or a designation, but is the result of a complete metallurgy and metalworking process including possible refinement and remelting steps. A material for a demanding application is also defined through specified non-destructive and destructive tests, and should be ordered and delivered with a complete certification. In some cases, a quality control plan should be defined between the producer and the customer and should be followed during the production. The cost to be foreseen for a material adapted to an envisaged application will depend on an adequate specification defined before the order. Taking as an example austenitic stainless steels, a general-purpose 304 L, issued from a simple electric furnace primary melting, can have a cost that is 25 times lower than an ESR product multidirectionally forged of a particular final size, properly specified for a vacuum application.

The severity of the specification will depend on the final application. Whenever leak tightness has to be guaranteed across a thin wall (see the example of the Velo windows, of the LHC, and experimental area interconnection components), an adequate microstructure in terms of grain size and

inclusion content has to be specified. A fabrication resulting in a proper texture (orientation of the fibres parallel to the walls, or if this is not achievable, absence of orientation through multidirectional forging) should be preferred.

In modern accelerator construction, several components have to be integrated in a vacuum system through delicate welding operations, introducing internal stresses that can bring to leak a previously tested leak-tight component.

Applications to cryogenic temperature, possibly associated with a requirement of non-magnetism, demand special care. A stainless steel such as 316LN, specified as fully austenitic in order to be non-magnetic at RT will show, at cryogenic temperature, a higher magnetic susceptibility that might be unacceptable for some components. For the beam screen of the LHC, a special austenitic stainless steel had to be developed to maintain low magnetic susceptibility in the base metal and the welds, in the temperature range between 10 K and 20 K. Strength and ductility requirements should not only be considered for the working temperature of a component, but defined as a function of all the temperature cycles undergone by the vacuum component (NEG activation, baking, etc.).

Proper selection of materials should take into account their availability (e.g., EN AW-2219 is scarcely available, special grades of stainless steels might require more than one year of lead time for delivery).

Corrosion issues should not to be neglected, especially for thin-walled components (example of the LHCb bellows). The use of halogen-activated fluxes should be strictly avoided in SS environment [76].

Material aspects should always be considered at the beginning of a project: issues of availability, interest of considering in due time alternative techniques, including near-net shaping, often make late selections more costly or incompatible with the tight schedule of an accelerator project. New projects will eventually result in less conservative solutions (bimetals by HIP-assisted diffusion bonding for the CLIC project, joints by explosion bonding being studied for both applications to CLIC and to the International Linear Collider).

Acknowledgements

The author wishes to thank N. Hilleret for his invitation to contribute to the special CERN Accelerator School, S. Mathot for helpful discussion on the material brazing topic, R. Veness for reading the manuscript, C. Saint-Jal for providing updated material prices.

References

[1] W. Espe and M. Knoll, Werkstoffkunde der Hochvakuumtechnik (Jul Springer, Berlin, 1936).

[2] W. Espe, Materials of High Vacuum Technology (Pergamon Press, Oxford), Vol. 1 (Metals), 1966; Vol. 2 (Silicates), 1968; Vol. 3 (Auxiliary Materials), 1968.

[3] W.H. Kohl, Handbook of Materials and Techniques for Vacuum Devices (Reinhold, New York, 1967).

[4] W.H. Kohl, Materials and Techniques for Electron Tubes (Reinhold, New York, 1960).

[5] C.J. Phillips, Glass, The Miracle Maker, 2nd ed. (Pitman, New York, 1948).

[6] H. Steinherz, Handbook of High Vacuum Engineering (Reinhold, New York, 1963).

[7] J. Yarwood, High Vacuum Technique: Theory, Practice and Properties of Materials, 4th ed. (London, Chapman & Hall, 1967).

[8] K. Diels, Werkstoffe und Werkstoffverbindungen in der Vakuum-Technik (Classen, Essen, 1968).

[9] J.F. O'Hanlon, A User's Guide to Vacuum Technology, 3rd ed. (Wiley, Hoboken, 2003).

[10] Metals Handbook Ninth Edition, Vol. 3, Properties and Selection: Stainless Steels, Tool Materials and Special-Purpose Metals (American Society for Metals, Metals Park, OH, 1980).

[11] S. Sgobba and G. Hochoertler, A new non-magnetic stainless steel for very low temperature applications, Proceedings of the International Congress Stainless Steel 1999: Science and Market, Chia Laguna, Italy, 6–9 June 1999, Vol. 2, pp. 391–401.

[12] C. Garion, B. Skoczen and S. Sgobba, Constitutive modelling and identification of parameters of the plastic strain-induced martensitic transformation in 316L stainless steel at cryogenic temperatures, Int. J. Plasticity **22** (2006) 1234–1264.

[13] D. Peckner and I.M. Bernstein, Handbook of Stainless Steels (McGraw–Hill, New York, 1977).

[14] P. Lacombe, B. Baroux and G. Béranger, Les aciers inoxydables (Les Editions de Physique, Les Ulis, 1990).

[15] Metals Handbook, Vol. 8, p. 291, 8th ed. (American Society for Metals, Metals Park, OH, 1973).

[16] J.W. Pugh and J.D. Nisbet, Trans. Am. Inst. Min. Metall. Petr. Eng. **188** (1950) 273.

[17] CERN technical specification No. 1003 - Ed. 3 - 02.08.1999, Round bar for vacuum applications, stainless steel type X2CrNi19-11 (1.4306, AISI 304L).

[18] CERN technical specification No. 1004 - Ed. 3 - 02.08.1999, Sheet steel for vacuum applications, stainless steel type X2CrNi19-11 (1.4306, AISI 304L).

[19] CERN technical specification No. 525 - Ed. 3 - 02.08.1999, Sheets and tubing for special cryogenic applications, stainless steel type X2CrNiMo17-12-2 (1.4404, AISI 316L).

[20] CERN technical specification No. 1000 - Ed. 3 - 02.08.1999, Forged round bar for vacuum applications, stainless steel type X2CrNiMoN17-13-3 (1.4429, AISI 316LN).

[21] CERN technical specification No. 1001 - Ed. 3 - 02.08.1999, Forged blanks for vacuum applications, stainless steel type X2CrNiMoN17-13-3 (1.4429, AISI 316LN).

[22] CERN technical specification No. 1002 - Ed. 3 - 02.08.1999, Sheet steel for vacuum applications, stainless steel type X2CrNiMoN17-13-3 (1.4429, AISI 316LN).

[23] EN 10088-2:2005 Stainless steels. Technical delivery conditions for sheet/plate and strip of corrosion resisting steels for general and construction purposes.

[24] C.J. Guntner and R.P. Reed, The effect of experimental variables including the martensitic transformation on the low temperature mechanical properties of austenitic stainless steels, Trans. ASM **55** (1962) 399.

[25] K. Couturier and S. Sgobba, Phase stability of high manganese austenitic steels for cryogenic applications, Proceedings of the 'Materials Week' European Conference, Munich, Germany, September 25–28 2000, http://www.proceedings.materialsweek.org/proceed/mw2000_432.pdf.

[26] A. Gérardin, Metallographic observations of 316LN leaking bellow, TS-MME-MM report no. EDMS 710706, 10/03/2006.

[27] ASTM E45-97ε2 (2002) Standard Test Methods for Determining the Inclusion Content of Steel.

[28] A. Gérardin, Metallurgical investigation of leaking Plug-In Module (PIM) QBBIB21R8, TS-MME-MM report no. EDMS 718986, 18/04/2006.

[29] A. Choudhury, Vacuum Metallurgy (ASM International, Metals Park, OH, 1990).

[30] ISO 15614-1:2004, Specification and qualification of welding procedures for metallic materials, Welding procedure test, Part 1: Arc and gas welding of steels and arc welding of nickel and nickel alloys.

[31] ISO 15614-11:2002, Specification and qualification of welding procedures for metallic materials, Welding procedure test, Part 11: Electron and laser beam welding.

[32] ISO 6520-1:1998, Welding and allied processes, Classification of geometric imperfections in metallic materials, Part 1: Fusion welding.

[33] ISO 5817:2003, Welding, Fusion-welded joints in steel, nickel, titanium and their alloys (beam welding excluded), Quality levels for imperfections.

[34] ISO 13919-1:1996, Welding, Electron and laser-beam welded joints, Guidance on quality levels for imperfections, Part 1: Steel.

[35] W.T. DeLong, Ferrite in austenitic stainless steel weld metal, Weld. J. **53**, Res. Suppl. (1974) 273s–286s.

[36] E. Folkhard, Welding Metallurgy of Stainless Steels (Springer, New York, 1988).

[37] N Suutala, Solidification studies on austenitic stainless steels, Acta Universitatis Ouluensis, Series C Technica, **23,** University of Oulu, Oulu, Finland, 1982.

[38] T. Takalo, N. Suutala and T. Moisio, Austenitic solidification mode in austenitic stainless steel welds, Metall. Trans. **10A** (1979) 1173–1181.

[39] D. L. Olson, Prediction of austenitic weld metal microstructure and properties, Weld. J. **64** (1985) 181–295.

[40] Ö Hammar and U. Svensson, Influence of steel composition on segregation and microstructure during solidification of austenitic stainless steels *in*: Solidification and Casting of Metals (The Metals Society, London, 1979), pp. 401–410.

[41] N. C. Anderson, Lectures in Ceramic Metal Joining, Amsterdam, 1996.

[42] Coors Porcelain Co., Golden, Colo., Tech. Data Bulletin 951.

[43] H. Vatter, Ceramic to metal brazes (in German), Glas und Hochvakuum Technik, **1** (1952) 79–85.

[44] H. Ishimaru, Developments and applications for all-aluminum alloy vacuum systems, MRS Bulletin **15** (1990) 23.

[45] H. Ishimaru, J. Vac. Sci. Technol. **15** (1978) 1853.

[46] Merkblatt DVS 3204, Elektronenstrahl-Schweiβeignung von metallischen Werkstoffe, Deutscher Verband für Schweisstechnik E.V., 1988.

[47] R. Crausaz, Cooling circuit of Atlas Pixel: qualification of edge laser welds between tubes and fittings, TS-MME-MM report n° EDMS 693669, 07/01/06.

[48] R. Liu, Z.J Dong and Y.M. Pan, Solidification crack susceptibility of aluminum alloy weld metals, Trans. Nonferrous Met. Soc. China **16** (2006) 110–116.

[49] Schelde Exotech report n° M97/044 Rev.0, Supplementary examination of weldments, Urenco Capenhurst, UK, 2003.

[50] C.E. Ransley and D.E.J. Talbot, J. Inst. Metals **88** (1959) 150.

[51] L.I. Sokol'skaya, Gases in Light Metals (Pergamon, Oxford, 1961).

[52] H. Schmitt and H. Wittner, Erzmetall **9** (1956) 417–421.

[53] Metals Handbook Ninth Edition, Vol. 2, Properties and Selection: Nonferrous Alloys and Pure Metals (American Society for Metals, Metals Park, OH, 1979).

[54] ANS H35.1/H35.1M (2006), Alloy and Temper Designation Systems for Aluminum.

[55] L. Kolla and E. Dulacska, Buckling of Shells for Engineering (Wiley, New York, 1984).

[56] J.G. Kaufman, Properties of Aluminum Alloys: Tensile, Creep, and Fatigue Data at High and Low Temperatures (ASM International, Materials Park, OH, 1999).

[57] A. Gérardin, Micro-optical investigation of a leak on Al welded assembly, TS/MME/MM report no. EDMS 681631, 29/11/2005.

[58] A. Gérardin, Microstructural analysis of EN AW 2219 forged blocks, TS/MME/MM report no. EDMS 696100, 25/01/2006.

[59] Metals and Alloys in the Unified Numbering System, Tenth Edition (SAE International, 2004).

[60] ASTM B601-02, Standard Classification for Temper Designations for Copper and Copper Alloys—Wrought and Cast.

[61] ASM Specialty Handbook, Copper and Copper Alloys, J.R. Davis ed. (ASM International, Materials Park, OH, 2001).

[62] ASTM B170-99(2004), Standard Specification for Oxygen-Free Electrolytic Copper—Refinery Shapes.

[63] CERN technical specification no. 2001 - Ed. 3 - 02.08.1999, Bar Copper OFE.

[64] P.K. Samal, Brazing and Diffusion Bonding of Glidcop®, technical information of SCM Metal Product Inc., Research Triangle Park, NC.

[65] Benjamin, US Pat. 3785801, 1974.

[66] K.M. Zwilski and N.J. Grant, Trans. AIME **221** (1961) 371–377.

[67] R.L. Crosby and D.H. Desy, U.S. Bureau of Mines Report 7266, 1969, pp. 1–17.

[68] Nadkarni, Mechanical Properties of Metallic Components, 1993.

[69] North American Höganäs, Inc., tech. data sheet no. 700A, rev. 2003.

[70] F. Zhu, L. Jiao, N. Wanderka, R.P. Wahi and H. Wollenberger, FIM atom probe study of an Al_2O_3 dispersion strengthened copper alloy, Surface Science **266** (1992) 337–341.

[71] S. Sgobba, PhD thesis EPFL no. 1215, 1994.

[72] CERN EST-SM-MB report no. 97/04/20, 1997.

[73] R. Kersevan, personal communication.

[74] S. Sgobba, F. Savary, J. Liimatainen and M. Kumpula, A powder metallurgy austenitic stainless steel for application at very low temperatures: Proceedings of the 2000 Powder Metallurgy World Congress, Kyoto, Japan, 2000, Vol. 2, pp. 1002–1005.

[75] G. Arnau Izquierdo, M. Taborelli, S. Calatroni, P. El-Kallassi, S. Sgobba, S. Perouse de Montclos, Activités MME pour CLIC, caractérisation de matériaux non conventionnels et leur mise en œuvre, Proceedings of the CERN-TS Workshop 2005, 24–26 May 2005, Archamps (F) http://indico.cern.ch/getFile.py/access?contribId=39&sessionId=6&resId=2&materialId=paper&confId=047 .

[76] S. Sgobba, J.P. Tock and G. Vandoni, Functional Specification, Soldering flux applied to the main bus-bar cables of the LHC, CERN EDMS document no. 334543.

Ultrahigh vacuum gauges

K. Jousten
Physikalisch-Technische Bundesanstalt, 10587 Berlin, Germany

Abstract
Ionization gauges exclusively are used for ultrahigh vacuum. After a brief history, the design, use, and accuracy of ionization gauges will be described in this article.

1 Introduction

In the ultrahigh vacuum (UHV) regime it is not possible to measure pressure as a force on a certain area as the definition of pressure indicates. Instead, it turns out that the only practical and economically reasonable indicator for pressure in UHV is the ionization rate produced by electrons hitting the neutral gas atoms in a UHV chamber (Fig. 1).

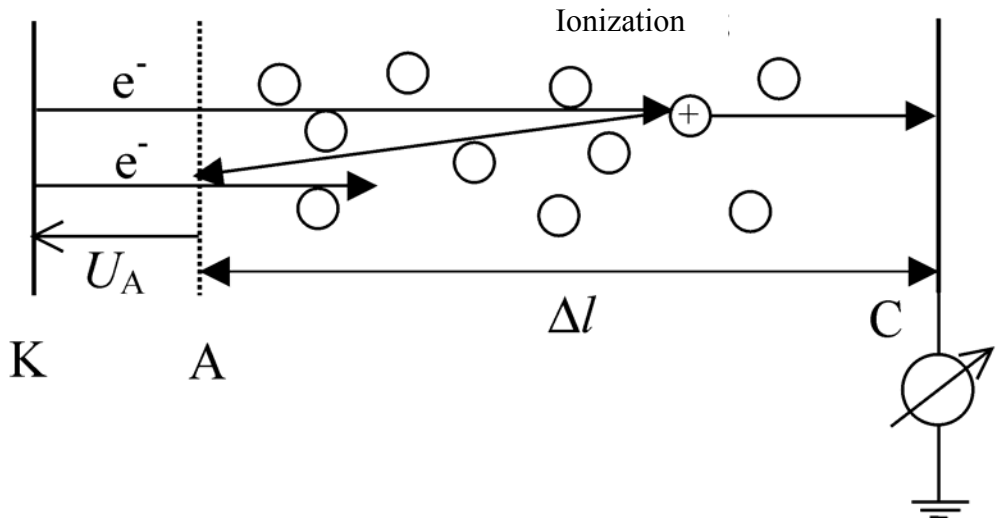


Fig. 1: The basic measuring principle of ionization gauges with electron emitting cathode K. Electrons hitting neutral molecules closely enough may ionize them. The ions are drawn to the collector C, the electrons finally reach the anode A.

In such ionization gauges (IG) the ionization rate is proportional to the particle density n in the gauge volume. 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; the temperature T of the gas also has to be known to indicate pressure with an IG.

Though, in principle, it is also possible to ionize neutral gas molecules by photons (high-intensity lasers) or ions, only the use of electrons is economically feasible. The production method of electrons, however, has generated two main types of ionization gauges: when the electrons are

generated by an electrical discharge, the gauges are usually called ‘cold cathode gauges’, and when the electrons are generated by a heated cathode, they are called ‘hot cathode ion gauges’.

In this article as in newer text books we prefer to call gauges where the electrons are produced by a discharge, crossed field gauges, and where an electron-emitting (i.e., hot) cathode is used, ionization gauges with emitting cathodes. The reason is that nowadays cold emitting cathodes also exist and may in the future come into practical use.

The left branch of Fig. 2 in the article “Gauges for fine and high vacuum” in the mentioned book gives an overview of the classification of the most well-known types of ionization gauges according to their measurement scheme, which will be explained step by step further on in this article.

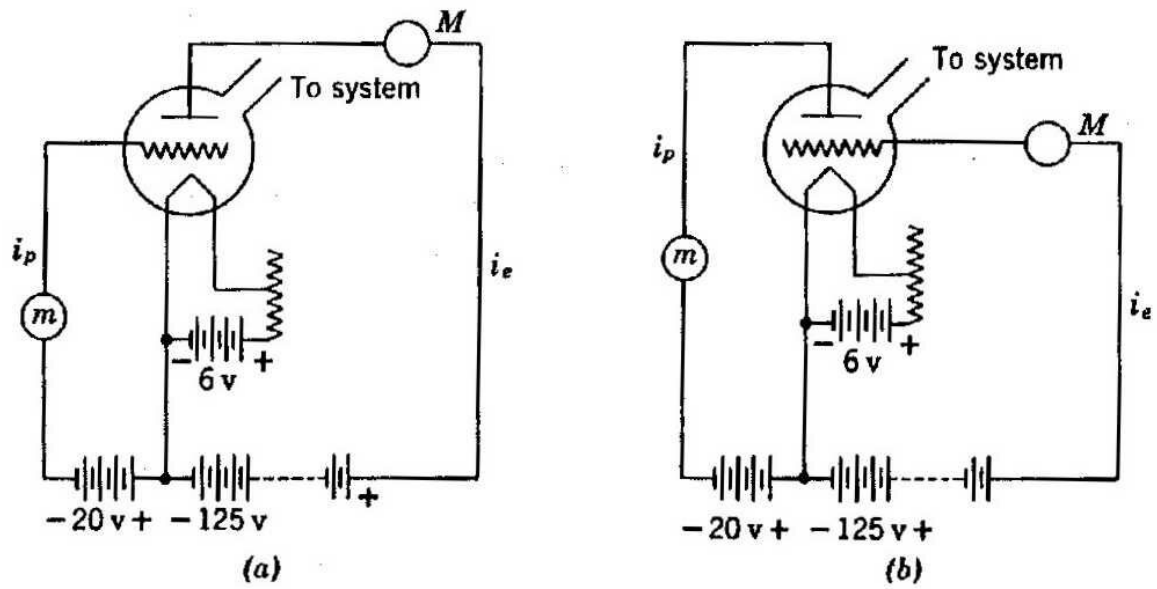


Fig. 2: Electrical circuits for historical triode ionization gauges: (a) Internal control type. (b) External control type. From Saul Dushman and J.M. Lafferty, *Scientific Foundations of Vacuum Technique*, 2nd ed., John Wiley & Sons, New York, 1962.

2 Brief historical review

The history of the IG dates back to 1909, when Baeyer [1] showed that a triode vacuum tube could be used as a vacuum gauge. However, Buckley [2] is usually recognized as the inventor of the triode gauge. He 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, Fig. 2(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 V. The reason is that the ionization 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 on Fig. 3.

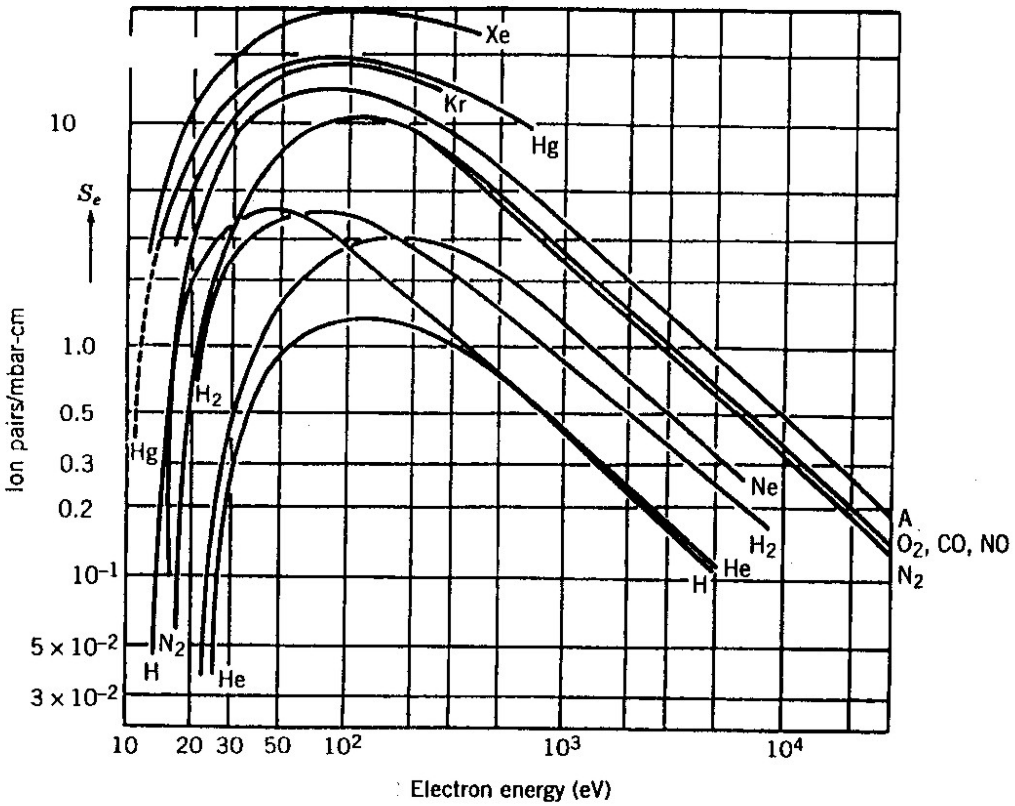


Fig. 3: Generated ions per centimetre electron path length per millibar at 20°C versus kinetic energy of incident electrons for various gases. From A. von Engel, *Ionized Gases*, AVS Classics Series.

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

The basic design of the triode gauge remained unchanged for more than 30 years, although 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.

At the 1st International Vacuum Congress (IVC) 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 [3] 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.

Penning is known as the inventor of the first crossed field gauge [4]. One of the crossed-field gauge types is named after him. His invention was based on an earlier patent by himself for coating by sputtering. It turned out that the discharge current was almost linearly proportional to the pressure in the gauge from 1 mPa to 0.1 Pa.

Figure 4 shows the electrode arrangement, fields, and trajectories in the Penning gauge 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.

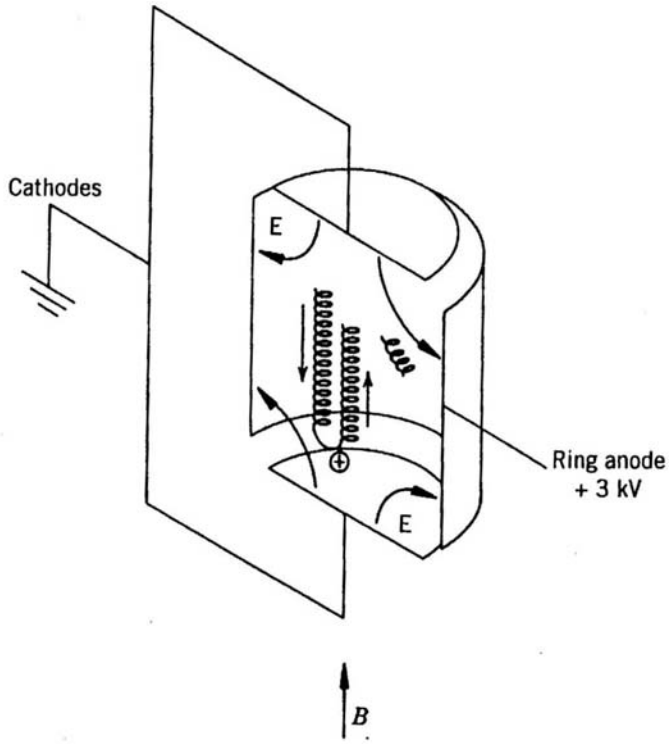


Fig. 4: Electrode arrangement, fields, and trajectories in the Penning gauge. From James M. Lafferty, *Foundations of Vacuum Science and Technology*, John Wiley & Sons, New York, 1998.

3 Crossed field gauges

3.1 The Penning gauge

The working principle of this type of gauge is to generate a discharge between two metal electrodes (anode and cathode) by applying a DC high voltage. The discharge current is pressure dependent and serves as measurand for pressure. The lower measurement limit lies around 1 Pa, since at lower pressures the gas density is too low to generate enough charge carriers to maintain the discharge.

To extend this limit, a magnetic field crossing the electrical field is used. This magnetic field greatly increases the path length of the electrons from cathode to anode, so that the electron can generate another electron by impacting on a gas molecule to maintain the discharge (Penning discharge). Owing to their higher mass the ions are only slightly affected in their trajectories by the magnetic field and travel directly to the cathode. Secondary electrons released when the ions hit the cathode (cathode sputtering) support the discharge.

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

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

where m depends on the type of gauge and varies between $m = 1 \ldots 1.4$.

The mode of operation in a Penning discharge is explained by Fig. 5 and Fig. 6. The case *G* of the Penning gauge is metal and on ground potential. Figure 7 shows a typical calibration curve of such a gauge. It can be seen that there exist two main modes of discharge with the transition around 10^{-4} mbar (10^{-2} Pa).

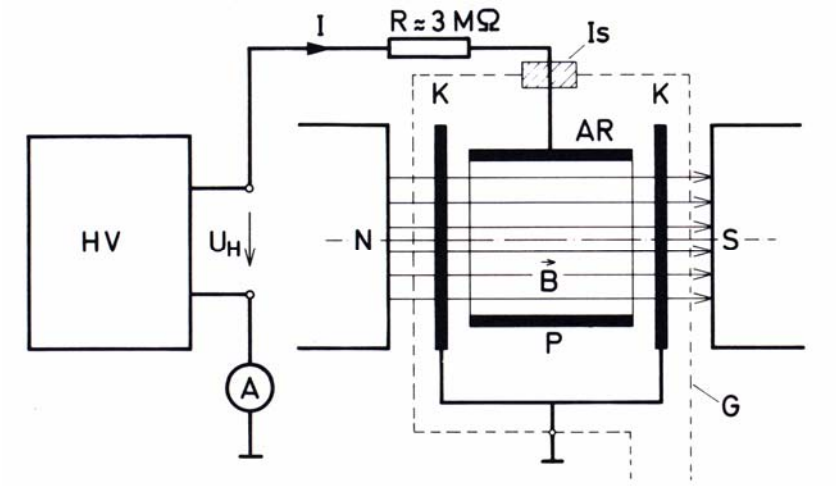


Fig. 5: Penning gauge: AR anode ring, K cathode, G case, N, S north and south pole of magnet, HV high voltage. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

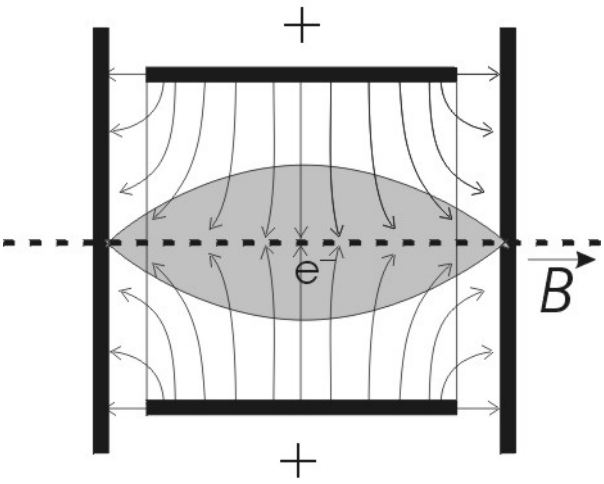


Fig. 6: Direction (not strength) of the electrical field in the Penning gauge as in Fig. 5. Grey: electron space charge. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

At low pressures, there is a rotating electron current of about 1 A symmetrical to the axis of the anode cylinder and perpendicular to the magnetic field (grey area in Fig. 6). Between this electron space charge and the anode there is a strong electric field and almost all of the full voltage drops between the space charge and the anode cylinder. For this reason the electrical field gets a strong radial component. Close to the axis of the cylindrical anode a plasma with equal negative and positive charges exists. The electron ring current would be completely stable, if no gas molecules were there. The electrons interact with them in two ways. They hit them with small energy and diffuse out of the electron space charge or they hit them with higher energy and ionize the molecule. In the latter case the new electron will be incorporated in the ring current, the ion will be accelerated by the electrical field and finally reach the cathode. Since both the ionization rate and the diffusion effect (diffusion coefficient) are proportional to the gas density *n*, in equilibrium the ring current will be such that the loss of electrons by diffusion is compensated by the generation due to ionization. This means that the

ring current will be roughly independent of *n* and *p*. For this reason, the outer discharge current will be proportional to *n* and *p*. The fact that the ring current does slightly increase with *n* has as consequence that *m* in Eq. (2) is > 1.

At higher pressures, the positive ion charge becomes so high that the ring current described above will no longer be stable. Instead, an equipotential plasma will build up in the whole space of the anode cylinder with respective space charges opposed to the two electrodes. Ions accelerated onto the cathode generate secondary electrons that compensate the loss of electrons by diffusion to the anode. This diffusion mechanism is amplified by plasma oscillations. In this regime the discharge current is no longer proportional to pressure (Fig. 7).

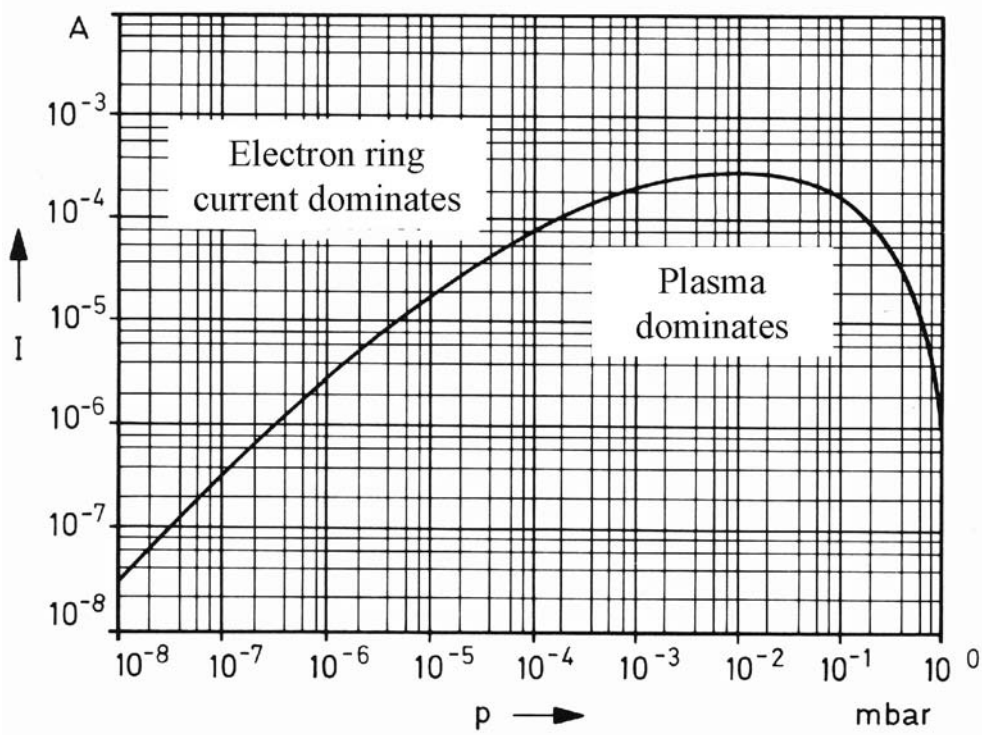


Fig. 7: Typical calibration curve of a Penning gauge. From *Wutz Handbuch Vakuumtechnik* by K. Jousten (ed.), Vieweg Verlag.

At high pressures the cathode erodes and the production of secondary electrons depends on the surface of the cathode and it has to be cleaned quite often to get reproducible results.

Since the ring current of a Penning gauge is very high (1 A or so), it has a high sensitivity and the discharge current may be inexpensively measured without an amplifier down to 10^{-4} Pa.

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 of increasing the active volume of the discharge and reducing discontinuities.

3.2 Magnetron and inverted magnetron

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

In the magnetron gauge [5] (Fig. 8) the anode is an open cylinder with the cathode on axis and as endplates, in the inverted magnetron gauge [6] (Fig. 9) the anode is a rod in the axis of an almost

closed cylinder as cathode. In the magnetron gauge, the end discs of the 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. Versions of the magnetron have been used in satellites and on the surface of the moon in 1969 on Apollo 12 and subsequently also on Apollo 14 to 16 [16].

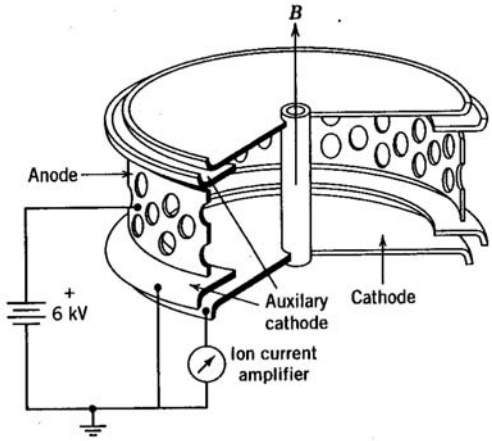


Fig. 8: Schematic diagram of the magnetron gauge. From Ref. [5].

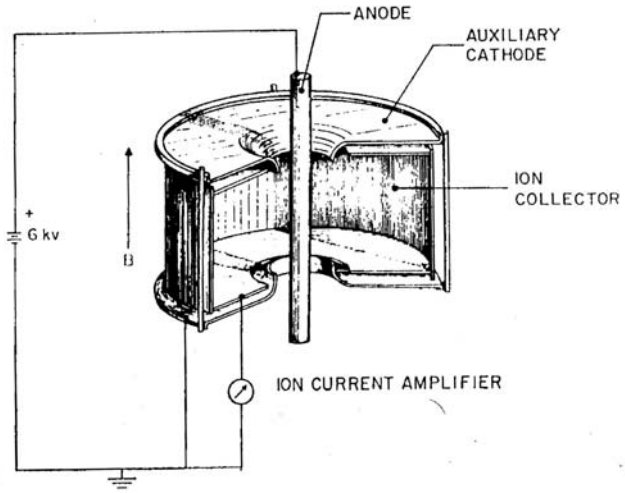


Fig. 9: Schematic diagram of the inverted magnetron gauge. From Ref. [6].

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 up to 6 kV with 0.2 Tesla.

In these gauges the electrons are trapped more efficiently than in the original Penning design. Because of 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 becomes a problem.

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 in extreme cases. 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 of 10^{-10} Pa this error may be as high as an order of magnitude.

More detailed theoretical descriptions of the characteristics of crossed field Townsend discharges including electron space charge, which controls the discharge, have been given than described in the previous section. However, Redhead [7] 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 which are associated with mode-jumping of the rf oscillations. Owing to interaction of the electrons with the produced AC fields, excess energy electrons are generated which easily come across the magnetic field and hit the cathodes (Penning gauge) or the cathode end plates (magnetron or inverted magnetron). They falsify the discharge current and ion current, respectively. 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 shall came back to this point.

Kendall [8] has designed an inverted magnetron gauge that is reduced in size and modified in the magnetic field (Fig. 10) in order to reduce the external magnetic field. This may be of interest wherever the magnetic field of a crossed field gauge has undesirable effects on its environment.

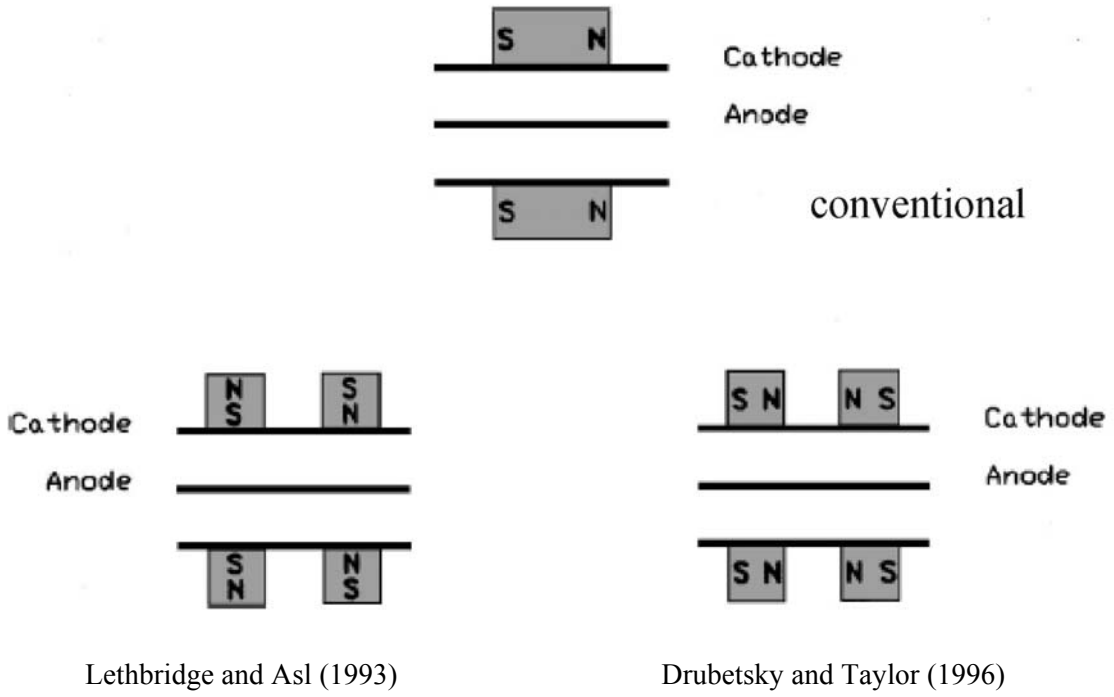


Fig. 10: Modified field configuration in an inverted magnetron. From Ref. [8].

4 Ionization gauges with emissive cathodes

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

One of the main problems in the beginning of the BA gauge (Fig. 11) 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 on a defined positive potential. Also it was soon noticed that ions could be lost through the open ends of the cylinder and 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 [9] at pressures of 10 mPa or more.

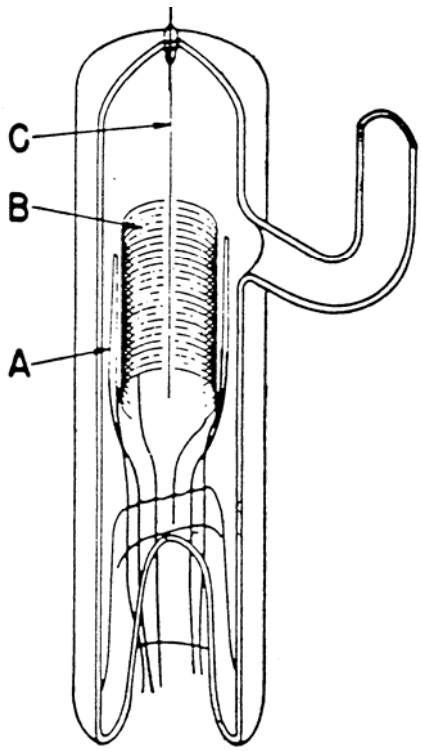


Fig. 11: The original design of the Bayard–Alpert gauge. From R.T. Bayard and D. Alpert, *Rev. Sci. Instrum*. **21** (1950) 571.

In order to further reduce the X-ray limit (Fig. 12) there was an attempt to reduce the thickness of the collector wire. For example Van Oostrom [10] reduced its diameter to about 4 μm. Although with this method the X-ray limit is reduced, theory [11] stated 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. Careful experimental investigations by Benvenuti [12] and Groszkowski [13], however, showed that the variation of collector efficiency with collector diameter was much less than theoretically predicted.

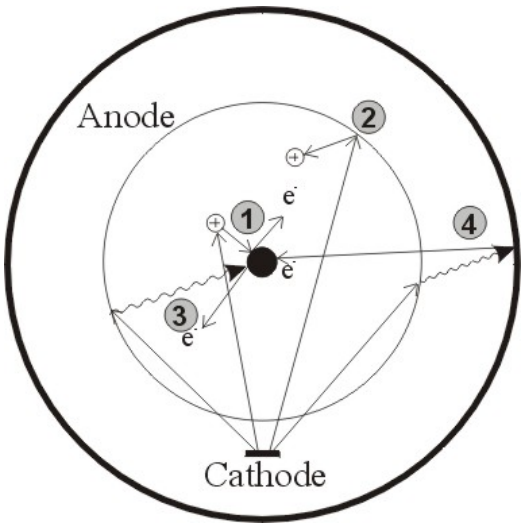


Fig. 12: Effects of ionization (1), electron stimulated desorption (2), X-ray effect (3), and inverse X-ray effect (4) in a BA gauge

When the X-ray limit was pushed down in this manner, another component to the background current became evident. Electrons hitting the anode may ionize molecules adsorbed on the surface with a subsequent release (Fig. 12). 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 be very clean. Two measures are usually taken to cure this problem: the grid is cleaned by electron bombardment after the gauge had been exposed to high pressures or the atmosphere, and the electron current to the anode should not be too small during operation so that the gauge is continuously 'self-cleaning'.

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 are made in the geometry of the gauge.
- The sensitivity is increased by several orders of magnitude without reducing the background level.

The first two techniques have been found 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.

By using the smallest practical diameter for the collector wire (50 μm), increasing the sensitivity of the BA gauge by maximizing the volume enclosed by the anode and using end caps, by optimizing the geometry, the materials, the voltages and the emission current, Benvenuti [12] was able to reduce the residual current of the BA gauge (mainly the X-ray induced limit) to the low 10^{-10} Pa regime. The sensitivity of this gauge is higher than 0.3 Pa^{-1}. Benvenuti ensured by choosing the right position, shape, and potential of the filament that the electrons cross the grid at right angles in order to maximize their path length and maximize the sensitivity.

The first technique evolved when Redhead [14] 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, so that the measured collector current is significantly reduced by the so-called modulation index. When this is measured at higher pressures, the residual current due to X-ray effects can be determined at lower pressures.

Most interestingly it was found also that the electron desorbed ions from the grid were modulated [15], [16] and the modulation effect can be used to measure and reduce the electrostatic discharge effect.

It turned out also that the residual current was modulated to a significant extent, because the electron trajectories were also modulated. Hobson [17] estimated that because of this, an error of $3 \cdot 10^{-10}$ Pa would be introduced in measuring pressure.

Benvenuti again [12], however, demonstrated with his BA gauge with a thin collector wire that a modulation index of 0.9 could be achieved and was able to reach a residual pressure limit $3 \cdot 10^{-11}$ Pa, one order of magnitude less than predicted. This modulated BA gauge is still in use at CERN and its design has apparently been commissioned first to a French and then to a Swiss company where it is commercially available [18].

Not mentioned so far was the inverse X-ray effect (Fig. 12) which occurs when X-rays hit the enclosure of a gauge and produce secondary electrons that may travel to the ion collector on the same potential and produce a negative current. There were attempts to cancel out the two X-ray effects, but

this is a very unstable situation on account of surface effects. An obvious solution to avoid this effect is to set the collector on a negative potential (–30 V).

The second method of realizing lower pressure measurement led about 30 years ago to the development of the so-called extractor gauge (Fig. 13). In this approach the ion collector is removed out of sight of the grid. A simple lens is introduced between the grid and the collector to pull out the ions to the collector. An ion reflector is used to reflect the ions onto the collector tip to increase the sensitivity similar to that of a conventional BA gauge. In this way pressures from about 10^{-10} Pa can be measured. Also in this type of gauge the electrostatic discharge (ESD) effect is greatly reduced. This is because ions released from the grid surface by electron bombardment have sufficient energy to reach the reflector electrode and are not collected.

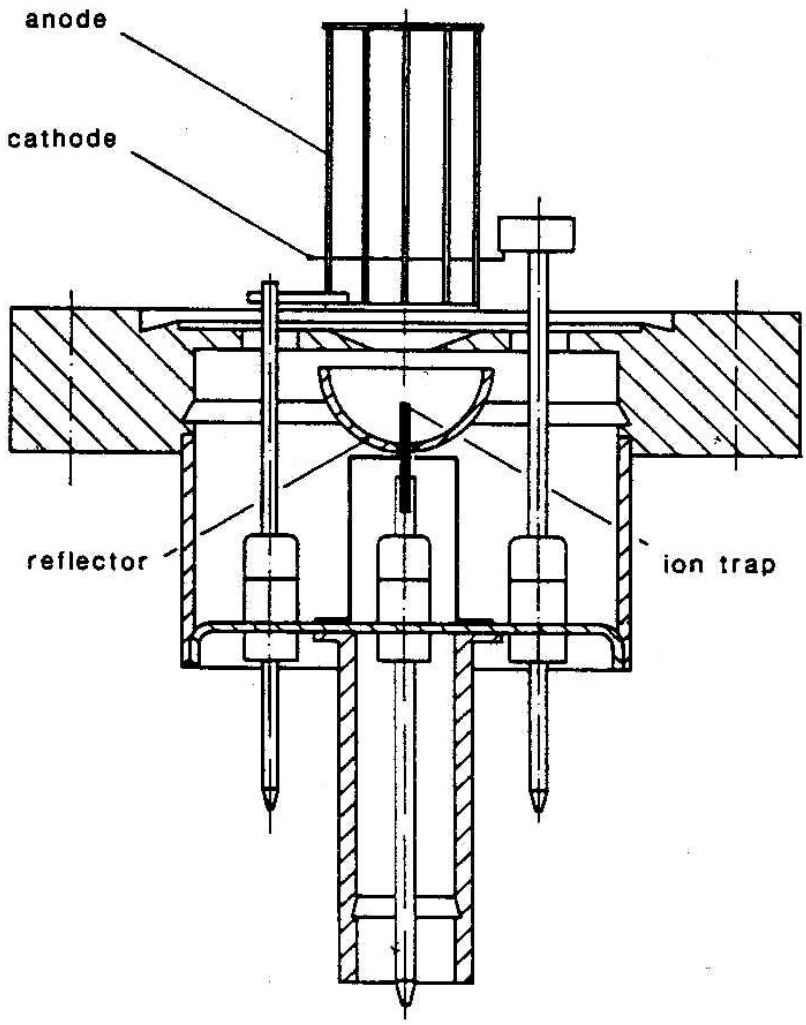


Fig. 13: Design of the extractor gauge manufactured by Leybold

This principle of extraction was further developed by Helmer [19]. By shaping the ion beam with a 90° deflector onto the collector there was no line of sight between anode grid and collector (Fig. 14) and the X-ray limit was further reduced to about $2 \cdot 10^{-11}$ Pa. Helmer used a fixed voltage and this was only possible because the energy spread out of a BA gauge (without collector wire in the centre) was found to be unexpectedly narrow (5 eV FWHM) [20]. Since inside the grid the potential varies by about 100 V, this is remarkable. Only due to this narrow energy width was the collection efficiency after the electrostatic analyser high enough.

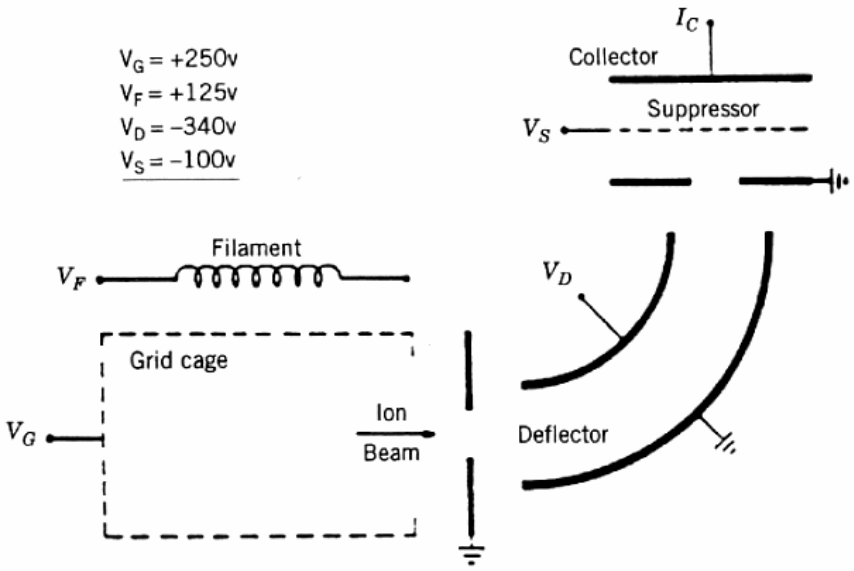


Fig. 14: Extractor gauge according to Helmer. From Ref. [19].

Benvenuti and Hauer improved the Helmer gauge by increasing the sensitivity of the ion source and optimizing some geometrical parameters of the extraction [21]. They obtained a residual pressure limit of $2 \cdot 10^{-12}$ Pa at a sensitivity of 0.3 Pa^{-1}. Jitschin [22] used a thoriated tungsten filament and also reduced Helmers limit. B. Lägel at CERN made the latest improvement to the Helmer gauge [23].

A very sophisticated ion gauge was invented by Watanabe in 1992 [24], which he called the ion spectroscopy gauge (Fig. 15).

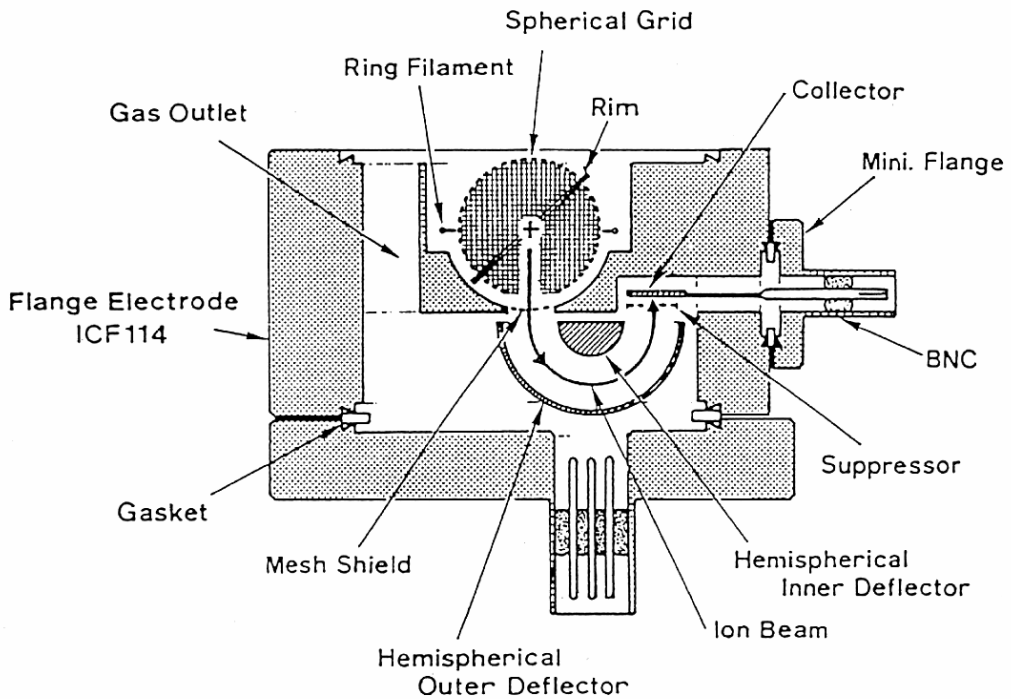


Fig. 15: Cutaway drawing of the ion spectroscopy gauge by F. Watanabe. From Ref. [24].

This gauge has so many features that only the most important ones can be mentioned. The gauge uses the extractor scheme, but with an hemispherical deflector so that the ion collector plate is completely out of sight of the grid. The collector is equipped with a suppressor electrode to inhibit electrons which are generated by reflected X-rays leaving the collector. With the hemispherical deflector where the inner electrode is on ground potential and the outer on 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 because of a potential gradient from the grid to the extractor and because of space charge effects. This effect was also used in the Helmer gauge, but in the ion spectroscopy gauge a spherical grid was used in order to increase the space charge of the electrons in its centre. By this means ESD ions and gas ions can be separated more efficiently than in the Helmer gauge (Fig. 16).

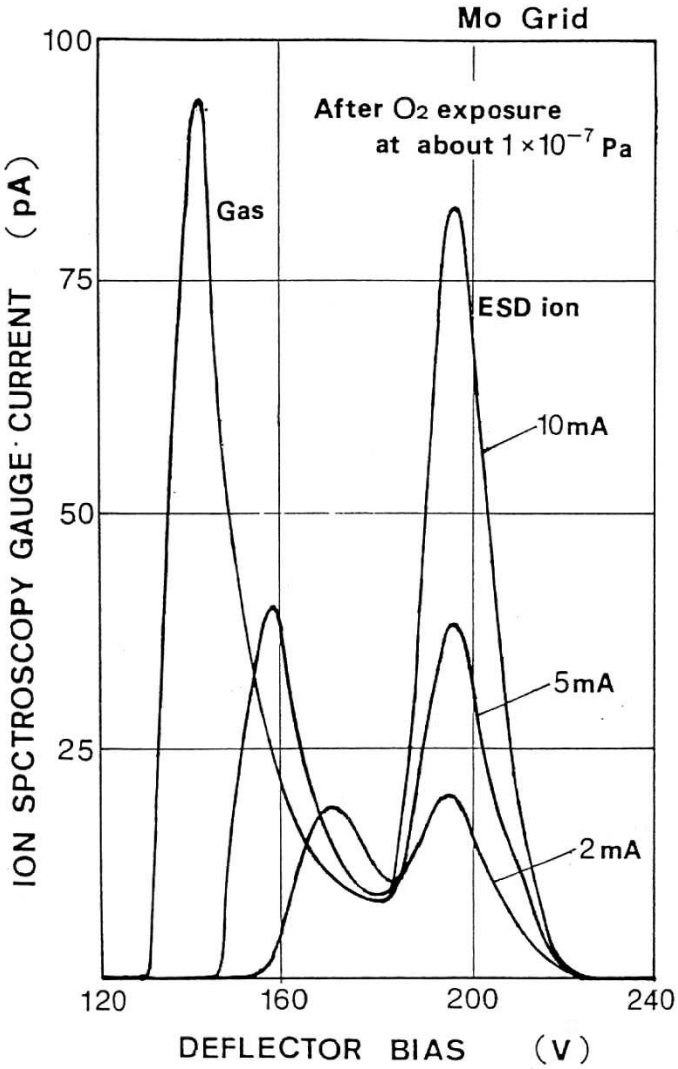


Fig. 16: Ion current vs. deflector bias voltage in the ion spectroscopy gauge after oxygen exposure at 10^{-7} Pa. From Ref. [24].

Those parts of the gauge close to the hot filament could be outgassed by resistive heating or electron bombardment. In addition the housing of the gauge was made of high thermal conductance materials such as copper or aluminium in order to reduce the warming up 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 best known and most successful example for the third method of approaching lower pressure limits is the so-called Lafferty gauge (Fig. 17) [25]. Lafferty adopted the diode magnetron principle by placing the filament along the axis of a cylindrical anode. An axial magnetic field provided by a magnet outside of the enclosure forces the electrons to follow circular paths and increases their path length by 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.

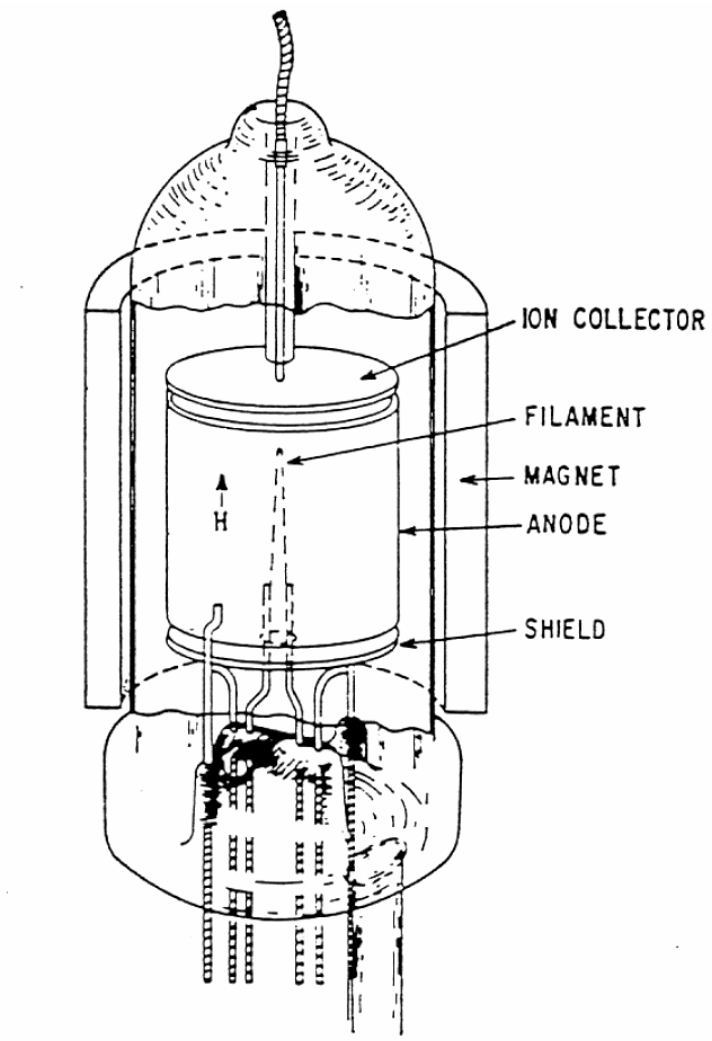


Fig. 17: Ionization gauge designed by Lafferty to increase the electron path length. From Ref. [25].

Most of the gauges designed for very low pressures and described above were not commercially successful and are no longer on the market. However, the so-called AxTran gauge (AXial symmetric TRANsmission gauge) [26] by the Ulvac Corporation (Fig. 18) is commercially available. In this gauge the separation between ESD ions and ions generated in free space is provided by an energy analyser called 'Bessel box' [27]. This type of energy analyser is of a straight cylindrical symmetrical design, which has the advantage that the construction of the ion gauge is less space consuming. For a given voltage V_{BE} only ions around a certain energy may pass the Bessel box and be detected by the secondary electron multiplier (SEM). By optimizing this voltage, ESD ions may be suppressed. In the centre of the Bessel analyser there is a disk with the same potential as the cylinder to avoid a direct line of sight between anode grid and SEM. As lowest measuring limit Akimichi [26] estimated $3 \cdot 10^{-12}$ Pa, for the commercial instrument it is specified as $5 \cdot 10^{-11}$ Pa.

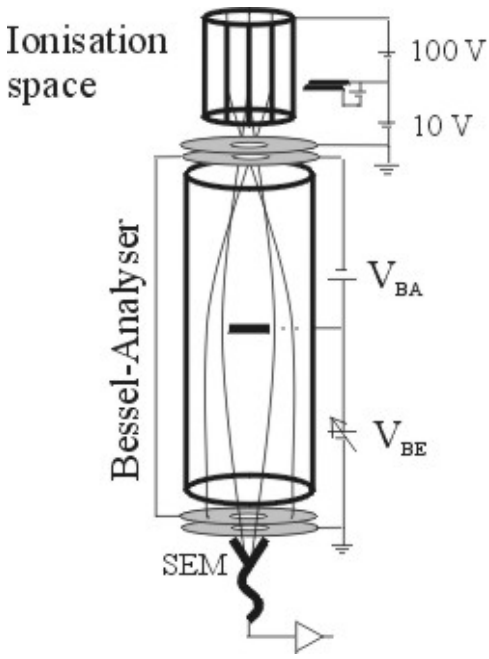


Fig.18: Design of the AxTran gauge by Ulvac Co.

There were more approaches to reaching lower pressure limits on all designs of ionization gauges, but the reader should refer to text books or review articles.

So far, only thermionic electron emitting cathodes (hot cathodes) have been described. In the past, field emitter tip arrays of molybdenum or silicon with 10^4 tips/mm² were developed and used in ionization gauges [28], but the current was only 20 μA. Today the work on cold emitters is still a hot topic and focuses on carbon nanotubes [29]–[33] which can produce current densities of up to 10^8 A/cm^2 and are commercially used in flat-screen TV sets and miniature X-ray generators. Perhaps carbon nanotubes can replace hot cathodes in the future.

5 Comparison of crossed field (CFG) and emitting cathode (ECG) ionization gauges

As a summary, the major types of ionization gauges have been schematically drawn by Redhead (Fig. 19). CFGs have the general advantage that they have no X-ray limit (the electron current producing X-rays is proportional to pressure) and electron stimulated desorption effects are small and cause few errors. Also, because they already have a strong magnetic field, their functionality is less affected by an outside magnetic field than an ECG. In case 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 an 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 ECG and cannot be controlled.

When an ion gauge pumps, this is a classical disturbance effect of a measuring device, because it changes the value of the quantity that 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 by field emitters of radioactive sources built in. It was found by Li [34], however, that the starting time of commercial gauges may largely exceed the manufacturer's specifications.

In ECGs the electron emission current can be controlled, stabilized and varied. Mainly for this reason, ECGs are more stable and accurate, when they are conditioned before measurement.

Li and Jousten [35] have performed a comprehensive study of the stability of CFGs and ECGs with hot cathodes and found that while is it difficult to calibrate CFGs because of the non-linearities and discontinuities, the reproducibility of CFGs is slightly worse than those of ECGs in nitrogen, argon, and helium, but better for hydrogen (Table 1).

Table 1: Maximum deviations in per cent from a first calibration run for several gauges (EXG extractors gauge, BAG Bayard–Alpert gauge, IMG inverted magnetron) in different gases over a period of 6 months [35].

	EXG	**BAG1**	**BAG2**	**IMG1**	**IMG2**
N_2	–2.5	–4.3	–3.2	–6.2	+5.9
Ar	–1.9	–3.8	+3.8	–2.4	+3.1
He	–5.9	–4.4	–3.6	+8.4	–5.0
H_2	+9.4	–1.9	–3.6	–1.0	–1.3

When measuring pressures in HV and UHV, one has to decide 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 gauge 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 ECG is significantly better at very low pressures. An order of magnitude error is easily possible below 10^{-8} Pa [36].

Table 2 and Table 3 give some recent published values [34] of the pumping speed and outgassing rates of some commercial gauges (both CFG and EFG) which were found to be quite consistent with other published data. The pumping speed of a EFG can be reduced by reducing the emission current, but then a complication may arise from the fact that the anode grid is not continuously cleaned by the electrons. As a consequence the ESD effect may increase and disturb measurement.

Table 2: Measured pumping speeds in ℓ/s in two inverted magnetrons and two BA gauges, all commercially available. From Ref. [34].

Gas	IMG1	IMG2	BAG1 at 4 mA	BAG2 at 1 mA	BAG2 at 10 mA
N_2	$4.5 \cdot 10^{-2}$	$6.5 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	–	$4.5 \cdot 10^{-2}$
Ar	$2.0 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$6.7 \cdot 10^{-2}$	$3.7 \cdot 10^{-2}$	$2.3 \cdot 10^{-1}$

Table 3: Measured outgassing rates in Pa ℓ/s of commercial extractor and BA gauges. From Ref. [34]. The outgassing rate of two inverted magnetron gauges was below the measurable limit.

EXG at 1.5 mA	BAG1 at 4 mA	BAG2 at 1 mA
$2.4 \cdot 10^{-8}$	$8.1 \cdot 10^{-8}$	$3.0 \cdot 10^{-8}$

Hot cathodes are extremely subject to disturbance when gases other than rare gases or nitrogen have to be measured. For in the sense of vacuum science so-called chemically active gases, CFGs should be used which can also be cleaned much easier than ECGs.

The price of a CFG is usually lower than that of an ECG.

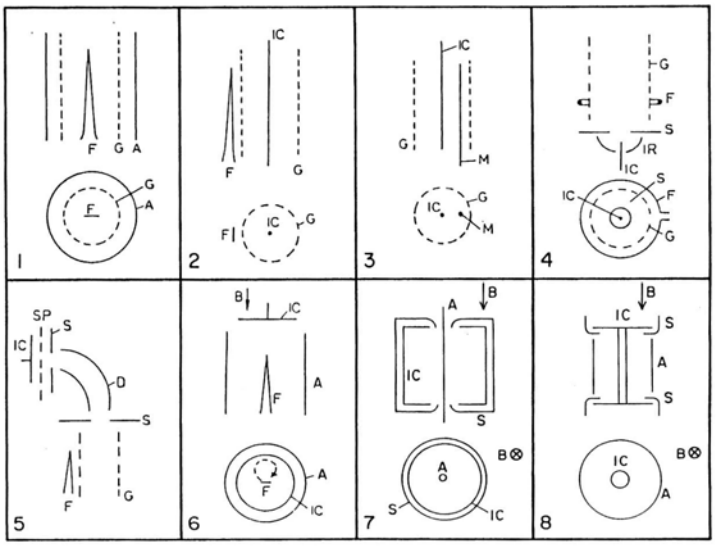


Fig. 19: Overview by Redhead of the major types of ionization gauges. 1 conventional triode gauge; 2 Bayard–Alpert gauge; 3 modulated Bayard–Alpert gauge; 4 extractor gauge; 5 bent-beam gauge (Helmer gauge); 6 hot-cathode magnetron (Lafferty gauge); 7 magnetron; 8 inverted magnetron. A-Anode, D-deflector, F-filament, G-grid (acts in 1 as collector), IC- ion collector, IR-ion reflector, M-modulator, S-shield, SP-suppressor.

6 Problems in applications of ionization gauges

Special to the application of ion gauges in accelerators are their interaction with radiation, strong magnetic fields, and EM radiation mainly in the radiofrequency range: radiation capable of ionizing molecules may contribute inside the gauge head to the ion current. Miertusova [36] found completely erratic pressure indications when an inverted magnetron gauge was installed very close to a photon absorber. The reason was the characteristic X-ray radiation from copper. Both CFGs and EFGs have to be shielded very carefully from strong magnetic fields in order to get reasonable pressure indications. Hysteresis effects are typical for an incomplete magnetic shielding of CFGs.

Suppose there is a sealed-off chamber at room temperature which is not pumped. An ionization gauge is installed on it to measure the pressure p_1 inside, which we assume as pure hydrogen. Now let us immerse the whole chamber in liquid nitrogen. The pressure will drop by

$$\frac{p_2}{p_1} = \frac{T_2}{T_1} = \frac{77}{300} = 0.257 \tag{3}$$

but the reading of the IG will be unchanged, because 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 according to room-temperature variations have to be considered when gauges are accurately calibrated [37].

In other cases, when a chamber is continuously pumped, the molecular flow will adjust such that the law of continuity holds. For example, installing a gauge with hot cathode in a tube (Fig. 20) results in the so-called thermal transpiration effect, where

$$\frac{p_1}{p_2} = \sqrt{\frac{T_1}{T_2}}\,. \tag{4}$$

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 $n_2 = n_1(T_1/T_2)^{1/2}$.

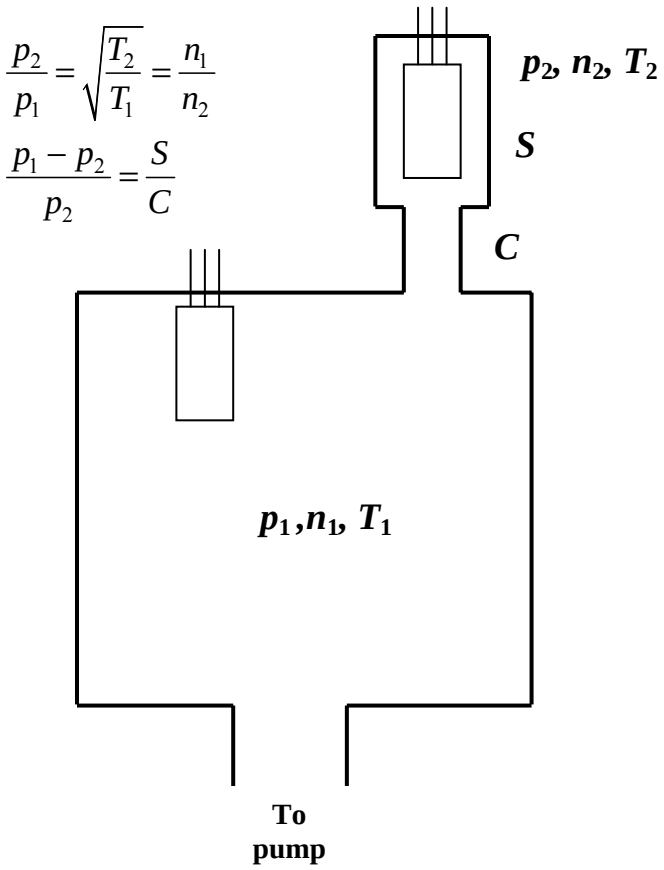


Fig. 20: Effects of tubulation of a gauge by conductance, internal pumping speed of the gauge, and thermal transpiration

Another disadvantage of installing IGs in tubes is problems associated with their pumping speed (Fig. 20). All IGs do pump, at least the ionized 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. 20) (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.

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. 21, where gas flows from the left to the right and suppose the right wall is a cryo surface with sticking probability = 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 equality of the rate of influx and the rate of return flow through the tubulation. Neither of these gauges represents the true pressure.

Problems with hot-cathode ionization gauges (HIGs) arise with dissociation and enhanced chemical reactions on the hot cathode surface. For example in tungsten filaments (2200°C), there is always carbon present on the surface which diffuses out of the bulk as impurity. Also oxygen is present on its surface. Some reactions which can take place after dissociation of hydrogen are shown in Fig. 22 [38]. It was also reported [39] that at high cathode temperatures hydrogen dissociates and adsorbs on the grid and other parts of the ion gauge. This will also change the sensitivity because of a different reflection coefficient of electrons at the grid. The use of thoriated tungsten or iridium cathodes with operating temperatures of 1200°C avoids this effect.

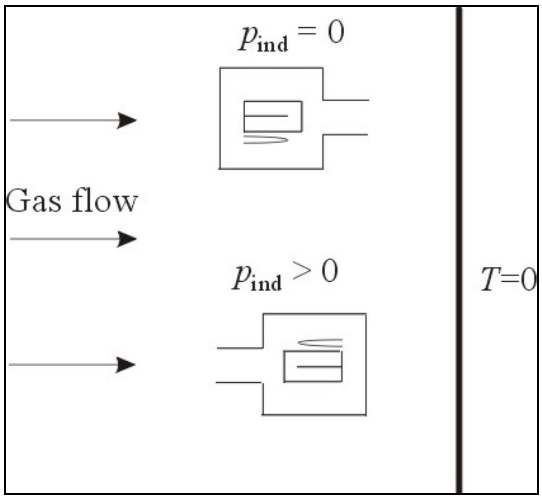


Fig. 21: Example of orientation effects when measuring gas pressures with vacuum gauges. $T = 0$ means in other words a sticking probability of 1.

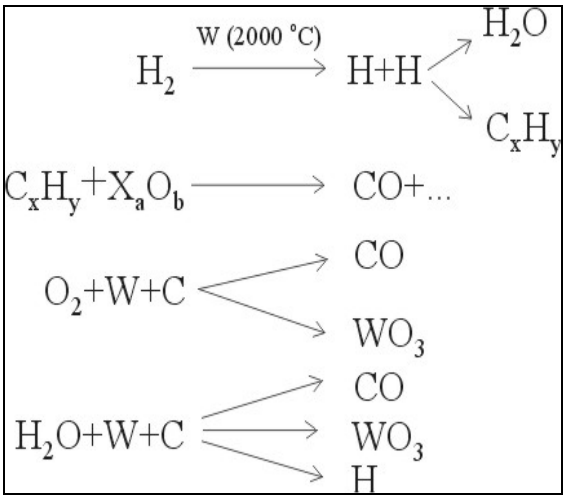


Fig. 22: Some chemical reactions which can occur at the hot tungsten filaments in ionization gauges [38]

Outgassing and re-emission of molecules previously pumped by the gauge, is a significant problem in IGs. 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 get a reliable and long-term stable gauge reading, the gauge electrode surfaces have to have a stable surface structure and composition. Not only does the secondary electron yield on the collector change with the surface composition, but also the number of secondary electrons generated by electrons hitting the anode grid is 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 ionization cross-sections failed due to other gas-specific effects like ion capture probability, dissociation effects and secondary electron generation. Values for relative ionization sensitivities (normalized for nitrogen = 1) presented in tables (Table 4, [40], [41]) can be applied with some confidence while jumping from one gas to another, but the level of accuracy is only 10–20%. Where greater precision is required, gauges must be calibrated individually and for the gas used in the application.

Table 4: Correction factors *CF* for different gases when an ionization gauge is set to a correct nitrogen reading. The uncertainty of these values (except nitrogen) is typically 10%, but may be higher in special cases.

Gas species	CF_i (N_2)
N_2	1
He	7.24
Ne	4.55
Ar	0.85
Kr	0.59
Xe	0.41
H_2	2.49
O_2	1.07
Air	1.02
CO	0.97
CO_2	0.70
J	0.17
CH_4	0.71
C_2H_6	0.37
C_3H_8	0.22
CF_2Cl_2	0.36
Oil vapours	0.1

As a final example of what effects have to be considered in a hot-cathode ionization gauge (HIG), calibration results for the sensitivity of H_2 and D_2 should be mentioned. 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 would be exactly 1. It was found that this is not true and the relative sensitivity varies from gauge to gauge. Moreover, the ratio was 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.

The reason for the difference in the sensitivity for H_2 and D_2 is that H_2 because of 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 that also explains why the ratio changes with treatments and history of the gauge.

If the secondary electrons are completely pushed back to the collector by applying a negative potential on a suppressor grid in front of the collector, as can be done in the ion spectroscopy gauge of Watanabe, the sensitivity ratio for H_2/D_2 indeed equals 1.

An ECG has to be outgassed when new and after each exposure to atmospheric pressure. This is best done by electron bombardment after a bake-out when the system is still warm. In addition, operation in argon at a pressure of about 1 mPa helps to clean the ion collector.

A safety precaution should be mentioned. During electron bombardment potentials as high as 1000 V are needed in ECGs and a glow discharge may develop and charge up electrodes in the vacuum chamber quite remote from the ECG. This may also happen when operating a CFG. Therefore all parts of a vacuum system (e.g. unused feedthroughs) should be effectively grounded at all times.

7 Accuracy and the calibration of ionization gauges

As far as is known to the author, manufacturers calibrate ion gauges in a rough manner for nitrogen before the gauge leaves the factory. This calibration gives you typically an accuracy of within 10% for nitrogen and good quality gauges, for other gas species the accuracy is worse. If better accuracy is required, especially over the lifetime of the ion gauge, it has to be calibrated with a primary standard or a secondary standard for vacuum pressures.

Table 5 and Table 6 list general and specific reasons for measurement uncertainties with ionization gauges. Some of the general reasons have also been mentioned in the section on fine vacuum gauges.

Table 5: General reasons for measurement uncertainties with ionization gauges

General reasons for measurement uncertainties
Uncertainties due to calibration chain
Uncertainties due to installation (or mistakes in installation)
Uncertainties due to operation (surface layers, corrosion, dust, ageing)
Inaccuracies caused by gas mixture
Uncertainties caused by the device itself

Table 6: Uncertainties that are caused by the individual ionization gauge

Gauge-specific reasons for measurement uncertainties
Offset due to X-ray, ESD, electronics, incomplete insulation
Offset instability (drift)
Resolution
Influences of environment (mainly temperature)
Non-linearity
Integration time (scatter of data), repeatibility
Reproducibility (stability of calibration constant)
Hysteresis (ESD)
Prior usage, cleanliness

The calibration constant of an emitting cathode ionization gauge is the so-called sensitivity of an ion gauge. This is defined by

$$S = \frac{I^{+} - I^{+}_{\mathrm{res}}}{I^{-}(p - p_{\mathrm{res}})} \ . \tag{5}$$

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. Simplifying equations like

$$S = \frac{I^{+}}{I^{-}p} \tag{6}$$

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 CFG the ionizing electron current cannot be measured and in this case the sensitivity is usually defined as [17]

$$S = \frac{I^{+}}{p^{m}} \tag{7}$$

where m is a numerical exponent. This equation for a CFG is more simple than that for ECG [Eq. (5)], 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. (5) is pressure independent. This is generally not true as mentioned in Table 6. In cases where high-precision current meters are being used to determine S, typical relative variations of S of a few per cent are found. In cases, where lower quality current meters as typical for built-in devices for ion gauge control units are used, 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 could be responsible: space-charge effects 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 [42]. Above about 10 mPa it can be expected that the sensitivity will be pressure dependent because of intermolecular collisions and ion-neutral collisions, but also because of changes in space charge [43].

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

Two basic calibration methods exist for the calibration of ionization gauges: the calibration by comparison with a reference gauge or the 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 and gas density are the same at the position of the test gauge and the reference gauge. In the review [44],

systems for calibration by comparisons have been described. If available, it is recommended that a spinning rotor gauge be used for the calibration of an ionization gauge between $3 \cdot 10^{-4}$ Pa and 10^{-2} Pa, because it is much more accurate than the calibration with an ionization gauge on account of the better stability of the spinning rotor gauge compared to the ionization gauge.

The calibration of an ionization 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 how the pressures can be generated have been reviewed in Ref. [44]. In the same book the procedures to calibrate ionization gauges have also been described. Primary standards for vacuum pressures are available in the major National Metrological Insititutes of the world, among them the Physikalisch-Technische Bundesanstalt (PTB, Germany), the National Institute of Standards and Technology (NIST, USA), and the National Physical Laboratory (NPL, England).

References

[1] O. von Baeyer, Phys. Z. **10** (1909) 168.

[2] O.E. Buckley, Proc. Natl. Acad. Sci. USA **2** (1916) 683.

[3] R.T. Bayard and D. Alpert, Rev. Sci. Instrum. **21** (1950) 571.

[4] F.M. Penning, Physica **4** (1937) 71, and Philips Tech. Rev. **2** (1937) 201.

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

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

[7] P.A. Redhead, J.P. Hobson and E.V. Kornelsen, *The Physical Basis of Ultrahigh Vacuum*, Chapman and Hall Ltd, London, 1968, p. 335.

[8] B.R.F. Kendall and E. Drubetsky, J. Vac. Sci. Technol. A **18** (2000) 1724–1729.

[9] R.N. Peacock and N.T. Peacock, J. Vac. Sci. Technol. A **8** (1990) 3341.

[10] A. Van Oostrom, *Transactions of the Eighth Vacuum Symposium and Second International Congress* (Pergamon, Oxford, 1962), p. 443.

[11] G. Comsa, J. Appl. Phys. **37** (1966) 554.

[12] C. Benvenuti and M. Hauer, Nucl. Instrum. Methods **140** (1977) 453–460.

[13] J. Groszkowski, Bull. Acad. Polon. Sci. Ser. Sci. Technol. **13** (1965) 2.

[14] P.A. Redhead, Rev. Sci. Instrum. **31** (1960) 343.

[15] P.A. Redhead, Vacuum **13** (1963) 253.

[16] P.A. Redhead, J. Vac. Sci. Technol. A **12** (1994) 904–914.

[17] J.P. Hobson, J. Vac. Sci. Technol. **1** (1964) 1.

[18] www.xtronic.ch (Mai 2006).

[19] J.C. Helmer and W.D. Hayward, Rev. Sci. Instrum. **37** (1966) 1652.

[20] J. C. Helmer, Vacuum **51** (1998) 7–10.

[21] C. Benvenuti and M. Hauer, Proc. IVC-8, Cannes 1980, Suppl. à la Rev Le Vide, les Couches Minces no. 201.

[22] S.W. Han, W. Jitschin, P. Röhl and G. Grosse, Vacuum **38** (1988) 1079–1082.

[23] B. Lägel, CERN Vacuum Technical Note 95-15, October 1995.

[24] F. Watanabe, J. Vac. Sci. Technol. A **10** (1992) 3333.

[25] J.M. Lafferty, J. Appl. Phys. **32** (1961) 424.

[26] H. Akimichi *et al*., Vacuum **46** (1995) 749–752.

[27] J.H. Craig and J.H. Hock, J. Vac. Sci. Technol. **17** (1980) 1360–1363.

[28] R. Baptist, Vacuum **48** (1997) 723–725, and R. Baptist and F. Bachelet, Vacuum **48** (1997) 947–951.

[29] B. Bushan (ed.), *Springer Handbook of Nanotechnology* (Springer, Berlin, 2004), p. 74.

[30] N.S. Xu and S. Ejaz Huq, Mater. Sci. Eng. R **48** (2005) 47–189.

[31] W. Knapp and D. Schleußner, Appl. Surf. Sci. **251** (2005) 164–169.

[32] P.G. Collins and A. Zettl, Phys. Rev. B **55** (1997) 9391–9399.

[33] A.N. Obraztsov *et al*., J. Vac. Sci. Technol. B **18** (2000) 1059–1063.

[34] Detian Li and K. Jousten, Vacuum **70** (2003) 531–541.

[35] Detian Li and K. Jousten, J. Vac. Sci. Technol. A **21** (2003) 937–946.

[36] J. Miertusova, Vacuum **51** (1998) 61–68.

[37] K. Jousten, Vacuum **49** (1998) 81.

[38] D. Alpert, Le Vide **17** (1962) 19.

[39] J.G. Werner and J.H. Leck, J. Sci. Instrum. E **2** (1969) 861–866.

[40] J.H. Leck, *Total and Partial Pressure Measurement in Vacuum Systems* (Blackie, Glasgow and London, 1989), p.73.

[41] R.L. Summers, *NASA Tech. Note* NASA TN D-5285 (1969).

[42] K. Jousten and P. Röhl, Vacuum **46** (1995) 9.

[43] A. Berman, *Total Pressure Measurements in Vacuum Technology* (Academic Press, Orlando, FL, 1985), p. 45.

[44] J.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, FL, 1985).

Saul Dushman, *Scientific Foundations of Vacuum Technique*, 2nd edition (John Wiley & Sons, New York, 1962).

Karl Jousten, *Wutz Handbuch Vakuumtechnik*, 9th edition (Vieweg, Wiesbaden, 2006), ISBN 3-8348-0133-X.

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, 1989).

P.A. Redhead, J.P. Hobson and 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.

Slowing-down processes, energy deposition, sputtering and desorption in ion and electron interactions with solids

J. Schou
Risø National Laboratory, DK-4000 Roskilde, Denmark

Abstract
The slowing-down processes of ions and electrons in matter are reviewed with special emphasis on the stopping cross-section. The relationships between elastic (knock-on) sputtering and nuclear stopping as well as between electronic sputtering and electronic stopping are discussed.

1 Introduction

Many phenomena which take place during ion, electron, and photon bombardment of solids are closely connected. These phenomena include energy loss of charged particles in solids, energy deposition and ejection of secondary particles during bombardment by charged particles. While the term sputtering usually comprises particles ejected as a result of momentum transfer to target particles or electronic transitions, desorption usually refers to the removal of less than a monolayer by electronic transitions at the very surface.

Sputtering as mass removal from the cathode during a gas discharge was discovered around 1850 [1], but the process was not understood, since the concept of 'charged particles' was not yet established. The discovery of radioactivity led to a big step forward towards well-defined ion sources [2]. Rutherford's famous work on the structure of the atom immediately stimulated work on particle slowing down and a number of papers on energy loss based on classical theory appeared during the following years. After quantum mechanical treatments became available, estimates on light ion and electron energy loss in matter could also be performed, but refinements of the theoretical framework are still continuing [2–3]. Sputtering of neutrals by keV ions was largely explained in the years 1960–1970 [4], whereas electronic sputtering (of insulators) is still a partly unresolved issue for non-elemental solids. Both types of sputtering could be explained by a relationship to the relevant part of the stopping cross-section (to be explained below). Secondary electron emission induced by ions was discovered around 1900, but the theoretical treatments did not appear until 1950–1980 [5]. Precise measurements of the desorption rates were not possible until reliable ultra-high vacuum systems were available. The leading theories were made around 1960 by Menzel, Gomer, and Redhead [6].

2 Stopping force (d*E*/d*x*)

2.1 Ion slowing down

The key quantity in energy loss considerations is the stopping force d*E*/d*x* (which in the past was named stopping power). This can be considered as the force that the medium exerts on the penetrating particle:

$$\mathrm{d}E/\mathrm{d}x = NS(E)\ , \tag{1}$$

where N is the number density of atoms in the medium and $S(E)$ the stopping cross-section, in which the dependence on the kinetic energy E of the primary particle is explicitly written. The stopping cross-section is a density-independent quantity except for ultrarelativistic energies and has the

dimension [energy × area]. In fact, many tables do not distinguish between stopping force (power) with the dimension [energy/length] and stopping cross-section, e.g., Ref. [7].

The theoretical framework for a general treatment of energy loss to the nuclei as well as to the electrons, i.e., the electronic system, was comprehensively treated by Bohr [8] and Lindhard *et al*., [9] in a number of important papers.

The collisions between the primary ions and the atoms in a solid can be divided into collisions between the primary particle and the nuclei, and those between the primary and the electrons. The first collisions take place for small impact parameters and lead to a large-angle scattering process, whereas the latter ones lead to energy loss without any significant deflection of the primary particle. The stopping cross-section $S(E)$ can be split up into

$$S(E) = S_n(E) + S_e(E) \quad (2) \tag{2}$$

where S_n is the nuclear stopping cross-section and S_e the electronic stopping cross-section. The term 'nuclear' is misleading, since the primary ion interacts with the screened nuclei rather than the bare nuclei.

The two stopping cross-sections drawn from the tabulations in Ref. [10] are shown in Fig. 1. The four regimes cover the low-energy (I), intermediate (II), high-energy (III) and ultrarelativistic regime (IV) (not shown in Fig. 1). The nuclear stopping is dominant in regime (I), but decreases gradually to less than a factor of the order $M_p/m \sim 2000$ than the electronic stopping (M_p is the proton mass) in regime (II). In regime (III) the electronic cross-section decreases monotonously.

An important scaling parameter is Lindhard's reduced energy ε:

$$\varepsilon = \frac{32.53 M_2 E}{Z_1 Z_2 (M_1 + M_2)(Z_1^{2/3} + Z_2^{2/3})^{1/2}} \tag{3}$$

where E is the energy in keV and the masses M_1 and M_2 are in amu. Lindhard and coauthors [11] showed that the nuclear stopping for all target–beam combinations could be expressed as

$$NS_n(E) = \frac{\pi a_L^2 \gamma N}{(\varepsilon/E)} s_n(\varepsilon) \tag{4}$$

where

$$\gamma = 4M_1M_2/(M_1 + M_2)^2 \tag{5}$$

is determined by the maximum energy transfer γE from a particle of mass M_1 with an energy E to a particle at rest with mass M_2 and by Lindhard's screening parameter

$$a_L = 0.8853 a_B \left(Z_1^{2/3} + Z_2^{2/3}\right)^{-1/2}. \tag{6}$$

The absolute magnitude is determined partly by the factor (ε/E) in the denominator, which means that for heavy atoms on a heavy target the factor becomes small and, in turn, leads to a large nuclear stopping. Equation (5) also demonstrates that ε decreases with increasing atomic number (and mass) of the projectile. The maximum of the nuclear stopping, which for the classical Thomas–Fermi model occurred at $\varepsilon \cong 0.3$, is therefore shifted to higher energies for heavy projectiles.

The reduced nuclear stopping power is usually calculated from a representative potential, the so-called KrC potential, for which several refined approximations exist [12]. The nuclear stopping shown in Fig. 1 is computed on the basis of the potential given in Ref. [12], which also is the underlying basis for Ref. [10].

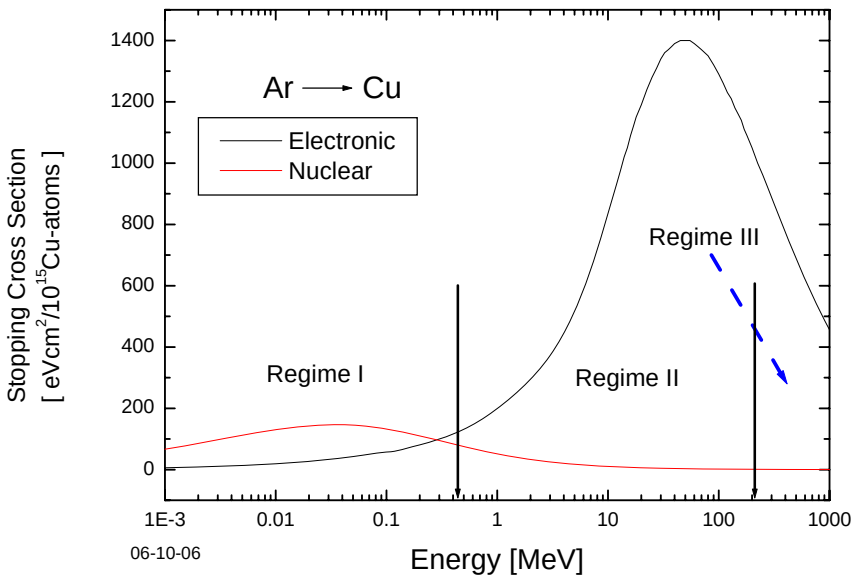


Fig. 1: Electronic and nuclear stopping cross-section for Ar ions incident on Cu. Data from Ref. [10].

The electronic stopping in the low-energy and intermediate regimes (I and II) can be determined by Lindhard–Scharff treatment [13]. The predictions from the Lindhard–Scharff model are usually correct within a factor of two, and can still be considered as a convenient reference standard. In reduced units the electronic stopping can be expressed as

$$S_e(\varepsilon) = k_L \varepsilon^{1/2} \tag{7}$$

where the constant of proportionality k_L depends on the atomic number and the mass of the beam atom and the target atom [9,13]. Except for very light ions on heavy targets, $k_L \cong 0.15$. The electronic stopping cross-section is thus proportional to the velocity and can be converted to real units with the same factor as the reduced nuclear stopping cross-section in Eq. (4). A number of treatments with improved accuracy have appeared [7,10,12,14], but the treatment of slow ions is still a difficult task [15].

With increasing velocity the projectile loses the electrons and at high velocities is completely stripped. At intermediate velocities in regime (II) the projectile electrons with velocities exceeding the projectile velocity v will stick to the projectile, while the slower ones, belonging to the outer shells, will be stripped. Sigmund [16] has estimated the broad maximum of the electronic stopping cross-section to be approximately at a velocity of $v_B Z_1^{2/3}$, where v_B is the Bohr velocity ($v_B \cong 2.188 \times 10^6$ m/s). For a proton ($Z_1 = 1$) this corresponds to a primary energy of 25 keV.

At high energies (regime III) the electronic stopping is determined by Bethe's or Bohr's treatment [2]. The decisive point is whether or not the problem can be treated by classical theory. This is possible for large values of Bohr's parameter [8]

$$k = \frac{2 Z_1 e^2 v_B}{(h/2\pi) v} \tag{8}$$

where h is Planck's constant. In the quantum mechanical limit $k \ll 1$ Bethe's formula based on a first-order perturbation treatment gives the high-energy expression in the non-relativistic case valid for light ions:

$$S_e(E) = \frac{4\pi Z_1^2 Z_2 e^4}{m_e v^2} \ln \frac{2 m_e v^2}{I} \tag{9}$$

and $I \cong Z_2$ 10 eV is the mean ionization potential. For relativistic corrections the reader is referred to Ref. [2].

One notes that the condition for quantum treatment cannot be expected to hold for slow projectiles or projectiles of high atomic number. In this case the logarithm in Eq. (9) is replaced by Bohr's expression:

$$S_e(E) = \frac{4\pi Z_1^2 Z_2 e^4}{m_e v^2} \sum_{\nu=1}^{Z_2} \ln \frac{C m_e v^3}{Z_1 e^2 \omega_\nu} . \tag{10}$$

The major difference is that Z_1 enters into the argument of the logarithm, and therefore influences the position of the stopping maximum. Here ω_ν is a characteristic frequency of each electron level of the target atom, since the derivation is performed for a harmonic oscillator [2].

A complicating feature is the charge state of the incoming ions. According to Bohr's estimate [8], the average charge state of heavy ions is

$$Z_1^* = Z_1^{1/3} \frac{v}{v_B} \tag{11}$$

for the velocity range where $1 < Z_1^* < Z_1/2$. It means that the average ion charge state increases with the ion velocity. However, regardless of the initial ion charge state, which may be very far from the equilibrium charge state, an ion beam will approach charge state equilibrium after having penetrated a few layers from the surface.

There are deviations from the elemental stopping powers, in particular for slow ions. The gas–solid difference is largest for low-elements, but exceeds rarely a 10% correction. For chemical components the stopping force is close to the weighted sum of the components, but also here only minor deviations may appear [2].

The printed tabulations, Refs. [7] and [12], have been replaced by the popular code SRIM [10]. SRIM is based on an effective charge state which can be adjusted to experimental values but is not supported by a firmly established theoretical basis. Also Paul [14] and Paul and Schinner [17] have published extensive tabulations based on experimental stopping cross-sections.

2.2 Electron slowing down

In contrast to ion slowing down, electron slowing down has not undergone any substantial improvement since the 1980s except for development of electron codes for microscopy [18]. A number of computations have been incorporated in Fig. 2. For the simple case of a positron with kinetic energy E the Bethe treatment gives

$$S = \frac{4\pi e^4}{m_e v^2} Z_2 \ln \frac{m_e v^2}{I} . \tag{12}$$

For an electron, where one cannot distinguish between a scattered electron and a 'true' secondary one, the stopping cross-section becomes slightly modified:

$$S = \frac{4\pi e^4}{m_e v^2} Z_2 \ln \frac{a m_e v^2}{2I} , \tag{13}$$

where a = 1.1658 in the non-relativistic limit.

The stopping force up to 10 keV has been compiled by Schou in Fig. 2 [19]. The Bethe limit is seen clearly at high energies at which $S \sim Z_2/E$. However, at low energy the Bethe formula (13) is no longer valid, and the stopping is largely determined by density of free electrons in the material (for example, aluminium has the largest stopping because of the high density of free (conduction) electrons).

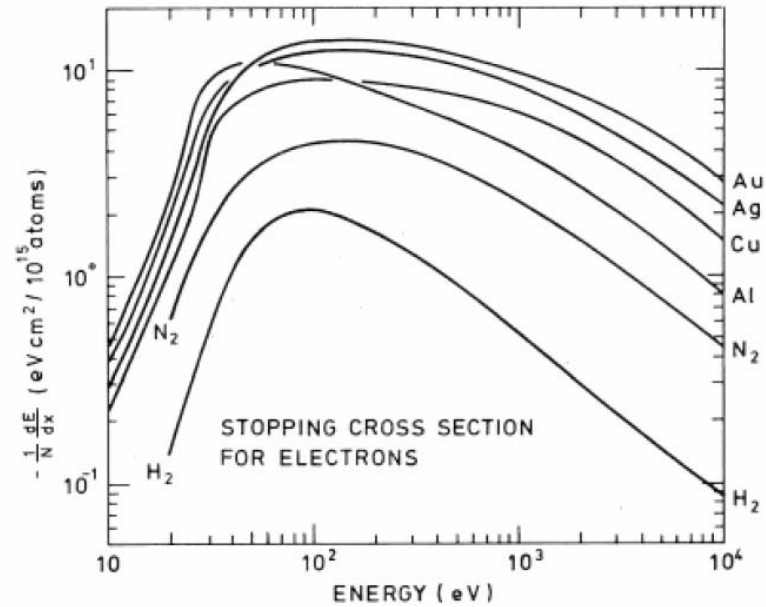


Fig. 2: Stopping cross-section for electrons in selected elements. From J. Schou [20]. (A part of the figure is published in Ref. [19].)

Range and stopping force calculations for electrons of energy higher than 10 keV have been performed by Berger *et al*. [21]. For energies above 10 keV, relativistic effects play a major role and the stopping cross-section falls off until about 1 MeV. Above 1 MeV the radiative component of the stopping force becomes dominant, and the total stopping increases strongly (Fig. 3). The energy loss by collisions is proportional to Z_2 and increases logarithmically with the energy, whereas the loss to radiation largely is proportional to Z_2^2 and increases practically linearly with the energy E [22].

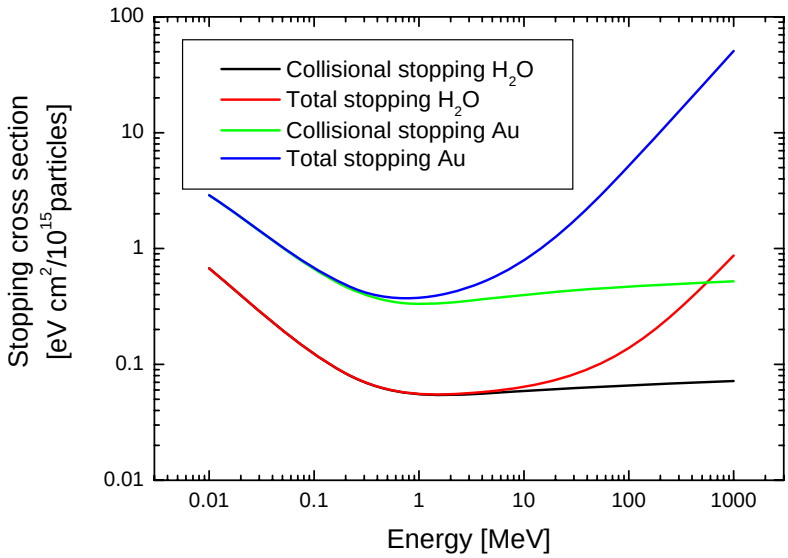


Fig. 3: Stopping of MeV electrons in liquid water and gold on the basis of the data in Ref. [21]

2.3 Energy deposition

The energy deposited by an ion is distributed in electronic excitations and kinetic energy that ends up in atomic motion and damage. Since the recoils may have a considerable range, the energy may be transported far away from the collision point. This means that the energy deposition can be substantially different from an energy-loss distribution. The theoretical work on the spatial distribution $F_D(E)$ of energy deposited in atomic collisions was initially performed by Winterbon *et al.* [23], and later published as extensive tabulations [24]. Distributions $D_e(E)$ for the electronically deposited energy by ions and their recoils were computed by Schou [25].

The energy deposited by electrons, e.g., in the keV range, is usually quite broad with a Gaussian profile, because of the pronounced scattering of the electrons. Deposition profiles have been published by Valkealahti *et al.*, [26]. Similar profiles for MeV electrons have been computed by Andreo [27].

3 Sputtering

3.1 Erosion processes

Material ejection from a solid can be induced by a number of processes which may even operate simultaneously (Fig. 4):

a) beam-induced evaporation,
b) collisional (elastic, knock-on) sputtering,
c) electronic sputtering, and
d) desorption of thin layers.

Sputtering is the erosion of material by single-particle impact in contrast to evaporation produced by many particles in beam-induced heating. Desorption is strictly speaking the removal of parts of a monolayer or full monolayers deposited on a different substrate. It is closely related to sputtering by electronic transitions, i.e., electronic sputtering. In the following only b) and c) will be treated. Beam-induced evaporation is treated by Schou in Ref. [28] and desorption in Ref. [6].

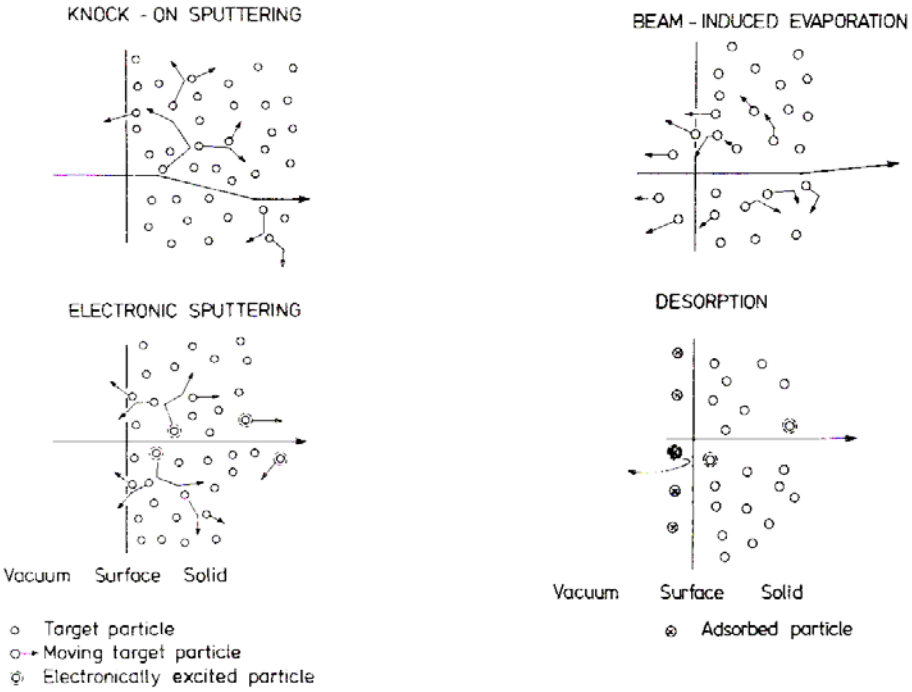


Fig. 4: A survey of different erosion processes. From Schou [29].

Reference [1] is still a classical reference for sputtering, while a more recent collection of papers on collisional and electronic sputtering can be found in Ref. [4].

The energy E deposited by a primary ion can be divided into the energy which is deposited into electronic excitations $\eta(E)$ and into atomic motion $\nu(E)$ according to Ref. [9]. These quantities can be expressed as the integral over the distributions $F_D(E,x)$ and $D_e(E,x)$:

$$E = \int F_D(E,x)\mathrm{d}x + \int D_e(E,x)\mathrm{d}x = \nu(E) + \eta(E) \ . \tag{14}$$

As mentioned above, the equation reflects the fact that the primary particle may generate recoils which undergo nuclear stopping as well, such that energy is transported away from the point of collision. It also means that the surface value $F_D(E,0) << NS_n(E)$ in many cases. In the case of very light projectiles the probability for backscattering increases strongly, and in this case $F_D(E,0) >> NS_n(E)$.

3.2 Collisional (elastic) sputtering

This type of sputtering is induced by the momentum transfer from the primary particle to the target atoms. Since the atoms are hit directly, the terms knock-on and ballistic sputtering have been used. A picture of the processes is shown in Fig. 4. The primary particle initiates a collision cascade via secondary or higher order collisions. Since the kinetic energy of the atoms set in motion is much larger than the binding energy of the material except for the last stage of the cascade, the standard treatment for metals can be largely extended to semiconductors and insulators.

The standard theory for sputtering is Sigmund's analytical theory [30] which is based on Boltzmann transport theory. The solution is obtained in the limit for high primary energy compared with the instantaneous energy of the cascade atoms. This corresponds to isotropic motion of the atoms in the solid. The treatment accounts for low collision densities in which the struck atoms in the cascade are at rest before the collision.

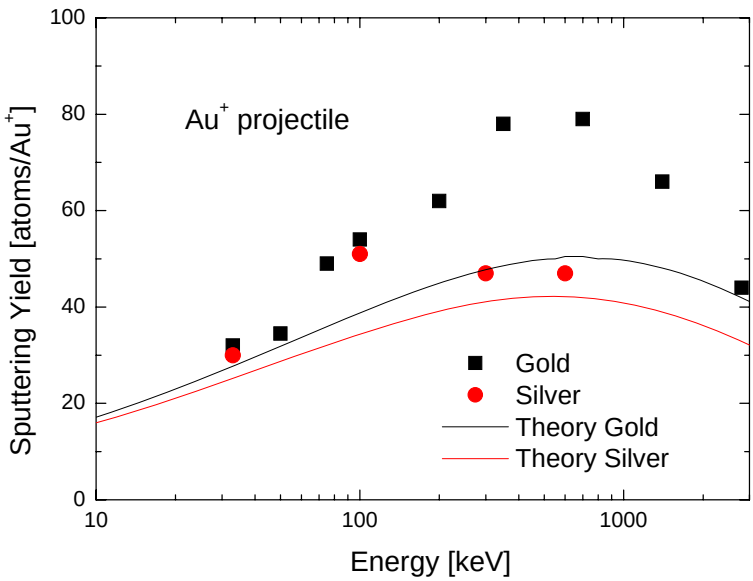


Fig. 5: Sputtering yield Y for Au ions on Au (black squares and line) and Ag (red circles and line). The solid lines are the predicted sputtering yield from Eq. (18). The deviations from theory occur in the so-called spike region, where the nuclear stopping is largest, in particular for the case Au+ ions on Au. Data from Bouneau *et al*. [31].

The (back)sputtering yield Y from a plane surface is given as

$$Y = \Lambda F_D(E,x) = 0 \ , \tag{15}$$

where $F_D(E,0)$ is the surface value of the spatial distribution of energy deposited into atomic motion. The constant Λ depends only on the material properties, the atomic density N, and the sublimation energy U_0 :

$$\Lambda = 3/(4\pi^2 N C_0 U_0) \ . \tag{16}$$

The cross-section C_0 ($\cong 1.81 \times 10^{-20}$ m^2) originates from a low-energy interatomic Born–Mayer potential and is common for all target materials in this approximation [1,30].

From dimensional arguments one may write

$$F_D(E, 0) = \alpha N S_n(E) \ , \tag{17}$$

where α is a function of the mass ratio M_2/M_1 (and of the angle θ of incidence which is not mentioned explicitly here). Here α is relatively insensitive for variations in the primary energy E, and can be well approximated by $\alpha \cong 0.17$ for most beam–atom target combinations, where $M_1 >> M_2$. This low value compared with unity also means that most of the energy deposited by the primary ion is transported away from the target surface by recoil atoms as discussed in the previous section. By combining Eqs. (16) and (17) one arrives at the well-known formula for the sputtering yield, in which no adjustable parameters enter [1,30]

$$Y = \Lambda \alpha N S_n(E) \ . \tag{18}$$

The linear-cascade theory has convincingly made it possible to predict important features of sputtering fairly well, and no other theoretical treatment has reached a comparable level [1,4,30]. Two beam atom–target atom combinations based on recent data are shown in Fig. 5.

An important exception from linear cascade theory occurs if the recoil cascades become too dense. This case, the spike regime, is characterized by a high collision density such that a moving atom in a cascade hits other atoms which have already been set in motion. The 'spike' yield is typically much larger than that from a linear collision cascade [1]. This is also seen in Fig. 5 for Au ions incident on the heaviest target material, gold, from 80 keV up to 2500 keV.

3.3 Electronic sputtering

Sputtering by electronic transitions, i.e., electronic sputtering, takes place primarily for insulating materials. A typical case is a volatile, frozen gas bombarded by light ions or electrons, but room-temperature insulators may also exhibit electronic sputtering [28,32]. The atomic motion is generated from repulsive potentials that arise during electronic de-excitation. The processes depend specifically on each material and can be extremely complicated for chemical compounds. The mechanism of electronic sputtering for the solid rare gases is now well understood, but the processes have not been fully identified even for a comparatively simple frozen material such as water ice [33].

In analogy to Eq. (16) the yield can be expressed by

$$Y = \Lambda (1/2) D_e(E, x = 0)(E_s/W) \ , \tag{19}$$

where $D_e(E, 0)$ is the surface value ($x = 0$) of the energy deposited into electronic transitions, E_s the energy produced by repulsive processes, typically a few eV, per ion generated by the primary ion, and W the energy required to produce an electron–ion pair (which is usually about 30 eV). Since the energy release from these non-radiative transitions is completely isotropic, one arrives at the factor (1/2) in Eq. (19). The equation is based on the assumption that the energy E_s is sufficiently large to generate a low-energy cascade. The factor (E_s/W) is usually much less than unity, since a considerable fraction of the energy is lost by luminescence from radiative transitions. Formula (19) cannot describe sputtering from a solid with mobile excitations, e.g., solid rare gases, but can give a qualitative trend for other frozen gases [28].

References

[1] *Sputtering by Particle Bombardment*, Ed. R. Behrisch (Springer, Berlin–Heidelberg, 1981).

[2] P. Sigmund, *Particle Penetration and Radiation Effects* (Springer, Berlin–Heidelberg, 2006).

[3] P. Sigmund, *Stopping of Heavy Ions* (Springer, Berlin–Heidelberg, 2004).

[4] *Fundamental Processes in Sputtering of Atoms and Molecules* (SPUT 92), Mat. Fys. Medd. Dan. Vid. Selsk. **43** (1993) 1, Ed. P. Sigmund.

[5] W. O. Hofer, in: *Fundamental Electron and Ion Beam Interactions with Solids for Microscopy, Microanalysis and Microlithography*, Eds. J. Schou, P. Kruit and D.E. Newbury (Scanning Microscopy International, AMF O'Hare – Chicago, 1990), p. 265.

[6] *Desorption Induced by Electronic Transitions, DIET1*, Eds. N. H. Tolk, M. M. Traum, J. C. Tully and T. E. Madey (Springer, Berlin–Heidelberg, 1983).

[7] H. H. Andersen and J. F. Ziegler, *Hydrogen Stopping Powers and Ranges in All Elements* (Pergamon, New York, 1977).

[8] N. Bohr, Mat. Fys. Medd. Dan. Vid. Selsk. **18**, no. 8 (1948).

[9] J. Lindhard, V. Nielsen, M. Scharff and P. V. Thomsen, Mat. Fys. Medd. Dan. Vid. Selsk. **33**, no. 10 (1963).

[10] J. F. Ziegler, SRIM (www.srim.org) 2006.

[11] J. Lindhard, V. Nielsen and M. Scharff, Mat. Fys. Medd. Dan. Vid. Selsk. **36**, no. 10 (1968).

[12] J. F. Ziegler, *Handbook of Stopping Cross Section for Energetic Ions in All Elements* (Pergamon, New York, 1980).

[13] J. Lindhard and M. Scharff, Phys. Rev. **124** (1961) 128.

[14] H. Paul, Stopping power graphs (www.exphys.uni-linz.ac.at/stopping) 2006.

[15] P. Sigmund, to be published.

[16] P. Sigmund, in: *Radiation Damage Processes in Materials*, Ed. C. H. S. Dupuy (Nordhoff, Leyden,1975), p. 3.

[17] H. Paul and A. Schinner, Atomic Data and Nuclear Data Tables, **85** (2003) 377.

[18] R. Gauvain, CASINO - (monte CArlo SImulation of electroNs in sOlids) (www.gel.usherb.ca/casino/) 2006.

[19] J. Schou, Scan. Micr. **2** (1988) 607.

[20] J. Schou, Thesis, Risø National Laboratory (1979) (unpublished).

[21] M. J. Berger, J. S. Coursey, M. A. Zucker and J. Chang (2005), *ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions* (version 1.2.3). [Online] Available: http://physics.nist.gov/Star [2006, September 25]. National Institute of Standards and Technology, Gaithersburg, MD. Stopping-power and range tables for electrons, protons and helium ions (physics.nist.gov/PhysRefData/Star/Text/) 2006.

[22] H. A. Bethe and J. Ashkin, in: *Experimental Nuclear Physics*, Ed. E. Segré (Wiley, New York, 1953), pp.166–357.

[23] K. B. Winterbon, P. Sigmund and J. B. Sanders, Mat. Fys. Medd. Dan. Vid. Selsk. **37**, no. 14 (1970).

[24] K. B. Winterbon, *Ion Implantation Range and Energy Distributions,* Ed. D. K. Brice (Plenum, New York, 1975).

[25] J. Schou, Phys. Rev. B **22** (1980) 2141.

[26] S. Valkealahti, J. Schou and R. M. Nieminen, J. Appl. Phys. **65** (1989) 2258.

[27] P. Andreo, Nucl. Instrum. Methods, B **51** (1990) 107.

[28] J. Schou, Nucl. Instrum. Methods, B **27** (1987) 188.

[29] J. Schou, in: *Stucture-Property Relationships in Surface-Modified Ceramics*, Eds. C. J. McHargue, R. Kossowsky and W. O. Hofer (Kluwer, Dordrecht, 1989), p. 61.

[30] P. Sigmund, Phys. Rev. **184** (1969) 383 and **187** (1969) 768.

[31] S. Bouneau, A. Brunelle, J. Depauw, D. Jaquet, Y. Le. Beyec, M. Pautrat, M. Fallavier, J. C. Poizat and H. H. Andersen, Phys. Rev. B **65** (2002) 144106.

[32] R. E. Johnson and J. Schou, Mat. Fys. Medd. Dan. Vid. Selsk. **43** (1993) 403.

[33] R.A. Baragiola, R. A. Vidal, W. Svendsen, J. Schou, M. Shi, D. A. Bahr and C.L. Atteberrry, Nucl. Instrum. Methods, B **209** (2003) 294.

The development of ultrahigh and extreme high vacuum technology for physics research*

H.F. Dylla
Thomas Jefferson National Accelerator Laboratory, Newport News, Virginia 23606, USA and Departments of Physics and Applied Science, College of William and Mary, Williamsburg, Virginia 23185, USA

Abstract
Over the last 50 years increasingly larger and more sophisticated devices have been designed and put into operation for the study of particle and nuclear physics, magnetic confinement of high-temperature plasmas for thermonuclear fusion research, and gravity wave observatories based on laser interferometers. The evolution of these devices has generated many developments in ultrahigh and extreme high vacuum technology that were required for these devices to meet their operational goals. The technologies that were developed included unique ultrahigh vacuum vessel structures, ultrahigh vacuum compatible materials, surface conditioning techniques, specialized vacuum pumps and vacuum diagnostics. Associated with these technological developments are scientific advancements in the understanding of outgassing limits of UHV-compatible materials and particle-induced desorption effects.

1 Introduction

The development of the science and technology of ultrahigh vacuum over the last 50 years has been strongly coupled to the development of increasingly larger and more sophisticated devices for physics research, such as particle accelerators, magnetic fusion devices and gravity wave observatories. This coupling has been bi-directional—sometimes innovations in vacuum technology have driven machine development, and sometimes the technology transfer has been reversed. In the last decade, several large physics research projects were initiated that are very dependent on frontier vacuum technology such as the Large Hadron Collider (LHC), currently under construction at CERN in Geneva, and the International Tokamak Engineering Reactor (ITER), which is continuing engineering design evolution at several international design centres, and several gravity wave observatories currently being commissioned across the globe. These are the current offspring in an evolution that began 80 years ago for particle accelerators, more than 50 years ago for magnetic fusion, and more than 30 years ago for gravity wave observatories.

During this evolution a wealth of ultrahigh vacuum technology has been developed for or incorporated into these machines, including unique vacuum vessel designs, vacuum sealing techniques, high performance vacuum materials, UHV and XHV cleaning and conditioning techniques, vacuum pumps, vacuum instrumentation and vacuum controls. This article highlights key scientific and technical advances driven by the necessity for appropriate vacuum environments for accelerating particle beams, for heating and confining high-temperature plasmas, and for maintaining laser interferometers sufficiently noise-free to detect the extremely weak vibrations of gravity waves.

* Originally published in J. Vac. Sci. Technol. A21 (2003) S25–S33.

2 Particle accelerators and storage rings

2.1 The evolution of particle accelerators and the need for UHV

A well-known diagram of the performance limits of various types of particle accelerators was first drawn by Stanley Livingston in his 1954 book *High Energy Accelerators* [1]. An extension of the original plot (Fig. 1) shows acceleration limits increasing by an order of magnitude every six years. The predicted trend remained surprisingly accurate until 1985, largely because of the incorporation of new acceleration schemes roughly once a decade and the innovation of colliding beams versus fixed-target experiments.

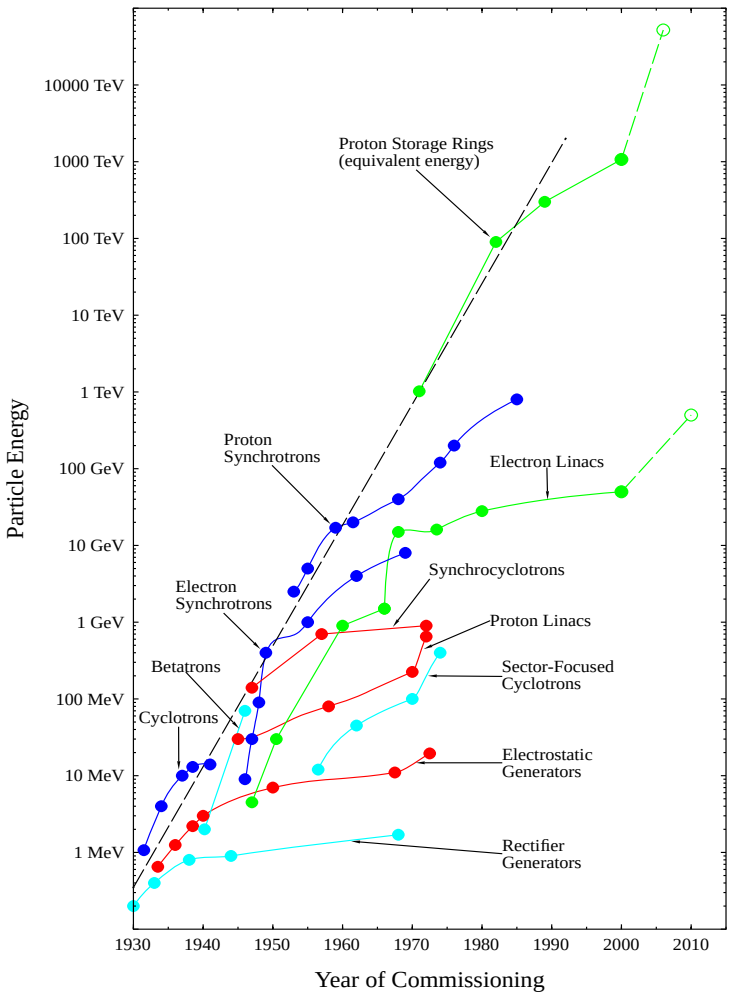


Fig. 1: The 'Livingston Curve' of the evolution of accelerator performance (1954), from Ref. [1] with an update by G. Krafft, Jefferson Lab (2003)

The family of accelerators that existed when Livingston first published his evolutionary diagram (i.e., electrostatic, cyclotrons, betatrons and synchrotrons) did not place significant demands on vacuum technology. As the machines grew larger, their design gave engineers the opportunity to incorporate cost-effective vacuum vessel fabrication techniques and vacuum pumping systems. The first incentive to incorporate UHV technology into accelerator designs came with the invention [2] of the storage ring based on a concept of Gerald K. O'Neill in 1956. O'Neill proposed a means of storing accelerated particles in a ring of strongly focused guiding magnets. Particles would be accelerated in a conventional synchrotron or linac and then injected into the storage ring. The full momentum transfer advantage of colliding beams could be obtained by intersecting counter-rotating beams from two

identical storage rings. The advantage of intersecting beams for achieving high centre-of-mass energies was suggested by Kerst et al. [3] at about the same time as O'Neill's description of the storage ring.

In O'Neill's first paper [2] on the storage ring concept, he estimated that storage times would be a few seconds in the typical (~10^{-7} torr) high vacuum environment. A follow-up paper published two years later by O'Neill and Woods [4] noted that if "vacuum technology already developed in thermonuclear power research [5] were employed, storage times would be hours." O'Neill began a collaboration with physicists at Stanford to construct two 500 MeV electron storage rings. The Stanford–Princeton storage ring completed in 1961, along with a contemporary machine—the 250 MeV electron storage ring AdA, built at the Frascati laboratories in Italy [6]—were milestones in accelerator machine physics. These were the first in a long succession of storage rings and colliders to be built over the next three decades. At least two important and related lessons were learned from the operation of the Stanford–Princeton collider which would affect the subsequent development of these machines. UHV conditions, and hence useful electron storage times, could not be maintained in the stainless-steel vacuum chamber because of contamination from the appendage oil diffusion pumps. Second, a new phenomenon was recognized: a large dynamic outgassing effect was observed in the presence of stored electron beam that was attributed by O'Neill [7] to a two-step, gas desorption process, whereby the incident synchrotron radiation generates photoelectrons responsible for the gas desorption. The effect has subsequently been well studied by numerous groups [8] because it is the dominant gas load to be dealt with in the design of electron or positron storage rings.

The next storage ring built at Stanford, SPEAR, the Stanford Positron–Electron Asymmetric Ring [9], and the first proton storage ring, the Intersecting Storage Rings (ISR) [10] built at CERN, are significant milestones in both particle physics and UHV development. In general, storage ring vacuum chambers, because of their lengths and narrow cross sections, are highly conductance-limited. Dynamic beam-induced vacuum problems could not be solved simply by increasing the number of appendage vacuum pumping systems. Innovations were necessary to meet both performance and realistic construction cost goals involving: (1) new vacuum vessel design and fabrication techniques, (2) vacuum vessel conditioning techniques, (3) distributed vacuum pumps, and (4) UHV compatible synchrotron-radiation absorbers.

The required dynamic pressures ($< 10^{-9}$ torr) are approximately the same for achieving stable high current (≥ 0.1 A) stored beams with reasonable storage times (~ hrs.) for both proton and electron rings—even though the offending gas-phase and surface interactions are different in the two cases. For proton rings, low gas pressures are required to minimize scattering of the beam and to minimize an instability that was first observed in the ISR. With insufficient vacuum and insufficiently cleaned vessel surfaces, ionized residual gas species repelled by the beam impact the surface, releasing more gas by ion-induced desorption. At pressures greater than 10^{-9} torr, the effect can avalanche, leading to quenching of the beam current [11]. Careful measurements of ion-induced desorption coefficients by Mathewson and Achard [12] show desorption coefficients greater than unity for unconditioned metal surfaces. With various combinations of vessel bakeout and discharge cleaning, ion-induced desorption coefficients can be lowered by two to three orders of magnitude.

The ISR, operated from 1971–1984, was a groundbreaking machine for both particle physics and the science and technology of ultrahigh vacuum. Proton beams with currents up to 20 A at energies up to 28 GeV were stored in a pair of 1 km diameter rings. The ISR incorporated the best UHV techniques known at the time and the device was well instrumented to characterize the vacuum environment [13,14]. The ISR vacuum group developed several vessel material conditioning techniques that were used on many subsequent machines at CERN and elsewhere in many other vacuum applications to optimize UHV performance. As a result of a study of the diffusivity of hydrogen in stainless steel [15], sheet stock used for vessel fabrication was baked in vacuum for 2 hours at 900°C, lowering the hydrogen content by an order of magnitude. Significant efforts were

devoted to developing and qualifying the effectiveness of chemical recipes for cleaning vacuum materials [16] and to developing discharge cleaning methods using argon and argon/oxygen mixtures [17]. These surface conditioning studies were extended to aluminium surfaces [18] in the 1980s as CERN focused on the design of the 27 km Large Electron–Positron (LEP) collider which operated from 1989 to 2000 [19].

At about the time the ISR device was being commissioned, the SPEAR device [20] at Stanford became operational (1972). Many of the design innovations developed for this collider were incorporated in subsequent electron–positron colliders and storage rings. Dynamic pressure requirements of 10^{-9} torr are also necessary for electron or positron storage rings, but in this case the primary gas scattering mechanism is bremsstrahlung losses in electron–molecule collisions [21]. The beam lifetime determined by this scattering mechanism scales inversely with molecular mass, thus putting a penalty on the higher molecular weight components of the residual gas. As mentioned earlier, the source of dynamic gas load is the synchrotron-radiation-induced desorption. This phenomenon is now well understood in terms of gas desorption by photoelectrons generated by synchrotron radiation. Measurements to study the phenomenon were first done with electron sources to simulate the photoelectron emission, including the early measurements by the SPEAR vacuum design team. Later, actual measurements of photo-induced desorption coefficients were obtained using beam lines on storage ring light sources by the vacuum groups at CERN [8, 19], Brookhaven [21], and KEK [22]. Measured desorption coefficients start out in the range of 10^{-2} to 10^{-1} depending on whether the substrate is stainless steel, aluminium or copper, and generally decrease inversely with the photon dose to 10^{-5} or 10^{-6} after a beam exposure that is typically 1 to 10 ampere-hours, corresponding to a photon dose of 10^{22} to 10^{23} photons per metre of vacuum vessel. This 'beam cleaning' effect has turned out to be the most efficient *in situ*, surface-conditioning technique for electron storage rings. The synchrotron emission which generates the photo-emitted electrons, and the resulting photoemission rate, can be calculated for a given set of electron beam parameters and knowledge of the vacuum vessel geometry. Combining the results of these calculations with the measured photo-induced desorption coefficients allows the designers of new machines to design the vacuum chambers and associated pumping systems. Some effort has been made to model [23] the dose dependence of the photo-induced desorption which is clearly rate-limited by the diffusion of H, C, and O species from the near-surface of the bombarded metal.

2.2 Storage ring vacuum vessel and pumping system developments

The large dynamic gas loads that are present in electron storage rings have driven the development of unique vacuum chambers and associated pumping systems in order to meet the UHV operating requirements. Solutions adopted for the SPEAR storage ring [24] were innovative and became prototypical for the next three decades of storage ring design. The use of a simple pipe-like geometry for accelerator vacuum chambers with appendage pumping ports spaced at convenient intervals between magnet gaps is sufficient for the modest vacuum requirements of most circular accelerators. Typically, pumping speeds per unit length of chamber of the order of 1–10 ℓ/s m can be obtained with appendage pumping. To cope with the synchrotron-radiation-induced gas load in an electron storage ring, so-called distributed pumps were placed within the beam chambers based either on sputter ion pumps placed within the bending magnet field or non-evaporable getters (NEGs).

The innovative SPEAR vacuum chamber was fabricated using an aluminium (type 6061) extrusion technique [24]. This was a cost-effective solution for producing a long (200 m), multi-chambered, non-concentric vessel. A cooling channel and ridged surface were extruded in the outer face of the chamber to deal with the localized heat load from incident synchrotron radiation. Many storage rings built after SPEAR, particularly those specifically designed as synchrotron light sources, have taken advantage of extruded aluminium chambers, because of the ease of fabrication and the ability to deal with the synchrotron radiation heat loads with co-extruded cooling channels (and the intrinsic high thermal conductivity of aluminium). Extruded chamber designs made after SPEAR often

included a multi-chamber arrangement that allowed the beam chamber to be isolated from a chamber that contained the distributed pumping elements in order to minimize beam impedance effects. Beam chambers with a uniform cross section present the minimal impedance for transmission of high-intensity particle beams. A large impedance can lead to heating of vessel components and induced wakefields that can affect beam trajectories [25]. The use of aluminium for storage ring vacuum chambers has generated a wealth of related development for the use of aluminium alloys in vacuum, such as effective surface cleaning methods [18] aluminium–stainless-steel bonding methods [24, 26], vacuum certified aluminium welding and aluminium ultrahigh and extreme high vacuum components [27].

As the synchrotron radiation power was increased in storage rings designed as dedicated light sources, specific synchrotron radiation absorbing structures had to be designed for the dipole magnet chambers to handle the localized heat loads. These structures have been as simple as copper bars brazed to the radiation plane of stainless-steel chambers and as sophisticated as water-cooled absorbers with Cu, C or Be as the radiation absorbing material [28].

The development of two classes of distributed pumps for storage rings has had a significant impact on the further development of both ion pumps and non-evaporable getters. The innovative use in SPEAR of the dipole magnet fields as the confining field for ion pump Penning cells was an extension of the studies by Schurrman [29] and originally Penning [30] of the magnetic field dependence of cross-field gas discharges. Malev and Trachtenberg of Novosibirsk [31], and later Hartwig and Kouptsidis of DESY [32] developed specific formulas applicable to the design and performance of distributed ion pumps in both low and high field situations.

For the 27 km Large Electron–Positron (LEP) storage ring at CERN, a more cost-effective distributed pumping scheme than the use of *in situ* ion pumps was required. LEP was the first to incorporate non-evaporable getter (NEG) pumps within the vacuum chamber as the primary pumping element. Impressive design analyses and prototype testing of the selected NEG system (SAES type 101, ZrAl alloy) were performed by CERN under the direction of C. Benvenuti [33]. As a result, the system has performed well for the entire lifetime of LEP (1989–2000) [19]. (See Fig. 2). Numerous other storage rings [34, 35] had incorporated NEGs as the distributed pumping element even before the start-up of LEP. The most recent developments on the design and use of innovative distributed pumping involve the use of getter materials (Ti,Vn and Nb alloys) that are evaporated onto most of the surface area of the vacuum pipes leading to a cost-effective, simple and high specific pumping speed solution to this engineering problem [36, 37].

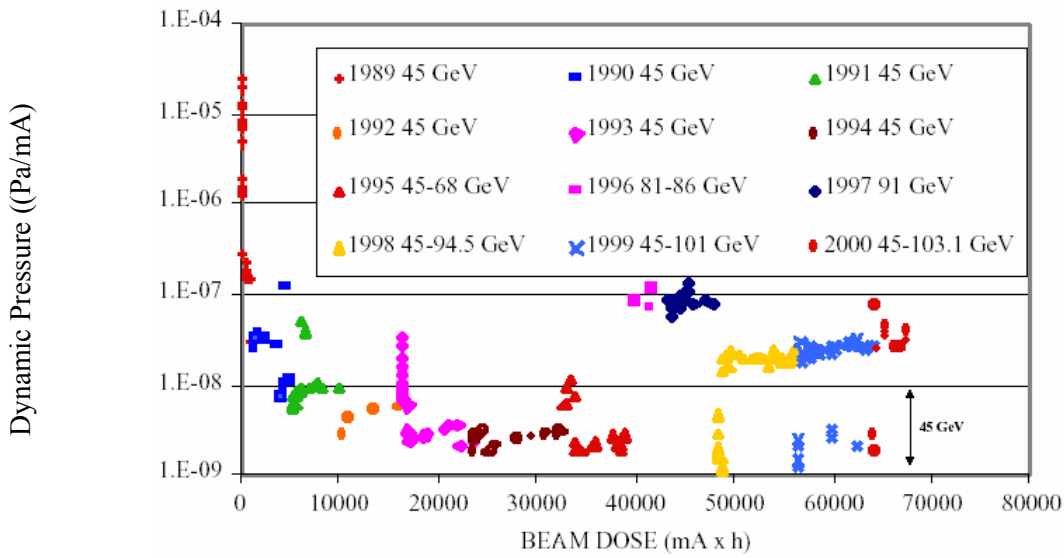


Fig. 2: The dynamic pressure evolution in CERN's Large Electron–Proton (LEP) Collider over the 10 year operation of the device (Ref. 19)

2.3 Cold-bore accelerators

What could very well be the last members of the family of colliders for high-energy particle physics research are represented by the Superconducting Super Collider (SSC) in the U.S. (a project cancelled in 1993), and the Large Hadron Collider (LHC) at CERN (which is under construction at CERN with projected completion in 2007). The enormous cost for these machines (> 11B$ for the SSC and > 4B$ for the LHC), which is the result of their size and need for thousands of state-of-the-art superconducting magnets, often outweighs discussions of the scientific benefits. The design work that has been accomplished for the SSC and LHC has already contributed to the UHV knowledge base because of the particular problems associated with storage ring vacuum chambers that must operate at liquid helium temperatures [38,39]. To minimize magnet costs the LHC beam chamber is relatively small in diameter (~5 cm) and is thermally shunted to the magnet temperature (1.8 K for the LHC).

Normally, the synchrotron radiation emission from the LHC proton beam would not be a problem in producing a dynamic gas load if the vacuum vessel was at ambient temperature. The photo-desorbed gases will not be readsorbed on nearby surfaces and would have a high probability of being removed from the gas phase by a modest array of distributed pumps. The case of photo-induced desorption and readsorption from surfaces near 4 K is quite different. All the desorbed species of interest (CO, CO_2, CH_4 and H_2O) except for H_2 are pumped well and have low equilibrium vapour pressures on 4 K surfaces. Isotherm measurements on clean stainless steel at low temperatures show that H_2 is pumped and the H_2 vapour pressure remains low ($<10^{-11}$ torr) only if the H_2 surface coverage remains below a monolayer [40, 41]. Dynamic pressures for the operation of LHC must remain below the 10^{-10} torr range in order to prevent excessive beam scattering and excessive heat load on the magnet cryostat as a result of the beam scattering. A design feature within the cold-bore vacuum chambers to prevent the dynamic pressure from exceeding the 10^{-10} torr limit is the incorporation of an intermediate temperature liner (@ ~ 20 K) between the beam aperture and the cold vessel wall. This liner would intercept the synchrotron radiation and be partially slotted or perforated to allow desorbed H_2 molecules into the interspace. The 1.9 K operating temperature of the LHC cold bore is sufficiently low to pump H_2 above monolayer quantities without exceeding the dynamic vacuum limits [42]. Related studies of photo-induced desorption phenomena at low temperatures (1.8–20 K) add to our basic understanding of this important process and have application to other scientific and technical problems such as the formation of molecules on interstellar dust grain surfaces [43], and the development of high-capacity cryopumps [44].

2.4 Superconducting accelerators

In addition to the use of superconducting magnets for saving power in accelerator operations, the other significant use of superconductivity in accelerator technology is RF acceleration cavities [45]. Linear accelerators are comprised primarily of an array of RF cavities, separated by occasional magnetic elements for correcting beam optics. In synchrotrons, the energy that the circulating beam loses due to synchrotron radiation and other processes is restored by RF cavities interspersed among the guiding magnets. For both applications, superconducting RF cavities save operating costs, offer lower impedance to the accelerating beam, and can be operated CW at high gradients. The disadvantages of superconducting RF cavities are the cost and complication of cryogenic cooling and the smaller operating experience. The latter concern has been alleviated with the operation of the CEBAF superconducting linac at Jefferson Lab in Newport News, Virginia [46] since 1993 and other large installations of operating superconducting RF systems at KEK [47], DESY [48], and LEP [49]. Presently, a 2 MW superconducting linac system is coming on-line for the Spallation Neutron Source at ONRL [50].

Over 30 years of development has been invested in superconducting RF technology, resulting in the current state of the art which is capable of fabricating production-scale cavity assemblies with accelerating gradients greater than 25 MV/m [51]. These high-performance cavities are also finding use in compact CW linacs for driving free electron lasers [52], which are being used for basic research

and industrial applications such as materials processing where high average power and energy-efficient UV and IR light sources are needed [53]. The successful development of superconducting RF cavities is strongly tied to UHV technology. High performance and reliable cavities require extremely careful cleaning and surface treatments, assembly in an environment free of particulate contamination, and vacuum sealing techniques which are UHV compatible [54,55].

2.5 Next generation accelerators

The evolution of accelerators as shown on the 'Livingston Plot' (Fig.1) was punctuated by new technologies that allowed performance to grow at affordable costs. Have the circular colliders reached this limit with the LHC? To provide alternative technologies for the next generation of particle accelerators, international collaborations are exploring designs options for the Next Linear Collider (NLC) project involving a pair of 250 GeV linacs for e^+e^- collisions. Several teams are investigating warm accelerating structures from 3 to 30 GHz [56], and a team at DESY has fabricated a prototype superconducting linac [57]. The primary problems with the warm structures, which are largely based on the success of the Stanford Linear Collider (which became operational in 1980 [58] shortly before LEP), are the extremely tight alignment specifications, and the observed erosion of prototype NLC vacuum chambers due to the intense, short-pulse electric fields. The primary problem with the superconducting option is the projected cost of the superconducting cavity/cryostat structures. As in most endeavours competition is healthy, and the design and prototyping exercises have already generated valuable results in vacuum vessel materials, RF and cryogenic engineering.

3 Magnetic fusion development

3.1 The roadmap

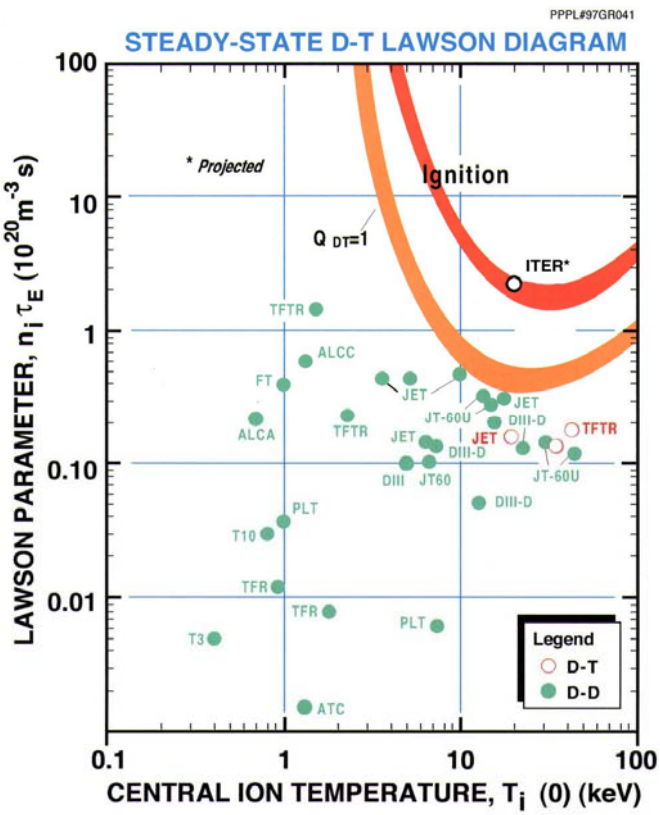


Fig. 3: The 'Lawson Criterion' for energy gain in D-T plasmas, courtesy D. M. Meade, Princeton Plasma Physics Laboratory, reprinted from Ref. [93] with permission

The development of devices to harness thermonuclear fusion for power generation has an analogous road map to the 'Livingston Plot' called the 'Lawson Plot' based on John Lawson's calculations [59] of the minimum conditions of plasma density, plasma temperature and energy confinement necessary for energy break-even in a D-T plasma (see Fig. 3). For the first two decades of fusion research, devices were designed to magnetically confine and heat low-to-moderate density, hydrogenic plasmas. In such magnetic confinement schemes, plasmas would have to obtain densities of 10^{19} m^{-3}, temperatures of 10 keV and energy confinement times > 1 s in order to reach the Lawson condition. This appeared to be a rather straightforward problem for the physicists and engineers who started this field of research over 50 years ago; however, it has turned out to be one of the most difficult scientific and technical challenges of our times. A sustained effort over the last five decades from practitioners in Europe, Japan, Russia and the US has seen the plasma triple-product rise over ten orders of magnitude from the primitive, small devices of the 1950s to within 50–80% of the Lawson criterion using the gigantic tokamaks at Princeton [60] and Abingdon, England [61] in the 1990s. UHV technology has played a significant role in the development of magnetic fusion devices.

3.2 First-generation magnetic fusion studies

The possibilities of fusion power or 'controlled thermonuclear reactors' were clearly in the minds of physicists who witnessed the first man-made thermonuclear reactions with the explosion of the first D-T weapons. In the U.S. a secret programme of fusion research was launched in the early 1950s code-named Project Sherwood [62]. This programme had proponents who championed three different schemes for magnetically confining plasmas. Lyman Spitzer from Princeton University, who is generally given credit for having first elucidated the basic physics of magnetic confinement, conceived a figure-eight confinement geometry called a stellarator. A solenoidal field would confine the plasma, an inductively driven current would heat the plasma, and the twist in the field geometry would compensate for the tendency for charged particles to drift out of a simple toroidal geometry. James Tuck, a British physicist working at Los Alamos developed a scheme with his colleagues at Oxford to confine and heat plasmas by 'pinching' the plasma to a small radius stream by rapidly increasing the toroidial field. A third team, led by Richard F. Post and Herbert York, at the Livermore branch of the University of California's Radiation Lab (the ancestor of Lawrence Livermore National Laboratory), proposed to open a solenoidal confinement scheme they termed 'mirror machines'.

All of the early schemes were fraught with difficulties: the plasma confinement appeared poor, plasma temperatures were low, and the primitive vacuum technology that was used guaranteed that the plasma species of interest was overwhelmed by impurities from the confining chamber walls. When Spitzer had conceived the stellarator concept, he developed a plan for moving the technology along an orderly route from the pioneering Model A, which was a table top demonstration, to a larger Model B, designed to push the plasma parameters, to an engineering prototype (Model C) which would be a scale model reactor, to finally a full-scale prototype reactor (Model D) [63]. This plan spells out many of the important subsystems that a fusion reactor would require: an auxiliary heating system to reach plasma ignition temperatures, a magnetic 'diverter' to remove plasma impurities, and a lithium blanket for absorbing the fusion product neutrons and extracting energy from the reactor. The disappointing results coming in from the experiments on the B-series of stellarators, and similar results from the plasma pinch and mirrors programmes, would considerably stretch out Spitzer's development plan for a reactor. Model C would not be a scale model reactor—it would be the next step in elucidating the plasma physics. When the Model C was commissioned in 1961, its stainless-steel vacuum vessel became the largest UHV system built to date [5]. Special double joint, gold wire flange seals allowed the system to be baked to 450°C. After bakeout, base pressures for the system were in the 10^{-10} torr range. Because of the concerns for hydrocarbon contamination, the Model C was pumped with two large mercury diffusion pumps that were isolated from the torus by lead-sealed valves and freon-cooled traps. Mercury was chosen as the pumping fluid because its accidental presence in the torus vacuum could be detected very sensitively by plasma spectroscopic techniques, and the high-

temperature vessel bakeout would easily remove the contamination. The performance of the Model C vacuum system was a huge success—the plasma performance on the other hand was disappointing. For most of the 1960s none of three confinement schemes being pursued in the U.S. or U.K. were making any significant forward progress toward the Lawson Criterion, and neither the measurements nor the theory could tell the experimenters why their plasmas remained relatively poorly confined and cold [64].

The cold war prevented any significant exchange of information between the Western and Soviet programmes except for brief interactions at plasma physics conferences organized by the International Atomic Energy Agency (IAEA). At the 1965 and 1969 [65] IAEA conferences, the Russians claimed that their toroidal magnetic chamber or 'tokamak' geometry was producing plasmas with confinement times of milliseconds and plasma temperatures near a kilovolt. After confirmation of the Russian results by a visiting British team [66], the race to build tokamaks was on.

In the early 1970s tokamaks were under construction at Oak Ridge [67], MIT [68], the General Atomics Co. [69] in San Diego, Princeton and in Europe. The confinement properties and stability limits of the tokamak geometry were confirmed and more importantly the energy confinement time appeared to improve with the square of plasma radius. Plans were quickly drawn up for a second generation of tokamaks by scaling up a factor of two in size, followed by another factor of two increase in size which would yield a device capable of reaching the Lawson condition. The Princeton Large Torus (PLT) [70] a tokamak with a 6000 ℓ vacuum vessel and 0.5 m plasma radius came on line late in 1975. With the use of high-current, hydrogen neutral beams developed at Oak Ridge National Laboratory, PLT became the first fusion device to exceed the Lawson Criterion minimum temperature of 10 keV in the summer of 1978 [71]. Scientific and technical progress came at an exhilarating pace for the next decade.

3.3 Plasma impurities and the vacuum vessel 'first wall problem'

The achievement of the high plasma temperatures in the PLT tokamak did not come without several key developments in controlling how the plasma interacts with the vacuum vessel wall. Spectroscopic and plasma resistivity measurements on the first generation of Western tokamaks were indicating that plasma impurities were a persistent problem. The ATC tokamak, at Princeton [72] achieved in 1974 the first pure hydrogenic plasma that was undiluted with carbon and plasma impurities desorbed from the vacuum vessel inner surface or 'first wall'. This result was achieved by evaporating titanium onto a large fraction of the vessel wall prior to a plasma discharge to suppress the impurities, combined with programmed injection of hydrogen during the discharge [73].

In 1975, the first in a series of compact, high-field tokamaks built at MIT, Alcator A, also achieved pure, hydrogen plasmas [74] by a combination of UHV vessel design, programmed hydrogen gas injection, and the pioneering use of a pulse discharge cleaning technique to condition the vacuum vessel wall prior to exposure to high-temperature plasmas. The discharge cleaning technique [75], developed by Robert Taylor, was rapidly adopted by all subsequent magnetic fusion experiments. Ironically, as a combination of the discharge cleaning and Ti gettering techniques was applied to the PLT device, low-*Z* (carbon and oxygen) impurities were reduced to the point that these ions no longer cooled the edge plasma by radiation. The resulting edge plasma was sufficiently energetic to sputter excessive quantities of metal impurities from the tungsten 'limiter' structure which defined the plasma radius and protected the vacuum vessel wall. A switch of limiter material to a high-purity graphite solved the problem for PLT [76] and helped launch a development programme for low-*Z* refractory materials which would be needed for 'first-wall' structures in the succeeding generation of tokamaks [77].

The problems of impurity generation and impurity suppression in tokamaks occupied significant portions of the experimental programmes on many machines and led to several dedicated devices devoted to impurity studies [78–81]. Large-scale tests of magnetic diverter structures for controlling

the transport of both hydrogenic and impurities species were performed on the PDX tokamak at Princeton [82] and the ASDEX tokamak at Garching, Germany [83]. These two tokamaks, which became operational in the early 1980s, required very large vacuum vessels (~ 36 m^3) to contain the internal magnetic structures for the diverter and a ballast volume for pumping the plasma particles that were neutralized on the diverter target plate. The diverter volumes were equipped with large arrays of titanium evaporation sources to provide up to 2×10^5 ℓ/s of hydrogenic pumping capacity. The PDX device was also used to test a prototype non-evaporable getter array [84] (based on the SAES Zr-Al alloy) and a large-area, toroidal limiter for first wall protection fabricated from graphite tiles [85]. Because of the large size of the PDX and ASDEX vacuum vessels compared to the plasma volumes (6:1), both machines became test beds for developing hydrogen glow discharge techniques for cleaning complicated vacuum vessel structures [86, 87] that took advantage of the concurrent studies of glow discharge cleaning of accelerator structures [17].

3.4 Toward the energy breakeven demonstrations

The rapid pace of the fusion programme in the mid to late 1970s compressed and overlapped the design and operational schedules of three generations of machines. This accelerated schedule was most evident in the construction of the large tokamaks built to reach the Lawson criterion for energy breakeven demonstrations. Three large devices became the flagship experiments of the worldwide magnetic fusion programme: the Tokamak Fusion Test Reactor [60] (TFTR) at Princeton began operating in 1982; the Joint European Torus [61], at the Culham Laboratories in the U.K., began operating in 1982; and the JT-60 [88] tokamak at JAERI in Japan began operating in 1984. TFTR and JET were both designed to operate with D-T mixtures, and therefore had the complication of dealing with radioactive gas handling [89, 90] and the additional activation of the vessel [91] induced by the 14 MeV D-T neutron fusion products. Initial operations in JT-60 were confined to hydrogen plasmas, and a subsequent upgrade of the machine (JT-60U) allowed the use of deuterium fueling [92]. D-T experiments in JET and TFTR began in 1992–93. Both machines achieved impressive plasma parameters on the Lawson plot (Fig. 3) [93] with the use of 20–30 MW of plasma heating [94–98] and frozen hydrogen pellet injection systems [99]. Plasma parameters within 80% of the minimum Lawson requirements were achieved in addition to pushing the plasma temperature records to 3–4 times the achievements of PLT [97]. Subsequent experiments at the upgraded tokamaks JT-60U [100] in Japan and the D III D [101] device (at General Atomics in San Diego) also achieved these high temperature regimes.

The performance of these large tokamaks required continued development of first-wall materials, first-wall structures, and conditioning techniques needed for edge-plasma particle and impurity control. The TFTR vacuum vessel was protected with a toroidal array of graphite tiles that is comprised of over 2000 tiles [102]. Special discharge cleaning techniques were developed to remove the large quantities of H_2O absorbed by this structure during vents to atmosphere [103]. The highest performance discharges were achieved in TFTR after the graphite first-wall was induced to behave like a plasma pump by conditioning the structure with a regimen of special He discharges [104]. Prior to initiation of D-T discharges on TFTR, extensive studies of tritium trapping and removal mechanisms were developed to recycle tritium trapped in the vessel [105]. After operational experience was gained initially with the use of fine-grained graphite for first-wall armour, JET replaced graphite in high flux areas with Be [106], and the TFTR, DIIID and JT-60U devices upgraded to carbon-fibre-composite materials [107]. A pioneering application of a large-scale vapour deposition process was demonstrated when the entire first-wall of the TEXTOR device in Germany was coated with thin layers of boron-carbide [108]. The technique was applied subsequently to many other tokamaks (ASDEX, TFTR, DIIID) as a means of *in situ* modification of first-wall properties.

The careful control of the surface properties of the first-wall structures in large tokamaks would not be possible if high reliability 'second-wall' or vacuum containment structures were not maintained. These complicated vacuum vessels must satisfy many constraints including UHV

standards for leak-tightness despite hundreds of access ports for diagnostic instrumentation and plasma heating equipment, structural integrity to withstand cyclic loading due to bakeout cycles, and large eddy-current-induced forces [109]. The need for high-capacity, non-contaminating vacuum pumping systems for these large vessels drove the design of large (> 3500 ℓ/s) turbo pumps [110], and the associated high-power neutral beam heating systems drove the design of large He cryocondensation pumps with speeds exceeding 10^6 ℓ/s for hydrogenic species [111]. As the magnetic fusion community continues with the engineering design of the next large fusion project, ITER, the International Tokamak Engineering Reactor [112], challenging vacuum technology problems will be encountered for the design of effective first-wall protection hardware, high-capacity divertor pumping systems, and tritium delivery and recovery systems [113–115].

4 Gravity wave observatories

After more than 25 years of development of high sensitivity laser interferometers, several large gravity wave observatories based on using this technique in large vacuum systems have recently come on line and are producing their first data. These observatories involve large diameter (1–1.2 m) stainless vacuum vessels arranged in km length, perpendicularly intersecting sections to house the multi-path, laser interferometers in a Michelson configuration. The largest installation is the LIGO project in the US comprised of identical 4 km by 4 km interferometer legs at two stations near Livingston, LA and Hanford, WA [116]. Data from LIGO will be combined from similar data coming from the 2 × 3 km VIRGO station near Pisa, Italy [117], the 2 × 600 m GEO-600 station in Germany [118], and the 2 × 300 m TAMA station in Japan [119]. The vacuum requirements for these observatories are determined by the maximum tolerable light scattering due to fluctuations in the density of residual gases that would interfere with the detection of the extremely small motions ($<10^{-18}$ m) of test masses located at the ends of the interferometer arms [116].

These derived specifications put maximum partial pressure limits for H_2 at $<10^{-8}$ torr and other residual gases such as H_2O at $<10^{-9}$ torr; lower limits are put on hydrocarbons such as CH_4 ($<10^{-14}$ torr) because of an additional concern for building up carbonaceous contamination on the surfaces of the sensitive optical elements. These partial pressure requirements and the large size of the vacuum vessels drove the need for developing a cost-effective means of producing low outgassing rates in stainless steel in order to minimize pumping system costs. A relatively simple air bake procedure, demonstrated by a number of the gravity wave observatory teams, appears to satisfy these demands. Hydrogen outgassing rates below 5×10^{-15} torr l/cm^2 s and hydrocarbon outgassing rates below 10^{-16} torr l/cm^2 s were obtained after a 400 ^{o}C bake of the stainless steel in air followed by an *in situ* 150ºC vacuum bakeout [117,118,120,121]. The mechanism for reduction of the outgassing is still not determined, however, thermal desorption measurements by Bradaschia *et al*. at VIRGO [117, 121] indicate that the air bake is nearly as effective as an equivalent high temperature vacuum bake for the reduction of the low binding energy state of H_2, which is presumably responsible for the primary share of hydrogen outgassing when the system is operated at room temperature.

In addition to the requirement to minimize intrinsic outgassing of the vacuum vessel material, careful attention had to be paid to the design and implementation of the vessel welding for the multi-kilometre arms of the interferometer vessels, in order to insure that the gas load due to residual leaks and entrained gas in the welds was below the limits set for intrinsic outgassing [120]. With the use of the dual temperature bakeout procedure to lower the vessel material outgassing, and careful qualification and inspection of the extensive vessel welding, relatively modest vacuum pumping capability (provided by commercial sublimation and ion pumps) was sufficient to maintain the specified partial pressures for these large systems. Thus, the first generation of these impressive machines has come on line meeting or exceeding their critical vacuum requirements while incorporating the largest ultrahigh vacuum vessels built for any purpose.

Acknowledgements

This work was supported by U.S. DOE Contract No. DE-AC05-84ER40150.

This paper is a reprint of a review paper given at the 50th National Symposium of the American Vacuum Society that was published in J. Vac. Sci. Tech. **A21**, (2003) S25–S33. (Reprinted with permission of the American Institute of Physics.) The author thanks P. Redhead, K. Welch, M. Hablanian, R. Weiss and G. Krafft for useful comments on the manuscript.

References

[1] M. S. Livingston, *High-Energy Accelerators* (Interscience, New York, 1954).
[2] G. K. O'Neill, Phys. Rev. **102** (1956) 590.
[3] D. W. Kerst *et al.*, Phys. Rev. **102** (1956) 590.
[4] G. K. O'Neill and E. J. Woods, Phys. Rev. **115** (1959) 659.
[5] G. Lewin and D. Mullaney, *1963 Vacuum Symposium Transcripts* (Pergamon, New York, 1964), p. 176.
[6] C. Bernardini, G. F. Corazza, G. Ghigo and B. Touschek, Nuovo Cimento **18** (1960) 1293.
[7] G. K. O'Neill, *1963 Super-High Energy Summer Study, No. 8* (Brookhaven National Laboratory, Brookhaven, NY, 1993).
[8] O. Gröbner, A. G. Mathewson, H. Störi and P. Strubin, Vacuum **33** (1983) 397.
[9] SLAC Storage Ring Group, *Proc. 8th Intern. Conf. on High Energy Accelerators*, CERN, 1971 (European Organization for Nuclear Research, Geneva, 1971), p. 145.
[10] CERN Report No. AR/Int. SG/64-9, 1974 (unpublished).
[11] R. S. Calder, Vacuum **24** (1974) 437.
[12] A. Mathewson and M. H. Achard, *Proc. 3rd Intern. Vacuum Congress,* Vienna, 1977, R. Dobrozemsky *et al.*, eds. (Berger, Vienna, 1977), Vol. 2, p. 1217; also see A. Mathewson, Vuoto **17** (1987) 102.
[13] E. Fischer, J. Vac. Sci. Technol. **2** (1965) 142.
[14] E. Fischer, J. Vac. Sci. Technol. **9** (1972) 1203.
[15] R. Calder and G. Lewin, Brit. J. Appl. Phys. **18** (1967) 1459.
[16] A. G. Mathewson, in *Proc. Intern. Symp. on Plasma Wall Interactions*, Julich, 1976 (Pergamon, Oxford, 1977), p. 517.
[17] E. Fischer, Jpn. J. Appl. Phys. Suppl. **2** (1974) 199.
[18] A. G. Mathewson, M. Andritschy, A. Grillot, O. Gröbner and P. Strubin in *Surface Conditioning of Vacuum Systems*, AIP Conference Proceedings No. 199 (AIP, New York, 1990),
pp. 110–123.
[19] J. C. Billy, J. P. Bojon, O. Gröbner, N. Hilleret, M. Jimenez and P. Strubin, in *Proc. European Particle Accelerator Conference*, 2000.
[20] J. LeDuff, Nucl. Instrum. Meth. **A239** (1985) 83.
[21] H. J. Halama, in *Vacuum Design of Synchrotron Light Sources*, Y. G. Amer, S. D. Bader, A. R. Kraus and R. C. Niemann, eds., AIP Conference Proceedings No. 236 (AIP, New York, 1991), pp. 39–51.
[22] T. Kobari, M. Matsumoto, T. Ikeguchi and S. Neda, in *Vacuum Design of Synchrotron Light Sources*, Y. G. Amer, S. D. Bader, A. R. Kraus and R. C. Niemann, eds., AIP Conference Proceedings No. 236 (AIP, New York, 1991), pp. 347–354.
[23] M. Andritschky, O. Gröbner, A. G. Mathewson, F. Schumann, P. Strubin and R. Souchet, Vacuum **38** (1988) 933.
[24] U. Commings, N. Dean, F. Johnson, J. Jorow and J. Voss, J. Vac. Sci. Technol. **8** (1971) 348.

[25] A. W. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific, River Edge, NJ, 1999).

[26] J. C. Schuchman, J. Godel, W. Jordan and T. Oversluizen, J. Vac. Sci. Technol. **16** (1979) 720; also see J. C. Schuchman, J. Vac. Sci. Technol. **A1** (1983) 196.

[27] H. Ishimaru, J. Vac. Sci. Technol. **A2** (1984) 1170.

[28] Y. Morimoto, T. Shirakura, K. Konishi, S. Takahashi, S. Yokouchi and S. H. Bee, in AIP Conf. Proc. No. 236 (AIP, New York, 1991), pp. 110–116.

[29] W. Schurman, Physica (Amsterdam) **36** (1967) 137.

[30] F. M. Penning and K. Nienhuis, Philips Tech. Rev. **11** (1949) 116.

[31] M. D. Malev and F. M. Trachtenberg, Vacuum **23** (1973) 403; **25** (1975) 211.

[32] H. Hartwig and J. Kouptsidis, J. Vac. Sci. Technol. **11** (1974) 1154.

[33] C. Benvenuti, Nucl. Instrum. Meth. **205** (1983) 391; C. Benvenuti and F. Francia, J. Vac. Sci. Technol. **A8** (1990) 3864.

[34] R. F. Niemann, R. Benaroya, M. Choi, R. J. Dortwest, G. A. Goeppner *et al*., AIP Conf. Proc. No. 236 (AIP, New York, 1991), pp. 84–101.

[35] S. H. Be, S. Yokouchi, T. Nishidono, Y. Morimoto, K. Watanabe *et al*., Phys. Rev. **102** (1956) 102–109.

[36] C. Benvenuti, P. Chiggiato, P. Costa Pinto, A. Escudeiro Santana, T. Hedley, A. Mongelluzzo, V. Ruzinov and I. Wevers, Vacuum **60** (2001) 57.

[37] C. Benvenuti, A. Escudeiro Santana and V. Ruzinov, Vacuum **60** (2001) 279.

[38] W. C. Turner, J. Vac. Sci. Technol. **A13** (1995) 2241.

[39] A. G. Mathewson, in *Proc. 1993 Particle Accelerator Conference* (IEEE, Piscataway, NJ, 1993), pp. 3828–3832.

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

[41] E. Wallén, J. Vac. Sci. Technol. **A15** (1997) 265; **A14** (1996) 2916.

[42] O. Gröbner, Vacuum **46** (1995) 767; Vacuum **53** (1999) 269.

[43] D. Hollenbach and E. E. Salpeter, J. Chem. Phys. **53** (1970) 79; Astrophys. J. **163** (1971) 155.

[44] M. G. Rao, P. Kneisel and J. Susta, Cryogenics **34** (1994) 377.

[45] Y. Kojima, in *Proc. 1991 Particle Accelerator Conference* (IEEE, Piscataway, NJ, 1991), p. 16.

[46] H. F. Dylla, L. R. Doolittle and J. F. Benesch, in *Proc. 1993 Particle Accelerator Conference* (IEEE, Piscataway, NJ, 1993), p. 748.

[47] S. Mitsunoba *et al*., in *Proc. of the 5th Workshop on RF Superconductivity*, Hamburg, Germany, 1992, DESY Report No. DESY M-92-01 (April 1992), p. 84.

[48] B. Aune *et al*., Phys. Rev. ST Accel. Beams **3** (2000) 092001.

[49] G. Cavallari *et al*., *LEP 2000 Status Report*, CERN-SL-98-011DI (March 1998).

[50] Spallation Neutron Source (SNS) Project : www.sns.gov

[51] A. Matheisen in *Proc. of the 10th Workshop on RF Superconductivity*, Tskuba, Japan, 2001 www.conference.kek.jp/SRF2001/ , paper TL007.

[52] G. R. Neil, S. Benson, G. Biallas, C. L. Bohn, D. Douglas *et al*., Phys. Rev. Lett. **84** (2000) 662.

[53] M. Kelley, Proc. of the SPIE, **3888** (2000) 598.

[54] J. Mammosser in *Proc. of the 6th Workshop on RF Superconductivity,* Newport News, VA, 1993 (Jefferson Lab, Newport News, VA, 1994).

[55] P. Kneisel, J. Vac. Sci. Technol. **A11** (1993) 1575.

[56] T. O. Raubenheimer, *Proc. of the 2001 Particle Accelerator Conference* (IEEE, Piscataway, NJ, 2001), p. 407.

[57] *TESLA Technical Design Report*, DESY Report 2001-001.

[58] M. Ross, in *Proc. 1991 Particle Accelerator Conference* (IEEE, Piscataway, NJ, 1991), p. 522.

[59] J. D. Lawson, Proc. Phys. Soc. **B70** (1957) 2.

[60] R. J. Hawryluk *et al*., in *Proc. 11th Intern. Conf. on Plasma Physics and Controlled Nuclear Fusion Research,* Kyoto, 1986 (IAEA, Vienna 1987), Paper No. IAEA-CN-47/A-I-3.

[61] E. Bertolini, in *Proc. 11th Symposium on Fusion Engineering* (IEEE, Piscataway, NJ, 1986).

[62] A. S. Bishop, *Project Sherwood – The U.S. Program in Controlled Fusion* (Addison-Wesley, Reading, MA, 1958).

[63] Lyman Spitzer *et al*., *Problems of the Stellarator as a Useful Power Source* (unpublished, June 1954).

[64] J. L. Bromberg, *Fusion: Science, Politics and the Invention of a New Energy Source* (MIT Press, Cambridge, MA, 1982), pp. 130–150.

[65] L. A. Artsimovich *et al*., *Experiments in Tokamak Devices,* Nuclear Fusion, Special Supplement 1969 (IAEA, Vienna, 1969).

[66] N. J. Peacock, D. C. Robinson, N. M. J. Forest, P. D. Wilcock and V. V. Sannikov, Nature **224** (1969) 488–490.

[67] G. G. Kelley, M. Roberts and J. F. Clarke, *Conceptual Design of the Ormak Facility*, Oak Ridge National Laboratory Report ORNL-TM-2821 (unpublished, Dec. 3, 1969).

[68] B. Coppie and J. Rem, Sci. Am. **227** (1972) 65.

[69] R. L. Hirsch, *Doublet II CTR Experiment at Gulf General Atomic* (unpublished, Jan. 21, 1970).

[70] H. P. Furth, Nucl. Fusion **15** (1975) 487.

[71] H. Eubank *et al*., in *Proc. 7th Intern. Conf. on Plasma Physics and Controlled Fusion Research*, Innsbruck, 1978 (IAEA, Vienna, 1979), p. 11.

[72] R A. Ellis, H. P. Eubank, R. Goldston, R. R. Smith and T. Nagashima, Nucl. Fusion **6** (1976) 524.

[73] P. E. Stott, C. C. Daughney and R. A. Ellis, Nucl. Fusion **15** (1975) 431.

[74] R. R. Parker, D. S. Pappas and E. Apgar, Bull. Am. Phys. Soc. **21** (1976) 1141.

[75] L. Oren and R. J. Taylor, Nucl. Fusion **17** (1977) 1143.

[76] E. B. Meservey, V. Arunasalam, C. Barnes, K. Bol, N. Bretz *et al*., J. Nucl. Mater. **93/94** (1980) 267.

[77] W. B. Gauster, J. A. Koski and R. D. Watson, J. Nucl. Mater. **122/123** (1984) 80.

[78] R. A. Langley, R. J. Colchin, R.C. Colchin, R. C. Isler, M. Murakami, J. E. Simpkins *et al*., J. Nucl. Mater. **85/86** (1979) 215.

[79] P. K. Mioduszewski, J. E. Simpkins, P. H. Edmonds, R. C. Isler, E. A. Lazarus *et al*., J. Nucl. Mater. **128/129** (1984) 884.

[80] P. K. Mioduszewski, R. E. Clausing and L. Heatherly, J. Nucl. Mater. **85/86** (1979) 963.

[81] M. A. Mahdavi, C. J. Armentrout, F. P. Blau, G. Bramson, N. H. Brooks *et al*., J. Nucl. Mater. **111/112** (1982) 355.

[82] D. M. Meade *et al*., in *Proc. Eighth Intern. Conf. on Plasma Physic and Controlled Fusion Research* (IAEA, Vienna, 1981), Vol. 1, p. 665.

[83] M. Keilhacker *et al*., Plasma Phys. Controlled Fusion **28** (1986) 29.

[84] J. L. Cecchi, R. J. Knize, H. F. Dylla, R. J. Fonck, D. K. Owens and J. J. Sredniawski, J. Nucl. Mater. **111/112** (1982) 305.

[85] R. J. Fonck, M. Bell, K. Bol, K. Brau, R. Budny *et al*., J. Nucl. Mater. **111/112** (1982) 343.

[86] H. F. Dylla, S. A. Cohen, S. M. Rossnagel, G. M. McCracken and Ph. Staib, J. Vac. Sci. Technol. **17** (1980) 286; H. F. Dylla, J. Vac. Sci. Technol. **A6** (1988) 1276.

[87] W. Poschenrieder, G. Standenmaier and Ph. Staib, J. Nucl. Mater. **93/94** (1980) 322.

[88] JT-60 Team, Plasma Phys. Controlled Fusion **28** (1986) 185.

[89] P. H. LaMarche, H. F. Dylla, G. D. Martin, S. Raftopoulos in *Proc. 13th Symposium on Fusion Engineering*, Knoxville, 1989 (IEEE, Piscataway, NJ, 1990), p. 96384.

[90] The JET Team, J. Nucl. Mater. **196/198** (1992) 3.
[91] E. E. Bloom, R. W. Conn, J. W. Davis, R. E. Gold, R. Little *et al*., J. Nucl. Mater. **122/123** (1984) 17.
[92] H. Horiike *et al*., Fusion Eng. Des. **16** (1991) 285.
[93] D. M. Meade, Fusion Eng., *Proc. 17th IEEE/NPSS Symposium* (IEEE Piscataway, NJ, 1997), Vol. 2, p. 752.
[94] The JET Team, Nucl. Fusion **32** (1992) 187.
[95] P. Andrew, C. J. Caldwell-Nichols, J. P. Coad, K. J. Dietz, J. Ehrenberg *et al*., J. Nucl. Mater. **196/198** (1992) 143.
[96] R. J. Hawryluk *et al*., Phys. Plasmas **5** (1998) 1577.
[97] J. D. Strachan *et al*., Phys. Rev. Lett. **58** (1987) 1004.
[98] M. Keilhacker *et al*., Nucl. Fusion **41** (2001) 1925.
[99] S. L. Milora, J. Vac. Sci. Technol. **A7** (1989) 925.
[100] K. Ushigusa and the JT-60 Team, Fusion Energy **1** (1997) 37.
[101] E. A. Lazarus *et al*, Phys. Rev. Lett. **77** (1966) 2714.
[102] J. L. Cecchi, J. Nucl. Mater. **93/94** (1980) 28.
[103] H. F. Dylla and the TFTR Team, J. Nucl. Mater. **145/147** (1987) 48.
[104] H. F. Dylla *et al.*, Nucl. Fusion **27** (1987) 1221.
[105] H. F. Dylla and K. L. Wilson, Eds, *Tritium Retention in TFTR,* PPPL Report No. 2523 and Sandia National Laboratories Report No. SAND-88-8212 (April 1988).
[106] The JET Team, J. Nucl. Mater. **176/177** (1990) 3.
[107] A. Miyahara and T. Tanake, J. Nucl. Mater. **155/157** (1988) 49.
[108] J. Winter, J. Nucl. Mater. **176/177** (1990) 14.
[109] W. G. Redan, J. Vac. Sci. Technol. **20** (1982) 1173.
[110] R. Frank and E. Usselmann, Vakuum Technik **25** (1976) 48.
[111] L. E. Valby, in *Proc. of the 7th Symposium on the Engineering Problems of Fusion Research*, Knoxville, 1977 (IEEE, Piscataway, NJ, 1977), p. 1040.
[112] ITER Project: www.iter.org .
[113] C. H. Skinner *et al*., J. Vac. Sci. Technol. **A14** (1996) 3267.
[114] C. H. Skinner *et al*., J. Nucl. Mater. **290/293** (2001) 486.
[115] P. Andrew *et al*., Fusion Eng. Des. **47** (1999) 233.
[116] A. Abramovici, W. E. Althouse, R. W. P. Drever, Y. Gursel, S. Kawamura, F. J. Raab *et al*., Science **256** (1992) 325.
[117] V. Brisson, J. Hoang, G. Lissilour, P. Marin, A. Reboux, M. Bernardini, A. Pasqualetti, Vacuum **60** (2001) 9.
[118] J. R. J. Bennett and R. J. Elsey, Vacuum **44** (1993) 647.
[119] Y. Saito, G. Horikoshi, R. Takahashi and M. Fukushima, Vacuum **60** (2001) 3.
[120] R. Weiss, *Proc. of the XHV Workshop*, Jefferson Lab., Newport News, VA, June 2000 (unpublished).
[121] H. F. Dylla, Vacuum **47** (1996) 647.

Partial pressure gauges

G. J. Peter and N. Müller
Inficon AG, LI 9496 Balzers, Liechtenstein

Abstract
The determination of partial pressures in vacuum systems is usually performed by mass spectrometers. The most common type is the QMS. Quadrupole mass spectrometers were developed decades ago and have been used by vacuum specialists as a diagnostic tool since then. In the first part of the paper the principles of these mass spectrometers are briefly reviewed together with the key features of the instruments. This is necessary to operate these instruments. In the second part the boundary conditions which arise from the application as residual gas analyser in UHV/XHV plants are described. These lead to special versions of mass spectrometers. Results obtained with these instruments and typical artefacts in mass spectra obtained in the UHV are discussed.

1 Introduction

The quadrupole mass spectrometer was invented in 1958 at the University of Bonn [1, 2]. The development for industrial applications started in the early 1960s. At the beginning these instruments were mainly used by specialized vacuum-scientists and engineers solely for vacuum diagnostic purpose. Later on the instruments found their way into other applications such as gas analysis in general, surface science, plasma diagnostics, and many other applications where partial pressure measurement is one of the basic tasks.

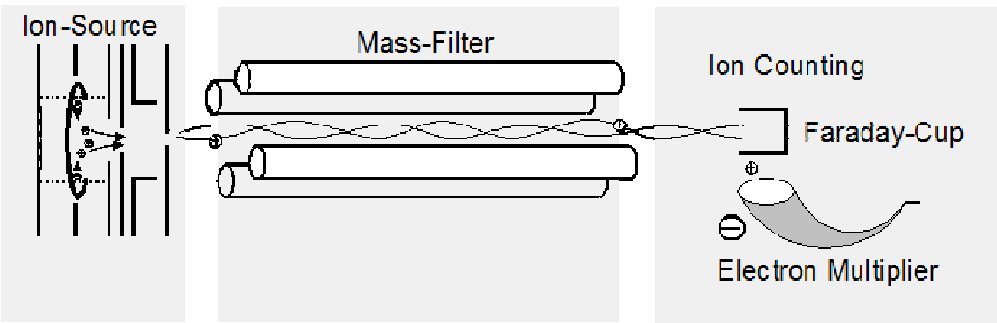
Today quadrupole mass spectrometers are also integrated into analytical systems for environmental applications and biological research and very often coupled to another analytical instrument (e.g., gas chromatograph, liquid chromatograph, ion mobility spectrometer).

A basic understanding of the principle is required to select the suitable instruments in terms of sensitivity, stability, mass range, and the like.

Ultra high vacuum applications result in high requirements on the instruments.

As the main differences compared to standard instruments, a low detection limit, extremely low outgassing of the gauge, and temperature and radiation resistivity must also be mentioned.

2 The functional units of a quadrupole mass spectrometer



2.1 Ionization of the gas

In a quadrupole mass spectrometer the particles are separated according to their *m*/*e* ratio in an electrical field, so the neutral atoms or molecules have to be ionized first. The most commonly used technique is electron-impact ionization, other methods such as chemical ionization, proton transfer reaction [3], ion attachment, and laser ionization etc. are applied for special analysis methods.

For electron-impact ionization, electrons are emitted from a heated filament and accelerated towards the anode (Fig. 1). In first order approximation the electron energy is given by the voltage drop between the filament and the anode.

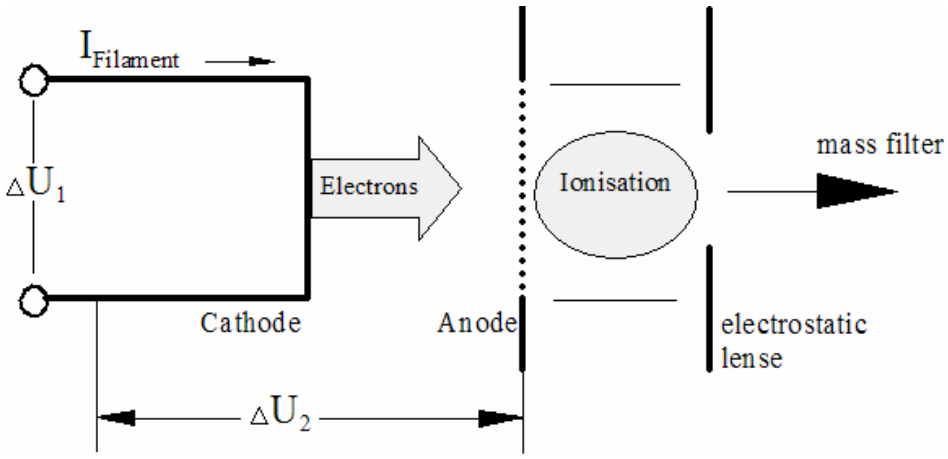


Fig.1: Electron impact ionization

Electrons which have reached the anode (here shown as a grid) have an energy of U_2 in eV.

Because they are decelerated by the electrical field of the electrostatic lens again, and ionization will occur at every space in the gas phase, this is only an approximation and spectra obtained by different instruments can not be compared totally.

As a well known fact the ionization cross-section for electron impact ionization depends on the atom or molecule to be ionized and on the electron energy as well.

This is shown in Fig. 2.

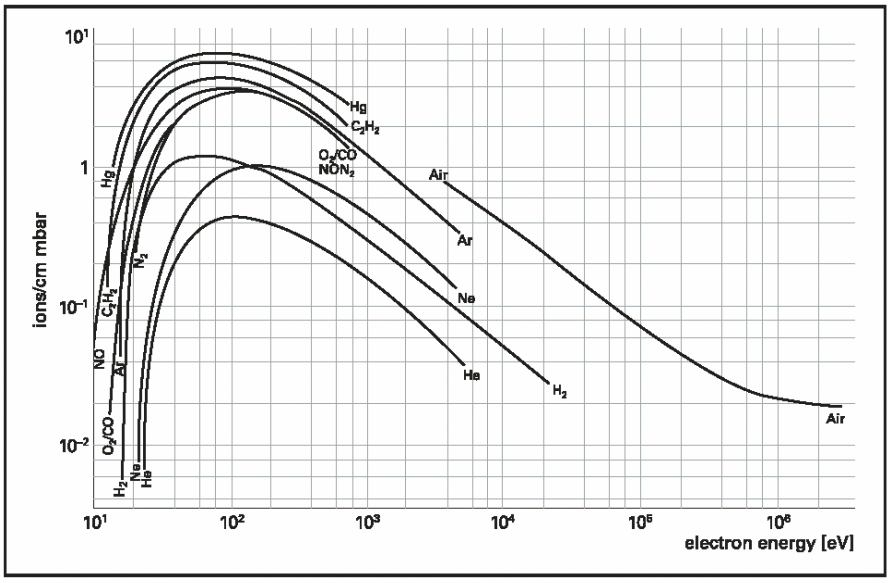


Fig. 2: Ionization cross-section as a function of the electron energy

From this it is evident that the sensitivity of a partial pressure gauge is gas specific and will also depend on the settings of the instrument (one essential parameter is the electron energy).

2.2 Mass separation

With the instruments described here a quadrupole mass filter is used.

Ions created in the ion source are injected with an energy of a few electronvolts into a high-frequency electrical field established among four metal rods, as shown in Fig. 3.

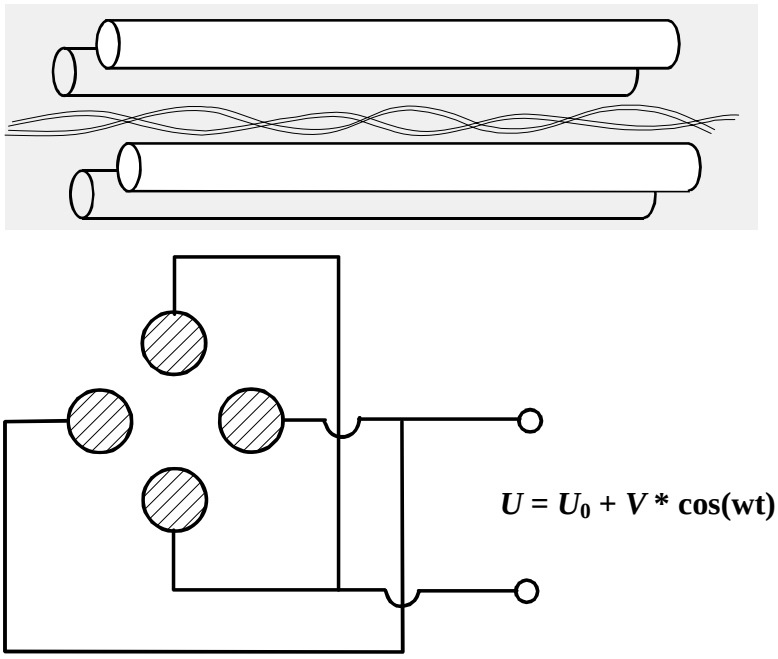


Fig. 3: The quadrupole mass filter

Only ions with a well-defined ratio of m/e, where m is their mass and e their charge, can travel on stable trajectories in this field.

The mass resolution of the instrument is adjusted by the ratio U_0/V and a mass scan is performed by increasing the absolute value of U_0 and V.

To use such an instrument this simple description of the filter is sufficient; those interested in the equations of motion, their solution, and the stability diagram of a quadrupole filter can find an introduction in the literature [4].

Although the principle of a quadrupole is rather easy, a very high mechanical precision of the filter and a stable RF frequency are required to ensure proper operation of the instrument.

For the user of such instruments it is important to know that impurities inside the filter—coatings or small dielectric particles—may strongly influence the performance of the instrument.

2.3 Detection

The primary quantity to be measured at the exit of the filter is an ion current. Depending on the application there are different ways of measuring this current.

Two types of detectors, a Faraday cup or a secondary electron multiplier SEM (discrete or continuous) are used (Figs. 4 and 5).

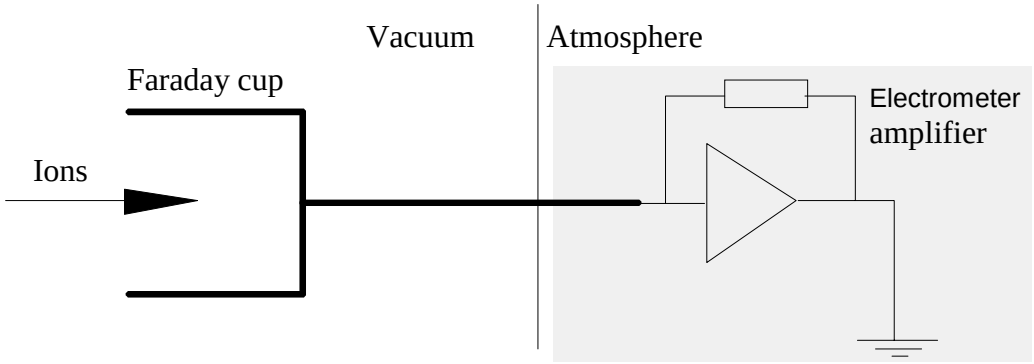


Fig. 4: Faraday cup in combination with an electrometer pre-amplifier

If a lower detection limit and a higher sensitivity are required, then the arriving ion current has to be amplified in the vacuum. Here an SEM or an C-SEM is used for an amplification up to 10^8.

Particles (ions, neutrals, electrons, photons) hitting a surface with high kinetic energy release several 'secondary electrons'. The amount of the secondary electrons released depends on the surface material. The use of several dynodes in combination with the appropriate dynode material allows for an amplification up to 10^8. One single ion impinging on the first dynode will cause an avalanche of electrons. This current can be converted to a transient in voltage and thus single-ion counting is possible with such a device too, provided the current pulse is short enough (10^{-8} s).

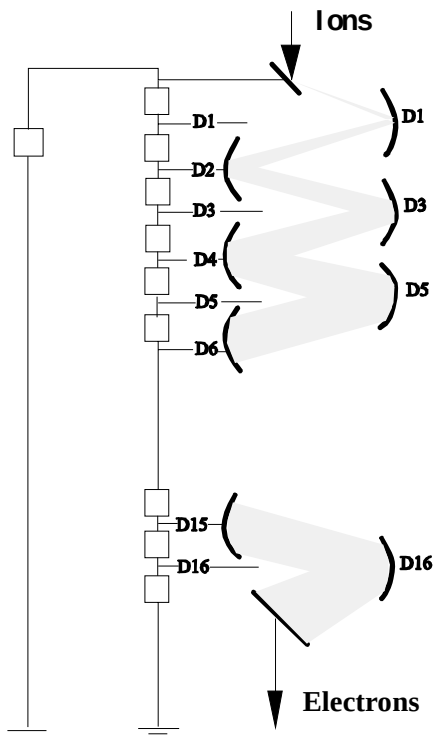


Fig. 5: The principle of a discrete SEM

A SEM can be used to amplify the initial ion current and as ion counter as well. In the latter the short voltage pulse caused by each ion arriving at the first is used to count the ions directly.

In Fig. 6 the response of the detector versus the SEM voltage at constant gas pressure is shown.

In this mode the SEM has a characteristic like a Geiger–Müller tube.

In the first region the signal increases rapidly with the increasing voltage, then it tends to saturate in a so-called plateau. (Note that Fig. 6 is a semi-logarithmic plot and the plateau achieved may be not as flat as with a Geiger–Müller tube.)

In this mode ion currents down to 0.1 CPS can be detected and a detection limit of about 10^{-16} mbar can be achieved.

The dynamic range of this method is mainly determined by the width of the pulse; for example a pulse width of 20 ns limits the upper linear range to 5×10^6 CPS.

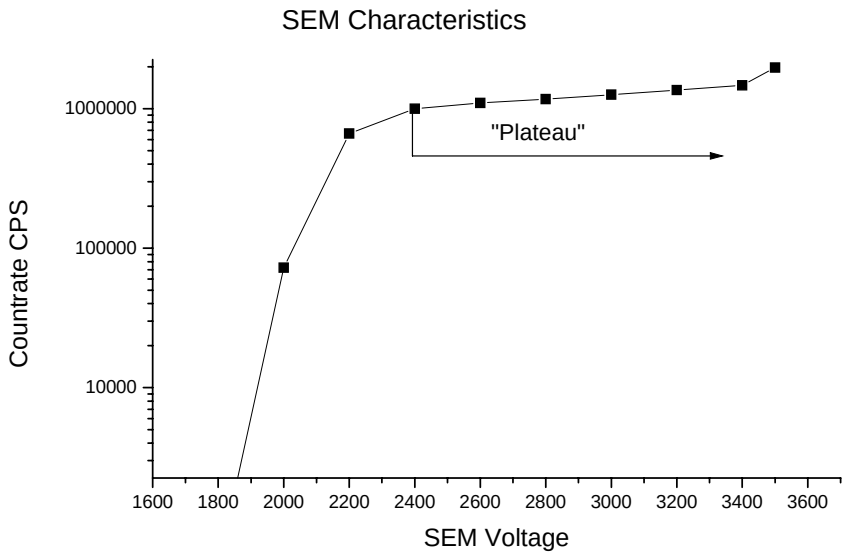


Fig. 6: Characteristics of a SEM operated in the counting mode

An SEM or a C-SEM is a passive amplifier which means that the amplifier undergoes ageing.

This effect is shown in Fig. 7. With long time-operation the characteristics will shift towards higher voltages. Therefore the characteristics have to be checked from time to time.

Because the ageing depends on many effects such as total pressure during operation, gas measured, signal height and the like, no general numbers about the lifetime of an SEM can be given.

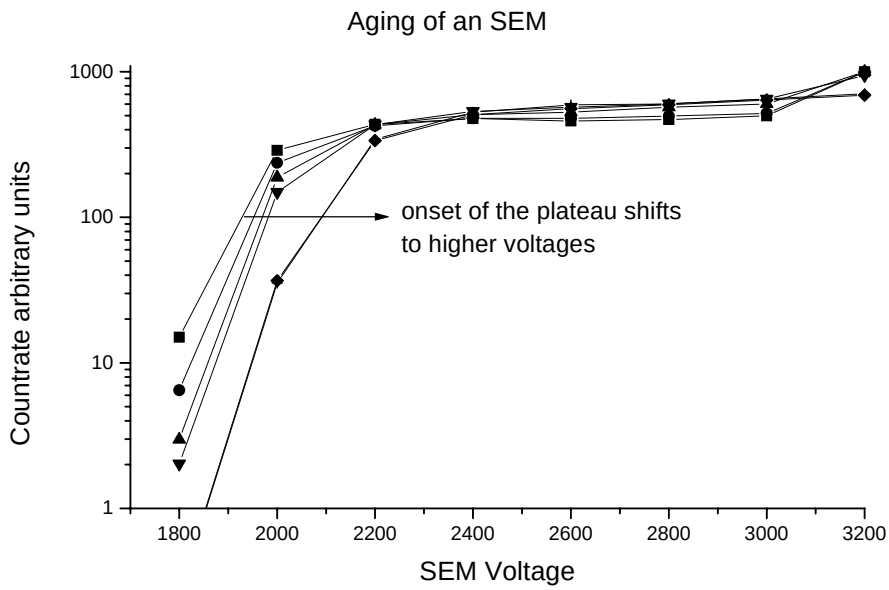


Fig. 7: Ageing of an SEM

In the ion count mode the plateau is shifted towards higher voltages during long time-operation. If the SEM is used for current amplification, the amplification decreases with time, which makes a re-adjustment of the SEM voltage necessary.

All parts of an SEM or a C-SEM are within the vacuum. Therefore the dynodes, insulators, and resistors used to build the SEM have to be UHV compatible. Other important features are the temperature resistivity during bakeout and the degas rate of the SEM under operation.

(Of course, also with an SEM used in UHV/XHV, gaps have to be avoided and threads if they are used at all have to be equipped with additional holes for easier degassing.)

2.4 Background reduction

In order to achieve a low detection limit, the background of the detector has to be as low as possible and on the other hand the sensitivity of the whole system has to be as high as possible.

One very effective method to reduce background is to use an SEM 90 degree off axis (Fig. 8). In this design photons and electrons from the ion source do not contribute to the signal and thus a very low background is achieved.

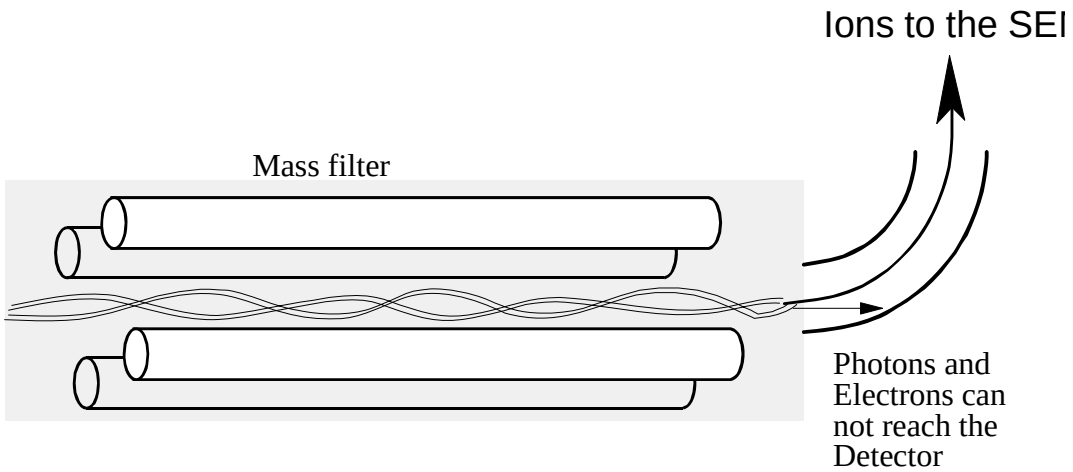


Fig. 8: SEM 90º off axis

Ions leaving the mass filter are deflected 90 degree onto the detector to suppress background.

The ions out of the mass filter have to bent 90 degree by an electric field to reach the detector in this configuration. Thus the sensitivity is lower compared to an in-line arrangement (about a factor of 2–5). However, the reduction of the background is more than one order of magnitude because photons and electrons will not reach the detector. Therefore the 90 degree off-axis geometry leads to a lower detection limit.

2.5 Interpretation of the measured spectra

With electron impact ionization not only the molecule is ionized. In most cases a fragmentation of the molecule occurs too. In addition, molecules can be multiple charged. If isotopic effects also have to be considered, the spectrum of one simple molecule can be rather complex.

For example CO_2 will show up at mass 44, 12, 16, 28, 44 and 46 amu corresponding to the ions CO_2^+, C^+, O^+, CO^+, CO_2^{++} and $^{12}C^{16}O^{18}O^+$, respectively.

As a consequence, different molecules may contribute to the signal at the same mass. For example, the water molecule will appear as H^+, H_2^+, O^+, OH^+ and H_2O^+, oxygen will appear as O^+ and O_2^+.

Thus at mass number 16 there will be an overlap of the two species. In Fig. 9 the situation for the main components of air is shown.

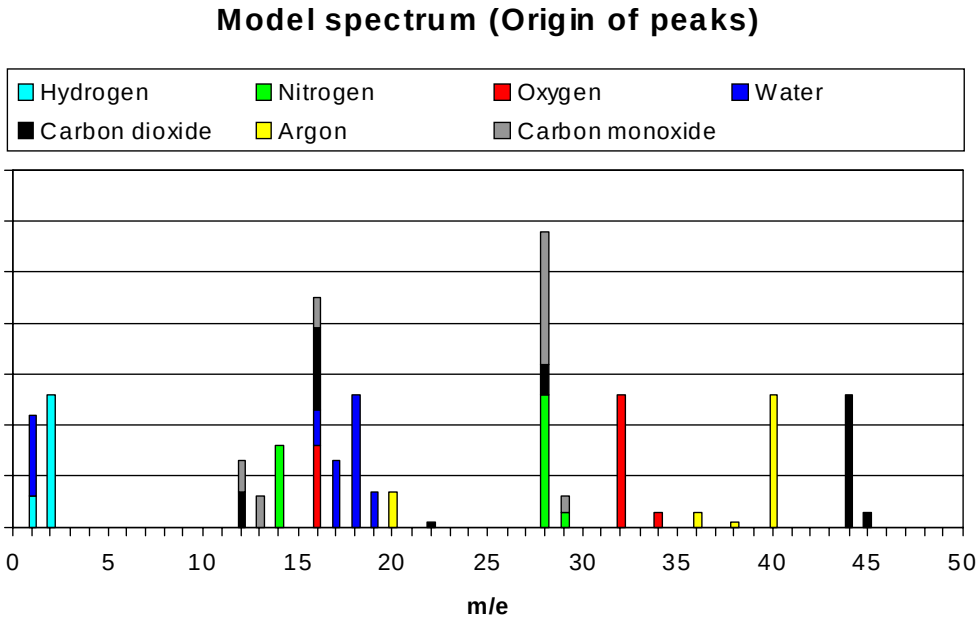


Fig. 9: Interpretation of mass spectrometric data

A quantitative analysis is possible if the cracking pattern and the sensitivity for the individual components are known. In this case the concentrations can be calculated via a set of linear equations.

For analytical purposes, spectra libraries and time-consuming calibrations by means of test gases are required.

A residual gas spectrum in the UHV or XHV region, however, can easily be interpreted; hydrogen, carbon-monoxide, carbon dioxide are the dominant gases together with water vapour (Fig. 10).

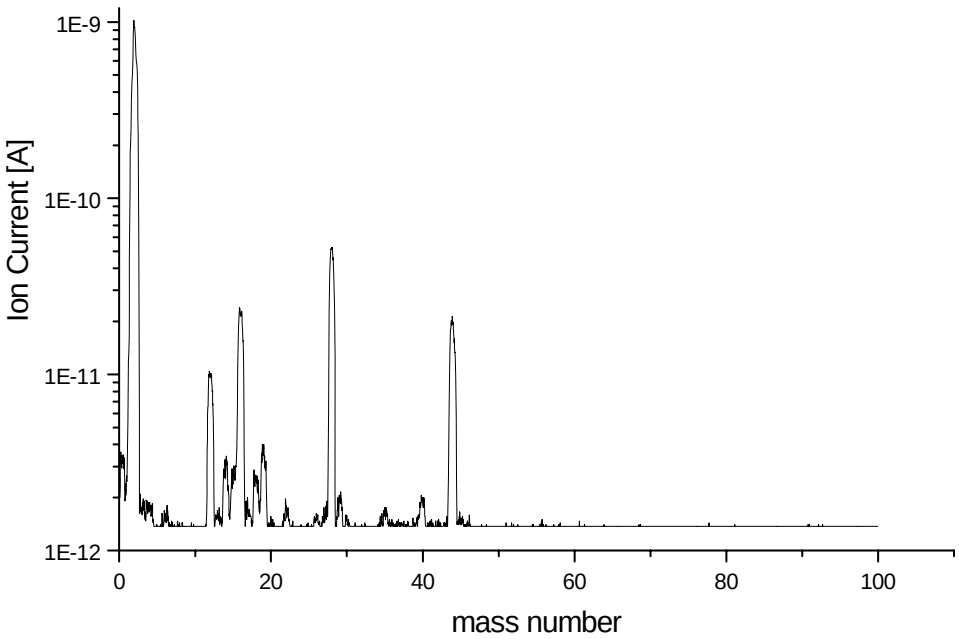


Fig. 10: Residual gas spectrum, total pressure about 10^{-10} mbar

3 Key features of quadrupole mass spectrometers

A large number of definitions and terms are used to describe these instruments.

For example, transmission, sensitivity, stability, reproducibility, and some other terms may be mentioned.

In order to get an overview, it is important to classify which of the parameters directly contribute to the measured result, which parameter just contributes to a key feature. (Of course only parameters which can be measured in common laboratory practice make sense.)

When beginning an experiment one has an idea what order of magnitude the quantity to be measured may be. Thus the basic question is the detection limit of the instrument or, more precisely, the experimental set-up. If more than one species has to be detected, as will be the case in most mass spectrometric applications, the next question is for the dynamic range of the instrument.

Of course, in many applications also the measurement speed, its influence on the accuracy and the like are essential.

3.1 Detection limit and dynamic range

However, let us stay first with the detection limit and the dynamic range and 'investigate' which of the other quantities contribute here. For this reason we describe the mass spectrometer again in brief:

- Gas is ionized in the ion source. The essential quantity is the sensitivity of the ion source in terms of ampere/mbar = $S_{\text{ionsource}}$.
- The ions generated in the ion source have to be transferred towards the mass filter. The important thing here is the transfer efficiency in per cent = T_{ransfer1}.
- Ions passing the mass filter. Here the transmission at a selected mass resolution is important = $TR_{\text{ansmission}}$.
- Finally the ions have to be transferred onto the detector, see transfer efficiency = T_{ransfer2}.
- The detector is characterized by its detection efficiency, dark noise and its dynamic range = D_{eff}.

The resulting signal at the detector can be described in an abstract formula:

$$S_{\text{ignal}} = P_{\text{artialPressure}} * S_{\text{ionsource (gas specific)}} * T_{\text{ransfer1}} * TR_{\text{ansmission}} * T_{\text{ransfer2}} * D_{\text{eff}} \ .$$

Looking more in detail, we see that the transmission and also the detection efficiency may depend on the mass number.

As a result of this elementary picture a high transmission of a filter, a high sensitivity of the ion source, or high detection efficiency alone do not tell us anything about the performance of such an instrument. All these quantities contribute to the sensitivity. However, the essential quantity is the total sensitivity at the output of the detector. This in detail depends on resolution settings and the gaseous component to be detected. In addition, all parameters contained in the equation above may be measured, in principle. The only quantity, however, which can be determined in normal laboratory practice is the sensitivity.

According to common practice in measurement technology a signal is accepted (species detected) if the signal is at minimum a factor of two of the background at the detector.

From the overall sensitivity S_{total} and the background b_{Detector} at the detector this detection limit can be calculated:

$$D_{\text{etection limit}} = (2 * b_{\text{Detector}})/S_{\text{total}} \quad \text{where} \quad [S_{\text{total}}] = \text{A/mbar} \quad [b_{\text{Detector}}] = \text{A} \ .$$

This is shown in Fig. 11. Here the sensitivity of the instrument (for argon) was determined to be 24.0 A/mbar. The noise band was measured with an integration time of 60 seconds ten times. This results in $b_{\text{Detector}} = 5.4 \times 10^{-14}$ ampere and a detection limit of 4.5×10^{-15} mbar.

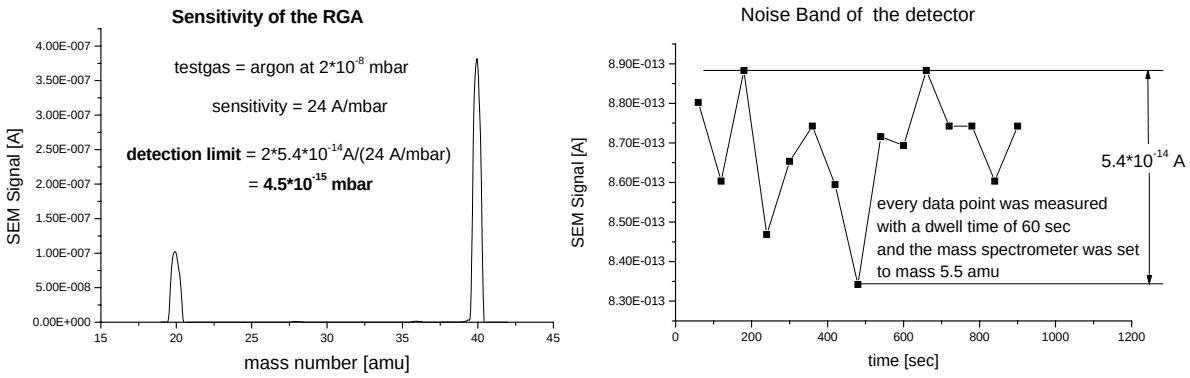


Fig. 11: Detection limit of a mass spectrometer

In some cases other gas components may interfere with the component to be detected. Then the detection limit is higher than calculated by the formula above. Because the noise band of the detector among other parameters depends on the integration time, the detection limit also depends on the desired measurement speed.

The dynamic range of the instrument is the span between the detection limit and the maximum detectable partial pressure. (If ion counting is applied, the pulse width of the SEM is an important parameter for the maximum detectable partial pressure.)

3.2 Mass resolution

The resolving power of the instrument is of vital importance especially if two gaseous components have to be detected which show up at neighbouring mass numbers.

There are some definitions which describe the mass resolution.

One according to AVS Standards is called **unit resolution**. A mass spectrometer is operated at unit resolution when the individual peaks are of equal width all over the spectrum and when there is 10% valley achieved between two neighbouring peaks at equal height.

This setting is most used for RGA applications.

Sometimes the full width at half maximum is specified. A more precise definition is the **contribution to the neighbouring mass**, because the full width at half maximum does not tell so much about the mass resolution.

This is shown in Fig. 12. Here two hypothetical peak shapes are compared and obviously the mass resolution is different. The contribution to the neighbouring mass provides more information than the full width at half maximum.

(Of course the contribution to the neighbouring mass depends on mass and mass range. And a high mass resolution always results in a lower sensitivity.)

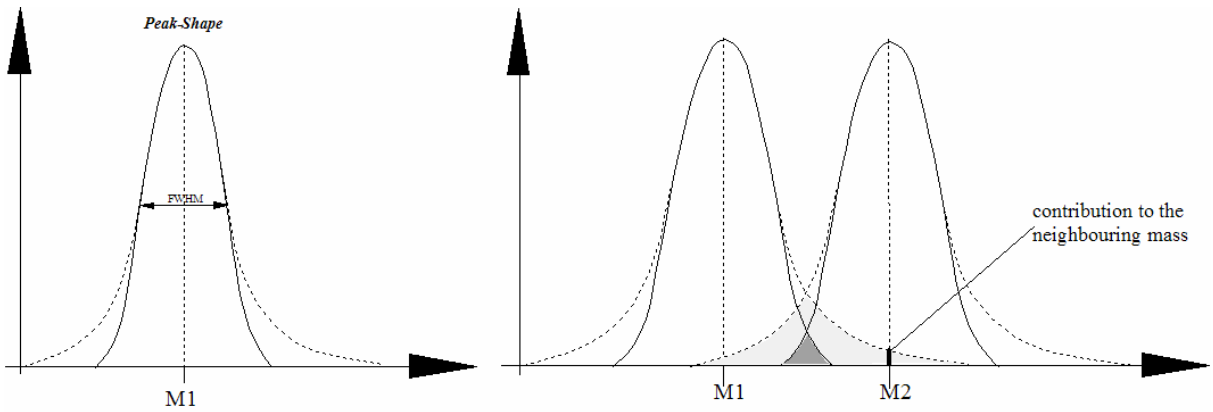


Fig. 12: Mass separation at different peak shapes with the same peak-width at half maximum

3.3 Stability, reproducibility, accuracy

Stability, reproducibility, and accuracy are well defined in measurement technology.

The latter is the band within which the measured signal stays for a given time. Reproducibility is the band in which the results are found when the measurement is repeated. The accuracy of an instrument can be specified, if a well-defined standard sample is measured.

In practice it turns out that the stability and reproducibility of the gaseous sample may have a large influence on the results too. Thus every specification of these terms should include a precise definition of the experiment.

3.4 Measurement speed

For some applications it may be important to collect data in the shortest time possible, no matter if a total mass scan or a sampling at pre-selected masses is required.

To achieve an appropriate mass resolution, the dwell time of the ions in the mass filter has to be *ad minimum* a few RF cycles. With a frequency of 3 MHz a dwell time of a few microseconds is reasonable, which results in a minimum possible response time of the same order of magnitude.

Apart from this, the electronics used to operate the mass spectrometer is the limiting factor.

One component is the RF generator used to drive the mass filter. The other component is the electrometer pre-amplifier (in combination with the SEM used which may in addition limit the measurement speed). Also the response time of an electrometer pre-amplifier depends on the type of amplifier and the range of the current to be measured. Obviously the dwell time to measure an electrical current of the order of 10^{-12} amperes differs in magnitude from the dwell time to measure a current in the order of 10^{-6} amperes provided the same accuracy is desired.

In general the electronics used determines the maximum possible measurement speed and again the maximum measurement speed depends on the accuracy and the dynamic range required.

State-of-the-art instruments offer a maximum speed of about 0.1 ms both in the sampling and the scan mode. (This may not be necessary just for residual gas analysis.)

Summary:

The key features of a quadrupole mass spectrometer are

– detection limit (depends on the setting of the instrument and in some cases on the vacuum system too)

– dynamic range of the analysis
– mass resolution, the contribution to the neighbouring mass is an appropriate definition
– the measurement speed.

All these can not be discussed as single parameters, they mutually influence each other.

For example, the specified detection limit may never be achieved at the maximum measurement speed.

4 Calibration of mass spectrometers

As already mentioned, the sensitivity and the fractionation pattern are gas-specific. Further, the sensitivity depends on the setting of the instrument, e.g., the resolution selected, the ion source parameters and the amplification by the SEM. Therefore all these parameters have to be filed together with the other calibration data, otherwise the calibration is questionable.

In general, one can use pure gas for a calibration or a gas imbedded in a carrier gas which shows no spectral overlap with the gas to be calibrated. A mixture of He, N_2, Xe in Ar as carrier gas is an example for the latter. (Such a gas mixture can also be used to check the mass scale and the resolution setting over a wide range.)

An appropriate dosing valve is required to introduce the gas mixture into the vacuum chamber (Fig. 13).

In order to discriminate against the background of the vacuum system it is recommended to introduce as much gas as the total pressure—measured by a calibrated gauge—at minimum two orders of magnitude higher than the base pressure of the system.

The measured quantity is an ion current versus mass number. Therefore the sensitivity of the instrument is specified in terms of ampere/mbar for the gas component under question.

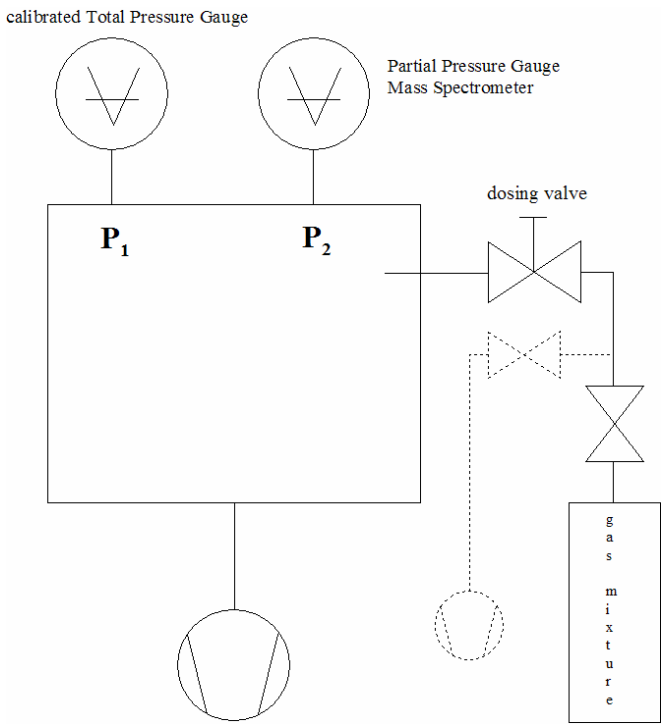


Fig. 13: Calibration of a mass spectrometer

Furthermore it is important to make sure that the pressure P_1 and P_2 at the gauge and at the mass spectrometer do not differ too much.

When a gas cylinder with calibration gas is attached to the dosing valve, it is important to avoid contamination with the air in the tubing when the gas supply is attached to the valve. A pump/purge method as indicated by the broken lines can be recommended for that.

General advice for calibration time intervals can not be given here, because the service-time of the instrument and the SEM depend strongly on the application.

In Fig. 14 the spectrum of a gas mixture is shown for different settings of the electron energy. Not only the sensitivity but also the cracking pattern changes with the electron energy.

The same is valid for changes in the settings of the resolution and electrical potential applied to the ion optical instruments.

Therefore calibration data for a RGA have to include every setting of the mass spectrometer, such as resolution, electron energy, emission current of the filament, the SEM voltage, and setting of the ion optical elements.

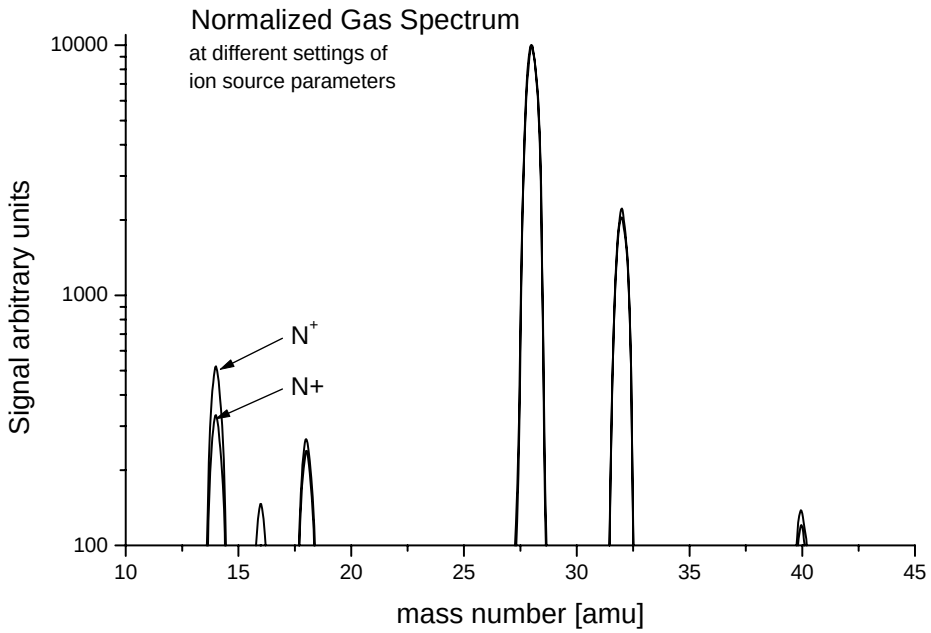


Fig. 14: Normalized gas spectra at different electron energies (70 eV and 50 eV), the sensitivity change for the nitrogen ion becomes clearly visible

5 General demands on a partial pressure gauge used in the UHV / XHV

5.1 Maximum allowed degas rate

Every material and instrument in a vacuum system will degas. Degas rates as high as 10^{-5} mbar l s^{-1} for a total plant in high vacuum technology are typical. Assuming an effective pumping speed of 1000 l s^{-1} the base pressure will be at 10^{-7} mbar which is a good number for a high vacuum system and thus indicates a low degassing rate of the whole system.

The question is how high the maximum tolerable degassing rate of a partial pressure gauge in an UHV system can be.

This can be calculated by the base pressure in the system and the effective pumping speed at the place of the total pressure gauge.

Assuming an effective pumping speed of about 100 l s^{-1}—not to be mixed up with the total pumping speed applied to the whole plant—and a base pressure 10^{-12} mbar, the measured pressure at the partial pressure gauge is also 10^{-12} mbar if the degas rate is not higher than 10^{-10} mbar l s^{-1}.

This simple calculation gives a limit for the maximum allowed degas rate of such instruments. At higher degas rates there would be a local higher pressure and the gauge would in principle measure its own gas atmosphere instead of the residual gas in the UHV system.

A total pressure gauge appropriate for high vacuum application may therefore be worthless for residual gas analysis in the UHV/XHV.

5.2 Influence of the gauge on the vacuum system

The gauge can influence the vacuum system in two ways. There is a hot filament which applies thermal radiation to the environment and can lead to desorption. Electrons are emitted from the filament which can hit the surface of the vacuum system and lead to gas releases or influence other instruments if the screening is insufficient.

These two effects have to be considered in the planning phase of an experiment, no general advice can be given here.

5.3 Bake-out of the vacuum system

In order to achieve a low degassing of the surfaces inside the vacuum chamber and thus a low base pressure, baking of the whole system is commonly applied. A pressure gauge suitable for UHV/XHV has to tolerate temperatures up to 400ºC.

It must be realized that the gauge is inside the system and during a short baking may not reach the same temperature as the other parts of the vacuum system. From this it is possible that material will condense at the colder parts of the plant—at the gauge.

Such cold spots can only be avoided if the bake-out time is sufficiently long.

Furthermore, under operation, the filament of the gauge is at 1000ºC or even above. Therefore it is strongly recommended to switch the filament on during bake-out or at least to switch it on during the cool-down phase.

5.4 Design criteria for (partial pressure) gauges in the UHV/XHV

For a partial pressure gauge to be used in UHV/XHV systems, the same design criteria as for the other components are mandatory. As mentioned before the gauge has to withstand a bake-out at up to 400ºC. Therefore, because of thermal expansion during bake-out, not every combination of materials can be used. If materials have to be in contact with each other their thermal expansion coefficients have to match, otherwise *ad minimum* a mechanical misalignment after bake-out may result.

Degassing of the device has to be kept at the lowest possible level. As a consequence, suitable materials have to be selected and the physical surface of the instrument has to be minimized by design. In addition, surface roughness has to be at a minimum. Further gaps with a low conductance are not allowed. Threads have to be avoided if possible and if they are unavoidable, additional holes for proper degassing are mandatory.

In particular, the surface of the ion source has to be minimized, otherwise this would be the largest source of desorption.

An example for a UHV design of an ion source is shown in Fig. 15.

Fig. 15: Grid ion source. The surface of this ion source is minimized by the grid design of the formation chamber.

6 Technical solutions

6.1 Choice of the material

Materials such as plastic can not be used at all, because of their low temperature stability, their low radiation stability and their high vapour pressure. Thus a whole class of materials for use as electrical insulator is ruled out.

The only materials allowed are metal and ceramics. Here the thermal expansion coefficients match sufficiently to allow bake-out temperatures up to 400ºC. The same is valid for an electrical feed through. In addition the junction between metal and ceramics has to be absolutely leak tight.

6.2 Pre-cleaning of the material

Even in high vacuum technology the material to be used has to be pre-cleaned. All materials such as lubricants from the machining process have to be removed from the surface. Cleaning in a supersonic bath with pure ethanol as the last step is mandatory. Thus one may be able to remove most of the impurities, however, water will stick in the form of a few monolayers on the surface and has to be removed by thermal treatment. Looking at a residual gas spectrum in the UHV/XHV, obviously H_2, CO, and CO_2 are the most dominant gases. The hydrogen partial pressure is most affected by the compression of the pump used. All three gases, however, are dissolved in the solid body in the metal. Therefore, if a very low degassing level is desired, the gases dissolved in the metal also have to be removed prior to the application in the vacuum. Firing these materials at temperatures up to 950ºC in a vacuum furnace is the most effective method here.

For XHV applications firing in a vacuum furnace is mandatory. (Only special types of high-grade steel can be fired at 950ºC without softening of the knife edges of the CF flange.)

6.3 Electrical design

So far methods keeping the degas rate of the gauge at the lowest level possible have been described. When such a gauge is operated—and the following is valid for every gauge in the UHV—the thermal

radiation of the hot filament and electrons emitted by the filament can contribute to undesired degassing of the other parts of the plant. To avoid most of the thermal degassing induced by the radiation of the hot filament, the position of the gauge has to be selected carefully.

Electron-induced degassing from the gauge, however, can only be avoided by electrical design. In Fig. 16 two electrical designs of gauges or ion sources are compared. Most common, the filament is electrically at negative potential, the cathode is positively biased and electrons mainly go from the anode (filament) towards the cathode. However, electrons may also get onto other surfaces in the system. Here electron-induced desorption will take place. To avoid this effect the whole gauge is positively biased. Thus electrons can end up only at well-defined surfaces of the ion source.

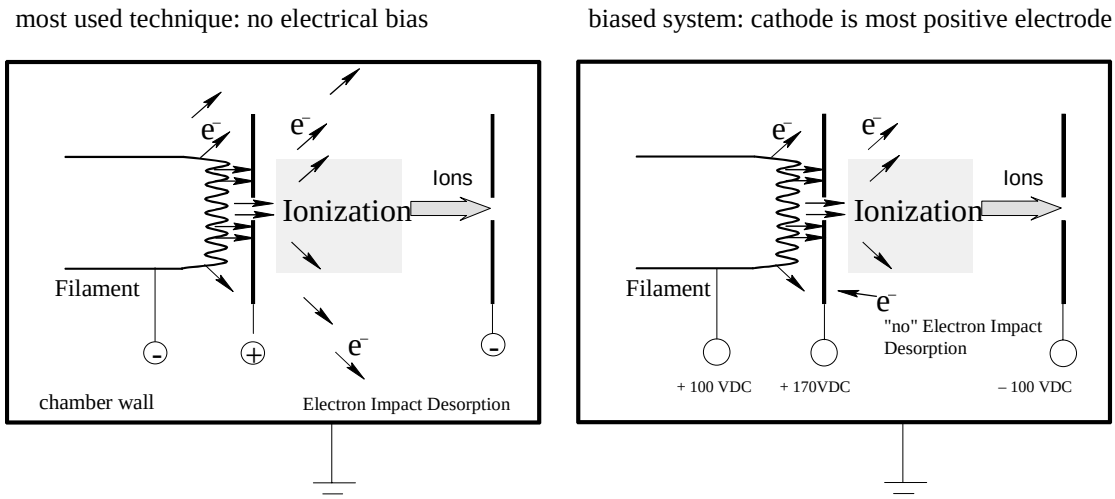


Fig. 16: Electrically biased ion source to further reduce EID

6.4 Radiation resistivity

The materials used to build up the analyser in the vacuum exhibit a sufficient resistivity against radiation. However, the electronics will suffer from radiation after some time. Therefore the electronics at best has to be separated from the analyser as indicated in Fig. 17.

Fig. 17: Analyser and electronics separated to apply shielding against radiation (not shown here)

In practice distances up to 6 metres between the analyser and the electronics and the electrometer pre-amplifier may be necessary.

7 Spectra obtained with RGAs in the UHV

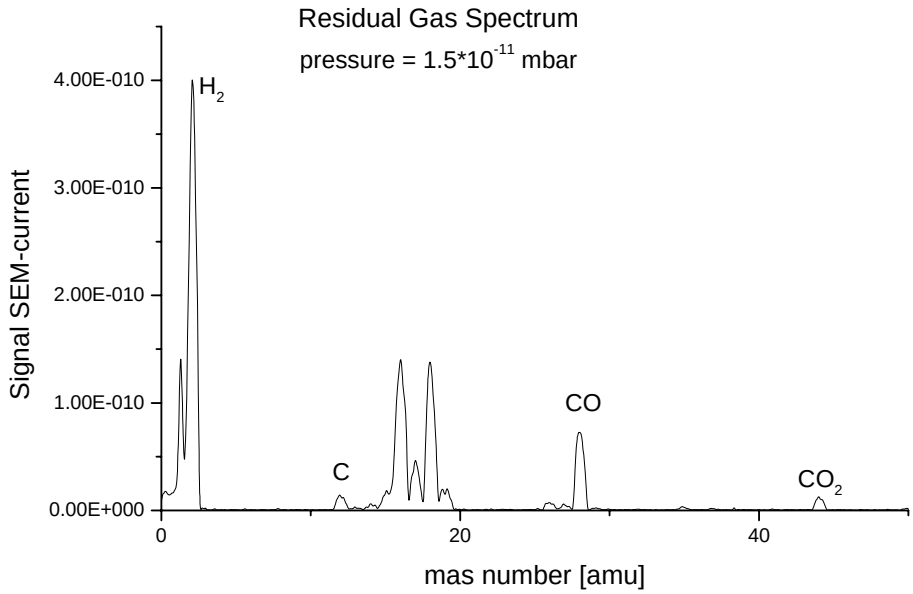


Fig. 18: Residual gas spectrum in UHV

7.1 Artefacts, EID ions

EID ions are often found in residual gas spectra from UHV/XHV plants.

In the residual gas spectra of HV plants these ions also may be present, however, as there are so many other gas components and the partial pressure gauge's sensitivity is not so high, they are just not detected or overlapped by others under most circumstances (Fig. 19).

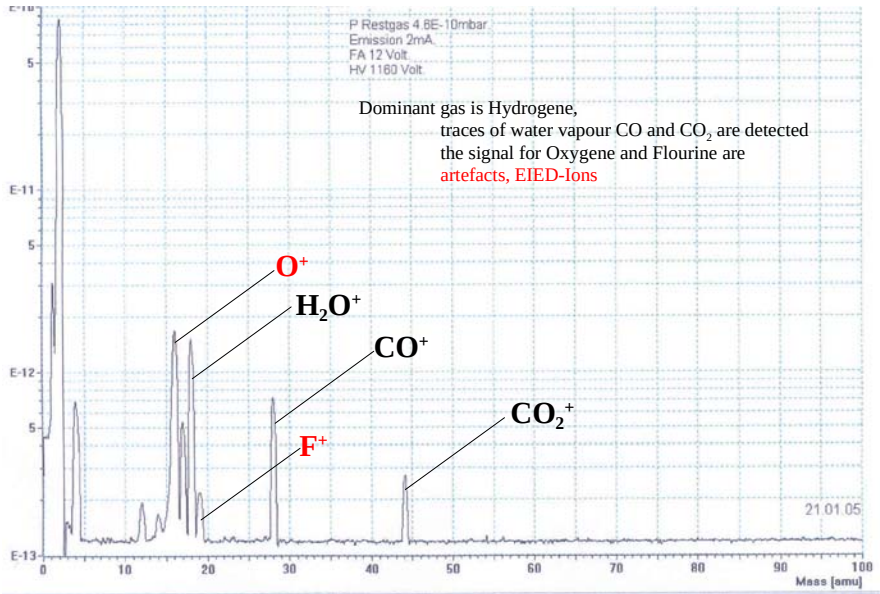


Fig. 19: EID ions in an UHV residual gas spectrum

EID ions originate from metallic surfaces. Depending on the cleanliness of the material, mostly the anode, they are set free as ions by electron bombardment. As these ions are not generated in the gas phase, their kinetic energy is higher than the energy of those generated in the gas phase.

Ions generated in the gas phase have to be extracted out of the ‘formation room’ and accelerated towards the mass filter. They travel with an energy of a few electronvolts through the filter. If the corresponding voltage is reduced, most of the ions out of the gas phase can no longer pass the filter. Only EID ions may pass the filter, because of their higher initial energy. Figure 20 shows a residual gas spectrum recorded at different ion energies (in the mass filter).

Clearly visible is that the signal generated by EID ions is less affected by the so-called ‘field axis voltage’ whereas the signal of the other ions strongly decreases.

(F^+ EID ions also show up in the spectrum below at a lower level than O^+.)

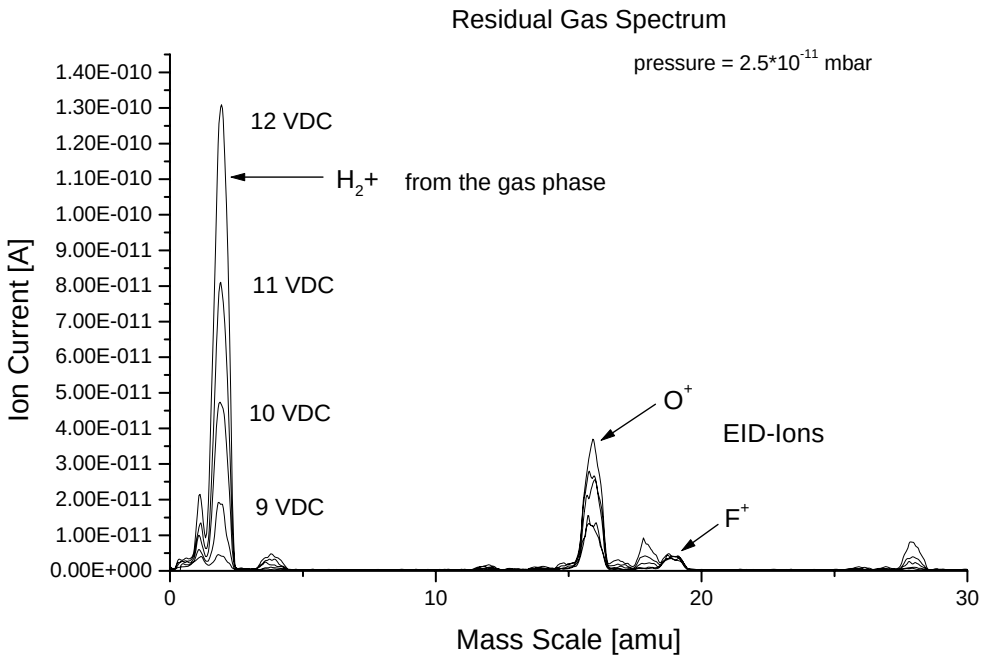


Fig. 20: EID ions and ions originating from the gas phase

8 Conclusion

Beneath the sensitivity in terms of ampere/mbar and a low detection limit required to detect critical gaseous components in the residual gas of an UHV system, a partial pressure gauge has to be of strict UHV design.

One effective method to achieve the low detection limit is the use of an SEM 90º off axis.

In the deep UHV and/or XHV region a UHV design of the gauge alone is not sufficient. The materials used to build the gauge have to be degassed at high temperatures in a vacuum furnace prior to assembly. Otherwise the gauge would measure its own gas atmosphere.

In order to avoid electron impact desorption from the vacuum system it is important to electrically bias the whole ionization chamber of the partial pressure gauge positively. This is important for UHV, XHV, and some analytical applications as well.

Even with this technique EID ions are detected in the mass spectrum, which result from the anode of the ion source.

Acknowledgements

We thank Hans Zogg who did most of the practical work, our other colleagues for many discussions, Edgar Mahner for his collaboration during the current project, and those scientists who started the development of UHV RGAs a long time ago.

References

[1] W. Paul and H. Steinwedel, Z. Naturforsch. **8a** (1953) 448.

[2] W. Paul, H.P. Reinhard and V. von Zahn, Z. Phys. **152** (1958) 143.

[3] A. Hansel and T.D. Märk, 2nd International Conference on Proton Transfer Reaction Mass Spectrometry and Its Applications, Obergurgl, Austria, 2005 (Institut für Ionenphysik der Leopold-Franzens-Universität Innsbruck, Innsbruck, Austria).

[4] P.H. Dawson (editor), Quadrupole Mass Spectrometry and its applications (Elsevier, Amsterdam, 1976).

Bibliography

M. Spiteller and G.Spiteller, Massenspektrensammlungen von Lösungsmitteln,

Verunreinigungen, Säulenbelegmaterialien und einfachen aliphatischen Verbindungen

(Springer Verlag, Wien, 1973) ISBN 0-387-81117-6.

Lifetime, cross-sections and activation

P. Grafström
CERN, Geneva, Switzerland

Abstract
The concept of cross-section and its relation to the beam lifetime in accelerators is being introduced. Some general properties of cross-sections for different particles and different interaction types are discussed. For some specific cases of elastic and inelastic reactions the conditions for beam losses and the corresponding lifetime is derived. The basics of activation or induced radioactivity in accelerators are also discussed at the end of lecture.

1 Introduce the concept of cross-section and its relation to the lifetime of a beam

To start with we shall introduce the concept of cross-section. The concept of cross-section in particle physics or in atomic physics concerns the interaction of elementary particles, nuclei, or atoms with each other. Given a particle approaching another particle, the cross-section for this process is the probability that the two particles interact with each other. This is the most general definition of cross-section. In a simple geometrical interpretation of the cross-section it can be thought of as the area within which a reaction will take place. Thus the units of a cross-section are the units of an area. In the early days of nuclear physics the following definition was introduced

$$1 \text{ barn} = 10^{-24} \text{ cm}^2 . \tag{1}$$

Since that time the cross-section has always been measured in 'barns'. Why this strange name of a unit? The explanation is historical. During the early experiments the physicists discovered that the interactions were far more probable than expected. The nucleus was 'as big as a barn'.

It is easy to deduce the relation between a given cross-section and the lifetime of a beam. We shall do this for a general case but in practice we have, of course, in mind particles in an accelerator interacting with the rest gas in the beam pipe.

Start with a beam of particles hitting a target (see Fig. 1).

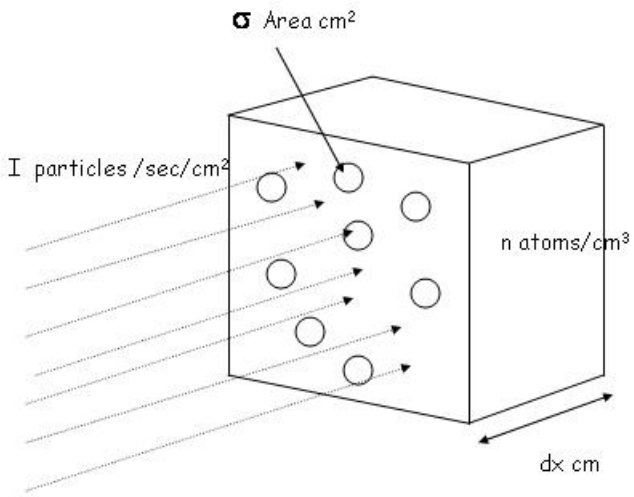


Fig. 1: The concept of cross-section

Assume that the beam flux is I particles/s/cm^2 and the target density is n target atoms/cm^3. If the target length is dx cm, the number of particles interacting in the target and thus disappearing from the beam, will be given by

$$\mathrm{d}I = -\mathrm{In}\ \mathrm{d}x\sigma\,. \tag{2}$$

Here we have introduced the symbol σ for the cross-section. In nearly all cases the symbol σ is used to indicate cross-sections. Given that the beam particles will move with a speed of v m/s the thickness traversed during a time interval dt will be

$$\mathrm{d}x = v\,\mathrm{d}t\,. \tag{3}$$

Combining Eq. (2) and Eq. (3) we have

$$\frac{\mathrm{d}I}{\mathrm{d}t} = -\mathrm{In}\ v\sigma \tag{4}$$

with the solution

$$I = I_0 e^{-\frac{t}{\tau}} \tag{5}$$

where

$$\tau = \frac{1}{n\sigma v}\,.$$

Thus we see that the intensity of the beam will fall off exponentially with a time constant inversely proportional to the cross-section. In the case of an accelerator beam interacting with the rest gas we normally have a mixture of H_2, CH_2, CO, CO_2 and other gases. In this case we have to replace $n\sigma$ in the formula above with $\sum_i n_i\sigma_i$ and then we get

$$\frac{1}{\tau_{\mathrm{total}}} = \frac{1}{\tau_{\mathrm{H_2}}} + \frac{1}{\tau_{\mathrm{CH_2}}} + \frac{1}{\tau_{\mathrm{CO}}}\frac{1}{\tau_{\mathrm{CO_2}}} + \ldots\ . \tag{6}$$

2 Classifications of cross-sections

In order to estimate the lifetime in a concrete case we thus need to know the different cross-sections entering in the calculation. What are typical values of cross-sections? Unfortunately there is no simple rule. There are no typical values. The cross-section depends on many factors. It depends of course on the target particle, i.e., the composition of the rest gas but also on the type of particle being accelerated. In addition there is often a strong energy-dependence and a dependence on the type of interaction that is involved. There are no simple guidelines.

To get an idea of the dependence on the target particle we can again rely upon some simple considerations. The typical size of the radius of an atom is of the order of one angstrom or 10^{-8} cm and

thus if we apply the simple geometrical rule $\sigma \sim \pi R^2$ we get a cross-sections in the order of megabarns. The typical radius of a nucleus is of order 10 fermi or 10^{-12} cm giving cross-sections in the barn region. The proton in itself has radius of about a fermi giving cross-sections in the millibarn range.

Similarly, there is a dependence on the incident particle. In accelerators we mainly deal with protons, electrons, or ions. They are fundamentally very different particles. The size and mass are very different but also the compositeness and the forces with which they interact. There are four fundamental forces in nature. The forces of gravity and electromagnetism are familiar in everyday life. Two additional forces are introduced when discussing nuclear phenomena: the strong and weak interaction. The strong interaction is what holds the quarks together to form a proton while the weak force governs beta decay and neutrino interactions with nuclei. The forces which are relevant for us when considering beam gas interactions are the strong force and the electromagnetic force. When two protons encounter each other, they experience both these forces simultaneously. However, the strong force dominates for head-on collisions and the electromagnetic forces dominate for peripheral collisions. Electrons do not feel the strong force at all and thus only the electromagnetic force is relevant.

As mentioned above the cross-section depends also on the energy of the incident particles. Just as an illustration of some typical energy dependence we show the energy dependence in proton–proton interactions in Fig. 2.

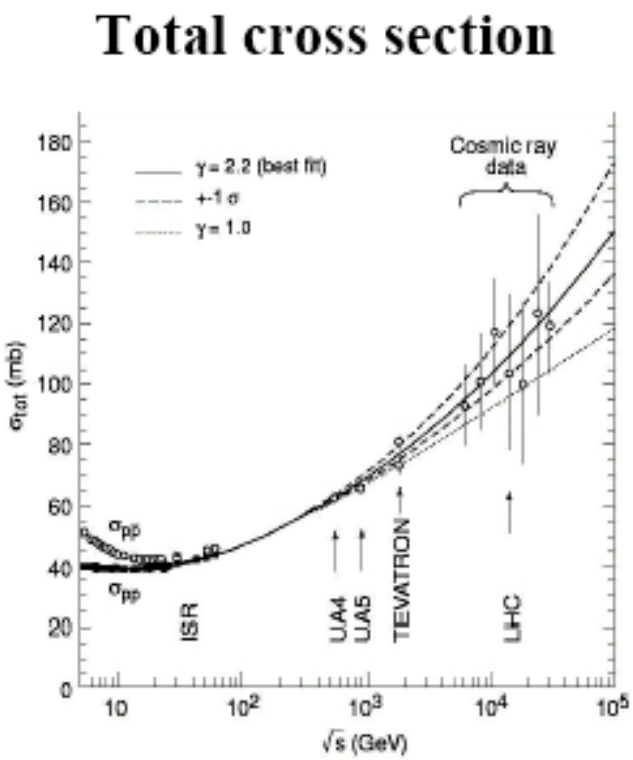


Fig. 2: Example of energy dependence of cross-sections

For future discussions it will be useful to classify the interactions between charged particles and the rest gas atoms. A collision is called *elastic* if the particles do not change identity during the interaction. This is just like the collision of billiard balls. It is possible to have elastic scattering both for the case of electromagnetic interaction and for the case of strong interaction. In the former case both particles must have a charge but in the latter case we can have elastic scattering independent of the charge. Moreover, as you will see, there can be both single and multiple elastic scattering.

All reactions that are not elastic are called *inelastic*. In this case there can be a change of nature of the particles and also new particles can be created. We can further classify the inelastic electromagnetic interactions as follows: bremsstrahlung, ionization, electron capture, and electron loss. In the case of strong interactions we have nuclear reactions, particle break up, and particle creation. In the following we shall look at some of the possible elastic and inelastic reactions in more detail.

3 Elastic scattering

3.1 Coulomb scattering

Elastic scattering via the electromagnetic force is called Coulomb scattering and sometimes Rutherford scattering. Actually it was Rutherford who realized while observing results from scattering of alpha particles off a thin foil that the atom must have a very tiny nucleus compared to the size of the atom itself. The scattering angle depends on the impact parameter as illustrated in Fig. 3. Small impact parameters give large scattering angles and vice versa.

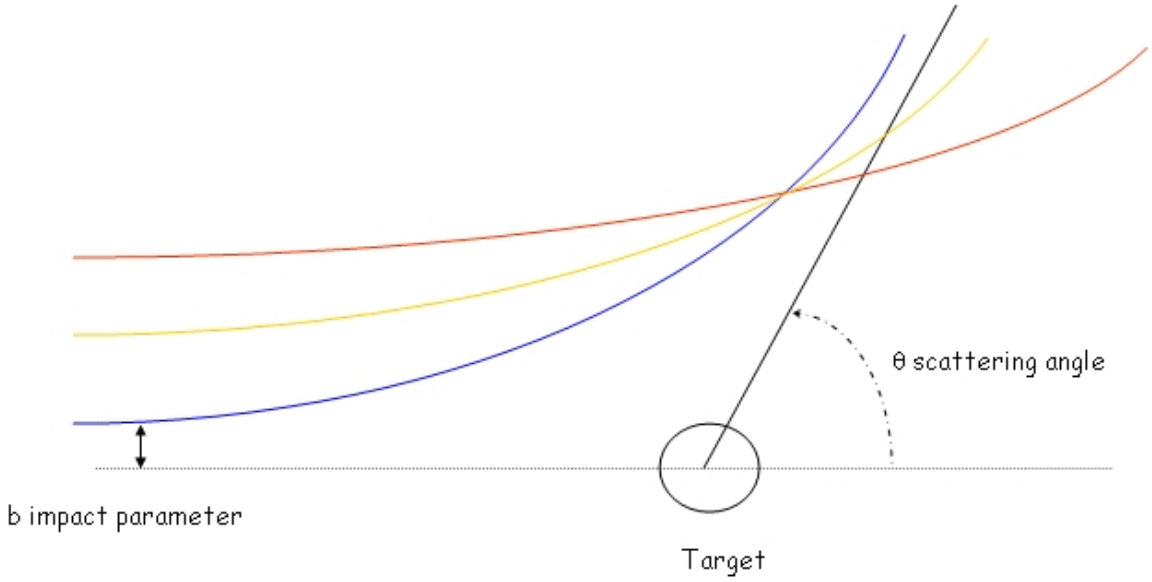


Fig. 3: Relation between impact parameter and scattering angle

It is rather straightforward to deduce a formula that gives the cross-section per unit solid angle as a function of scattering angle θ

$$\frac{d\sigma}{d\Omega}(\theta) = \left(\frac{ZZ'e^2}{4E}\right)^2 \frac{1}{(\sin\theta/2)^4}\,. \tag{7}$$

Here Z and Z' are the charge of the projectile and the target, respectively, and E is the kinetic energy of the moving particle.

There are deviations from this simple formula as illustrated in Fig. 4.

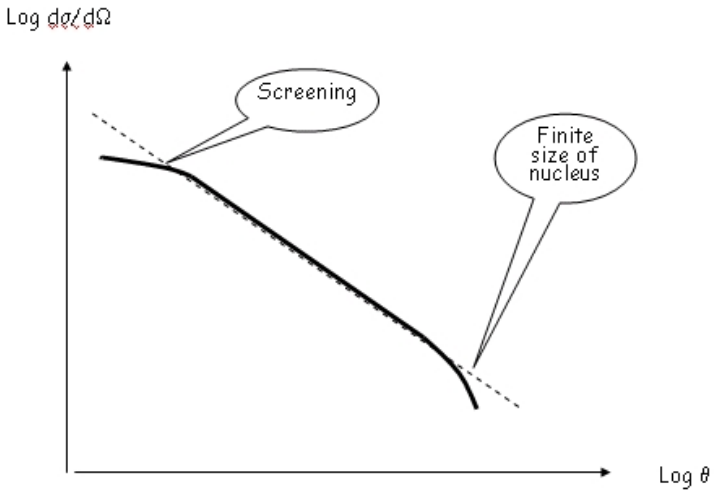


Fig. 4: Deviations from simple Coulomb scattering

For small scattering angles or large impact parameters, the screening effect of the surrounding electrons starts to play a role. For the very big angles or small impact parameters, the finite size of the nucleus affects the result.

3.2 Multiple Coulomb scattering

Many small-angle events will lead to an accumulation effect which will affect the emittance of the beam. The cumulating effect of many small deviations will lead to an emittance growth that in turn affects the lifetime of the beam. The r.m.s. value of the scattering angle can be approximated by a simple formula [1]:

$$\theta_{\text{r.m.s.}} \propto \frac{1}{p}\sqrt{\frac{L}{X_0}} \tag{8}$$

where L/X_0 is the amount of material seen by the particle expressed in radiation lengths and p is the momentum of the incoming particle. We shall come back to the definition of radiation lengths in a moment.

3.3 Elastic scattering via strong interaction

As briefly mentioned before, elastic scattering can also occur via strong interaction. This part of the elastic scattering dominates at large angles as seen in Fig. 5.

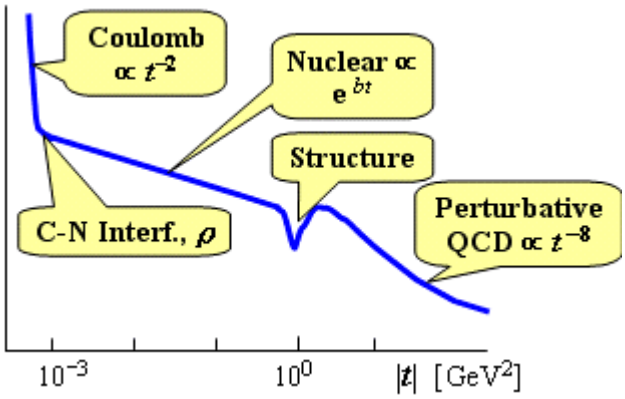


Fig. 5: Differential elastic cross-section as a function of t

In the figure the differential cross-section is given as a function of the variable t. This variable is often used in particle physics and is proportional to the square of the scattering angle. All the effects at large t values are due to the strong interaction. However, as can be seen, the cross-section has dropped considerably and thus is of little importance for beam losses.

3.4 Beam losses from single Coulomb scattering

The important question for us is of course to understand when elastic scattering with the beam–gas atoms implies a loss of beam particles. We shall look at single Coulomb scattering as an illustrative example. It can be shown that giving a circulating beam particle an angular kick θ_i at a point i around the ring will result in an oscillation [2]

$$u(s) = \theta_i \sqrt{\beta(s)\beta_i} \sin\left[\varphi(s) - \phi_i\right]. \tag{9}$$

Here $u(s)$ is the amplitude at a point s around the ring, and $\beta(s)$ and $\phi(s)$ the beta function and phase at that point. Similarly β_i and φ_i are the beta function and phase at the point i. Let A be the half aperture of the dynamic aperture at the position of minimum aperture and β_A the value of the β function at this point. At the position of minimum aperture the particles will thus be lost if

$$\theta_i \sqrt{\beta_A \beta_i} \geq A. \tag{10}$$

Averaging over the circumference of the machine we get

$$\theta_{\max} = \frac{A}{\sqrt{\beta_A \beta_{\text{average}}}}. \tag{11}$$

To get the loss cross-section σ_{loss} we just need to integrate $\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\theta)$ from Eq. (7) from $\theta_{\max}$ to π and we get

$$\sigma_{\text{loss}} \propto \frac{Z'^2}{E^2} \frac{1}{\theta_{\max}^2} = \frac{Z'^2 \beta_A \beta_{\text{average}}}{E^2 A^2}. \tag{12}$$

This is the cross-section to be used for lifetime calculation. We see that the loss cross-section is a strong function of Z' of the gas and that for a given aperture the loss cross-section decreases with higher energies and becomes small with small beta values.

4 Inelastic cross-sections

4.1 Electron beams—bremsstrahlung

In general, when a charged particle is accelerated it emits electromagnetic radiation. If a charged particle is accelerated by the field of an atomic nucleus it will emit electromagnetic radiation in the form of photons. This phenomenon is called bremsstrahlung. The origin of the word is German and it simply means ‘braking radiation’. The energy that is emitted by the accelerated particle is proportional to the inverse mass squared of the particle and thus bremsstrahlung is very important for light particles like electrons and positrons. The energy loss of the particle is roughly proportional to the energy of the

particle and thus bremsstrahlung becomes important at high energies. The constant of proportionality is normally written as $1/X_0$ where X_0 is called the radiation length and is a characteristic of the material (it depends on the square of the atomic number of the material and the density of the material).

We thus have

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{E}{X_0} \tag{13}$$

and

$$E_{\mathrm{ave}} = E_0 e^{-\frac{x}{X_0}} . \tag{14}$$

From this we see that the radiation length can be interpreted as the distance a particle traverses in a medium until its energy has gone down a factor $1/e$.

The energy distributions of photons which are emitted in the process follow the so-called Bethe–Heitler spectrum [1]:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\varepsilon} \propto \frac{4}{3}\frac{1}{X_0}\frac{F(\varepsilon,E)}{\varepsilon} \tag{15}$$

where $F(\varepsilon,\, E)$ is given by

$$F(\varepsilon,E) \cong 1 - \frac{\varepsilon}{E} + \frac{3}{4}\left(\frac{\varepsilon}{E}\right)^2 . \tag{16}$$

Here E is the nominal energy of the radiating particle and ε the energy of the emitted photons. Observe that $F(\varepsilon,\, E)$ is a rather slowly varying function.

In the case of bremsstrahlung it will be the energy acceptance of the accelerator that will determine the losses. Assume that all energies from the nominal E down to $E - \varepsilon_{\mathrm{m}}$ are accepted. This means that energy losses bigger than ε_{m} lead to particle losses.

To get σ_{loss} we need to integrate $\mathrm{d}\sigma/\mathrm{d}\varepsilon$ of Eq. (15) from ε_{m} to E and we get

$$\sigma_{\mathrm{loss}}^{\mathrm{brems}} \propto \frac{4}{3X_0}\left(\ln\frac{E}{\varepsilon_{\mathrm{m}}} - \frac{5}{8}\right) \quad (\varepsilon_{\mathrm{m}} \text{ is small relative to } E) . \tag{17}$$

We see that σ_{loss} has a strong dependence on the atomic number of the residual gas via the $1/X_0$ factor but a rather weak dependence on the maximum energy acceptance. It should be pointed out that this is a rather simplified treatment excluding the effect from electrons of the atom and screening effects.

4.2 Ion beams

Ion beams are in general more complicated than electron or proton beams. There are two more degrees of freedom to take into account. There is the charge of the nucleus being accelerated and there is the total charged state of the atom (i.e., number of electrons attached minus the charge of the nucleus). At low energies, the interaction with the rest gas atoms will lead to either a capture of additional electrons or stripping (loss) of existing electrons. Both the capture cross-section and the loss cross-section are functions of the charge of the nucleus, the total charge state of the atom, the energy of the ion being accelerated, and the charge of the rest gas nucleus.

At low energies we can compare in a simple and intuitive picture the velocity of the accelerated ion with the velocity of its outermost electrons. If the velocity of the ion is close to the velocity of the outermost electrons, the capture cross-section will be approximately balanced with the loss cross-section. Electron capture will dominate if the velocity of the ion is significantly lower than the electron velocity. On the contrary, for higher energies with velocities of the ion significantly higher than the electron velocity, the electron loss will dominate. The notion of equilibrium charge is important in this context. The equilibrium charge ($\bar{q}$) is reached after many collisions in a given gas and it is approximately the charge state for which the orbit velocity of the outermost electrons is equal to the ion velocity. If the charge of the ion is not too far away from the equilibrium charge, there are some simplified scaling rules for the cross-section. Below we give an example of a simple scaling rules [3],

$$\sigma_{\mathrm{c}}(q) = \sigma_{\mathrm{c}}(\bar{q})\left(\frac{q}{\bar{q}}\right)^{a} \tag{18}$$

$$\sigma_{1}(q) = \sigma_{\mathrm{c}}(\bar{q})\left(\frac{q}{\bar{q}}\right)^{b} \tag{19}$$

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. Observe that other scaling rules also exist.

At very high and relativistic energies like RHIC (100 GeV/A) and LHC (2.76 TeV/A) we deal with bare ions with high atomic number and there are different mechanisms coming into play. For the peripheral collisions with large impact parameter the electromagnetic cross-section is large and of the order of several hundred kbarn. Because of the high atomic number of the ion there is a strong electromagnetic field and e^+e^- production dominates. This reaction in itself is a rather harmless inelastic reaction because there is no significant change of momentum of the ion. However, in some cases, the produced electron is captured by the ion and then the charge state of the ion changes and it is of course lost from the beam. In this high-energy regime there is also the mechanism of electromagnetic dissociation of the nucleus. The photon exchange between the bare nucleus and the rest gas nucleus leads to a break-up of the ion which is then again lost from the beam. These processes also occur between the ions in the two colliding beams. The predicted values of the cross-section for those processes at LHC and for Pb–Pb collisions are given in Table 1 [4].

Table 1: Cross-sections for Pb–Pb collisions at the LHC [4]

Cross-sections for Pb–Pb collisions	Symbol	(barn)
Hadronic	σ_{h}	8
EM dissociation	σ_{emd}	225
e^- capture	σ_{ec}	204

In addition to the cross-section for e^+e^- production followed by electron capture and the cross-section for electromagnetic dissociation, the strong hadronic cross-section is given. Observe that due to the high charge of the ions the hadronic cross-section is smaller than the electromagnetic.

The large cross-section seen in Table 1 means that we get a short lifetime from the beam–beam interaction and normally not from the beam–gas interaction. Beam lifetimes at LHC and RHIC from beam–beam interactions are in the range of hours for high atomic numbers of the colliding ions and vary very quickly with the atomic number. Some examples of how the cross-section varies with the atomic number of the ion species colliding are given in Table 2 [4].

Table 2: Cross-sections for different ion species colliding at the LHC [4]

Ion Species	σ_h **(barn)**	σ_{emd} **(barn)**	σ_{ec} **(barn)**	σ_{total} **(barn)**
Pb	8	225	204	437
Sn	5.5	44.5	18.5	68.5
Kr	4.5	15.5	3.0	23.0
Ar	3.1	1.7	0.04	4.84
O	1.5	0.13	1.6×10^{-4}	1.63

4.3 Proton beams

The proton is about 2000 times heavier than the electron and this implies that the bremsstrahlung of protons is heavily reduced as compared to electrons. Compared to heavy ions, the charge of only one unit for the protons also implies lower rates for electromagnetic reactions like electron capture. As a consequence, for proton beams, the strong interaction is of importance. There are basically two types of reactions involving the strong interaction; diffractive and non diffractive interactions. Diffractive interactions imply that no quantum numbers are exchanged between the incoming proton and the rest gas nucleus. However, in contrast with elastic scattering, some of the kinetic energy of the incoming proton is converted to mass and several new particles are created. The non-diffractive interactions are more violent. Here the quarks in the proton and the quarks of the rest gas nucleus collide head on and both the proton and the rest gas nucleus are typically destroyed. From the point of view of beam losses in an accelerator both the diffractive and non-diffractive reactions in general mean total loss of the proton.

The proton–proton cross-section depends strongly on energy as can be seen from Fig. 2. In general, at high energies, the proton–nucleus cross-section has the same energy dependence as the proton–proton cross-section and as a rule of thumb the proton–nucleus cross-section can be deduced from the proton–proton cross-section according to

$$\sigma_{pA} = \sigma_{pp} A^{0.7} \quad . \tag{20}$$

Table 3 gives an example of the expected proton–nucleus cross-section at LHC for some nuclei corresponding to frequently present gases [5].

Table 3: Expected proton–nucleus cross-section at LHC for some gases [5]

Gas	H_2	He	CH_4	H_2O	CO	CO_2
Cross-section (mb)	94	130	568	554	840	1300

4.4 Ionization of the rest gas

When a charged particle traverses a medium it suffers repeated collisions with the electrons of the atoms. The electron may be kicked out of its orbit and we have what is called ionization energy loss. The cross-section is large but the energy losses are very small and thus the beam particles are hardly affected. The energy loss is given by the well-known Bethe–Block formula [1]:

$$\frac{\mathrm{d}E}{\mathrm{d}X} = K z^2 \frac{Z}{A}\frac{1}{\beta^2}\left(\frac{1}{2}\ln\frac{2m_e^2 c^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta}{2}\right). \tag{21}$$

Here K is a constant, β and γ the usual relativistic variables describing the accelerated particle ($\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$), z is the charge of the incoming particle, Z/A is the charge mass ratio of the rest gas atom, I is the ionization potential, $T_{\max}$ is the maximum kinetic energy which can be imparted to a free electron in a single collision, and δ is a relativistic correction factor. The energy dependence is seen from this formula. At low energies the $1/\beta^2$ term dominates and at higher energies the dependence on the logarithmic term takes over. This is seen in Fig. 6.

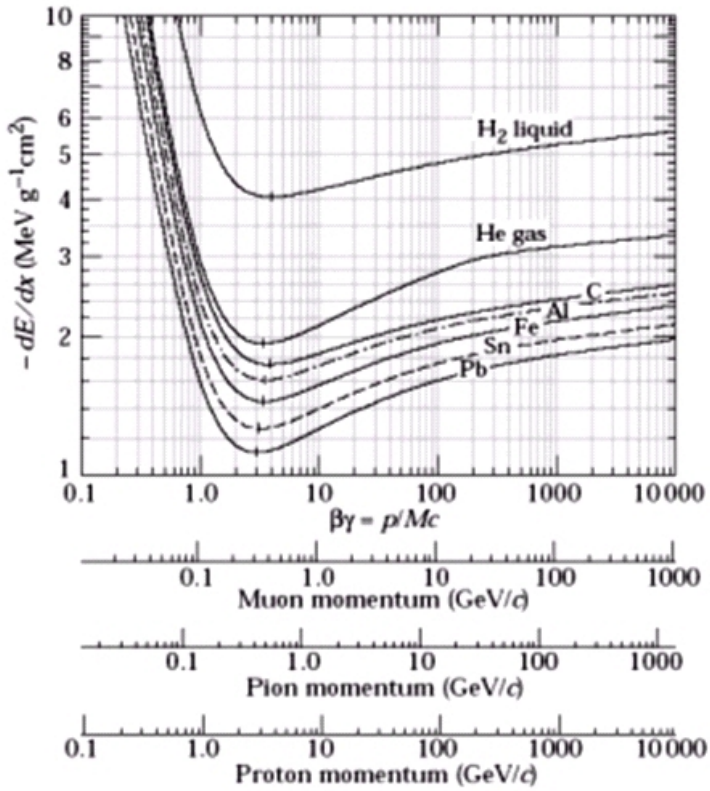


Fig. 6: dE/dX as a function of energy

What is important for us is the z^2 term. The energy loss depends on the square of the charge of the accelerated particle. Even if the energy losses from ionization in general are small, we can have some more substantial energy losses for bare ions with high atomic number. As the cross-section for ionization is large there might thus be some beam losses associated with ionization for ions with high atomic number but ionization is in general not an important beam loss factor.

5 Activation

With activation or induced radioactivity we mean the process of making a material radioactive by bombardment with particles or radiation. It basically means the transformation of a stable nucleus of an atom to one or several unstable nuclei.

There are several origins of activation in accelerators. The dominating source is objects that are directly hit by the primary beam, like dumps, septa, collimators or other beam obstacles. Another source is related to localized beam losses associated with some malfunctioning piece of equipment. Bad vacuum is in general not the main source of activation. Observe that the problem of activation is significantly less severe in electron machines as compared to hadron machines and this is of course due to the different interactions and the different processes that are involved.

There are basically two distinct phenomena that are involved in activation. There is the reaction that creates the unstable nuclei and there is the radioactive decay of the unstable nuclei created. The various nuclear decay modes are well known. The most frequent decay modes are alpha and beta decays, gamma transitions, and spontaneous fission. In the first two cases the nucleus emits an alpha particle and an electron, respectively. The latter cases deal with photon emission and spontaneous break-up of the nucleus. In the following we shall no longer discuss the decay modes but give a couple of examples of reactions that create unstable nuclei. A typical example is neutron capture. Here all cross-sections are well known but they may vary from the kbarn range to the mbarn range (see Fig. 7 [6]).

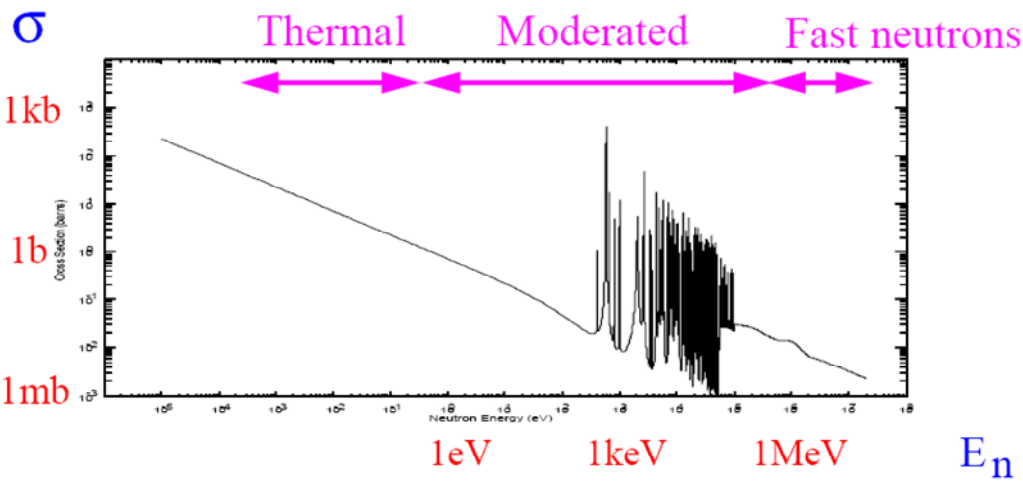


Fig. 7: Neutron cross-section as a function of energy

The activation depends very strongly on the material in the case of neutron capture. As an example we can compare the neutron capture cross-section for Ag^{109} which is 20 kbarn with that of Pb^{208}, being only 3 barns.

The activation by high-energy hadrons has different properties. The origin is so-called 'stars' or violent inelastic reactions. Here there are a large number of possible final states and not all reaction cross-sections are well known. Also the activation depends more weakly on the material. An example of the material dependence is given in Fig. 8 [7].

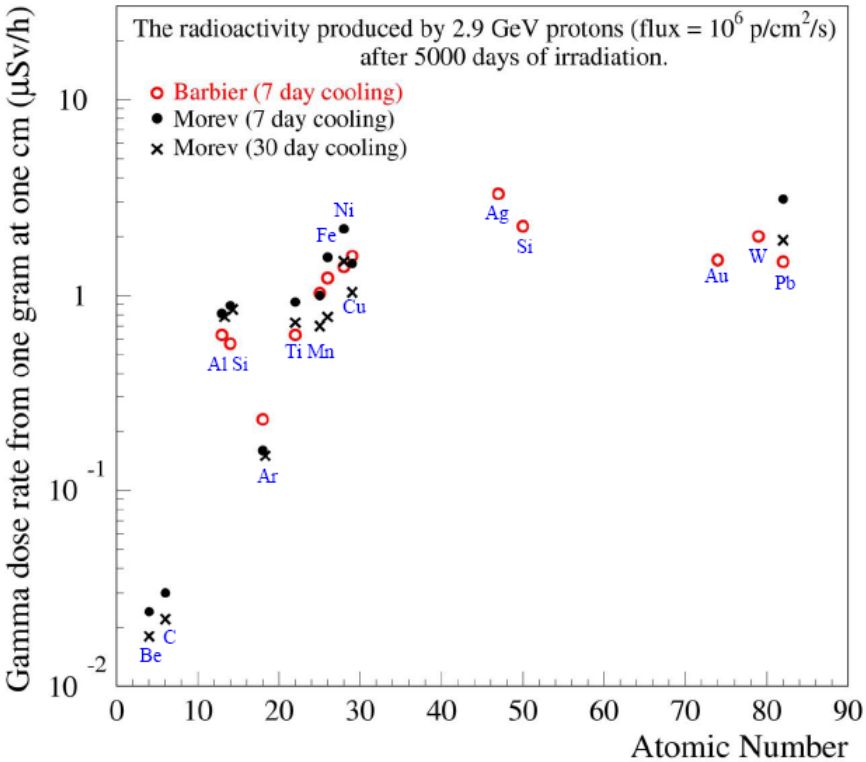


Fig. 8: Activation dose rate as a function of atomic number

As an example of considerations relevant for choice of material of a beam pipe we shall take the beam pipe of the ATLAS experiment. Observe that here the dominating source of activation will be the beam–beam interactions at the intersection point and not the beam–gas interaction. Looking again at Fig. 8 we see that a beryllium or a carbon beam pipe would hardly get activated at all compared to other materials. However, there are several other inconveniences associated with such materials and thus it is also interesting to compare the difference between an aluminium pipe and a steel pipe. In comparing induced activity from different materials one has to consider both the exposure time and the cool-off time. An example for aluminium and steel is given in Table 4 [7].

Table 4: The ratio of the dose rates from a steel and an aluminium beam pipe

Cooling time/running time	5000 days	1000 days	100 days	30 days
1 day	9	13	23	23
5 days	9	15	76	180
30 days	4	7	22	39

We see that depending on the conditions there can easily be a factor 10 gained by using aluminium instead of steel.

The ATLAS beam pipe was in the end constructed with a beryllium pipe close to the intersection point and stainless-steel pipe further away. The choice of beryllium was dictated by the necessity of minimizing multiple Coulomb scattering in the walls of the beam pipe and not by activation considerations. However, the stainless-steel pipes might be exchanged for aluminium pipes at a later stage because of the favourable activation properties.

Acknowledgement

I want to thank the organizer for a very pleasantly organized school and I also thank Daniel Brandt for having invited me to give this lecture.

References

[1] See, for example, S. Eidelman *et al., Review of Particle Physics,* Phys. Lett. **B592** (2004).

[2] See, for example, E.J.N Wilson, *An Introduction to Particle Accelerators* (Oxford University Press, Oxford, 2001).

[3] B. Franzke, *Interaction of stored ion beams with residual gas* (CERN Accelerator School, Fourth Advanced Accelerator Course, CERN 92-01, p. 100).

[4] D. Brandt, *Review of the LHC ion program*, LHC Project Report 450 (2000).

[5] K. Eggert, K. Honkavaara and A. Morsch, *Luminosity considerations for LHC*, CERN AT/94-04, LHC Note 263.

[6] S.A. Cetin and A. Romaniouk, *Induced radioactivity in the ATLAS inner detector*, ATL-INDET-2000-013 and references therein.

[7] S. Baranov *et al*., *Estimation of radiation background, impact on detectors, activation and shielding optimization in ATLAS,* ATL-GEN-2005-001 and references therein.

Bibliography

S.P. Moeller, *Beam-residual gas interactions* (CERN Accelerator School, Vacuum Technology, CERN 99-05, p. 155).

A. Wrulich, *Single-beam lifetime* (CERN Accelerator School, Fifth General Accelerator Physics Course, CERN 94-01, p. 409).

C. Bochetta, *Lifetime and beam quality* (CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, CERN 98-04, p. 221).

B. Franzke, *Interaction of stored ion beams with residual gas* (CERN Accelerator School, Fourth Advanced Accelerator Course, CERN 92-01, p. 100).

Leak detection

K. Zapfe
Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

Abstract
This paper will give an introduction to the leak detection of vacuum systems. Various methods to detect leaks as well as the most widely used helium leak detectors and their different applications are presented. Practical examples in the context of accelerator vacuum systems will illustrate the topic.

1 Introduction

Ideally a vacuum chamber should maintain the achieved vacuum pressure forever after switching off the pumps. However, without active pumping the pressure in a real system will rise with time. This pressure rise is produced by outgassing—the spontaneous evolution of gas molecules from the walls—and gas molecules penetrating through leaks and entering by permeation from the outside into the vacuum system. Figure 1 shows three typical curves for the behaviour of the pressure increase as a function of time on a linear scale: outgassing starting with a linear pressure increase and levelling off once outgassing and vapour pressure reach equilibrium, the linear increase due to a leak, and the combined curve of the two effects.

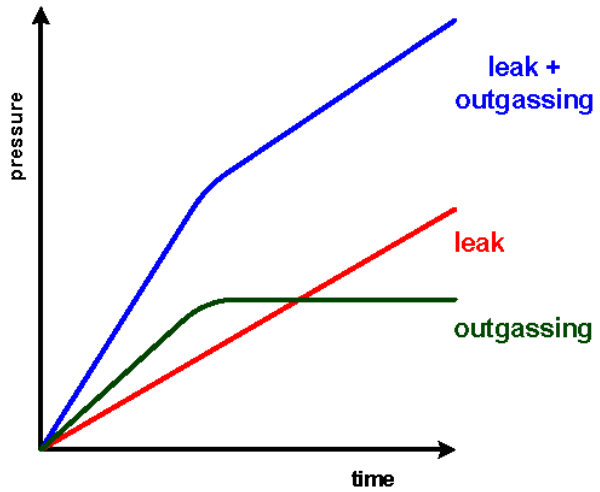


Fig. 1: Pressure increase due to outgassing, a leak, and the combination of outgassing and a leak

In practice it is impossible to build a completely leak-tight vacuum system. And this is not even necessary. On the other hand, the leak rate must be small enough to allow the required pressure level to be reached. Therefore it is important to specify an acceptable leak rate for each vacuum system. After manufacturing of a vacuum vessel it must be proven that the tightness specifications are fulfilled. Further checks are necessary during as well as after assembly and installation to locate possible leaks created during the previous steps. Thus leak detection is an important step in the production of vacuum to guarantee that the required pressure and gas composition conditions of a vacuum system can be reached. Adequate methods and leak detection equipment have been developed over the past decades following the ever-increasing demands by industry.

This lecture will start with some introductory remarks about leak rates and leak types followed by various leak detection methods. The most widely used helium leak detectors and their different applications will be described in detail. This will be illustrated by practical examples in the context of accelerator vacuum systems.

2 Leak rate

The leak rate is defined as the pV throughput of a gas through a leak. It is a function of the type of gas, pressure difference, and temperature. In a system of volume V the leak rate Q_l is given by

$$Q_l = V \cdot \frac{\Delta p}{\Delta t} \ . \tag{1}$$

Here Δp is the pressure rise during the time interval Δt. Commonly used units for leak rates and their conversion factors are given in Table 1.

Table 1: Conversion factors for leak rates in various system units [1]

	mbar · l/s	**Torr · l/s**	**Pa · m^3/s**	**cm^3/s***
mbar l/s	1	0.75	0.1	0.99
Torr l/s	1.33	1	0.133	1.32
Pa m^3/s	10	7.5	1	~10
cm^3/s*	1.01	0.76	0.101	1

* STP - standard temperature and pressure (0°C, 1 atm)

As an example, for a high vacuum (HV) system one can take the leak rates given below as a rule of thumb:

- $Q_l < 10^{-6}$ mbar l/s: very tight system
- $Q_l < 10^{-5}$ mbar l/s: tight system
- $Q_l < 10^{-4}$ mbar l/s: leaky system.

The following examples illustrate the relationship between the size of a pore, the corresponding leak rate and the amount of gas entering into a vacuum system. For simplification it is assumed that the pore is a straight channel of circular shape. A diameter of 0.01 mm, e.g. a hair, corresponds to a leak rate of 10^{-2} mbar l/s. For a pore with a leak rate of $Q_l = 10^{-10}$ mbar l/s = 10^{-10} cm^3/s, an amount of 1 cm^3 of gas needs 317 years to flow through the leak channel.

3 Leak types

Leaks may appear because of various defects within the material and/or at connecting areas:

- fixed connections by brazing, welding, or gluing—especially transitions between different materials like glass-metal, ceramics-metal, etc;
- pores and hair cracks due to mechanical—or thermal stress, which to some extent are always present and therefore must be small enough in size and number not to cause disturbance;

– flanged connections;

– cold/warm leaks opening up at extreme temperatures—often being reversible;

– virtual leaks, where gas evaporates from inner excavations, dead holes etc., leading eventually to long pumping times (see Sub-section 8.1) and

– indirect leaks from supply lines, for example, for cooling water or gas/liquids (He, N_2) of cryogenic systems.

A leakage, but not a leak due to a defect is the

– *permeation*, i.e., the natural porosity of material.

The permeation could be quite significant and even limit the detectable leak rate. For example, at HERA large amounts of Perbunan® rubber seals with a permeation rate of $2 \cdot 10^{-2}$ mbar l mm/s/m^2 are used at the insulating vacuum tank resulting in a total permeation rate of $2 \cdot 10^{-5}$ mbar l/s/m [2].

This long list shows that during design and manufacture great care is necessary to avoid potential leaks.

4 Leak detection methods

The aims of a leak search are to localize a leak and/or to determine the total or local leak rate. Depending on the size of the leak, various effects can be used for leak detection. All methods are based on the variation of a physical property measured on one side of the vessel while the pressure or the nature of the gas is changed on the other side. Large leaks can generate mechanical effects as described in Sub-section 4.1, while smaller leaks require more sophisticated methods (see Sections 4.2 and 4.3). A detailed list of possible leak detection methods, their sensitivities, and references can be found in Ref. [3]. A comprehensive review of leak detection methods and apparatus is given in Refs. [1], [4] and [5].

4.1 Mechanical effects

Methods applying measurable mechanical effects are limited to large leaks. For example ultrasound detectors may be used to monitor the oscillations produced by the gas in the vicinity of a leak. The detection limit of this method is limited to leak rates of 10^{-2} mbar l/s. Somewhat more sensitive is the formation of bubbles when water or even soaped water is spread on a leak with the vessel being pressurized to over pressure. A sensitivity of 10^{-4} mbar l/s can be reached.

These methods are simple, very quick to carry out, cheap and one can locate a leak. However, owing to the limited sensitivity they are restricted to the high-pressure region.

4.2 Pressure increase

In approaching the problem of a supposed leak larger than the admissible value, it is necessary to determine first if such a leak actually exists. Taking curves of the pressure versus time as shown in Fig. 1 will assist in determining if there is a leak and estimating the actual leak rate. First the system is evacuated to a stable minimum pressure. When no further improvement in the pressure is evident, the pumps are switched off or separated from the system. There will always be an initial pressure rise due to outgassing, thus one will have to wait sufficient time until outgassing has come to equilibrium with the vapour pressure in the gas phase. In case of a leak, the pressure will continue to increase linearly with a different gradient instead of levelling off. The pressure curve should be taken till the shape of the curve becomes evident. More details might be found in Refs. [3] and [6].

However, it requires a lot of experience to reliably analyse such curves during initial pump down with still a lot of outgassing or in the presence of virtual leaks. Here the application of a liquid nitrogen trap might be useful to minimize the effect of the vapours present in the system. Small leaks might not be visible with this method. In any case additional methods as described in the following sub-section need to be applied for further diagnosis.

4.3 Modification of the physical properties of the residual gas

By changing locally the air composition in the vicinity of a leak by adding a gas (*the tracer gas*), the composition of the residual gas and thus its physical properties are modified. As accurate and sensitive measurement methods for many physical gas properties are available, measuring these alterations can be used to determine the position and size of a leak. The sensitivity of such methods is sufficient to detect even quite small leaks. In the following the most widely used methods are briefly described.

The variation of the heat conductivity can be detected using a Pirani gauge and alcohol, CO_2, or helium as tracer. The pressure will rise in case of helium, otherwise drop down. Using (heavy) noble gases, the change in ionization cross-section can be monitored by the signal of an ion gauge or even a sputter ion pump. A nice example of this method is described in Ref. [7].

Analysing the mass of the residual gas is the most sensitive and widespread method in locating leaks. Using an optimized mass spectrometer and helium as tracer gas, leak rates down to 10^{-12} mbar l/s can be detected. A more detailed description will follow in Section 5.

4.4 Tracer gas

The tracer gas should have the following properties:

- unambiguous signal in the mass spectrum of the residual gas;
- chemically and physically inert, non explosive and cheap;
- very low content in air and
- easily removable by pumping and does not contaminate the system.

Helium is most commonly used since it fulfils all conditions mentioned above. Therefore the leak rate is usually given as helium standard leak rate (He Std) assuming a pressure difference of 1 bar from air to vacuum. The small diameter of the helium atoms allows the detection of very small leaks. As the speed of helium is three times higher than the speed of air, the amount of helium entering through a leak and thus the sensitivity is increased by a factor of 3 compared to air.

Alternatively, argon is frequently used as tracer gas. However, the signal in the mass spectrum of the residual gas has a mixed signature of mass 40 for Ar^+ and hydrocarbons and mass 20 for Ar^{++}, ^{20}Ne and $^{18}OH_2$. In addition the natural content of argon in the ambient air is significantly higher than that of helium.

5 Helium leak detectors

In principle any type of residual gas analyser can be used as mass spectrometer helium leak detector. The most sensitive and safe device, however, is a mass spectrometer with 180° magnetic sector field optimized for the detection of mass 4.

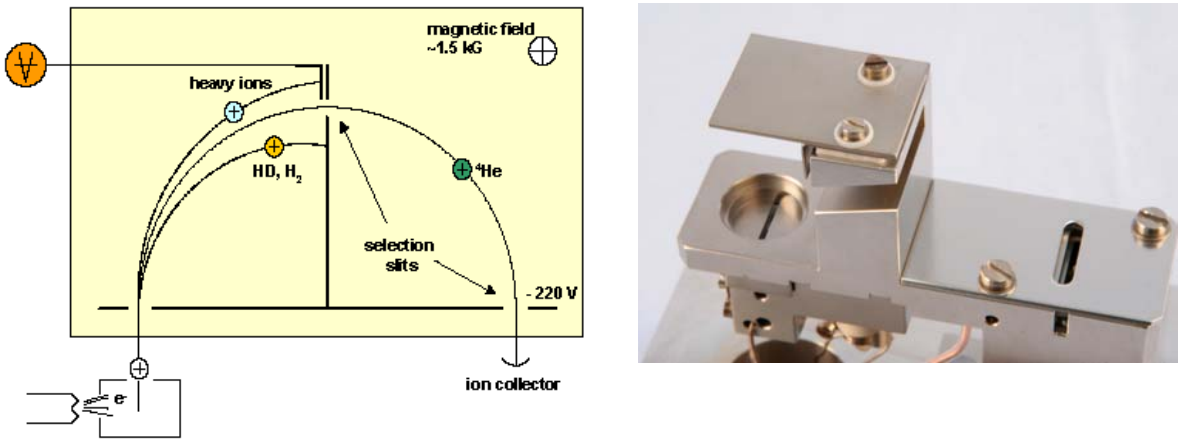


Fig. 2: Schematic drawing of a helium leak detection cell with 180° magnetic sector field (left) and photo of the slit system (right)

Figure 2 shows a schematic drawing and a photograph of such a helium leak detection cell. The central parts are an ion source, a separation system, and an ion collector system. The residual gas is ionized by bombardment with electrons. The resulting ions are accelerated into a magnetic field, where they are deflected. A system of slits allows the helium atoms to pass to a collector unit located at a deviation of 180°. All other ions are blocked. The magnetic field of about 1.5 kg G is typically produced by permanent magnets. A second collector at the middle plate is used to measure the ion current of heavy ions, which is then converted into a total pressure value.

To detect tiny leaks the currents to be measured are extremely small. At the highest sensitivity of 10^{-12} mbar l/s currents as low as 10^{-15} A have to be measured. This is achieved thanks to the use of electron multipliers in the most modern detectors.

In a mass spectrometer the typical path length of the ions to be collected is about 15 cm. As the ions should pass without collisions with other gas molecules, the operating pressure should be below 10^{-4} mbar corresponding to a mean free path of 60 cm.

6 History of helium leak detectors

In 1910 J.J. Thomson described the first mass spectrometer used to detect neon isotopes. For several decades such huge apparatus, easily filling a room, were operated by specialists only. The idea of using mass spectrometers for leak detection goes back to the Manhattan Project in 1942/43 [8]. The leak-tightness requirements needed for the apparatus of the uranium enrichment plants triggered the need for very sensitive and reliable leak detection systems. This led finally to the choice of a mass spectrometer tuned on the mass of helium designed by Dr. A.O. Nier [9]. The first publication dates back to 1947.

During the 1950s and 1960s the original choice of glass as material was replaced by a metal solution since it was much more robust and adequate for industrial use. Also the first compact units including a pump station—weighing about 200 kg—came onto the market. While today the leak detection cell is still very similar to the original design, the pumping systems have considerably improved replacing the original diffusion pumps by turbomolecular or even dry pumps. The sensitivity of helium leak detectors has increased from an initial value of 10^{-6} mbar l/s by 6 orders of magnitude down to 10^{-9} mbar l/s around 1970 and 10^{-12} mbar l/s today.

7 Helium leak detection

When searching for leaks at vacuum vessels the system to be tested is continuously evacuated. The tracer gas penetrating from the outside into the system is pumped through a leak detector where its concentration is measured. For leak location the tracer gas is sprayed locally, while for total leakage measurement, the vacuum vessel is completely enclosed by the tracer gas as shown schematically in Fig. 3.

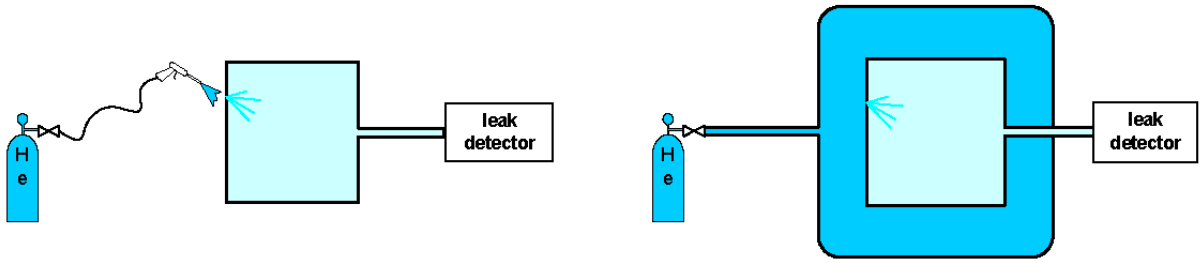


Fig. 3: Leak detection of vacuum vessels using the tracer method – left: leak location, right: total leakage measurement

For pressurized systems the detector method is usually applied. The device to be investigated is pressurized with the tracer gas. For leak location a sniffer probe is connected to a conventional helium leak detector by a capillary tube or needle valve to reduce the pressure from atmosphere to the maximum admissible pressure of about 10^{-4} mbar (see Fig. 4, left). The detection limit of the 'sniffer method' of about 10^{-7} mbar l/s is determined by the natural He content of air. For total leakage measurement the vessel is placed in an evacuated test chamber with detector unit as shown in the right part of Fig. 4.

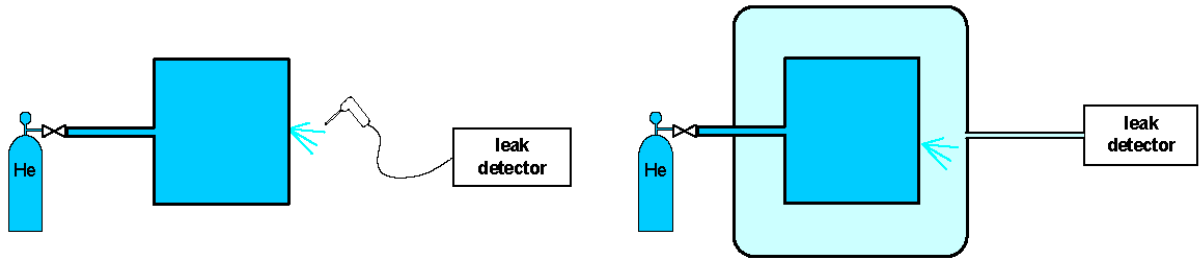


Fig. 4: Leak detection of pressure vessels using the detector method – left: leak location using a sniffer, right: total leakage measurement

The sniffer method can also be applied in case of very large leaks of vacuum systems, when the use of any classical leak detection method is impossible as the pressure can not be sufficiently lowered. In that case the sniffer can be used to detect the presence of helium in the exhaust line of a roughing pump while spraying helium onto the system.

7.1 Direct flow method

In principle the leak detector can be connected to a vacuum system such that the tracer gas flows completely or only partially through the detection cell. Using the direct-flow method, both the leak detection cell and pumps are directly connected to the vacuum system as indicated in Fig. 5. In this case the helium leak rate Q_{He} is given by the effective helium pumping speed $S_{eff,\,He}$ of the high-vacuum pump and the helium partial pressure p_{He} within the cell:

$$Q_{He} = p_{He} \cdot S_{eff,\,He} \quad . \tag{2}$$

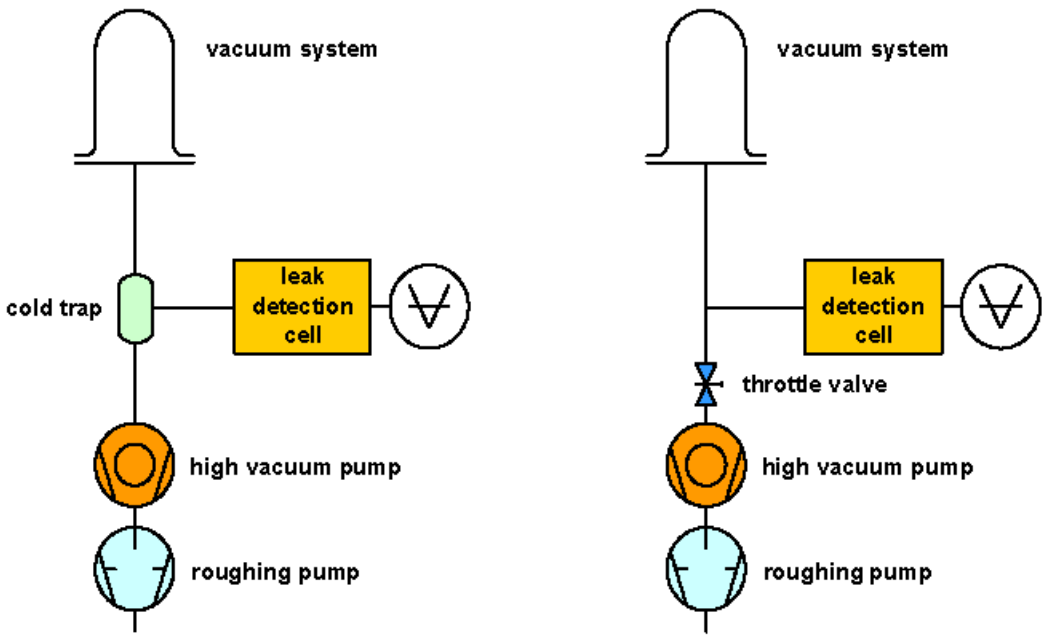


Fig. 5: Layout for the direct flow method with cold trap (left) and throttle valve (right)

The ratio between the partial pressure increase in the leak detection cell and the helium flow through the leak defines the intrinsic sensitivity of a leak detector. In order to increase the sensitivity, the pumping speed for the tracer gas has to be reduced [see Eq. (2)]. This must be done without diminishing the pumping speed for the other gases in order to keep an appropriate operating pressure for the filament emitting the ionizing electrons. As leak detection usually takes place in unbaked systems, selective pumping is therefore needed to provide a high pumping speed for water and a low pumping speed for helium.

Originally diffusion pumps were most commonly used to produce high vacuum. They were combined with a liquid nitrogen trap to effectively condense water (left part of Fig. 5). The trap also impeded back-streaming of oil vapour from the diffusion pump to the leak detection cell. This arrangement has been very successful for many years, although their operation had several disadvantages. The cold trap needed to be refilled periodically during operation thus requiring an easy access to a liquid nitrogen source. The diffusion pumps had to be operated in quite well defined ways as they were sensitive to misuse such as inadequate venting. Since the 1980s turbomolecular pumps started to replace the diffusion pumps and cold traps.

In order to increase the sensitivity, a throttle valve could be installed as indicated in the right part of Fig. 5 to reduce the gas flow to the pumps. According to Eq. (2), decreasing the pumping speed by a factor of 10 decreases the minimum detectable leak rate by the same factor. The direct flow method is quite fast and very sensitive. Nevertheless, in most commercial leak detectors it is nowadays replaced by counter-flow detectors.

7.2 Counter-flow method

In 1968 W. Becker proposed a different arrangement of the pumps and leak detection cell [10]. As shown schematically in the left part of Fig. 6 the leak detection cell is no longer connected directly to the vacuum system, but to the inlet of a high-vacuum pump. However, the realization of the counter-flow method still took several years. Only in the middle of the 1970s were adequate turbomolecular pumps available on the market [11, 12].

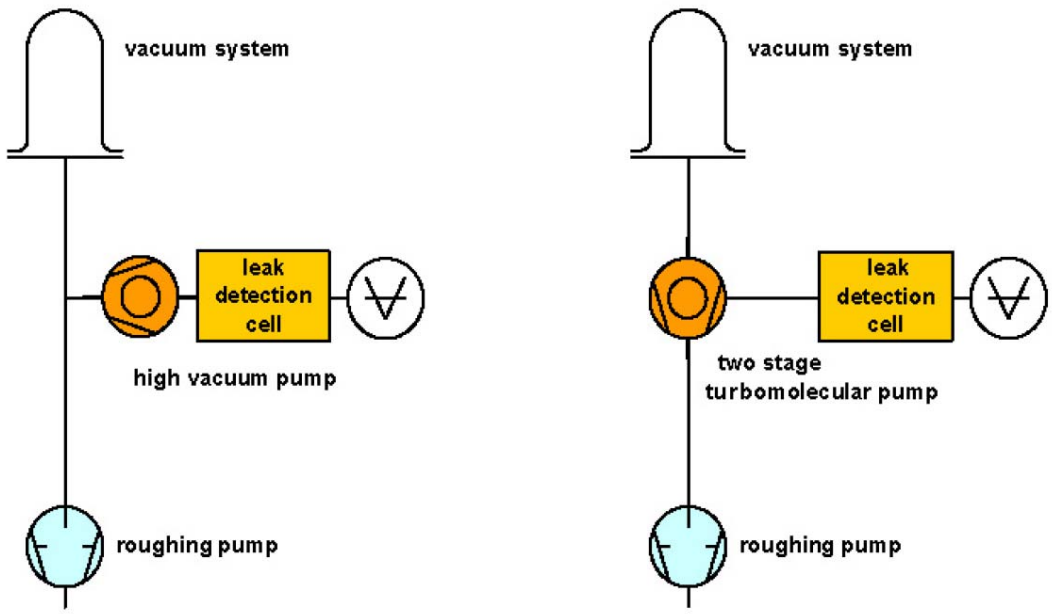


Fig. 6: Layout for the counter-flow method: single-stage high-vacuum pump (left) and two-stage turbomolecular pump (right)

The method is based on the fact that the compression ratio K of turbomolecular (and diffusion) pumps increases rapidly with the mass of the pumped gas. Hence by injecting the gas from the tested vessel at the exhaust of the pump, it is possible to obtain at its inlet a backstreaming flux largely enriched for lighter gases. Typical values for the compression ratio of a turbomolecular pump running at full speed are $K_{\mathrm{He}} = 50$, $K_{\mathrm{H_2O}} = 4000$ and $K_{\mathrm{N_2}} = 30\ 000$ resulting in a suppression by a factor of 80 for water and 600 for nitrogen. Compared to the direct flow method the helium partial pressure within the detection cell is reduced by the turbomolecular pump by the factor K resulting in a leak rate Q_{He} of:

$$Q_{\mathrm{He}} = p_{\mathrm{He}} \cdot S_{\mathrm{eff,\,He}} \cdot K \quad . \tag{3}$$

A major drawback of this simple method is the direct connection of the tested vacuum vessel to the roughing pump and thus the risk of its contamination by oil vapour. In addition, the stability of the pumping characteristics is important to ensure the necessary stability for accurate leak detection. The typically small pumping speed of the roughing pump significantly increases the time constant for leak detection.

As a remedy to the problems described above, more sophisticated commercial leak detectors have been developed using specially designed turbomolecular pumps. For example, installing a second turbomolecular pump between vacuum vessel and roughing pump ensures clean pumping with high pumping speed and thus a short time constant. Nowadays two-stage turbomolecular pumps are most widely used having an outlet flange between the two stages to the leak detection cell (right part of Fig. 6). Alternatively, a simple counter-flow leak detector can be connected to the outlet of the turbomolecular pump in a pump station. With the recent developments of dry pumps, more and more counter-flow leak detectors using dry pumps are available on the market, thus avoiding contaminations from oil vapour.

The counter-flow detectors offer several advantages compared to the direct-flow detectors. They do not need cold traps. The tracer gas no longer flows directly through the detection cell, but just a part of it by backstreaming through the high vacuum pump. Hence for the same flow, the pressure in the cell is lower and consequently the detection can start earlier, i.e., at a higher pressure level of

about $p \cong 10^{-1}$ mbar. In addition the leak detection cell and thus the filament are better protected in case of a sudden pressure increase. As a consequence, however, they are less sensitive ($1 \cdot 10^{-10}$ mbar l/s) than direct-flow detectors, the latter one being still unbeatable in the case of very small leaks.

7.3 Characteristics

Various parameters characterize a leak detector. Most important are the sensitivity, detection limit, response time, and maximum allowable inlet pressure.

According to Eqs. (2) and (3) the intrinsic partial pressure sensitivity s for the direct flow (DF) and counter flow (CF) detector are given by the helium partial pressure and the helium pumping speed at the connecting flange of the vacuum system:

$$s_{\mathrm{DF}} = \frac{p_{\mathrm{He}}}{Q_{\mathrm{He}}} = \frac{1}{S_{\mathrm{eff}}} \tag{4}$$

and

$$s_{\mathrm{CF}} = \frac{p_{\mathrm{He}}}{Q_{\mathrm{He}}} = \frac{1}{S_{\mathrm{eff}} \cdot K_{\mathrm{He}}} \quad . \tag{5}$$

For the overall sensitivity the intrinsic sensitivity and the sensitivities of the mass spectrometer and the amplifier need to be taken into account [1].

The smallest detectable leak or detection limit is given by the ratio of the minimum detectable signal and the overall sensitivity. This means that in practice the total pressure in the detection cell needs to be taken into account. Thus using the same leak detector, the minimum detectable leak is different when testing a small vessel or a 100 m long accelerator section. The detection limit could also be limited due to a significant helium level in the atmosphere surrounding the leak detector. This could lead to a permanent high helium background signal from, for example, helium entering by permeation through O-rings, small internal leaks in the detector, or via the exhaust line into the roughing pump and thus being stored in the pump oil. To counteract the latter case the roughing pump could be flushed by running with gas ballast for some period.

Another important characteristic is the time-dependent behaviour of the helium signal. This depends not only on the leak detector itself but also on the system under investigation. After spraying helium onto a vessel of volume V the helium partial pressure p_{He} will start to increase to its equilibrium value $Q_{\mathrm{He}}/S_{\mathrm{eff}}$ according to

$$p_{\mathrm{He}} = \frac{Q_{\mathrm{He}}}{S_{\mathrm{eff}}}\left(1 - e^{-\frac{t}{\tau}}\right) \quad . \tag{6}$$

Here τ is the time constant

$$\tau = \frac{V}{S_{\mathrm{eff}}} \quad . \tag{7}$$

Usually the response time is defined as the time to reach 95% of the equilibrium value. Three examples for the time-dependent behaviour of the detector signal are shown in Fig. 7. In the case of curve 1 the signal quickly reaches its equilibrium value. Increasing the volume by a factor of 10 the response time will be 10 times longer (curve 2). Increasing the pumping speed will decrease the response time, but according to Eq. (2) also reduce the signal by the same factor as shown in curve 3. One has to note that the signal will start to rise after a dead time only.

When removing the test gas from the system, the recovery behaviour for the three cases is similar as plotted on the right side of Fig. 7.

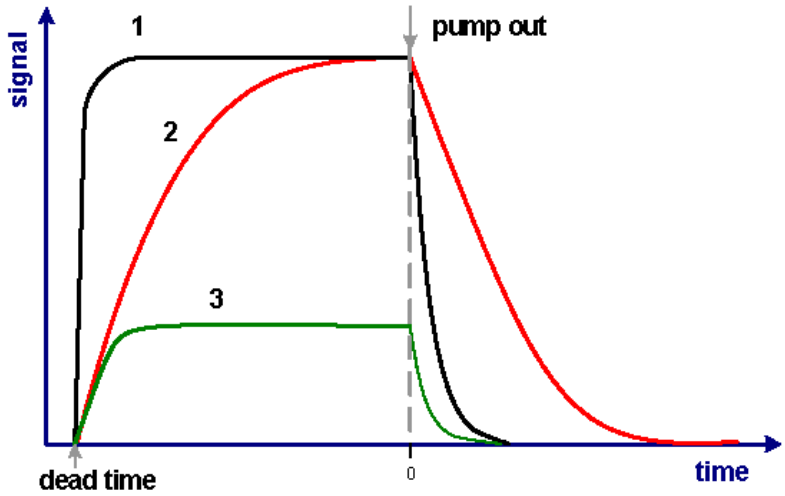


Fig. 7: Time-dependent behaviour of the helium signal for a leak (left) and during pump out (right)

Before leak detection is started, the system must be pumped down such that the maximum inlet pressure of the leak detector has been reached. The warm-up or start-up time defines the period needed till stable operation of the mass spectrometer is possible.

In order to perform quantitative measurements, helium leak detectors need to be calibrated with a standard leak. This needs to be repeated regularly as the long-term stability is usually insufficient, especially when working on large vacuum systems. Drifts from the electronics of the detector need to be checked and corrected regularly where necessary. Details of the procedures and calibration equipment may be found in Refs. [5] and [13].

8 Practical experience and examples

For practical applications the choice of a proper leak detector is important. Depending on the volume to be checked (e.g., small set-ups versus accelerator sections), vacuum level (e.g., O-ring-sealed versus metal-sealed systems), tolerable/expected leak rates (e.g., HV versus UHV systems) or frequency (mass production versus individual measurement), various leak detector systems are available on the market.

A leak search of larger systems requires systematic work. Checking the proper functioning and leak tightness of the leak detector first are mandatory. Working with reduced gas flow from helium bottles, the system should be searched from top to bottom and against an existing air flow when working, for example, in a tunnel. One should also keep in mind that small leaks might initially be blocked by water and only open up after sufficient pumping time or even after bake-out. This may require several leak checks at various vacuum levels.

Recording the pressure and other relevant parameters during pump-down and operation of a vacuum system can save a lot of time in analysing and diagnosing the behaviour of the system. Any deviation from the usual pump-down curve is a hint of a possible leak. If the system is pumped for the first time, more experience is necessary to evaluate the pressure curve. As indicated in Fig. 8, the pressure of a tight system decreases linearly with time in a log-log scale, while in the presence of a leak the pressure tends to level off at a value given by the ratio of the leak flow to the pumping speed [7]. Similar information can be taken from the pressure increase versus time when isolating the vacuum system from the pumps as described in more detail in Section 4.2 and shown in Fig. 1.

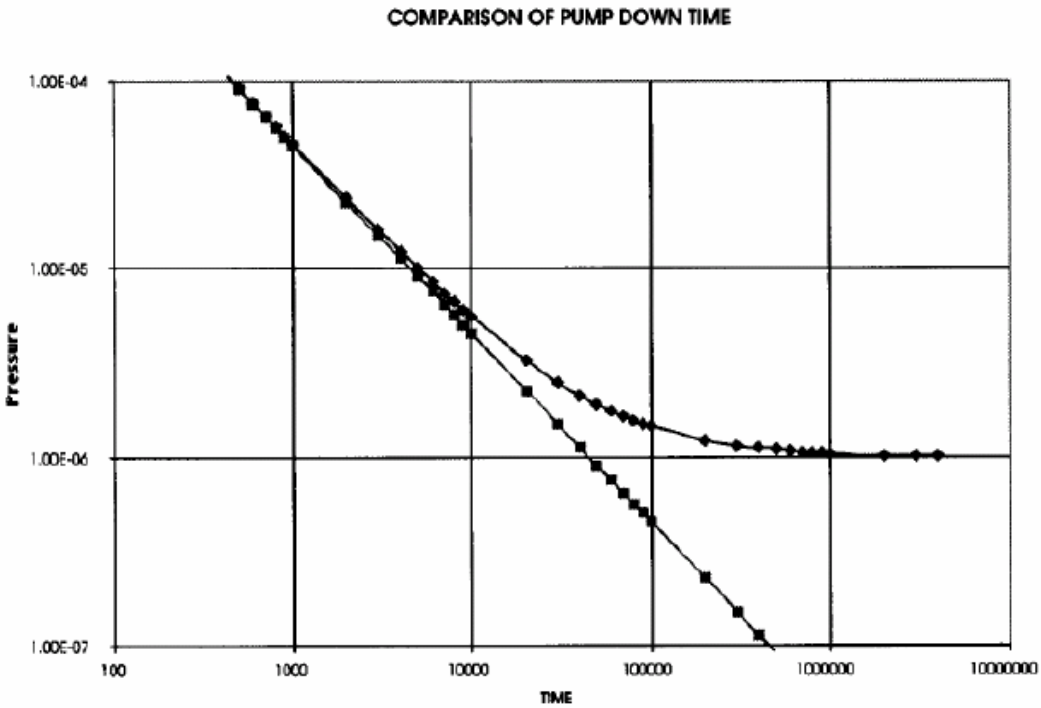


Fig. 8: Pressure versus time during pump-down with and without a leak [7].

As it is much faster and easier to find and repair leaks in a laboratory than in an accelerator, a careful check of all components prior to installation is strongly recommended. This should include the use of metallic gaskets on all sealing surfaces.

8.1 Virtual leaks

A virtual leak looks like a real leak in a pump-down or pressure rise curve. However, there will be no signal in a leak detector and it cannot be located from outside. Usual causes of such leaks are a poor design and/or poor manufacture of the vacuum vessel, e.g., excavations, dead holes for screws, air enclosures in a weld or narrow slits leading to excessive outgassing. Diagnosis could be done using a residual gas analyser and measuring the gas composition before and after venting with argon. After such a venting, the argon replaces the nitrogen in the virtual leak leading to an argon enhancement in the residual gas after the second evacuation.

8.2 Cold leaks

When operating a vacuum system at extreme temperatures (hot or cold), additional leaks could open up. A typical arrangement for testing a vacuum vessel, e.g., a superconducting cavity at temperatures below 4.5 K, is shown in Fig. 9. The vessel is immersed in liquid helium, while in general pumps, vacuum gauges, leak detector, and residual gas analyser are operated in the room temperature part of the system. In the presence of a ‘cold leak’ He atoms will directly enter from the liquid into the vacuum vessel and at least be partially absorbed onto the cold walls. Thus the delay between the atoms entering the system and a significant He signal in the detectors might be very long. In practice such systems are warmed up to about 10 K after a certain time period to release the He atoms from the walls to measure the total leak rate. Leak location requires the system to be warmed up to room temperature; however it is not exceptional that cold leaks close reversibly and/or are not detectable after warm up.

Alternatively, hot extractor gauges can be operated at liquid helium temperatures as indicated in Fig. 9. As these gauges are very sensitive to helium, they can be used for *in situ* leak detection as described in more detail in Ref. [14].

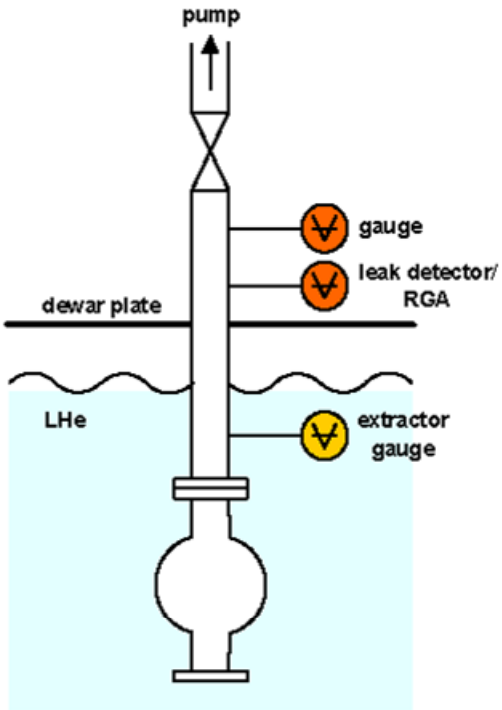


Fig. 9: Set-up for testing a superconducting cavity in liquid helium (LHe) at temperatures below 4.5 K

8.3 Total leakage measurement

As described in Section 7, for total leakage measurement the vacuum vessel is completely enclosed by a tracer gas. By leaving the system without active pumping and thus accumulating the tracer gas, very small leaks can be detected, especially on big volumes. For example, in the HERA proton ring large parts of the beam pipe are surrounded by liquid helium from the superconducting magnets. The integral leak rate of a section of about 1.4 km was determined by looking for helium during a warm-up of the beam pipe after a longer period at liquid helium temperature. Using a residual gas analyser no helium signal could be detected resulting in a total leak rate below 10^{-15} mbar l/s at 4.5 K [15].

Examples for total leakage measurements of pressurized systems are given elsewhere in these proceedings [16].

9 Helium leak detectors - state of the art

There is a wide range of commercial leak detectors on the market. Very compact and even portable units are available. Triggered by the semiconductor industry, an increasing number of oil-free leak detectors are available.

Standard detectors use the counter-flow method. A flexible system of valves allows the injection of the helium flow into the detection cell from various locations of the pumping system, thus giving a huge range of sensitivity from quite large to very small leaks. The lowest quoted values for the minimum detectable leak are $5 \cdot 10^{-12}$ mbar l/s and $5 \cdot 10^{-8}$ mbar l/s for sniffers. The mass selection system is optimized for ^{4}He as the standard tracer gas, mostly including the option to use masses 2 (H_2) and 3 (^{3}He).

The electronics is usually equipped with semiconductors and thus sensitive to radiation. For applications requiring a radiation-hard solution, there is no alternative to old-fashioned tubes. Typical start-up times are of the order of a few minutes, response times less than one second. A high degree of automation is standard, usually including calibration, tuning, and regular correction of the zero line.

10 Concluding remark

Accelerator vacuum systems do have demanding requirements for leak tightness. Adequate leak detection methods, appropriate leak detectors, and well trained personnel are required to prove that the leak tightness specifications are fulfilled and to detect and locate the leaks. The various methods presented during this lecture are well suited to measure and localize the full range of leaks from very large to tiny. A variety of suitable commercial leak detectors are available on the market. There has been great progress in operational availability, performance, and size since 1950 when the first devices came onto the market. During the past 20 years improvements have been mainly in handling and automation. Today's leak detection equipment is robust, easy to use, and offers a large range of sensitivity. Nevertheless, finding and repairing leaks is a time-consuming process, especially in the case of emergency leak testing of a running accelerator system. It still requires well-trained and experienced technicians with good knowledge of the system to be checked. Therefore it is most important to prevent leaks during all stages of design, manufacture, and assembly.

Acknowledgements

I would like to thank my colleagues of the vacuum group MVP at DESY for their patience in introducing me to the field of vacuum techniques and skills of leak detection. Special thanks are due to N. Mildner and A. Wagner for preparing part of the drawings, S. Holm for taking some of the photographs as well as performing measurements for this lecture, and N. Hilleret of CERN for careful reading of the manuscript.

References

[1] J.M. Lafferty, Foundations of Vacuum Science and Technology (Wiley, New York, 1998).

[2] D. Trines *et al*., Proc. XVth Int. Conf. on High Energy Accelerators, Int. J. Mod. Phys. A (Proc. Suppl.) **2A** (1993) 347.

[3] A. Roth, Vacuum Technology (Elsevier, Amsterdam, 1990).

[4] F.F. O'Hanlon, A User's Guide to Vacuum Technology (Wiley, New York, 1989).

[5] M. Wutz, H. Adam and W. Walcher, Handbuch Vakuumtechnik (Vieweg, Braunschweig, 1997).

[6] C.A. Schalla, J. Vac. Sci. Technol. **12** (1975) 430.

[7] N. Hilleret, Leak detection, S. Turner (Ed.), Proc. CAS Vacuum Technology, Snekersten, 1999, CERN 99/05 (CERN, Geneva, 1999), p. 203.

[8] A. Nerken, J. Vac. Sci. Technol. **A9** (1991) 2036.

[9] A.O. Nier, C.M. Stevens, A. Hustrulid and T.A. Abott, J. Appl. Phys. **18** (1947) 30.

[10] W. Becker, Vak. Techn. **8** (1963) 203.

[11] M.H. Hablanaian and W.E. Briggs, Proc. 7th Int. Vac. Congress, Vienna, 1977, p. 199.

[12] W. Becker and W. K. Huber, Proc. 7th Int. Vac. Congress, Vienna, 1977, p. 203.

[13] http://www.iso.ch ISO 3530:1979.

[14] M.G. Rao, Adv. Cryog. Eng. **41** (1996) 1783.

[15] D. Trines *et al*., Proc. XVth Int. Conf. on High Energy Accelerators, Int. J. Mod. Phys. A (Proc. Suppl.) **2A** (1993) 344.

[16] K. Zapfe, Large systems commissioning, these proceedings.

Basics and applications of cryopumps

C. Day
Forschungszentrum Karlsruhe, Institute of Technical Physics, 76344 Eggenstein-Leopoldshafen, Germany

Abstract
This report introduces the wide field of cryogenic pumping, from commercial cryogen-free refrigerator cryopumps to very special, tailor-made cryopumps. It starts with a short historic overview of the field of cryopumping. Then, the principles of cryogenic pumping via sublimation/condensation and physisorption are introduced, and we illustrate how they are exploited to derive a design for a usable cryopump. In the third part, typical characteristics of cryosorption pumps are discussed. The report finishes with a few examples of applications.

1 Introduction

This report is aimed at giving a general overview of cryopumping. Usually, the cooled surface is at least partly covered with a porous sorbent material. The design of a cryopump has to combine cryogenic aspects and technological vacuum considerations in a unique manner, as the generation of low temperature itself presupposes the existence of vacuum conditions and vice versa. Thus, the cryopump is by its physical principle a high-vacuum pump. Besides straightforward parameters, such as pressure and temperature, the performance cryopump is very much governed by the complex interaction between gas particles and cooled sorbent surface, which will be discussed in detail in this report. The cryopump is the pump type that provides the highest pumping speeds, especially when operated in situ of the vacuum recipient.

The present report contains an extensive list of references, a mixture of textbooks and specialized articles, to provide the reader with the opportunity to pursue any further details he might be interested in.

2 History of cryopumping

Not surprisingly, the development of cryopumping went in parallel with the advances made in liquefaction (liquid nitrogen, 77.3 K, Dewar, 1874; liquid hydrogen 20.4 K, Dewar, 1898; liquid helium, 4.2 K, Kammerlingh Onnes, 1908). It was soon found that vacuum can be produced by charcoal kept at these cryogenic temperatures. The successful use of liquid-nitrogen-cooled traps to prevent oil backstreaming from diffusion pumps was reported in the 1910s by Gaede and Langmuir.

However, the first large-scale applications of cryopumping were triggered by the first space projects in the 1950s. They often included operation with liquid hydrogen. In the second half of the last century, liquid helium bath-cooled cryopumps were very popular for laboratory applications, when oil contamination was requested (so that diffusion pumps could not be used) and manual re-fill was not a problem.

The next big step forward was the development of cryogen-free regenerative refrigerators by Gifford and McMahon in 1960. These devices have seen a significant progress towards higher cooling power at lower end temperatures, and their development continues.

Nowadays, Gifford–McMahon cooled cryopumps are the high-vacuum pump of choice in many industrial processes, such as the semiconductor industries, especially in the USA. The company CTI-Cryogenics (now with Brooks Automation Inc.) is the leading global cryopump manufacturer. In Europe, turbomolecular pumps have widely replaced cryopumps in vacuum process plants.

3 Cryopumping fundamentals

By international classification, a cryopump is defined as a vacuum pump which captures the gas by surfaces cooled to temperatures below 120 K [1]. To achieve vacuum in a closed volume means, simply speaking, to remove all molecules in the gaseous phase within this volume. According to the different physical principles, which are exploited to create, improve or maintain vacuum, there are different types:

- Positive displacement or mechanical pumps, which provide volumes to be filled with the gas being pumped. The volumes are cyclically isolated from the inlet, the gas is then transferred to the outlet. In most types of positive displacement vacuum pumps the gas is compressed to atmosphere before the discharge at the outlet. Positive displacement pumps work independently of the gas species to be pumped.
- Kinetic pumps, which impart momentum to the gas being pumped in such a way that the gas is transferred continuously from the pump inlet to the outlet.
- Entrapment or capture vacuum pumps, which retain gas molecules by chemical or physical interaction on their internal surfaces.

Together with the getter and the sputter ion pump, the cryopump is the most prominent representative of the latter group. As for all high vacuum pumps (i.e., pumps working below the 10^{-3} Pa range) an appropriate medium forepressure has to be provided by a mechanical forepump, before the cryopump can be started. The forepump is also needed to exhaust the gas to the atmosphere. Whereas this gas transfer is done continuously in the case of kinetic and positive displacement pumps, entrapment pumps do this in a batch-wise manner as they accumulate the gas during pumping and must be regenerated from time to time. During actual pumping, the gases within the entrapment pump systems are instantaneously immobilized and no outlet is required at all.

3.1 Cryopumping mechanisms

The cryopumping effect is produced by intimate interaction between the gas particles to be pumped and a cold surface provided to them in the cryopump. The forces involved are relatively weak (primarily van der Waals dispersion type) and do not include chemical bonds, as is the case for chemisorption with the chemically active alloys in getter pumps. Consequently, cryopumps do not require such high temperatures for regeneration. Cryopumps can pump all gases including noble gas, if the temperature is sufficiently low.

The amount of molecules that can be accumulated depends on a number of physical factors such as temperature of gas and surface, physicochemical properties of gas and surface (surface energy distribution), microscopic roughness of the surface, etc. The different categories are treated in the following sections.

3.1.1 Cryocondensation

In this case, the surfaces must be cooled to such a temperature as to keep the corresponding saturation pressure equal to or below the desired vacuum pressure in the chamber. The achievable pressure is determined by the saturation pressure at the temperature chosen for the cold surfaces. This principle is the most elementary of all forms of capture pumping. For many gases, the pressure range of

cryopumps is below the triple point pressure of the individual gases to be pumped. Thus, the relevant saturation curve becomes identical with the sublimation curve, i.e., during pumping, the gas particles undergo direct phase transition from the gaseous to the solid phase without any liquid phase. Within this report, as is customary in cryopumping, the term 'condensation' is used for both types of transition out of the gaseous phase, and combines re-sublimation (gaseous → solid) and condensation (gaseous → liquid). It must be noted that standard data libraries do not include the sublimation regime, but often have the triple point temperature as lower limit [2], as typical problems are encountered in the low temperature region. For example, the 'correct' vapour pressure values are only relevant for surface coverages greater than one monolayer [3, 4]. Thus, for use in the low temperature regime needed for cryopump design, carefully selected sources have to be quoted (^{3}He [5, 6], ^{4}He [7], hydrogen isotopes [8]). Figure 1 shows the sublimation equilibrium curves of some relevant gases. In principle any gas can be pumped, provided the surface temperature is low enough. It can be seen from this figure that neon, hydrogen isotopes, and helium require the lowest temperatures.

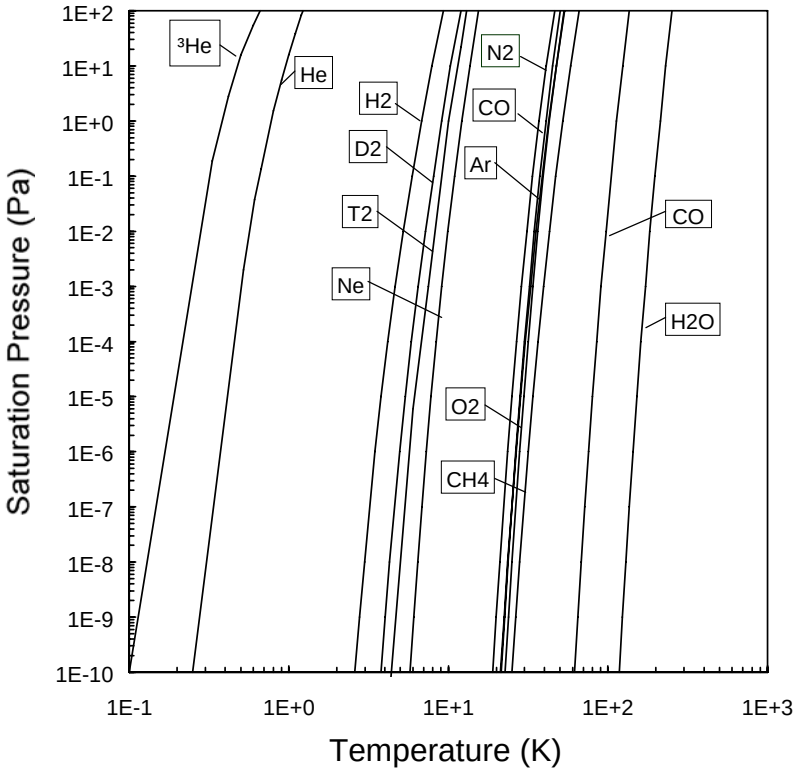


Fig. 1: Saturation curves of common gases

The temperature of about 100 K would be sufficient to condense water and all hydrocarbons, e.g., oil vapours. To condense the air components, the 20 K level is needed. The 4 K level is sufficient to condense the hydrogen isotopes and neon. By further reduction of the temperature, the ultra-high vacuum range (10^{-5} to 10^{-8} Pa) is fully accessible.

When using Fig. 1 as a basis for cryopump design, one must be aware that the saturation curve is associated with thermodynamic equilibrium at the phase boundary. This means a net particle flux of zero, or: zero pumping speed. For practical applications, it is recommended to provide for oversaturation of the gas by two orders of magnitude in pressure. Figure 2 exemplifies the consequences of that procedure. In order to establish a pressure of 10^{-4} Pa during pumping of H_2 (corresponding to an equilibrium temperature of 4.2 K), one needs to provide a cold surface temperature of 3.6 K (the equilibrium temperature of 10^{-6} Pa).

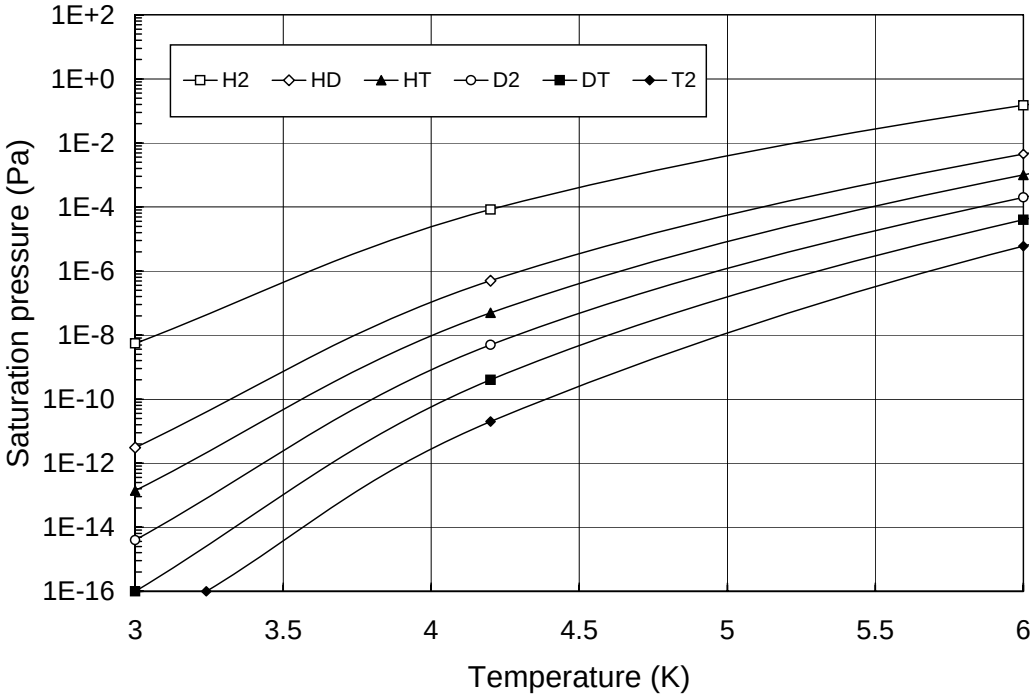


Fig. 2: Sublimation pressures of the six hydrogen isotopomers [8]

During cryocondensation, condensate layers are formed by crystal growth out of the gas phase. This growth process takes place over three stages: Firstly, thermal accommodation of the impinging particle and adsorption on the surface, then, diffusion to the growth site on the crystal lattice, and, thirdly, atomic incorporation. It has been shown that the second step governs the overall condensation velocity [9]. The efficiency of the condensation process is expressed by means of the condensation coefficient α_c, defined as the number of condensed particles related to the number of particles incident upon the cryosurface [10, 11]. Assuming high supersaturation, a condensation coefficient of unity is achievable without any problem. In most cases, the cryocondensation growth starts by forming one or a few monolayers and then changes to form islands on top of them. Thus, the cryodeposits obtained during pumping may be quite non-uniform. The growth of condensed layers is not limited in principle. But for regions with increased layer thickness, the surface temperature of the condensate rises, thus leading to higher surface mobility of the pumped particles. This results in a higher risk of spontaneous transpositions and should be avoided in any case.

3.1.2 Cryosorption

Gas particles impinging on a surface of sufficiently low temperature lose so much of their incident kinetic energy that they stay attached to the cold surface by weak intermolecular forces, resulting in significantly higher molecular concentration on the surface than in the gas phase. This phenomenon is called physical adsorption or physisorption. The reader interested in going into details is advised to check the pertinent textbooks on adsorption energies and theory on intermolecular forces [12–15]. Cryosorption denotes the physical adsorption process under vacuum conditions and low temperatures. The equilibrium pressure of adsorbed gas particles is significantly lower than the corresponding saturation pressure for cryocondensation. This is due to the fact that the dispersion forces between the gas molecule and the surface are greater than between the gas molecules themselves in the condensed state. Hence, gas can be retained by adsorption even in a subsaturated state, i.e., at considerably higher temperatures than would be required for condensation. This fact is essential in cryopumping helium, hydrogen, and neon, which are difficult to condense.

However, the cryosorption process is quite complex and very much determined by the actual nature of the surface (chemical, mechanical, physical), not only by its temperature, as is the case for cryocondensation. Porous materials with high sorption capacity, such as molecular sieves or activated carbons are most often used as sorbent materials. However, layers of condensed gas frost (Ar, CO_2, SF_6) may also be applied. The characteristic of cryosorption is given by the respective sorption isotherm. It is possible to bind helium or hydrogens in the 5 K temperature range and to achieve equilibrium pressures in the 10^{-7} Pa region without any problem. Figure 3 shows the typical sorption isotherm curves for helium, hydrogen and deuterium, which cannot be pumped by condensation at the standard available 4.2 K. As customary in adsorption technology, the sorbed amount is given in standard cm³ (referenced at 1 atm and 273.15 K), related to the sorbent mass. Contrary to condensation, where a simple saturation *p–T* plot is sufficient to estimate the performance (Figs. 1 and 2), cryosorption requires the knowledge of a 3-D relationship between *p*, *T*, and the sorbed amount.

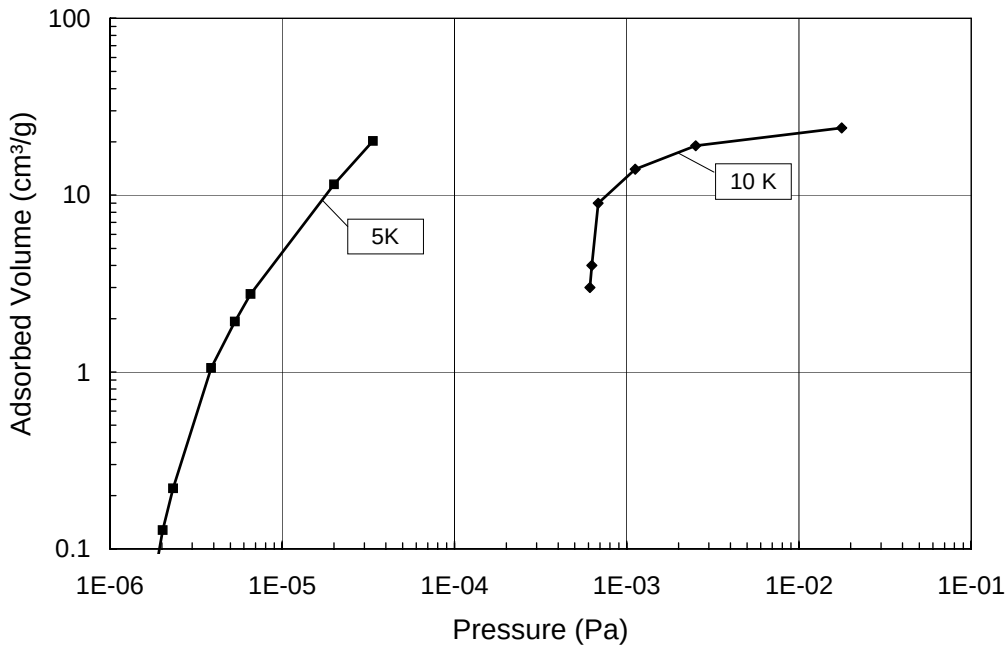


Fig. 3: Sorption equilibria for helium at a microporous activated charcoal

The sticking coefficient (the sorption analogy to the condensation coefficient), i.e., the number of molecules which are sorbed on a surface in a given interval of time divided by the number of molecules impinging on that surface, may be considerably below unity, depending on the material–gas pair. Due to the physical principle, cryosorption pumping is limited to some monolayers of gas coverage on the surface. Then the effect of the surface becomes negligible and the resulting pressure starts to increase rapidly.

3.1.3 Cryotrapping

This is the concurrent pumping of two or more gases by entrainment of gas particles which are not condensable at the prevailing temperatures and pressure conditions. For the purpose of entrainment, a condensing gas is used, so that a mixed condensate is formed. Usually, the small molecules of the gas to be pumped are caught in the open lattice of the cryodeposit of a more abundant species and are quickly buried by subsequent layers. As most of the bonded gas particles are occluded in the condensate layer and only a certain percentage directly interacts with the surface, the achievable equilibrium pressures are even lower than those achievable by cryosorption. The entrapment ratio (i.e., the number of pumped molecules of one species related to the number of deposited molecules of the

condenate species) is higher than with pumping by layers of pre-condensed frost (see, for example, pumping of He by condensed Ar). However, the main disadvantage of this technique is the additional gas input into the system which has to be evacuated. This helps during pumping, but has to be coped with during regeneration and, thus, requires larger forevacuum systems.

3.2 Heat load assessment

Cryopumping means the generation of vacuum by low temperatures. Thus, the generation and availability of low temperatures is a prerequisite for the operation of cryopumps. Cryogenics is dealt with in detail in a separate report of this summer school. Therefore, the following section is limited to the specific cryogenic aspects of cryopumping.

The cryopump housing will usually be at ambient temperature, whereas the cryopanel must be refrigerated. So, there will always be heat flux onto the panel due to the different temperatures. It is essential for establishing stably low temperatures to reduce heat influx on the cold surface. The total thermal load $\dot{Q}$ transmitted to the cryosurface consists of the sum of heat flow rates produced by thermal conductivity of the solids, gas heat conduction, and thermal radiation. For working cryopumps the thermal loads due to cryosorption and/or cryocondensation still have to be added. This results from the enthalpy change between particle temperature and phase transition temperature $\Delta\dot{H}_{\mathrm{cool}}$, and the phase change enthalpy $\Delta\dot{H}_{\mathrm{Ph}}$ itself:

$$\dot{Q} = \dot{Q}_{\mathrm{S}} + \dot{Q}_{\mathrm{G}} + \dot{Q}_{\mathrm{R}} + \Delta\dot{H}_{\mathrm{cool}} + \Delta\dot{H}_{\mathrm{Ph}} \rightarrow \min . \tag{1}$$

Under steady-state conditions, the total heat load becomes equal to the refrigeration power by the cold source. At increasing inlet pressures, the load also increases and the immobilization gas load becomes predominant. The different contributions and the strategies of how to minimize them are an important issue in cryopump design.

3.2.1 Solid heat conduction

The heat transmitted by solid conduction (e.g., via piping and instrumentation, fastenings, supply lines, cables) is proportional to the cross-section A and the thermal conductivity λ at the applied temperature difference (between warm wall and cold surface), and inversely proportional to the conductor's length L:

$$\dot{Q}_{\mathrm{S}} \sim \frac{\lambda}{L} \cdot A . \tag{2}$$

This energy input must be kept at an adequately low level by the constructional geometry and by the correct choice of materials. To avoid parasitic heat flows, stainless steel is a good material owing to its low conductivity. However, one has also to compromise the need for materials with high thermal conductivity at cryogenic temperatures (such as copper) to provide good thermal contact with the cold surface.

3.2.2 Gaseous heat conduction

The influence of residual heat conduction in the gas is due to the energy transfer within collisions between the molecules and any surface. This effect is associated with the mean free path, which denotes the average distance which a molecule travels between two successive collisions. It can be practically eliminated, when the mean free path Λ between two particle hits is considerably larger than the characteristic vessel dimension d, which begins to hold at lower pressures. In this case, the collisions between molecules become less frequent and the molecules collide predominantly with the

vessel walls. Then, the heat transfer depends only on the number of molecules and a linear relationship develops between gas heat conductivity and pressure. At higher pressures, the gaseous heat transfer is much higher and owing to convective bulk motion of the gas and the thermal conductivity of the gas does not change significantly with pressure.

To characterize this effect and for division into different regimes of flow, the Knudsen number is used, which is defined as the ratio of the mean free path to a characteristic dimension of the system (like, for example, the vessel diameter):

$$Kn = \frac{\Lambda}{d} . \tag{3}$$

According to Fig. 4, there are four regimes of gaseous heat transfer.

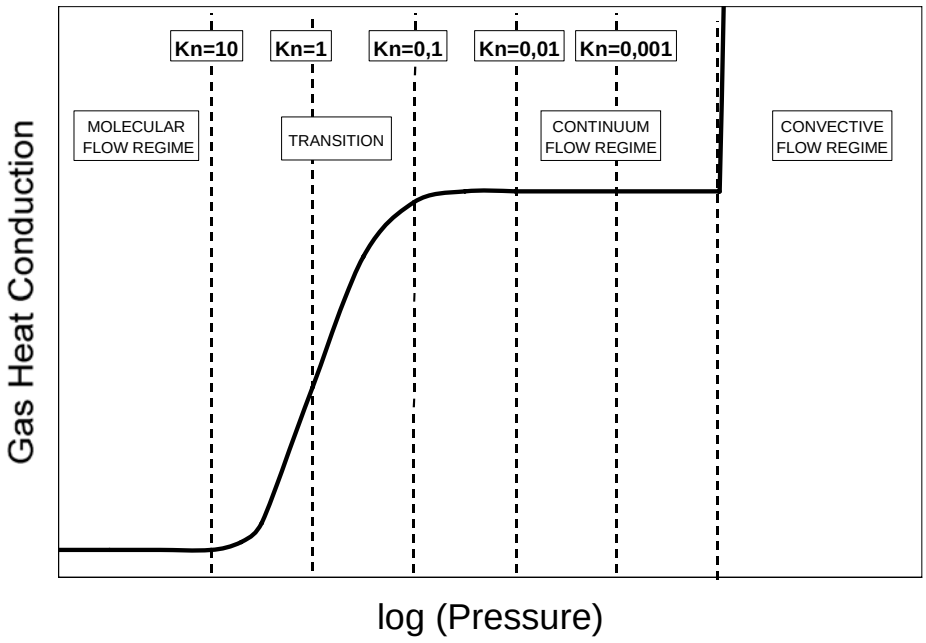


Fig. 4: Schematic gas heat conduction (arbitrary units) as a function of pressure at different *Kn* numbers

The regime with *Kn* >> 1 is called the molecular flow regime, where the influence of gas heat conduction can be neglected [16]. A cryopump is normally operated under molecular flow conditions. However, depending on the application, and especially for tailor-made cryopumps for which cryogenic supply is not the limiting factor, one may benefit from the fact that the pumping speed of the cryopump increases with decreasing *Kn* number [17]. Table 1 gives a hint as to what pressures are needed to establish molecular flow conditions for different gases at different temperatures.

Table 1: Gas pressures (in Pa) corresponding to *Kn* = 10 for a vessel with 1 m diameter

Temperature (K)	He	H_2	N_2	CO_2
300	1.9×10^{-3}	1.3×10^{-3}	6.7×10^{-4}	4.4×10^{-4}
80	4.4×10^{-4}	2.6×10^{-4}	1.1×10^{-4}	
4	1.2×10^{-5}	CONDENSATION		

3.2.3 Radiative heat transfer

Under molecular flow conditions, radiant heat from the process side in the vacuum chamber and the pump body is the primary heat load on the panels. For a pump to be heated by radiation, there are two requirements. Firstly, the heat has to be emitted from the chamber, and, secondly, the pump must absorb the incident radiation. Both possibilities have to be minimized by an appropriate design.

The radiation heat exchange between two diffuse, grey surfaces A_1 and A_2 with emissivities ε_1 and ε_2, respectively (for example 1 is the chamber wall, 2 the cryosurface), is given by [18]:

$$\dot{Q}_{12} = C_{12} \cdot A_1 \cdot (T_1^4 - T_2^4)\,, \tag{4}$$

with the exchange number C_{12} given by:

$$C_{12} = \frac{\sigma \cdot \varepsilon_1 \cdot \varepsilon_2}{1-(1-\varepsilon_1)\cdot(1-\varepsilon_2)\cdot\varphi_{21}}\,, \tag{5}$$

where $\sigma = 5.678 \times 10^{-8}$ W/(m^2 · K^4) is the Stefan-Boltzmann constant and φ denotes the view factor. The view factor (or shape factor) φ_{12} is defined as the fraction of the radiation leaving 1 that is intercepted by surface 2 and determined by the geometric situation only. The analytical treatment to assess these values is quite complicated, so compilations of these data are usually employed [18, 19], whenever possible. For complicated geometries, the (Test Particle) Monte Carlo method or the angular reflection coefficient method are very appropriate tools to calculate the thermal radiation heat transfer due to photons scattered on the existing obstructions in the pump [20, 21]. These types of calculations directly yield the (thermal) transmission coefficient *t*, which denotes the percentage of the total power of incident thermal radiation finally arriving at the cryosurface.

The minimization of radiation heat influx is usually done by reducing the emissivity ε_1 of the internal pump surface. By very clean electropolishing, emissivities of 0.01 are obtained. But this effect is often masked by water humidity adsorbed or condensed on the walls, which leads to radiation almost equal to black body radiation [22]. So, the thermal load on the cryosurface coming from the chamber walls at 300 K is still much too high. Therefore, the cryosurface is even more protected from the radiation heat load. To do that, an optically dense and blackened (emissivities ε greater 0.9), cooled radiation shield is installed around the cryopanel to adsorb the entering stray radiation from the vacuum chamber source. The thermal transmission coefficient *t* associated with such array structures is of the order of 0.01 [23]. By installing baffle structures, the resulting overall radiation onto the panel has two contributions. Firstly, the residual radiation from the walls at ambient temperature [reduced by factor ($\varepsilon_2 \cdot t$)], and, secondly, the direct radiation from the installed shields and baffle structures (at about 100 K). The disadvantage of any baffle from the vacuum pumping point of view, is a significant conductance reduction on the route of the particle from the cryopump inlet to the cryosurface (see Section 3.5). If the radiation heat load must be further decreased, the cryopump must be removed from the direct line of sight by designing some bends and elbows. But the main disadvantage of this method is that it again reduces conductance and the available net pumping speed.

The radiation shield is operated as a second cryopanel, but at a higher temperature level. In the entrance direction, the radiation shield is open, but constructed in such a way that no particle can traverse without colliding with the wall. Consequently, all gases impinge first on the cold baffle surface and become thermalized before reaching the actual cryosurface. This will already remove some high boiling gases by being pumped via cryocondensation on the baffle before they reach the sorbent. Thus, full capacity of the sorbent is available for removing the remaining gases. Good designs of the entrance baffle have to compromise efficient radiation shielding of the cold cryopanel with as high a conductance as possible for the incoming gas flow. Usually, louvre or chevron baffles are used, providing maximum (molecular) transmission probabilities *w* of about 40% and 25%, respectively [21, 24, 25].

3.2.4 Enthalpy transfer

During pumping, the cryopanel must absorb the energy of the incoming particles, i.e., the enthalpy difference for cool-down between baffle temperature and panel temperature and the energy connected with the phase change in the subsequent immobilization step. In this respect, the baffle is an essential help.

3.3 Cryopump types

In this section we address the different designs for cryopumps, which have been developed to meet the cryogenic and vacuum aspects discussed above. Each cryopump is made up of the three basic parts, namely, the entrance baffle, cryopanel and refrigeration unit. In most cases, the cryopanel is, at least partly, designed as a cryosorption panel. In the following, the cryopump designs are classified according to the cooling principle which is involved to cool the cryopanels.

3.3.1 Bath cryopumps

These cryopumps take the pattern of well-known dewar and cryostat designs. The pumping surfaces facing the vacuum vessel are directly cooled from their back side with liquid cryogens, i.e., boiling at ambient pressure, which are stored in dewar vessels. Thus, temperatures of 77.3 K (liquid nitrogen, LN), 27.1 K (liquid neon), 20.3 K (liquid hydrogen H_2), 4.2 K (liquid helium LHe) can be attained. The pumping surface in any case has to be shielded by a LN-cooled stage against thermal radiation.

To achieve long holding times with continuous operation, an automatic refill device must be used. To achieve further fractionation of the gases to be pumped, the 4 K LHe cryosorption panel may be shielded by an additional baffle, which is also cooled with LHe. In this three-stage version, a 80 K (LN temperature) condensation stage is obtained for high molecular gases, a 4 K condensation stage for all remaining gases passing the 80 K stage except for helium and hydrogen, and the final 4 K cryosorption stage. However, the overall conductance of such a configuration is reduced. The typical set-up of such a cryopump is shown in Fig. 5.

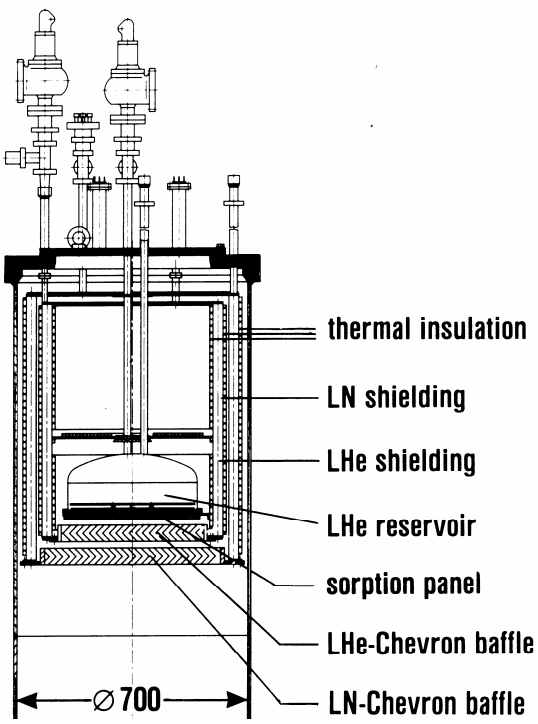


Fig. 5: Typical set-up of a bath cryopump in three-stage configuration. The LHe baffle is removable to obtain a two-stage configuration.

On account of clumsiness in handling liquid cryogens and the high capital and operational costs involved, bath-cooled cryopumps, with a few notable exceptions, are nowadays almost exclusively used in research laboratories and to a great extent replaced by refrigerator-cooled cryopumps which are described in the following.

3.3.2 Refrigerator-cooled cryopumps

Most modern cryopumps today are cooled by closed-loop mechanical refrigerators, using helium as a working fluid. Nowadays, refrigerator-cooled cryopumps are almost exclusively based on the two-stage Gifford and McMahon (GM) process [26, 27]. The gaseous helium is circulated continuously between a compressor and an expander, which are connected via flexible hoses. Two stages of these machines are required to achieve temperatures low enough to pump all gases except neon, helium, and hydrogen by cryocondensation. The first stage usually operates in the range of 50 to 75 K. It is used to cool the structures that provide for the radiation shielding and cools the baffles across the inlet of the pump. The second stage, which usually operates at about 10 K, is used to cool the inner cryopanel that is used to immobilize the gases which pass the baffle structures. Often, the second stage is covered with a sorbent material (at the back side) to allow for cryosorption pumping.

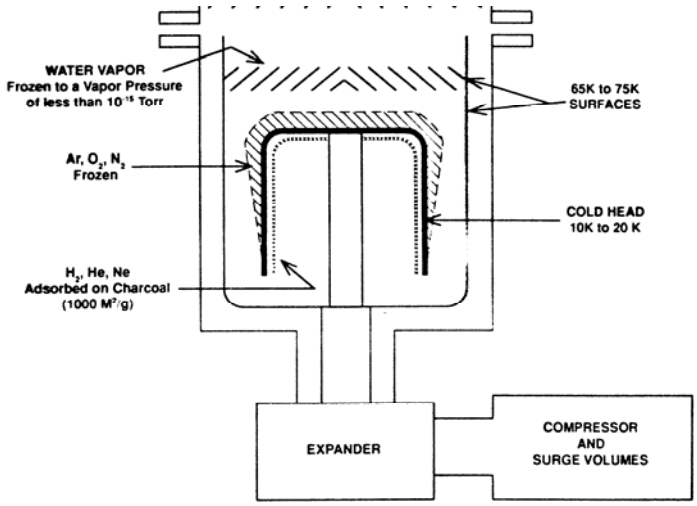


Fig. 6: Typical set-up of a two-stage refrigerator-cooled cryopump

Fig. 7: Cryopumps of the 3000 l/s class: Oerlikon Leybold Coolvac 3000 (left) and CTI On-Board 10 right)

As the refrigeration power for the regenerative cryoprocesses is limited, they are used as small-size cryopumps. From a pragmatic point of view the practical cross-section is often given by the availability of an appropriate upstream valve, which is difficult/expensive for sizes above DN250CF. Nevertheless, GM cryopumps at moderate size have been installed in many industrial processes in the semiconductor and coating industries [28, 29]. A typical set-up is shown in Fig. 6. Figure 7 gives two photos of commercially available pumps.

In former years, the pumping speed for He was very low, as the lowest temperature at the second stage was about 8–10 K, at which temperature the sticking coefficient of He on charcoal is too small. However, this disadvantage has been overcome within the last decade; excellent progress has been made in extending the available temperature range at the second stage to lower temperatures by integration of new rare-earth regenerator materials into the cold heads, providing up to 2 W at 4 K. The system-inherent vibration of refrigerator-cooled cryopumps, which has been detected to be a problem for some severe applications in nanotechnology, can be handled by sophisticated set-ups [30].

The refrigerator-cooled pumps are easy to apply and allow for an easy and complete automation of the pumping process, as storage or transfer systems for liquid cryogens are no longer needed. Standard off-the-shelf cryopumps have been available since the mid 1980s, with nominal aperture diameters of up to 1000 mm (e.g., for equipment of space simulation chambers). Technical data of a typical representative are listed in Table 2 (see also Section 3.4.2 for explanation of the categories). It becomes obvious that, even though pumping speed data for helium are comparable to those of other gases, the capacity is smaller by two orders of magnitude and, hence, becomes one of the pump's performance limits.

Table 2: Listed performance data (average taken from the catalogue of different manufacturers) for a typical 3000 l/s class two-stage GM cryopump

Parameter	**Value**
Pumping speed (l/s) water	9000–10500
Air	3000–3250
Hydrogen	4500–5200
Argon	2500–2700
Helium	1500–2300
Maximum throughput (Pa·m³/s) argon	1.0–2.5
Hydrogen	1.2
Pumping capacity (Pa·m³) argon	1.5×10^5–3×10^5
Hydrogen	1500–5000
Helium	10–100
Ultimate pressure (N_2 equivalent) (Pa)	$10^{-9} – 10^{-10}$
Cool-down time (h)	1.5–2.5
Crossover (Pa·m³)	35–50
Weight (kg)	30–50

3.3.3 Supercritical helium-cooled cryopumps

In this class of cryopumps the cryosurface is cooled with a continuous flow of cold helium, pressurized to supercritical pressures (such as 0.4 MPa). By appropriately selected heat exchangers to warm up the 4 K helium gas, all temperatures are adjustable. These cryopumps are recommended in cases where large helium reservoirs are available anyhow, as, for example, in a number of large-scale scientific research areas, like space simulation chambers, particle accelerators and storage rings, or thermonuclear fusion experiments.

3.3.4 Alternative cryopump concepts

The development of novel cryogenic pump concepts is a vital field of continuous efforts. Some promising concepts are mentioned below. Shigi tried to improve the drawbacks of the turbomolecular pump performance with respect to limited pumping speed for water and back-diffusion of hydrocarbon vapours by adding a cryogenically cooled baffle to the turbopump [31]. Perin *et al*. at CEA, France, continued the development of a turbomolecular pump under cold conditions, involving cryocoolers but without any sorbent stage [32]. Hemmerich used the diffusion pump principle [33]. His cryopump is successfully in use in the JET Tokamak, Culham, UK. For cyclotron vacuum chambers, a special heat transfer concept between the refrigerators and the cryopanel based on cryogenic heat pipes was developed and installed [34]. Foster developed an Ar frost based cryopump with a special device to regenerate the cryosurfaces; this pump can be operated continuously as the cryodeposits are removed mechanically by a snail head riding over the cryopanel surface [35].

3.4 Cryopump performance parameters

3.4.1 Choice of cryosorbent

Depending on the actual pumping task, a pure condensation pump without any sorbent may be fully sufficient. However, at the temperature level of ~ 4 K, which is the minimum level for standard cryopumps (cooled by liquid helium, or refrigerator-cooled), cryosorption is needed to pump helium; helium pumping is a function that is favourable to have, because helium is the standard gas used for leak detection. Moreover, the great improvement of using cryosorbent materials against cryocondensation is due to the achievement of much smaller equilibrium pressures. As the effect of cryosorption pumping is surface-generated, it tends to vanish with increasing surface coverage. Therefore, candidate materials must provide very high internal surfaces to have enough sorption capacity. The physical characterization of a porous sorbent material involves parameters such as

- geometry (granules, powder, spheres, pellets, fibres) and particle size distribution,
- surface structure,
- density,
- total specific surface area,
- pore size distribution and from that pore specific volume and mean pore size,
- pore shape.

The three latter aspects can be derived from measured sorption isotherms. Under high vacuum and small load conditions it is often justified to use very simple expressions for the isotherms.

The simplest adsorption isotherm is Henry's law, in which the amount of adsorbed gas Q varies linearly with the pressure (cf. Fig. 3):

$$Q = c \cdot p \,, \tag{6}$$

with c being the proportionality constant. This equation holds in the boundary case of perfect gases with no interactions between the adsorbed gas molecules (however, all correct model equations should possess the Henry dependency at the low end of the pressure range).

Another equation which is often used for cryosorption is the empirical Freundlich isotherm

$$Q = c \cdot p^{1/n} . \tag{7}$$

It is known that the cryosorption properties of a sorbent do not linearly correlate just with the total surface area (usually between 1000 m²/g and 2000 m²/g): the pore size distribution is the more fundamental property [36]. The pore width denotes the characteristic diameter of a pore, e.g., the diameter of a cylindrical pore. It is customary to differentiate between [37, 38]

- ultramicropores: pore width < 0.7 nm,
- supermicropores: pore width between 0.7 and 2 nm,
- mesopores: pore width between 2 and 50 nm,
- macropores: pore width > 50 nm.

The most popular technique for characterization of porous solids by means of the sorption isotherm is based on the 77 K nitrogen isotherm. The two most interesting properties to derive from experimental sorption isotherms are the pore-size distribution and the surface area A of the solid, for which there exist standard procedures, based on the isotherm model of Brunauer, Emmett and Teller (BET) [39]. However, the BET model is not applicable to microporous solids, which is the material of choice for cryopumping. Consequently, the total surface area can only be regarded as a fingerprint for intercomparison. One should be very careful to take these values as real. Moreover, this procedure is based on the interrelation of nitrogen molecules with the pores at 77 K. It is easy to understand that the situation may be very different for hydrogen at 20 K, or helium at 4 K. Therefore, in most cases, the choice of a cryosorbent material must be supported by characterization experiments.

In the literature, several programmes have been carried out to investigate and benchmark the suitability of different materials for use in cryosorption pumping. They all originated from world-wide R&D programmes launched in the early 1990s for nuclear fusion development. Perinić and co-workers at Forschungszentrum Karlsruhe, Germany, performed extensive comparative sorption screening tests involving materials such as sintered metal, porous ceramics, metal fibre fleece, alumina, silica gel, zeolite (molecular sieve), activated carbon and condensed Ar frost [40–42]. A parallel benchmark study was carried out at Los Alamos, USA, by Sedgley *et al.* [43–47]. Similar work, including thin metal films was done at the Efremov-Institute in Russia by Gurevich and co-workers [48, 49]. The use of metal structures as cryosorbent has successfully been demonstrated [50–52]. Recently, carbon fibres have been characterized for the LHC at CERN [53, 54]. However, the three sorbent types zeolite, activated carbon and condensed gas are the most important representatives and they will be treated in detail in this section. It was unanimously revealed that activated charcoal, especially of the coconut type, exhibits the best dynamic performance characteristics.

A special aspect of the benchmark tests was the determination of the optimum bonding technique to fix the sorbent on a flat panel structure. The influence of the material used for attachment was investigated comprehensively (mechanical fastenings, organic cements (epoxy resins), inorganic cements (silicate, metallic), braze bonding, soldering, plasma spraying) [42, 43, 49], but the results found differed very strongly. Thus, it is believed that this influence is very much due to the manufacturing know-how available and it will therefore not be discussed broadly in this report. One must ensure that the material used for bondage is compatible with UHV conditions and does not

produce outgassing problems. In almost all commercial cryopumps with sorption stage, activated charcoal is used as sorbent and it is bonded to the cryopanel by means of epoxy resins. At Forschungszentrum Karlsruhe, a special technique based on an inorganic cement was developed within the fusion programme, characterized in much detail as reference solution for ITER, and successfully employed in several tailor-made large-scale cryopumps [55].

3.4.1.1 Zeolite molecular sieves

Zeolites are hydrated aluminosilicates with well-defined structures. In most cases they are produced synthetically and their properties may be fine-tuned by the ratio of cations and anions; natural zeolites are formed from the alteration of volcanic ash. A wealth of different zeolites are used commercially because of their unique adsorption, catalytic, and ion-exchange properties. Because of their regular and reproducible structure they behave in a very predictable fashion, based on the defined monodisperse pore size distribution which relates strongly with the kinetic diameter of the molecule to be adsorbed. Therefore they can work as molecular sieves, allowing smaller molecules in while excluding larger ones.

Numerous papers have been published on the use of 3 Å, 4 Å, and 5 Å zeolite as filling material of sorption columns. Nevertheless, sorption isotherm data on molecular sieve material in the sub 77 K range are rarely found [56–60]. Some results on cryosorption pumping by means of zeolites are reported in Refs. [61–65]. Good cryopumping results for helium have been reported for a LHe-cooled cryopump with a molecular sieve coated cryopanel [66, 67]. However, for the application in cryopumping, the main disadvantage of molecular sieve materials is the hydrophilic character of polar sieves, necessitating high regeneration temperatures [68–70]. In the literature, a significant diffusion-rate limitation has been described for pumping of hydrogen isotopes and helium at molecular sieves, resulting in unstable pumping speed at high flow rates. Correspondingly poor results for co-pumping of mixtures have been reported [71–74]. This is why the current R&D on cryosorption materials is focused mainly on carbon materials.

3.4.1.2 Cryocondensates

The dynamic and static characteristics of sorption by condensed gases depend on several factors, like the physical properties of the sorbent (interaction potentials), the structure of the frost layer (grain size, number of defects), temperature during layer deposition, and temperature at which the adsorption occurs. These many influences render it more difficult to achieve a reliable pump performance than with other sorbent materials. However, vacuum cryocondensates of gases formed under optimal conditions are not essentially inferior to zeolites or activated charcoals. Typical adsorbate/condensate combinations are He/Ar, He/SF_6 or H_2/CO_2.

There are two different techniques to involve cryocondensates for cryopumping (see Section 3.1), namely the pre-frost and the concurrent method. Both methods have been applied successfully for pumping helium on argon in several large fusion machines. The underlying physics of helium cryosorption on solid argon has therefore been investigated comprehensively [75], as this is the only application which gained practical importance. However, lots of other cryocondensates to pump hydrogens, neon and/or helium were investigated, especially by Yuferov *et al*. [76–80].

According to the pre-frost technique, the condensate layer is predeposited on a blank cryosurface prior to the actual pumping operation. In this case, the gas being pumped interacts with a static sorbent, which gets saturated layer by layer [81–85]. The condensate layer must be built up in a defined way to have a sufficiently porous structure and is usually about some μm thick. With respect to the pumping mechanism, this is based on standard cryosorption, but the achievable equilibrium pressures are lower than for other sorbent materials.

However, the pumping efficiency achievable for pure gases is often very much decreased when mixtures are pumped due to a blockage of the frost pores by the other gas species. Very important for obtaining a porous structure that can be penetrated by the gas pumped is the temperature of condensation of the adsorbent gas [83, 86]. Each combination of adsorbent and adsorbate has its own optimum condensation temperature, which is between 4 and 10 K for hydrogen or helium as adsorbates. At lower temperatures, the adsorption process is limited to sites near the surface of the condensate layer. Only when the temperature is high enough, can the adsorbate molecules be transferred to sites with more favourable energy levels inside the condensate layer. Thus, the active surface is in the order of some 100 m^2/g. The efficiency of the pumping process is determined by the ratio of the number of molecules of the sorbent gas to the number of molecules of the gas to be pumped, which is necessary to obtain appreciable pumping speeds. This ratio is quite high, for example of the order of 20 Ar atoms to pump one He atom at 4.5 K. Nevertheless, Ar is very much suited as cryosorbent, as it has a very high thermal conductivity and effectively removes the heat released during adsorption of helium.

Alternatively, the adsorbent can be injected continuously with the process gas to be pumped; this is the classical mechanism of how to realise cryotrapping pumping. Ar trapping was comprehensively investigated by Hengevoss and Boissin [87–89]. The latter developed the Ar spray technique with a special gas injection manifold. It is convenient to carry out when a cryocondensation pump is readily available for the vacuum system. The advantage of concurrent injection is the availability of fresh frost throughout the pumping process, whereas in the case of pre-frosting, deleterious effects leading to considerably reduced pumping speeds have been observed, e.g., due to icing of other gases present over the frost surface and thus blocking the pathways into the bulk. For cryotrapping, of H_2 on Ar at 4 K, the ratio of the number of adsorbed particles to the sorbent particles becomes equal to 1 and even higher.

However, the biggest disadvantage of using cryocondensates is the considerable increase in the amount of gas which has to be dealt with during regeneration by correspondingly greater sized forepumps.

3.4.1.3 Activated charcoal

Activated charcoal sorbents are complex products, as they are manufactured from a wide variety of natural precursors, such as coconut shell, wood, lignite, bitumen, pitch, or peat. The structure of activated carbon is characterized by amorphous regions embedded in well-ordered graphite-like microcrystalline regions, which are formed by regular hexagonal rings. Two different ways are used to activate the charcoal, i.e., increase porosity thus leading to many functional groups on the surface: chemical activation involves the addition of compounds such as phosphoric acid to the parent feedstock prior to carbonization; physical activation refers to gasification of the carbon (introduction of oxygen). The type of raw material, the specific additives, the carbonization and activation temperatures are of great importance in the preparation of activated carbons with respect to their porous structure. Carbons are usually categorized according to the particle shape and size into powders, granules, and formed material (spheres, pellets or fibres). The suitability of an activated carbon for a particular application depends on the ratio in which pores of different sizes are present. The study of carbon materials and their characterization is a lively scientific field of its own.

The broad benchmark studies for different cryosorption materials (references above) unanimously revealed that activated charcoals, especially coconut-shell-based materials, have optimum pumping characteristics. Consequently, numerous investigations on cryosorption pumping involving charcoal as sorbent have been reported and will be used for comparison in the following sections of this paper (e.g., Refs. [90–96]). Nevertheless, sorption equilibrium data for different cryogenic gases in the sub 77 K range are still scarce [56–58, 97–103]. An additional advantage of charcoal compared to molecular sieves is the moderate temperature of about 400 K required to achieve

complete regeneration, whereas molecular sieves would have to be heated up to 300°C for reactivation. This high temperature is incompatible with the use of vacuum seals and gaskets. Carbon has a greater capacity than molecular sieves and is less sensitive to impurity accumulation and water pre-loading.

Thus, we recommend charcoal as the material of choice in operating with the large flow rates and impure gases encountered.

For application as sorbent for cryopanels, the sorbent has to be bonded to a surface. In order to develop the optimum panel set-up, material aspects (sorption isotherms, pore distribution) have to be combined with application aspects (compatibility with bonding agent, thermal cycling, bonding technique). A special bonding technique developed at Forschungszentrum Karlsruhe has been benchmarked within comprehensive thermal cycling tests and no deterioration or thermal degradation has been detected [55]. Some sample panels were manufactured, and the pumping speed for helium was measured. The results are shown in Fig. 8 which illustrates the different significance of the capacity of charcoal when related to unit mass or to the available panel surface. Thus, the material yielding the lowest mass specific capacity for He provides the highest gas loads, when related to the panel surface. This is due to the fact that the available capacity for the cryopanel is not limited by the sorption performance itself, but by the coating process, yielding thin layers for powders and thick layers for charcoal granules.

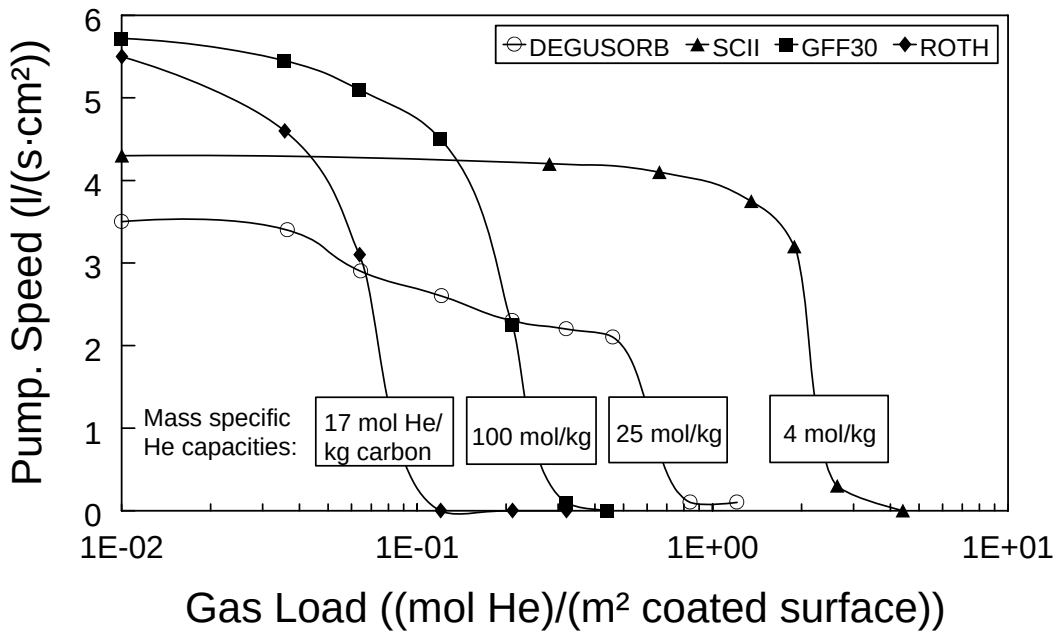


Fig. 8: Pumping speed values for helium at 5 K with different activated charcoal materials, plotted in surface-related coordinates. The materials are pellets (DEGUSORB, 1.5 mm diameter, 5 mm long), granules (SCII), and powders (GFF30, ROTH). DEGUSORB is very close to the materials used in commercial cryopumps.

At Forschungszentrum Karlsruhe, a stage-wise programme was performed in the 1990s to identify the 'best' activated charcoal + bonding technique for nuclear fusion applications (concurrent pumping of helium and hydrogen isotopes at high throughputs). Recently, the class of novel super-high surface area carbon fibre materials (above 2000 m^2/g) has been included. Finally the SCII material (Fig. 8) has become the reference cryosorbent. It is a coconut-shell-derived, granular (US mesh size 12 × 30), microporous (peak at ~ 0.8 to 0.9 nm) activated charcoal material. The reference coating leads to a surface related carbon coverage of ~ 450 g/m^2.

The test programme also showed that the charcoal with the largest surface area was not that which yielded the best performance data. This is in excellent agreement with what has been reported by Sedgley *et al*. [43]. Consequently, all the tests described in the following parts of this report were performed with this kind of carbon material, unless indicated otherwise. Figure 9 presents photographs of the SCII carbon material, taken at different microscopic scales.

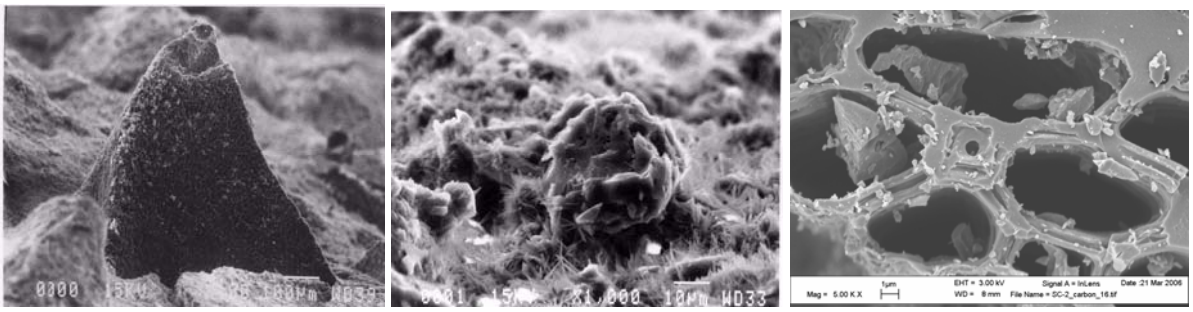


Fig. 9: SEM photographs of the SCII material (100 μm, 10 μm, 1μm scale)

3.4.2 Standardized criteria for refrigerator-cooled cryopumps

There is a set of parameters to quantitatively characterize the performance of cryopumps and, thus, the characteristics of the sorbent material therein, see also Table 2. They are listed and discussed in the following section. For refrigerator-cooled cryopumps, which are the only class of cryopumps with large-scale industrial application, the practices of their measurement are fixed by recommendations of the American Vacuum Society (AVS) [104].

1. *Pumping speed.*

The pumping speed is the central scaling criterion for a cryopump and defined as the quotient of the throughput Q of a gas and the working pressure p.

$$S = \frac{Q}{p}. \tag{8}$$

This equation holds, as long as the ultimate pressure of the pump is negligible against the working pressure, which is usually the case. It is equivalent to a volume rate of flow and is expressed in l/s, referenced to $T^* = 273.15$ K. To convert to number of particle or amount of substance based units, the ideal gas law can be used:

$$p \cdot V = n \cdot R_0 \cdot T^*, \tag{9}$$

where V denotes the volume of the vacuum chamber, n the number of moles of gas present, and R_0 the universal gas constant $R_0 = 8.3145$ J/(mol·K).

For easy intercomparison of different pumps, the pumping speed is often related to a typical area, such as the cryopanel area or the entrance cross-section.

2. *Maximum throughput.*

This is the maximum constant gas flow rate which can be pumped by a cryopump at temperatures below 20 K in the second stage. The throughput is usually expressed in (Pa·m^3)/s, taken at T^*.

3. *Pumping capacity.*

The pumping speed decreases with an increasing amount of pumped gas. For adsorbed gases in particular, the pumping speed asymptotically reaches zero. To have a comparable and reproducible measure, the pumping capacity is defined as the quantity of gas that has been pumped until pumping

speed has been reduced to 50% of the initial value. The test for determining the capacity must be run at constant throughput, so that a decrease of 50% in pumping speed is indicated by a corresponding increase of 100% in pressure. The capacity is expressed in ($Pa \cdot m^3$).

4. *Ultimate pressure*.

This denotes the minimum achievable base pressure of a cryopump which has been preconditioned (typically minimum 6 h vacuum bake-out), and which is thereafter operated at minimum temperature for at least 24 h.

5. *Cool-down time*.

This denotes the elapsed time between turning on the refrigerator of a cryopump at room temperature and achieving temperatures of 100 K in the first stage and 20 K in the second stage.

6. *Crossover*.

The crossover is defined by the maximum amount of nitrogen gas that can be admitted into the pump in a time interval of 3 s, while the second stage remains at temperatures below 20 K. However, this temperature level has been chosen quite arbitrarily from a physical point of view.

To design a vacuum system properly it must be emphasized that the different properties listed above are strongly interrelated among themselves and must in any case be coupled with the performance characteristics of the backing pump. To prevent overloading during and immediately after switching from pre-evacuation to high vacuum pumping, the crossover must be performed when the gas mass flow from the vacuum chamber is less than the maximum throughput of the cryopump. Moreover, the maximum throughput for the crossover condition must be correlated to the period of regeneration.

3.4.3 Regeneration and safety aspects

Cryopumps retain pumped gases as long as their capacity is not exceeded and the arrays are maintained at appropriate temperature. This condition may arise when the cryodeposit condensed at the baffle blocks the louver aperture or when the surface temperature of the panels covered with frozen gas becomes too high due to the temperature gradient through it. However, the cryopump must be regenerated at the latest when the capacity limit has been reached. The capacity is very large for cryocondensation pumps (only limited by the temperature gradient within thick condensate layers) and smaller for cryosorption pumps (limited by full coverage of the active sorption sites, which may be partly clogged). The correct choice of the sorbent is therefore a vital point in cryopump design.

Regeneration is regularly done by heating the cryosurfaces up so that the captured gas deposits are released. As a result, the pressure in the pump increases strongly so that the gas can be pumped away by mechanical forepumps. Before heating, the isolation valve (most appropriate are gate valves which provide full opening) between the pump and system is closed. As the pumping speed starts to decrease before the maximum saturation capacity is reached, a certain safety margin with respect to the maximum operating time must be ensured. After pumping out the release gas, the pump is cooled down again and, thus, the cryopump is ready to be used for the next pumping period. The isolation valve to the vacuum system must remain closed during cool-down, because contaminants resident in the system might be pumped by the second-stage sorbent material, which may result in the plugging of this material. Thus, a regeneration cycle combines five basic steps:

– Closing of the valve to the system.

– Heating of the panel.

– Pumping down of the released gas.

– Cool-down of the panel.

– Opening of the isolation valve.

If the regeneration times needed cannot be complied with by the process (e.g., in high-throughput fusion plasma processes), a quasi-continuous pumping system can be achieved by combination of several cryopumps operated staggered by a certain time interval.

Different methods have been used to achieve the heating of the cryosurfaces [105]. Often, the regeneration comprises partial stages at intermediate temperatures sufficient to release low molecular gases and total stages, which are more time consuming, to release all remaining gases. The regeneration process can also be separated geometrically, e.g., in the case of three-stage pumps (cf. Fig. 5), by confining the heating to parts of the pump. Thus, very efficient fractionation of the different gases can be achieved by intelligent control of the regeneration process. The simplest regeneration method is the natural warm-up. The refrigeration is switched off and the warm-up of the panels occurs by thermal radiation from the environment. After reaching the equilibrium temperature of phase transition or desorption, the captured gas is liberated and the pressure rises. Then, heat transfer via convection and gas conduction starts and leads to a faster warm-up. In most commercial cryopumps, this principle is supported by injection of inert purge gas to accelerate the natural heating by initiation of gas heat conduction. This combined regeneration technique can be performed in a fully automated mode controlled by means of an electromagnetic purge valve and cryogenic temperature sensors fastened to the cryosurfaces. However, it results in increased roughing pump requirements to cope with the increased gas amounts. Therefore, electric regeneration systems based on resistance heaters are offered. Normal regeneration times are of the magnitude of some tens of minutes.

In some cases, the point when regeneration is initiated is not given by capacity and speed aspects, but by safety considerations. For example, when hydrogens are pumped at high throughputs, formation of explosive gas mixture compositions must be avoided in case of a sudden air in-rush due to a leak when the complete pumped gas load has just been liberated [106]. In this case the minimum pressure limit for combustibility may be exceeded. The hydrogen inventory is directly proportional to the period between successive cryopump regenerations, which indicates a need to have a greater number of pumps and to regenerate them at shorter intervals and more rapidly (of the order of some minutes). For these applications, faster heating techniques have been developed, including infrared, glow discharge, microwave and synchrotron radiation (photodesorption) heating. Certain precautions must also be taken, since the cryopumped gases may be released species by species according to their desorption/sublimation curves, rather than mixed.

3.5 Cryopump design

The vacuum technological design of a cryopump aims to maximize the pumping speed S within the existing cooling supply conditions and space limitations. The basic design Eq. (10) shows that apart from the usual parameters such as temperature T, inlet cross section A, and the properties of the gas being pumped (mass M and gas constant R_0), the only parameter left to adjust during design is the capture coefficient c, to which the pumping speed directly scales. The capture coefficient is given by the ratio of the actual pumping speed of the cryopump to the theoretical black hole pumping speed S_{id} and indicates the efficiency of the pump:

$$S = c \cdot S_{\mathrm{id}} = c \cdot A_{\mathrm{inlet}} \cdot \sqrt{\frac{R_0 \cdot T}{2\pi \cdot M}} \,. \tag{10}$$

Thus, the pumping speed of a cryopump is a function of the geometry of its internal structures, the molecular weight M of the gas being pumped, and temperature. However, as outlined above, a

reference temperature of $T = T^* = 273.15$ K is usually chosen. It should also be noted that it is not correct to compare the pumping speed without referring to the gas. So, a pump with the same pumping efficiency differs, for example, by factor 2.6 in pumping speed between He and N_2, with the lighter gas He having the higher speed.

The ideal speed is reduced, firstly on the path of a particle to the cryosurface due to the limited transmission probability w of the installed baffles etc., and, secondly, by the incomplete adhesion on the cryosurface due to the different types of gas–surface interaction. The latter effects are summarized by an overall sticking probability α (the number of particles sticking to the cryosurface related to the number of particles impinging on it) characterizing the overall gas–surface interaction. By this phenomenological approach, the first-order influences of geometry (summarized by w) are separated from the second-order influences of gas–surface interaction (summarized by α). For the simplified case of parallel arrangements of the panel and the baffle, as typical for commercial refrigerator cryopumps, we have the following simple expression:

$$\frac{1}{c} = \frac{1}{\alpha} + \frac{1}{w} - 1 . \tag{11}$$

Because of this geometry influence, it is not easy to scale the pumping speed from one geometry to another.

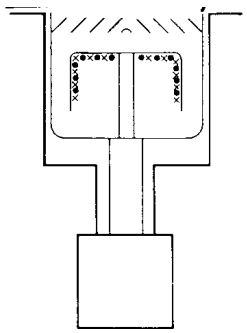
Values for the transmission probability can be derived from compilation textbooks (see the bibliography at the end of this report). The sticking probability itself is determined by many parameters, with temperature, type of gas, and, for adsorption pumping, sorbent pore size distribution and pumped gas amount being the main influential factors. Because of these many influences and because the data retrievable from the literature are scarce and inconsistent, a parametric characterization has been performed for the ITER type charcoal under ITER-relevant conditions, the results of which are summarized in Table 3. As expected, the sticking probability of He, which is the most difficult gas to pump, is a strong function of temperature, whereas the temperature dependence for H_2 is less pronounced. The sticking probability for particles pumped by re-sublimation is higher (up to unity) and less dependent on temperature.

Table 3: Comparison of initial (zero gas load) sticking probabilities for nuclear fusion relevant gases at different temperatures

Temperature (K)	**He**	**H_2**	**D_2**	**T_2**
5	0.35 [107] 0.25 [108]	0.6 [107]	0.9 [107]	1.0 [109]
7	0.17 [107, 110]	0.5 [111]	–	1.0 [109]
12	0.03 [107]	0.3 [112]	0.85 [112]	–

As an example, based on Eq. (11), Table 4 gives some numerical values for a typical commercial refrigerator cryopump, assuming a transmission probability for the inlet baffle of 0.4, which is typical for a louver. It becomes obvious that the capture coefficient depends strongly on the type of gas being pumped. There results more than an order of magnitude between the gas being pumped best and worst.

Table 4: Characteristic parameters for a standard refrigerator cryopump (first stage at 70 K, second stage at 10 K)

Gas	*W*	α	*C*	Where and how pumped	Reference set-up
H_2O	1.0	1.0	1.0	At the inlet baffle by condensation	
T_2, N_2	0.4	1.0	0.4	At the cryopanel front side by re-sublimation	
H_2	0.25	0.6	0.21	At the cryopanel back side by cryosorption	
He	0.25	0.05	0.04	At the cryopanel back side by cryosorption	

One of the advantages of cryopumps is that they can be designed as *in situ* pumps tailor-made according to the available space, dedicated to the specific application. The arrangement of cryopanels and shields is then rather complex so that the application of the simplified Eq. (1) to estimate the capture coefficient is no longer justified. In this case, Monte Carlo simulations of the pump interior have to be performed to calculate the capture coefficient. The pump designer, who has to decide upon the arrangement of the individual pump components, can use Monte Carlo simulations as a study tool. As a design goal, one should aim to come as close as possible to the ideal capture coefficient curve shown in Fig. 10. A pump exhibiting such a performance curve, i.e., with constant capture coefficient over a wide range of sticking probability, would pump practically all gas species at the same pumping speed.

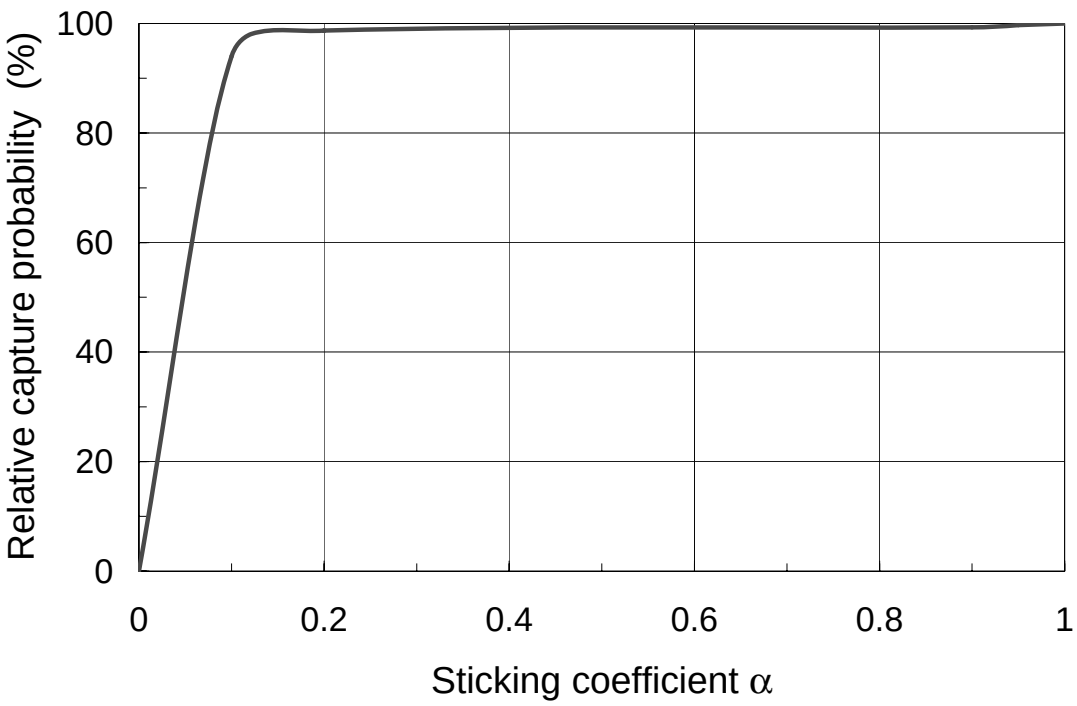


Fig. 10: The 'ideal' curve of the capture coefficient, related to its maximum value at a sticking probability of unity

3.6 Summary of advantages of cryopumping

The practical design of a cryosorption pump is a compromise between several parameters, some of which are mutually contradictory. It is, therefore, evident that to arrive at an optimum configuration is not an easy matter and the final design is possibly best achieved by a combination of theory and experiment. Having reviewed the advantages and limitations of cryosorption pumping, the following list summarizes briefly the advantages.

1. Cryopumping offers unparallelled performance where absolute cleanliness and oil-free conditions are required. Consequently, it is used as a reference for benchmarking contamination introduced from the process side or from other vacuum pumps.

2. Cryopumping is the method of choice for reproducible pumping of water vapour or very humid gas mixtures, as it provides black hole pumping speed (capture coefficient of unity).

3. Cryopumping provides increasing pumping speed with lighter gases.

4. Cryopumping is the most economic solution if highest pumping speeds are required: up to 60 m^3/s (nitrogen) via DN1200 for the biggest commercial pumps. Customized cryopumps can provide biggest pumping speeds via smallest cross-sections (e.g., 100 m^3/s via DN750).

5. Cryopumping based on liquid cryogens does not involve any electric supply (i.e., inherent safety against power breakdowns, full compatibility with applications in magnetic fields) and is totally free of vibration.

6. No mechanically moving parts are present in the vacuum chamber (i.e., highest reliability, minimal maintenance, no bearing and shaft seal problems, no problems with dust contamination, highest safety level).

7. The free orientation of the cryopump allows for installation at places with limited access.

8. Cryopumping provides the possibility of selective pumping.

Besides the conventional refrigerator-cooled cryopumps, cryopumping is mainly used for large-scale R&D applications, such as space simulation [113], plasma physics and thermonuclear fusion [111, 114, 115], particle beam and accelerator systems [116–119]. These pumps are almost exclusively prototypes and dedicated developments.

4 Special regeneration aspects of a cryosorption pump

The operation of a cryosorption pump includes some specialities due to the gas–sorbent interaction, some of which will be discussed in the following sections. A major aspect is the cryopumping mechanism. The fact that a sorbent is present in the pump does not necessarily mean that all gases are really being pumped by adsorption. There are cases where gas is being pumped by condensation on top of the charcoal. In these cases, the charcoal pores may be clogged and the pumping speed for gases being pumped by sorption only may be drastically reduced. Moreover, if it comes to regeneration, the behaviour for all gases is given by the charcoal inside the pump.

Figure 11 shows the result of a defined regeneration test under a constant heating rate with active forepumping. As a result, typical pressure peaks are obtained which indicate gas releases.

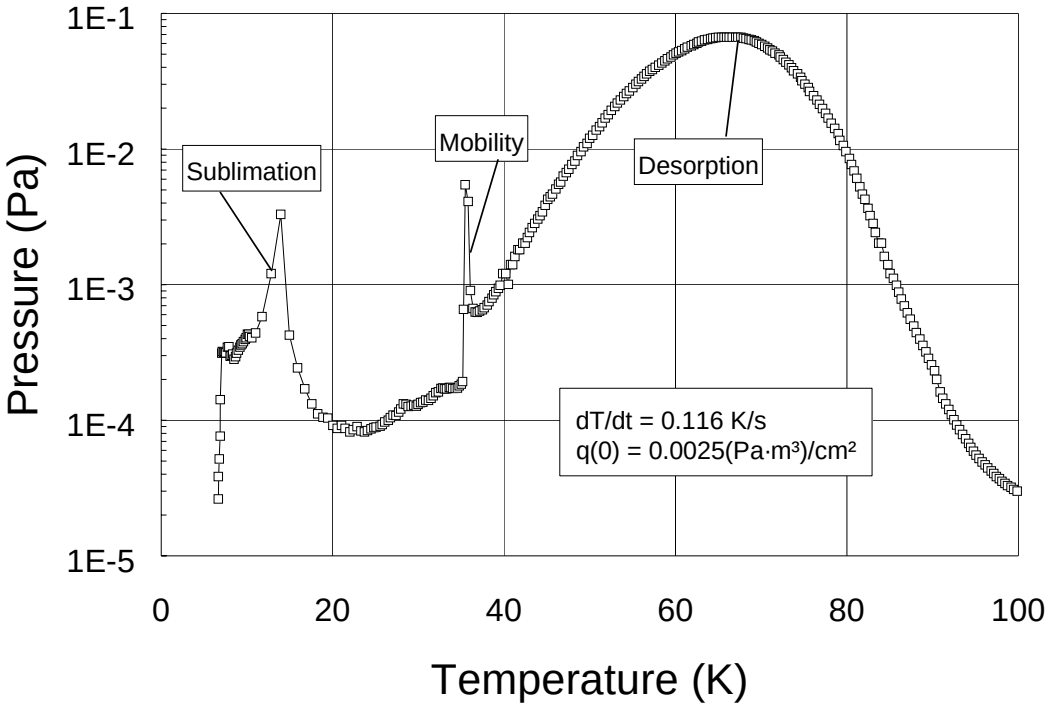


Fig. 11: A temperature programmed desorption experiment for D_2, illustrating mobility effects

The curve in Fig. 11 exhibits three clearly distinguishable peaks. The largest peak in the higher temperature range indicates desorption. The low temperature peak for deuterium can be attributed to the sublimation pressure curve. However, there is a further peak at the intermediate temperature. This peak is a kinetic effect due to dynamic loading, which was described first by Lessard [90], typical for quantum gases. The quantum delocalization of surface molecules follows a strong, exponential functional relation between the residence time of a molecule impinging and the sorbent temperature. At a certain threshold temperature, the surface mobility of the adhered molecules is high enough so that they can clear the surface. In this activated state, the molecules can migrate deeper into the sorbent pores, but are at the same time liberated more easily. Consequently, a pressure peak can be observed. This boundary temperature also depends on the gas load situation (intermittent or continuous throughput) [91] and can be exploited to increase the loading capacity of the sorbent by heating beyond this temperature limit, which helps the molecules to disperse rapidly into the sorbent bulk, and thus, offers new active surface sites for cryopumping. In this respect, the static sorbent capacity, which is known to increase with decreasing temperature, is not necessarily the correct measure to judge a real dynamic cryopumping process. To evaluate the holding time of a cryopump, capacity aspects and mobility aspects have to be combined.

Figure 12 is another temperature programmed desorption experiment to illustrate the fractionation of the gases pumped during controlled heating. The first peak is formed by helium desorption and partial release (to a small extent) of deuterium due to mobility and sublimation. The second peak is produced by desorption of the remaining deuterium. Thus, a fine separation between the two species can be achieved by heating to an intermediate temperature range of about 30 K. Alternatively, the sorbent can be used for selective pumping, if operated at this temperature, i.e., only the hydrogens will be pumped, whereas helium will not. This application is of vital interest for trace analysis and leak detection devices for large vacuum systems [120].

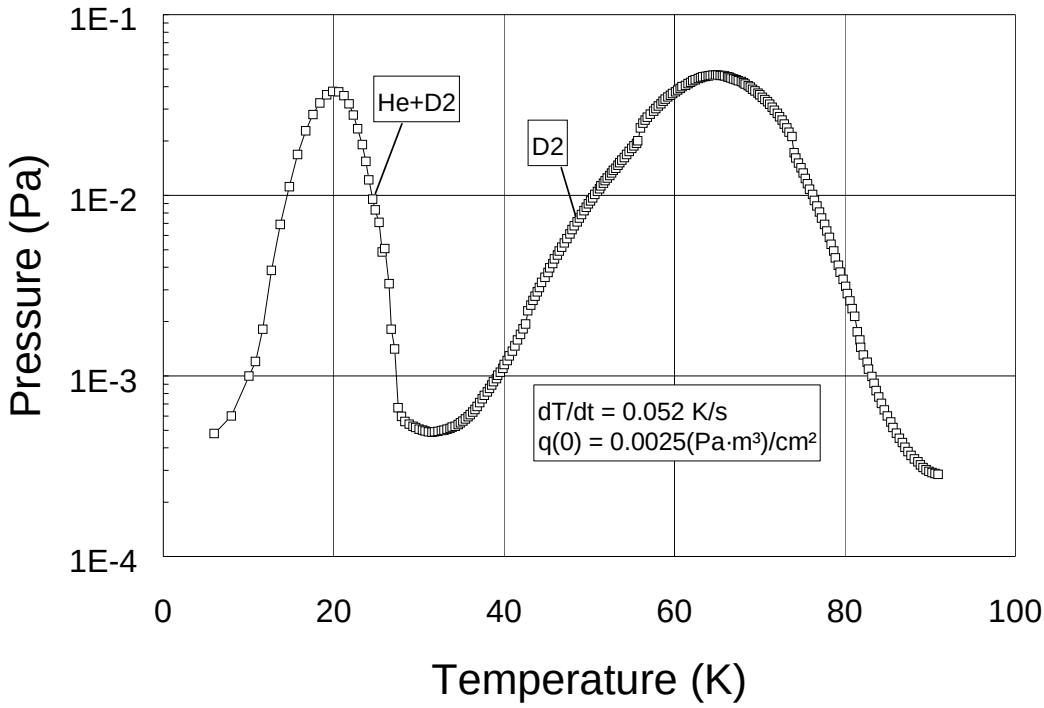


Fig. 12: A temperature programmed desorption experiment of a mixture with 10% helium in deuterium, illustrating fractionation effects

As illustrated in Fig. 13 (and Table 3), helium, which is the gas that is most difficult to pump, and has by far the lowest sticking probabilities, is also the most sensitive to temperature increase. The strongest helium release will take place in the temperature range between 10 and 20 K. Figure 13 illustrates the 3D relationship between pumping speed, temperature, and capacity limit for helium. The capacity is given related to the coated surface (at 450 g charcoal/m^2).

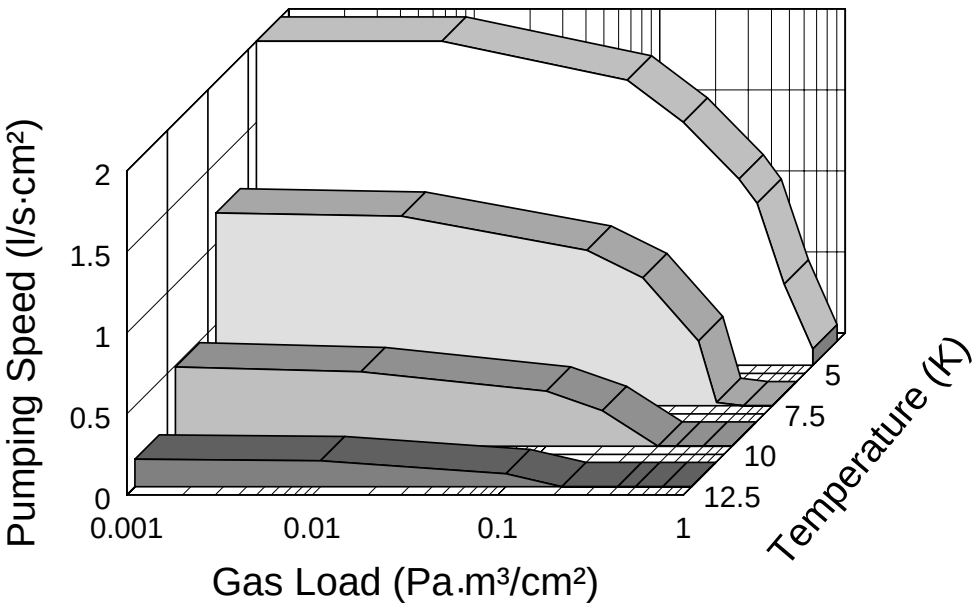


Fig. 13: Performance map of helium croyosorption on activated charcoal

Hydrogen H_2 also shows some sorption capacity limitation, but this is at least a magnitude higher.

As outlined above, species being pumped by sorption and species being pumped by condensation are pumped on different surfaces. The latter are cryocondensed at the 80 K baffle arrays and, in this respect, will not immediately influence the sorption performance of the panels downstream. However, if regenerated, the condensed species will evaporate as soon as the temperature becomes equivalent to the boiling temperature at the corresponding partial pressure and may re-adsorb on the charcoal, as the temperatures sufficient to sublimate the cryodeposits are still sufficient to provide for a sorption at the charcoal. By this mechanism, all gas species, no matter where pumped at first, will have a principal impact on the sorption panels. An example of this effect is illustrated in Fig. 14, which shows the interaction of sublimation from the heated baffle stage and subsequent adsorption at the sorption stage.

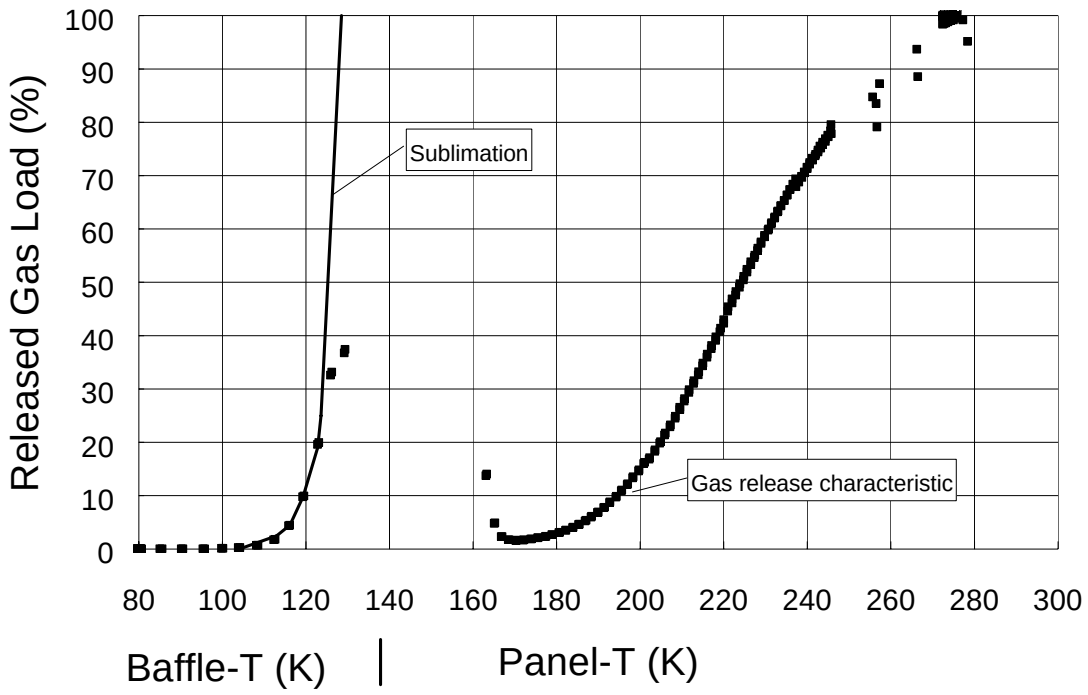


Fig. 14: Interrelation of sublimation of the condensed CO_2 pumped at the baffle and re-adsorption on the charcoal coated panel

This measurement was performed during regeneration by heating up the panel directly and, via radiation and starting gas heat conduction, the baffle indirectly. Therefore, there is a certain driving temperature difference between panel and baffle. Nevertheless, the molecules released from the baffle, will instantaneously stick to the panel, even at the higher temperature there. Hence, the performance of each cryopump with a sorption stage is governed by the ad-/desorption process, even for those gases which at first are not pumped by cryosorption, but condensed at the baffle. This has major implications on cryopump operation, because there will often be a pumping situation involving a cryosorbent which is pre-loaded with a certain amount of condensables. Figure 15 gives typical desorption temperature curves needed to achieve efficient regeneration of the charcoal. Of course, desorption from a sorbent can also be stimulated by pressure reduction. However, this is a much slower process and is therefore only used to increase sorption efficiency in parallel to the usual temperature increase.

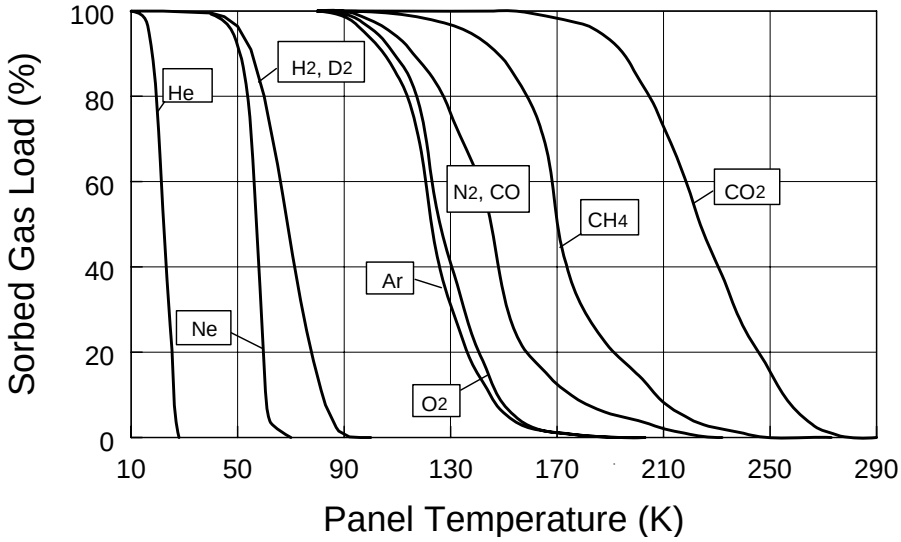


Fig. 15: Temperature-induced desorption curves

This leads to the detailed investigation of the dynamics involved in the sorption pumping process, which is characterized by limitations due to saturation and poisoning. The saturation limit is reached when the maximum gas load has been pumped. The reduction of pumping speed due to accumulation of trace components in the gas mixture to be pumped, which are not being released during normal regeneration procedures, is another important feature of cryopump operation [120].

5 Application examples

This section features two selected application examples for customized cryopumps.

5.1 Combined refrigerator/liquid cryogen pump

Figure 16 shows a three-stage cryopump which was developed for space simulation chambers. These applications usually require a vacuum in the order of about 10^{-4} Pa. The inlet baffle is cooled with liquid nitrogen to 77 K. Inside the pump, there is a two-stage GM cooler, featuring a 50 K condensation stage and a 20 K low temperature sorption stage (charcoal covered). The regeneration can be done via electrical heaters attached to the cryosurfaces, or with warm gas purging, alternatively. The pump features an opening diameter of 1.25 m and has a nominal pumping speed of 50 m^3/s for nitrogen (capture coefficient $c = 0.35$) and hydrogen (capture coefficient $c = 0.09$).

Fig. 16: Large cryopump for space simulation chambers

5.2 ITER cryopumps

ITER is the next generation thermonuclear fusion device of the tokamak type, and represents the experimental step between today's studies of plasma physics and tomorrow's electricity-producing fusion power plants. It is based on a deuterium–tritium plasma operating at over 100 million K, and will produce 500 MW of fusion power. Control of the gas throughput, especially the helium ash produced by D–T fusion reactions, is one of the key issues affecting the performance and achievable burn time of a fusion reactor.

The ITER machine includes three large cryogenic high-vacuum pumping systems. One is for evacuation and maintenance of the required pressure levels in the torus (1350 m^3); the second is for generation of the required vacuum conditions in the neutral beam injectors (NBI) (570 m^3), which are used to heat up the plasma by injection of highly energetic accelerated neutral H and D particles; and the third is for provision of the insulation vacuum in the cryostat (8400 m^3), which houses the superconducting coil system. The typical pressures inside the torus and the NBI systems are in the range of 1 to 10 Pa. Thus, the rationale to use cryogenic pumping is not given by the need to establish low pressures, but by the need to process very high gas throughputs necessitating highest pumping speeds. As ITER will be a burning deuterium–tritium plasma experiment, the vacuum system must not only be designed for the high magnetic and electric fields but also withstand disruption events creating high mechanical loads, radiation, and be compatible with tritium. The latter excludes the use of any organic material in direct or sealing contact with the process gas. These requirements can be met best with cryogenic pumps without any moving parts, backed by a multi-stage forepump train based on roots pumps.

All large high-vacuum pumping systems on ITER are based on charcoal-coated cryopanels forced-cooled by 4.5 K supercritical helium at 0.4 MPa. Extensive preparatory tests were run in the past to find an optimum combination of sorbent type and bonding cement, resulting in a ~ 1 mm thick layer of specific granular activated charcoal, bonded by an inorganic tritium-compatible cement. The ITER type charcoal is a microporous, granular, highly activated carbon, derived from coconut shells. An automatic coating facility using a special spray technique has been developed at Forschungszentrum Karlsruhe. Figure 17 shows a photo of the typical coated cryopanels which are used as modules to build up the sorption cryosurface in any required geometry.

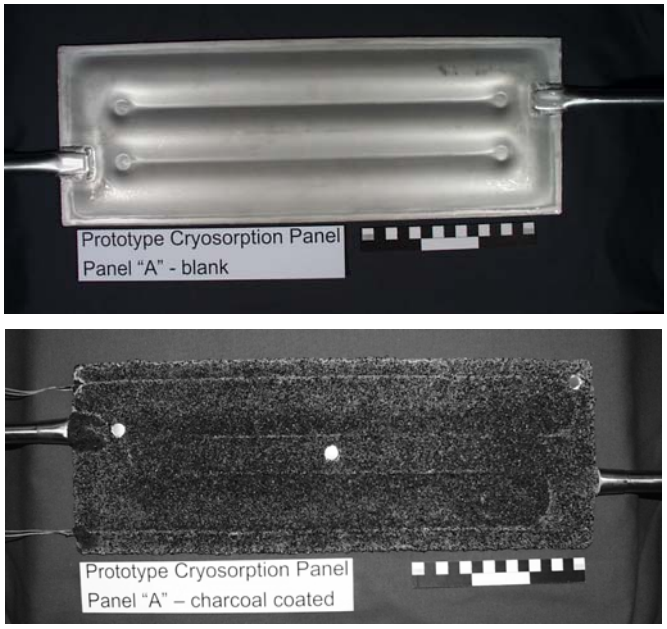


Fig. 17: Photographs of a prototype cryosorption panel (150 mm wide, 450 mm long); in blank state (top) and ready coated (bottom, here also equipped with heaters and prepared spots for temperature sensors)

Although being tailor-made pumps, the cryopumps share the same basic features as every cryopump, namely a low-temperature circuit (the 4.5 K cryosorption panel system) and an intermediate temperature circuit (the 80 K thermal shielding and inlet baffle system) to reduce thermal loads on the low-temperature circuit. The temperatures are given by the available cooling conditions at ITER. The system of the eight ITER torus cryopumps is the most complex on ITER. A speciality of the ITER cryopumps is the inclusion of an inlet valve, to control the throughput and to isolate the pump from the torus during regeneration. The inlet valve adds an additional degree of freedom to the pump control. Figure 18 illustrates the principle circular design of the ITER torus cryopumps, featuring 28 cryosorption panels [121]. Figure 19 shows two photographs of a half-scale model pump version which was manufactured some years ago. The pumping speed results of the model pump performance are summarized in Fig. 20.

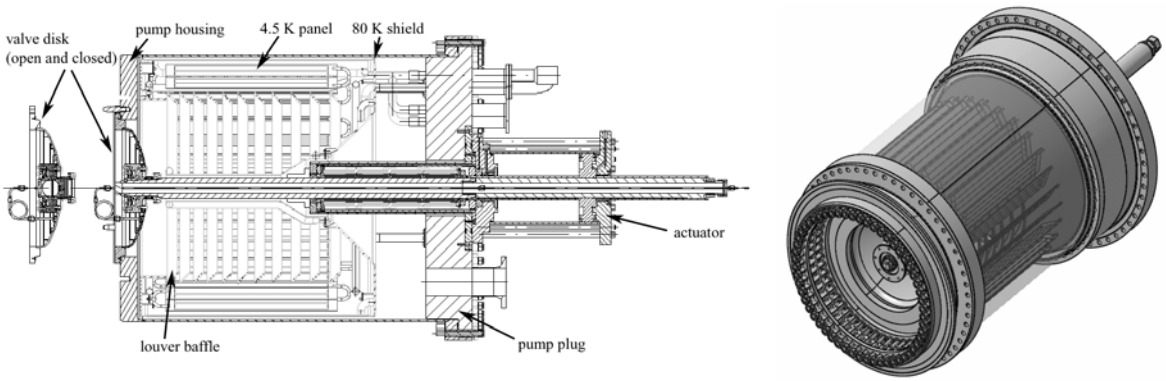


Fig. 18: Design of the ITER torus cryopump (nominal free molecular pumping speed of 50 m^3/s, DN 800). 2D cut (left) and a 3D view (right, without outer 80 K shield and housing)

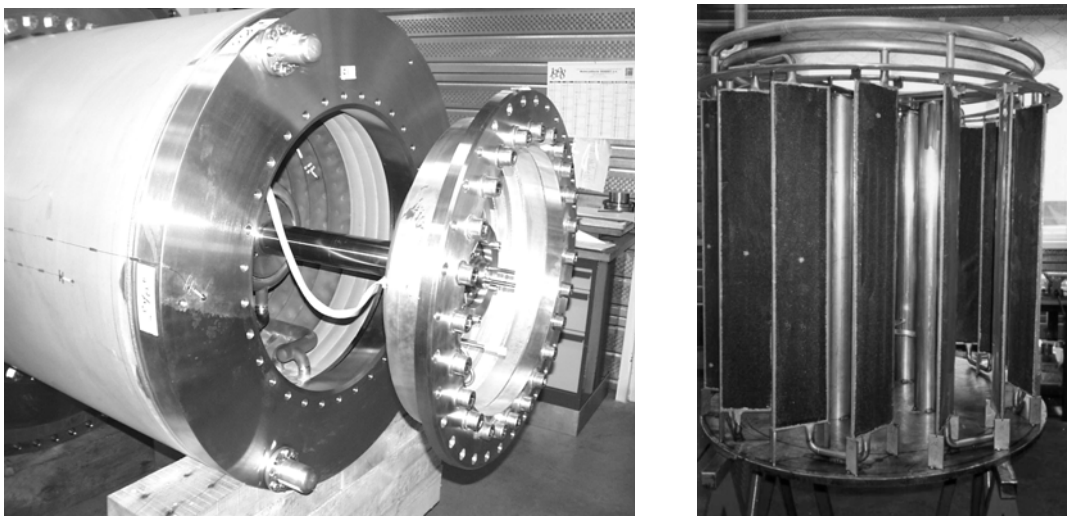


Fig. 19: Photos of the model pump, showing the opened inlet valve (left) and the interior cryopanel circuit (right)

Figure 20 illustrates the valve control effect. The pump performance under high throughput conditions is a good fingerprint of the robustness of the pump design, as this involves maximum heat loads on the cryopanel system. The measured pumping speed values are plotted twice, on the left-hand side as a function of the pressure inside the pump, on the right-hand side as a function of the pressure at the pump inlet. The difference of the 100% curves indicates the pressure loss of the fully opened valve, which is about 1 decade over the whole range of investigated flowrates. The vertical interconnecting lines combine points of the same throughput. For increasing flowrate, the pressure is

also increased. However, the pressure situation inside the pump depends only on the throughput and is completely decoupled from the valve position. Whereas the pressure at the pump inlet, which is the decisive one since it is closely linked to the torus pressure, can be controlled by appropriate adjustment of the valve position. Thus, the valve is a helpful tool to make the pump more flexible and suited to establish the specified inlet pressure range.

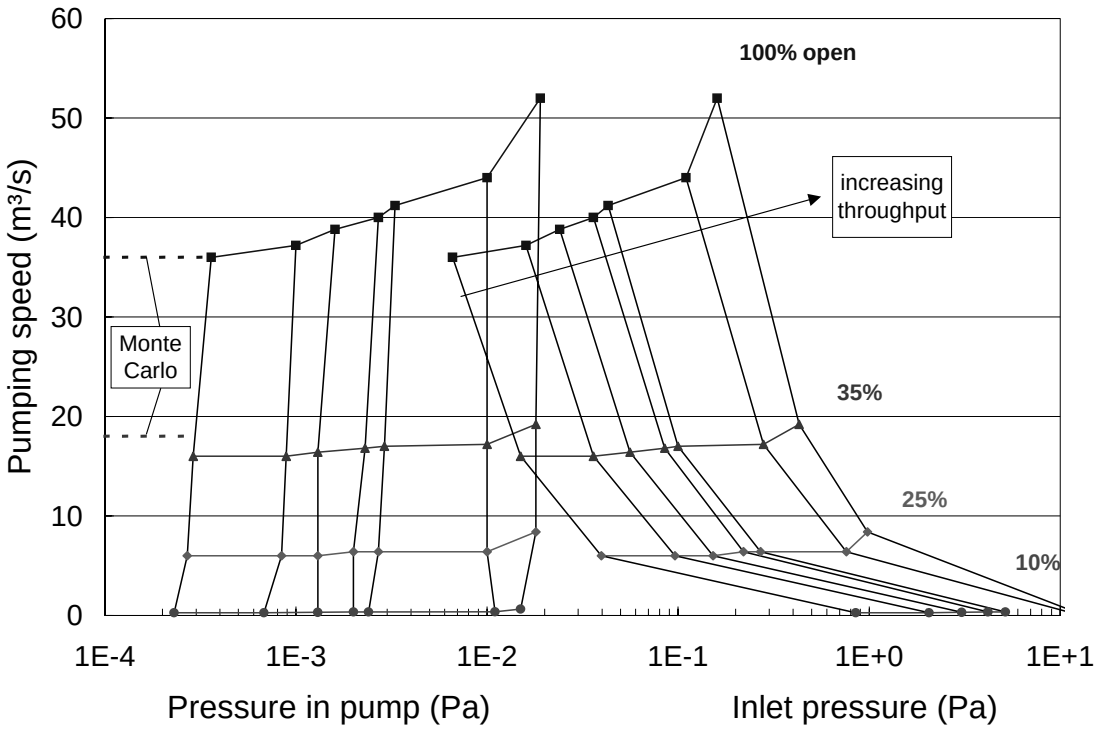


Fig. 20: Pumping characteristics of the model pump. Results for a typical fusion exhaust model gas, consisting of 87% D_2, 10% He, and 3% impurities (CO, CH_4, CO_2, O_2). The throughput was varied [between $6 \cdot \times 10^{-6}$ Pa · m^3/(s · cm^2) and 4×10^{-4} Pa · m^3/(s · cm^2)] at constant valve position (100%, 35%, 25% and 10% open).

The increase of pumping speed with increasing throughput indicates that the pump will be under transitional flow regime inside the pump during regular operation.

The same design principles, but without an inlet valve, will be applied to develop the NBI cryopumps. The required integral pumping speeds are 3800 m^3/s for H_2, and 2600 m^3/s for D_2, respectively [122].

6 Conclusions and prospects

This review articles gives an introduction to the underlying basics and different aspects of practical cryopumping, and reviews the state of the art. The modelling of tailor-made cryopumping is a demanding field because it cross-links many different disciplines such as vacuum science, cryogenic engineering, surface and physical chemistry. It further brings together in a unique way technological and physics aspects. In view of the large R&D projects coming up in the next few years, further development work and fundamental investigation of cryopumps will be necessary. This is the main guideline for the research to be done in the future.

References

[1] ISO 3529/1–3, *Vacuum Technology: Vocabulary* (International Organization for Standardization, Geneva, Switzerland, 1981).

[2] R.T. Jacobsen, S.G. Penoncello, and E.W. Lemmon, *Thermodynamic Properties of Cryogenic Fluids* (Plenum Press, New York, 1997).

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

[4] A.G. Mathewson, Vacuum **44** (1993) 479.

[5] S.G. Sydoriak, T.R. Roberts and R.H. Sherman, J. Natl. Bur. Stand. **68A** (1964) 559.

[6] B. Fellmuth and G. Schuster, Metrologia **29** (1992) 415.

[7] V. Arp, R.D. McCarty and J. Fox, *HEPAK Code* v. 3.40 (Cryodata Inc., Louisville, CO, 1997).

[8] P.C. Souers, *Hydrogen Properties for Fusion Energy* (University of California Press, Berkeley, 1986).

[9] K. Erents and G.M. McCracken, Vacuum **21** (1971) 257.

[10] J.P. Hirth and G.M. Pound, *Condensation and Evaporation* (Pergamon Press, Oxford, 1963).

[11] M.M. Eisenstadt, J. Vac. Sci. Technol. **7** (1970) 479.

[12] V. Ponec, Z. Knor, and S. Černy, *Adsorption on Solids* (Butterworths, London, 1974).

[13] G.C. Maitland, M. Rigby, E.B. Smith and W. Wakeham, *Intermolecular Forces* (Clarendon Press, Oxford, 1981).

[14] A.W. Adamson, *Physical Chemistry of Surfaces*, 5th ed. (John Wiley, New York, 1990).

[15] L.W. Bruch, M.W. Cole and E. Zaremba, *Physical Adsorption: Forces and Phenomena* (Clarendon Press, Oxford, 1997).

[16] M.G. Kaganer, *Thermal Insulation in Cryogenic Engineering* (IPST Press, Jerusalem, 1969).

[17] Chr. Day, B. Kammerer and A. Mack, Adv. Cryog. Eng. **43** (1998) 1327.

[18] J.R. Howell, *A Catalog of Radiation Configuration Factors* (McGraw-Hill, New York, 1982).

[19] R. Siegel and J.R. Howell, *Thermal Radiation Heat Transfer*, 2nd ed. (Hemisphere Publishing, New York, 1981).

[20] J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods* (Chapman and Hall, London, 1979).

[21] G.L. Saksaganskii, *Molecular Flow in Complex Vacuum Systems* (Gordon and Breach, New York, 1988).

[22] S. Tsujimoto, A. Konishi and T. Kunitomo, Cryogenics **22** (1982) 603.

[23] C. Benvenuti, D. Blechschmidt and G. Passardi, J. Vac. Sci. Technol. **19** (1981) 100.

[24] G.S. Ash, Calculation of pumping speeds of high vacuum systems in molecular flow, *Proc. 37th Annual Technical Conference Society of Vacuum Coaters*, 1994, pp. 100–109.

[25] W. Obert and G. Perinić, Vacuum **43** (1992) 163.

[26] H.O. McMahon and W.E. Gifford, Adv. Cryog. Eng. **5** (1960) 354.

[27] P.D. Bentley, Vacuum **30** (1980) 145.

[28] J. Harvell and P. Lessard, Semiconductor International, June 1991.

[29] B. Engers and H.U. Bauer, Surface and Coatings Technol. **116** (1999) 705.

[30] Y. Shiokawa, T. Kikuchi and M. Ichikawa, Vacuum **47** (1996) 803.

[31] T. Shigi, M. Nakamura, M. Yamaguchi, M. Azuma and M. Yanai, Cryogenics **30** (Supplement) (1990) 523.

[32] J.P. Perin, D. Henry, R. Vallcorba, J.J. Cordier and F. Samaille, Vacuum **48** (1997) 743.

[33] J.L. Hemmerich and E. Küssel, J. Vac. Sci. Technol. **A 8** (1990) 141.

[34] A. Horbowa and P. Dolegieviez, Vacuum system performances on spiral facility, *Proc. 15th Int. Conf. on Cyclotrons and Their Applications*, Caen, France, 1998.

[35] M. Gouge and C. Foster, *Argon Frost Continuous Cryopump for Fusion Application*, Report DOE/ER/81236-6 (Department of Energy, Washington, USA, 1995).

[36] Chr. Day and V. Hauer, Adv. Cryog. Eng. **50** (2004) 75.

[37] D.D. Do, *Adsorption Analysis: Equilibria and Kinetics* (Imperial College Press, London, 1998).

[38] F. Rouquerol, J. Rouquerol and K. Sing, *Adsorption by Powders and Porous Solids* (Academic Press, London, 1999).

[39] S.J. Gregg and K.S.W. Sing, *Adsorption, Surface Area and Porosity*, 2nd ed. (Academic Press, London, 1982).

[40] D. Perinić, H. Haas and A. Mack, Fusion Eng. Design **18** (1991) 79.

[41] D. Perinić, H. Haas and A. Mack, Cryogenics **30** (Supplement) (1990) 509.

[42] D. Perinić, H. Haas and A. Mack, Adv. Cryog. Eng. **39** (1994) 1553.

[43] D.W. Sedgley, G. Tobin, T.H. Batzer and W.R. Call, J. Vac. Sci. Technol. **A 5** (1987) 2572.

[44] D.W. Sedgley, A.G. Tobin, T.H. Batzer and W.R. Call, Nuclear Eng. Design **4** (1987) 149.

[45] D.W. Sedgley, T.H. Batzer and W.R. Call, J. Vac. Sci. Technol. **A 6** (1988) 1209.

[46] A.G. Tobin, D.W. Sedgley, T.H. Batzer and W.R. Call, J. Vac. Sci. Technol. **A 5** (1987) 101.

[47] D.W. Sedgley, Fusion Eng. Design **10** (1989) 217.

[48] L.S. Gurevich, I.N. Moreva, V.V. Petrovsky, A.V. Pustovoit and A.S. Shelukhin, Plasma Devices and Operation **2** (1994) 287.

[49] L.S. Gurevich, V.V. Petrovski and A.V. Pustovoit, Plasma Devices and Operation **1** (1990) 97.

[50] M. Xu and Y. Matsui, J. Vac. Sci. Technol. **A 13** (1995) 132.

[51] J. Ma, D.L. Kingsbury, F.-C. Liu and O.E. Vilches, Phys. Rev. Lett. **61** (1988) 2348.

[52] V.V. Anashin *et al*., Vacuum **76** (2004) 23.

[53] V.V. Anashin *et al*., Vacuum **75** (2004) 293.

[54] V.V. Anashin *et al*., Vacuum **72** (2004) 379.

[55] Chr. Day, H. Haas, N.T. Kazakovsky, A. Mack, D.K. Murdoch, D. Röhrig and G.L. Saksagansky, Summarized results of the cryosorption panel test programme for the ITER cryopumping system, *Proc. 17th IAEA Fusion Energy Conference*, Yokohama, Japan, 1998, vol. 3, pp. 1073–1076.

[56] R. Yanik, U. Yanik and C. Heiden, J. Low Temp. Phys. **45** (1981) 443.

[57] D. Basmadjian, Can. J. Chem. **38** (1960) 141.

[58] D. Basmadjian, Can. J. Chem. **38** (1960) 149.

[59] J.G. Daunt and C.Z. Rosen, J. Low Temp. Phys. **3** (1970) 89.

[60] M. Jäckel and R. Bartel, Exp. Tech. Phys. **28** (1980) 153.

[61] P.J. Gareis and G.F. Hagenbach, Ind. Eng. Chem. **57** (1965) 27.

[62] S.A. Stern, J.T. Mullhaupt, R.A. Hemstreet and F.S. diPaolo, J. Vac. Sci. Technol. **3** (1965) 165.

[63] G.E. Grenier and S.A. Stern, J. Vac. Sci. Technol. **3** (1966) 334.

[64] S.A. Stern, R.A. Hemstreet and D.M. Ruttenber, J. Vac. Sci. Technol. **3** (1966) 99.

[65] L.C. Pittenger, Adv. Cryog. Eng. **23** (1978) 648.

[66] R.J. Powers and R.M. Chambers, J. Vac. Sci. Technol. **8** (1971) 319.

[67] H.J. Halama and J.R. Aggus, J. Vac. Sci. Technol. **11** (1974) 333.

[68] D. Cheng and J.P. Simson, Adv. Cryog. Eng. **10** (1965) 292.

[69] S.A. Stern and F.S. diPaolo, J. Vac. Sci. Technol. **4** (1967) 347.

[70] D. Burghardt, A. Mack, D. Perinić, H. Reichert and D. Röhrig, Vacuum **45** (1994) 547.

[71] H.J. Halama, H.C. Hseuh and T.S. Chou, Adv. Cryog. Eng. **27** (1982) 1125.

[72] H.J. Halama and J.R. Aggus, J. Vac. Sci. Technol. **12** (1975) 532.

[73] P.W. Fisher and J.S. Watson, J. Vac. Sci. Technol. **16** (1979) 75.

[74] C.F. Dillow and J. Palacios, J. Vac. Sci. Technol. **16** (1979) 731.

[75] G.V. Kachalin, A.P. Kryukov and S.B. Nesterov, Low Temp. Phys. **24** (1998) 97.

[76] V.B. Yuferov, Low. Temp. Phys. **19** (1993) 413.

[77] V.B. Yuferov and P.M. Kobzev, Sov. Phys. Tech. Phys. **14** (1969) 427.

[78] V.B. Yuferov and F.I. Busol, Sov. Phys. Tech. Phys. **11** (1967) 1518.

[79] V.B. Yuferov and P.M. Kobzev, Sov. Phys. Tech. Phys. **14** (1969) 1261.

[80] V.B. Yuferov and P.M. Kobzev, Sov. Phys. Tech. Phys. **15** (1971) 2041.

[81] M.M. Menon, G.J. Laughon, R. Maingi, M.R. Wade, D.L. Hillis and M.A. Mahdavi, J. Vac. Sci. Technol. **A 13** (1995) 551.

[82] J. Kim, K.M. Schaubel and A.P. Colleraine, J. Vac. Sci. Technol. **A 8** (1990) 3084.

[83] J.H. Kamperschroer, M.B. Cropper, H.F. Dylla, V. Garzotto, L.E. Dudek, L.R. Grisham, G.D. Martin, T.E. O'Connor, T.N. Stevenson, A. von Halle, M.D. Williams and J. Kim, J. Vac. Sci. Technol. **A 8** (1990) 3079.

[84] T.H. Batzer, R.E. Patrick and W.R. Call, J. Vac. Sci. Technol. **18** (1981) 1125.

[85] M.P. Larin, Sov. Phys. Tech. Phys. **33** (1988) 455.

[86] E. Wallen, J. Vac. Sci. Technol. **A 14** (1996) 2916.

[87] J.C. Boissin, J.J. Thibault and H. Richardt, Le Vide **157** (1972) 103.

[88] J. Hengevoss and E.A. Trendelenburg, Vacuum **17** (1967) 495.

[89] J.J. Thibault, J.M. Disdier, G. Arnaud, J.C. Boissin and C.B. Hood, J. Vac. Sci. Technol. **18** (1981) 1140.

[90] P.A. Lessard, J. Vac. Sci. Technol. **A 7** (1989) 2373.

[91] R.C. Longsworth, Adv. Cryog. Eng. **23** (1978) 658.

[92] R.C. Longsworth and Y. Lahav, J. Vac. Sci. Technol. **A 5** (1987) 2646.

[93] H.C. Hseuh and H.A. Worwetz, J. Vac. Sci. Technol. **18** (1981) 1131.

[94] R.I. Shcherbachenko and V.N. Grigor'ev, Low Temp. Phys. **24** (1998) 831.

[95] B. Liu, J. Xie, J. Ren and X. Cui, J. Vac. Sci. Technol. **A 5** (1987) 2577.

[96] J.R. Coupland, D.P. Hammond, W. Bächler and H.H. Klein, J. Vac. Sci. Technol. **A 5** (1987) 2563.

[97] A.J. Kidnay, M.J. Hiza and P.F. Dickson, Adv. Cryog. Eng. **13** (1968) 397.

[98] C. Johannes, Adv. Cryog. Eng. **17** (1972) 213.

[99] B.P.M. Helvensteijn, A. Kashani, R.A. Wilcox and A.L. Spivak, Adv. Cryog. Eng. **39** (1994) 1561.

[100] B.P.M. Helvensteijn, A. Kashani and R.A. Wilcox, Adv. Cryog. Eng. **37** (1992) 1129.

[101] I. Vazquez, M.P. Russell, D.R. Smith and R. Radebaugh, Adv. Cryog. Eng. **33** (1988) 1013.

[102] R. Yanik, Vacuum **47** (1996) 205.

[103] H. Parish, Adv. Cryog. Eng. **35** (1990) 1541.

[104] K.M. Welch, B. Andeen, J.E. deRijke, C.A. Foster, M.H. Hablanian, R.C. Longsworth, W.E. Millikin, Y.T. Sasaki and C. Tzemos, J. Vac. Sci. Technol. **A 17** (1999) 3081.

[105] H.-J. Mundinger, H.U. Häfner, M. Mattern-Klosson, H.H. Klein and U. Timm, Vacuum **43** (1992) 545.

[106] M. Iseli, Fusion Eng. Design **54** (2001) 421.

[107] Chr. Day, Use of porous materials for cryopumping, *in*: H.S. Nalwa, *Handbook of Surfaces and Interfaces of Materials* (Academic Press, San Diego, 2001), vol. 5, p. 265.

[108] I. Özdemir and D. Perinić, J. Vac. Sci. Technol. **A 16** (1998) 2524.

[109] Chr. Day *et al*., Fusion Sci. Technol. **48** (2005) 29.

[110] Chr. Day and A. Schwenk-Ferrero, Vacuum **53** (1999) 253.

[111] Chr. Day et al., Vacuum **81** (2007) 738.

[112] Chr. Day, Colloids and Surfaces **A 187–188** (2001) 187.

[113] H.E. Nuss and I. Streuff, Vacuum **47** (1996) 391.

[114] W. Obert, C. Mayaux and G. Perinić, Adv. Cryog. Eng. **39** (1994) 1569.

[115] Y. Oka, T. Takanashi, R. Akiyama, O. Kaneko, K. Toi, H. Morimoto, M. Terashima and T. Kuroda, Fusion Eng. Design **31** (1996) 89.

[116] P. Lebrun, IEEE Trans. Appl. Supercond. **10** (2000) 1500.

[117] J. Miertusova, Vacuum **48** (1997) 751.

[118] H.C. Hseuh, R. Davis, D. Pate, L. Smart, R. Todd and D. Weiss, Vacuum **53** (1999) 347.

[119] H. Dombrowski, D. Grzonka, W. Hamsink, A. Khoukaz, T. Lister and R. Santo, Nucl. Instrum. Methods Phys. Res. **A 386** (1997) 228.

[120] Chr. Day *et al*., Fusion Eng. Design **58–59** (2001) 301.

[121] V. Hauer *et al*., Design of the ITER torus prototype cryopump, accepted for publication in Fusion Engineering and Design.

[122] M. Dremel *et al*., Adv. Cryog. Eng. **51** (2006) 583.

Bibliography

Vacuum technology in general

R.A. Haefer, *Kryo-Vakuumtechnik* (Springer, Berlin, 1981).

M. Wutz *et al*., *Handbuch Vakuumtechnik*, 8th ed. (Vieweg, Braunschweig, 2004).

J.M. Lafferty (Ed.), *Foundations of Vacuum Science and Technology* (Wiley, New York, 1998).

A. Roth, *Vacuum Technology*, 3rd ed. (North-Holland, Amsterdam, 1990).

J.F. O'Hanlon, *A User´s Guide to Vacuum Technology*, 3rd ed. (Wiley, Hoboken, 2003).

N.S. Harris, *Modern Vacuum Practice* (McGraw-Hill, London, 1989).

Cryopumps in particular

R.A. Haefer, *Cryopumping* (Clarendon Press, Oxford, 1989).

K.M. Welch, *Capture Pumping Technology*, 2nd ed. (North-Holland Elsevier, Amsterdam, 2001).

M.H. Hablanian, *High-Vacuum Technology*, 2nd ed. (Dekker, New York, 1997).

L.N. Rozanov, *Vacuum Technique* (Taylor & Francis, London, 2002).

Cryogenics

K.D. Timmerhaus and T.M. Flynn, *Cryogenic Process Engineering* (Plenum Press, New York, 1989).

R.F. Barron, *Cryogenic Systems*, 2nd ed. (Oxford University Press, New York, 1985).

J.G. Weisend II (Ed.), *Handbook of Cryogenic Engineering* (Taylor & Francis, Philadelphia, 1998).

Commissioning of large vacuum systems

K. Zapfe
Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

Abstract
This paper will give an overview of the various steps of the commissioning of large vacuum systems for accelerators. Following some introductory remarks, the pump-down, leak check, bake-out as well as component, interlock, and safety checks will be covered in detail. Special emphasis will be given to beam vacuum systems in combination with cryogenic systems. Finally, the transition from the commissioning phase to beam operation will be treated. Practical examples will illustrate most of the topics.

1 Introduction

Once the manufacture and testing of all individual components of a complex system are successfully finished and the installation comes to completion, the last step before starting the routine operation is the system commissioning. During the commissioning phase it should be shown that the required performance is accessible with the installed equipment and that the system meets the needs and expectations of the user. The commissioning phase should also be used to prepare a system for standard operation with respect to functionality, safety, and remote operation. If necessary, additional equipment could be installed, and integrations, modifications, or extensions of control parameters, for example, could be performed.

Today's high-energy accelerators such as the 6.3 km long proton–electron collider HERA at DESY [1–3] or the 27 km long electron–positron collider LEP at CERN [4] (now dismantled) certainly contain large vacuum systems; the latter having about 1900 sputter ion pumps, 700 gauges, 2700 feedthroughs, 2200 pick-up connections, 130 sector and 520 roughing valves as well as 13 000 LEP-type flange pairs. However, systems are considered to be large not only due to their length or number of components, but also due to their complexity like the vacuum system of the 200 m long superconducting linear accelerator FLASH at DESY [5] containing various alternating sections operated at room temperature and 4 K, and an extremely large number of diagnostic elements.

For an accelerator vacuum system the required beam vacuum levels as well as beam pipe aperture, choice of material, RF shielding etc. are defined and strongly influenced by the beam parameters aimed for such as beam intensity, beam lifetime and beam stability. Thus during the commissioning phase the vacuum systems performance should be verified with respect to pressure, residual gas composition, outgassing rate, pump down and/or conditioning time, etc.

In order to carry out a 'smooth' commissioning the following premises should be fulfilled:

- a proper planning with respect to the overall system design, choice of materials, manufacturing techniques and vacuum equipment as well as specifications and vacuum controls;
- a thorough quality control during design and manufacture of components, handling procedures, cleaning, testing and installation;
- the availability of a continuous monitoring and data logging system.

2 Pump-down and leak check

Following the installation of the vacuum components into an accelerator, the next important step is the pump-down of the vacuum sections. In general, mobile pump stations are temporarily connected to a sector for this purpose. Before starting the pump-down, the proper attachment of the components should be carefully checked, especially for bellows. Vacuum forces can easily shift poorly fixed components and even damage them. Once the pressure is sufficiently reduced, the permanently attached high-vacuum pumps can be started.

Observing and recording the pressure during pump-down allows experienced staff a first analysis of the system's behaviour. Any deviation from a usual pump-down curve is a hint of a possible leak as described in more detail in Ref. [6]. In case of such an indication, further checks using a residual gas analyser or the pressure rise method might be useful before starting a proper leak check.

As the initial pump-down of an accelerator vacuum section often takes a while, the leak check is usually done in several steps starting with a gross check at a pressure below 10^{-4} mbar followed by another check at a pressure value below 10^{-6} mbar. Systematic work is necessary so as not to forget any of the numerous flanged connections, ceramics, feedthroughs, windows etc. For leak detection either a residual gas analyser or a leak detector connected to a mobile pump station could be used. Although sputter ion pumps are normally switched on as soon as possible during pump-down to accelerate the decrease in pressure, they should be switched off during leak check so as not to reduce the helium flow to the leak detection cell and even more important to avoid spoiling the sputter ion pumps with helium in case of a leak.

The activation of further pumps like titanium sublimation or non-evaporable getter (NEG) pumps requires sufficiently long pumping times in the order of days or even weeks for NEG pumps. Practical examples are shown in Fig. 1 clearly indicating a significant pressure drop after the pump's activation. A careful observation of the pressure behaviour should be continued and a leak check added if necessary. Finally the gas composition should be measured using a residual gas analyser.

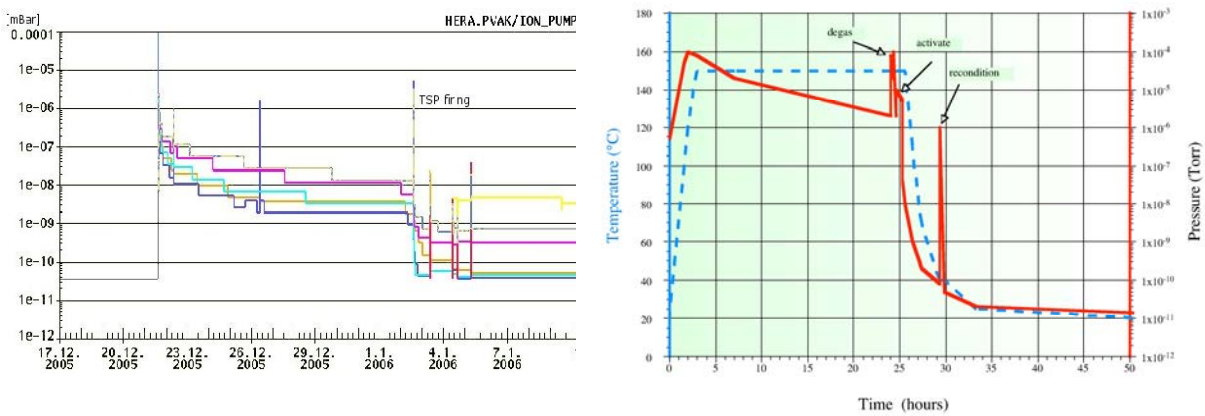


Fig. 1: Pressure drop due to activation of titanium sublimation pumps (left) and NEG pumps [7] (right)

3 Bake-out

The bake-out of a vacuum system to temperatures of typically 150–300°C is a delicate procedure requiring careful planning and execution. A lot of additional equipment such as heaters, thermocouples, insulating material, cabling etc. is necessary. The preparation should include thorough checks of the individual components and their functionality as well as of the heating process and interlocks in case of power failure, failure of the equipment, or unexpected pressure rise.

During bake-out large temperature gradients should be avoided and the movement of components due to the heating checked. Temperature-critical components such as magnets or monitors must be protected by sufficient insulation or active cooling. Figure 2 shows an example of the temperature versus time for a vacuum chamber inside a quadrupole magnet during a bake-out of the HERA proton ring [8].

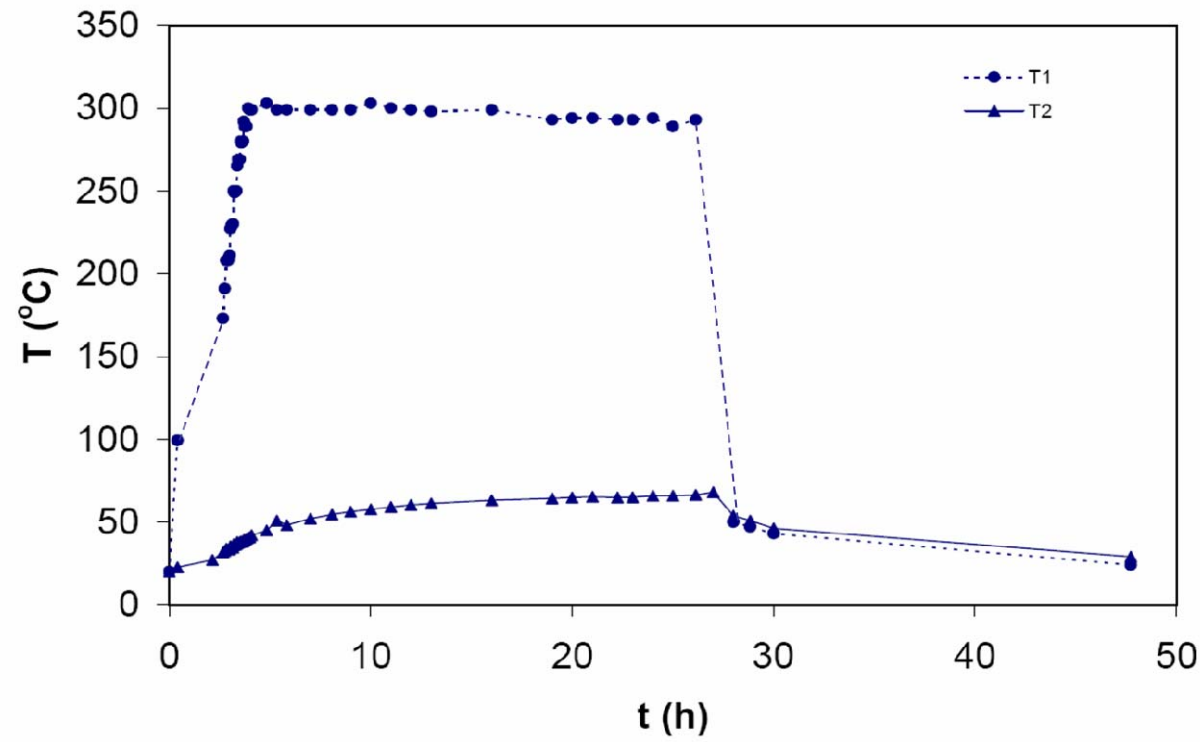


Fig. 2: Temperature versus time of a vacuum chamber (●) and a magnet coil of a quadrupole magnet (▲) during a bake-out cycle of the HERA proton ring [8]

4 Components check

An important task of a systems commissioning is the check-out of the components. This includes checking the correct allocation of the vacuum equipment, cables and corresponding electronics as well as testing the functionality of pumps, gauges, valves, etc.

Following the more hardware-oriented work the vacuum controls need to be checked as well. This should include verifying the correct display of the vacuum components within the various vacuum and accelerator control programs, checking the remote operation of components, and testing the functionality of the control programs. Wherever necessary, handling options should be improved or added and additional parameters be integrated.

The check-out of the vacuum system should also include components of other subsystems and special equipment of experiments such as gas inlet systems, which are installed in the vacuum system. For example, the correct path length and proper functioning of end switches of movable elements like collimators and diagnostic units as well as their remote operation should be tested. Pressure increases during initial movements of such elements are not exceptional and should be observed carefully.

5 Interlock and safety checks

In order to protect the vacuum system against severe damage in case of failure of vacuum components or other subsystems, a proper interlock system is an essential part of the vacuum system and thus requires a thorough testing of the various interlock conditions. In general the operation of the vacuum system should be designed in a self-safe way such that failing of equipment like pumps or pressure gauges will not require manual operations to avoid damage of the system. Failing of single components like gauges or pumps should not affect the proper functioning of the overall system and thus the beam operation of an accelerator.

Testing the foreseen vacuum interlock conditions and reactions of the system is an essential step towards a safe system operation. Adequate tools for the check-out like complete accessibility to the status and timing information will ease this important step. Examples are checking the proper closing of gate valves or fast shutters in case of significant pressure increases or failure of certain pieces of equipment like pumps or gauges being part of the interlock signal chain, as well as switching off equipment under certain circumstances. In addition interlock conditions allowing or stopping beam operation as well as allowing different modes of beam operation need to be carefully tested. Last but not least there are usually interlocks connected to other subsystems like diagnostic units or experiments.

The final tests of interlock systems require the completeness and correct conditions of the systems to be tested. On the other hand, the time available before the start of beam operation is often substantially reduced due to major shifts within the installation schedule. Nevertheless, taking the required testing time for the vacuum system will finally pay off with respect to proper functionality and reliability during beam commissioning and routine operation.

6 Vacuum systems in cryogenic environments

In this section special emphasis will be given to the commissioning of vacuum systems in combination with cryogenic applications. Such systems are generally complex arrangements requiring substantially more effort and time for the commissioning phase. In addition to the beam vacuum system a second vacuum system is required as insulation of the cold components against the ambient air. Although the required pressure level within an insulating vacuum system is moderate—typical pressure values are around 10^{-3} mbar before cool down—such systems comprise rather large volumes containing huge numbers of components. In the following some examples of cryogenic applications at accelerators will be given and the typical commissioning procedures for the beam and insulating vacuum system described in detail.

The most prominent accelerator application of cryogenic systems is the use of superconducting magnets as in several of today's high-energy proton or heavy-ion storage rings like the TeVatron (Fermilab), HERA (DESY), RHIC (Brookhaven) and the LHC (CERN). The iron yoke of the magnets is directly assembled onto the beam tube. When filling the housing of the magnets with liquid helium for cooling, part of the beam pipe is also directly surrounded by liquid helium and thus cooled to temperatures around 4 K. In this way the beam pipe itself will act as a huge cryo pump. Long chains of magnets and interconnecting elements are inserted into large vacuum vessels which are connected and evacuated for insulation.

Another application requiring cryogenic systems is the operation of superconducting accelerating structures, so-called cavities. Here the superconducting cavity itself, surrounded by liquid helium, forms the beam pipe. During recent years the interest in this technology has greatly increased on account of the huge progress in performance reached with the TESLA technology [9]. Figure 3 shows schematically the layout of the TESLA vacuum vessel containing the superconducting cavities and all necessary cryogenic supply lines. Superconducting cavities are currently in use on a larger

scale, for example, at the accelerators TRISTAN at KEK, CEBAF at Jefferson Lab and FLASH at DESY and were applied at the now dismantled LEP ring at CERN.

Other cryogenic applications include superconducting wigglers as insertion devices at synchrotron radiation sources and special superconducting magnets within experimental set-ups.

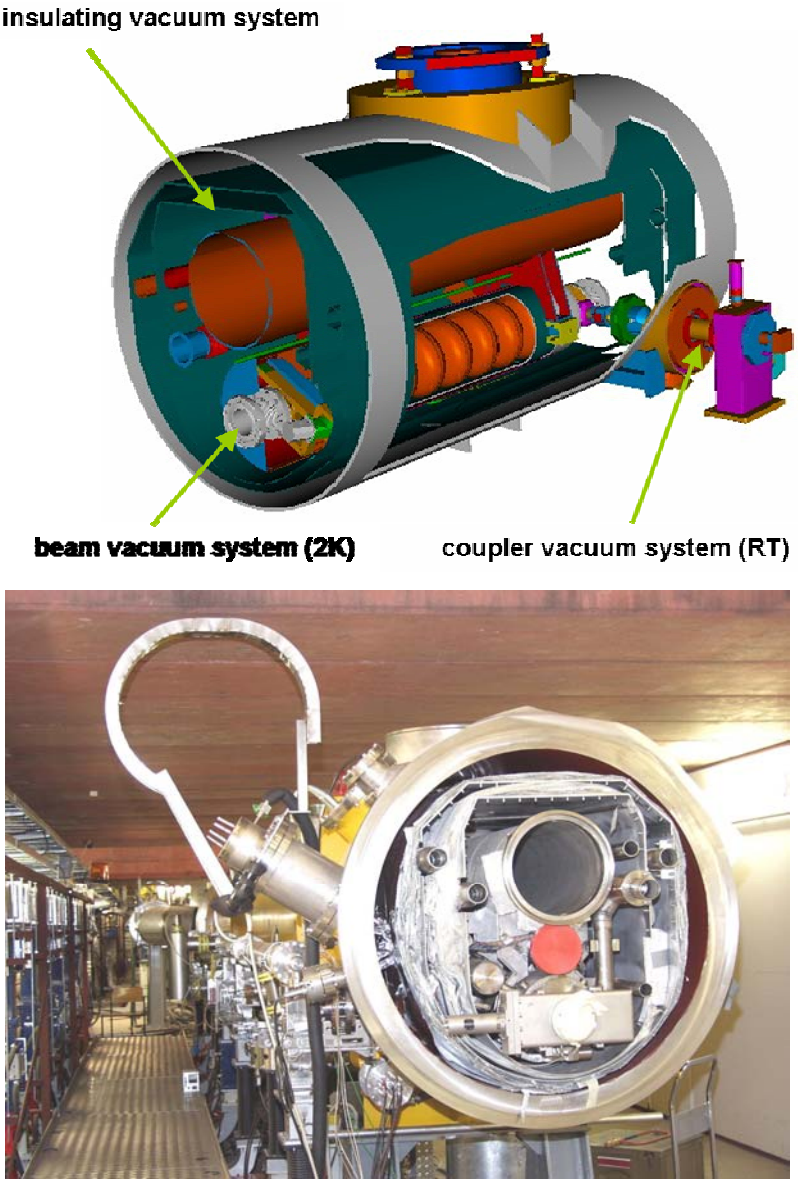


Fig. 3: TESLA vacuum vessel containing the superconducting cavities and all necessary cryogenic supply lines as used at the FLASH linear accelerator at DESY [9]

6.1 Cold beam vacuum systems

Following the installation of the beam line vacuum components a leak check of the final connections like flanges and *in situ* welds is mandatory. One has to keep in mind that a local leak check once the insulating vacuum vessel is closed and thus under cold conditions will not be possible. Therefore the final checks need to be taken with extreme care by applying integral leak tests, for example. In addition a careful and systematic pre-checking of the components is necessary.

Before starting to cool down the beam pipe, sufficiently stable pressure conditions should be reached. The pressure level depends on the specific system. For example at HERA [1] one requires a pressure of 10^{-6} mbar using sputter ion pumps, while at RHIC there is no active pumping of the beam pipe after pump-down by mobile pump stations to a value of 10^{-4} mbar [10]. During cool-down to final temperatures below 4.5 K, the pressure will drop significantly once the beam pipe starts to act as a huge cryo pump as shown in Fig. 4.

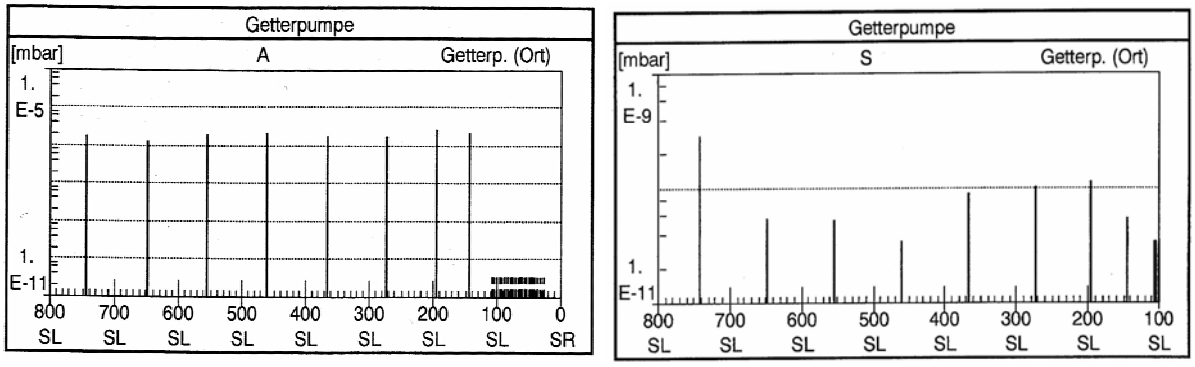


Fig. 4: Pressure distribution within the beam pipe section SL of the HERA proton ring before (left) and after (right) cool down to 4.5 K. The pressure scale in the right plot is enlarged [11].

Once the system has been exposed to very low temperatures and partially to liquid helium for some time, the integral leak rate as well as the wall coverage could be measured. Therefore the system needs to be warmed up to sufficiently high temperature such that the accumulated gas will be released and could be measured using a residual gas analyser as described in more detail in Ref. [6].

6.2 Insulating vacuum systems

The application of large numbers of superconducting magnets or cavities in accelerators results in rather large insulating vacuum systems up to several kilometres in length. A proper segmentation of such systems is necessary for commissioning as well as for reliable operation in case of failure of equipment and leaks. Technically the segmentation is realized by massive barriers and bypass lines with sufficient conductance for pumping, which can be closed off by valves. The distance of such barriers within the insulating vacuum system has increased with time. While at HERA a very conservative distance of 20 m was chosen, this number was increased to 500 m at RHIC and 200 m at the LHC. Major benefits of the enlarged distances are financial and reduced heat load, however, one has to face an increased complexity for the commissioning phase, especially for leak checking.

Before pumping down the insulating vacuum, the connections of the process lines, usually *in situ* welded joints should be leak checked. Therefore the lines could be pumped out and sprayed with helium gas. Alternatively they are pressurized and the leak check is done using a sniffer or a vacuum-jacketed fixture. While the last method results in a much better sensitivity, one has to invest a significant effort for assembly and pump-out of the fixtures at the numerous joints. At RHIC, for example, there are about 25 000 *in situ* helium line welds [10].

The initial pump-down of such large insulating vacuum systems is usually very slow. The gas load is dominated by huge amounts of water from the superinsulating foils. Sufficient mobile equipment must be added to the permanently installed pump stations starting initially with large, if necessary, roughing pumps. Operating the roughing pumps under air ballast and, if necessary, flushing the system with nitrogen will assist the process. In addition one should be prepared for frequent maintenance of the roughing pumps.

The leak check for air leaks should start as early as possible. A full leak check of a section is often an iterative process. Initially, large leaks are not exceptional and often prevent the detection of smaller leaks at high pressure levels. Thus one needs to repair large leaks first and repeat the complete leak check of the section thereafter. Owing to the size of the systems and the number of connections, sufficient granularity of the leak detection equipment is of utmost important. Vacuum gauges could complement mobile leak detectors and/or leak detection cells attached to the mobile pump stations. The base pressure is usually not sufficient for the use of residual gas analysers directly attached to the insulating vacuum vessels. An example of a pump-down curve of a section of the RHIC insulating vacuum system including venting to repair air and helium leaks is shown in Fig. 5 (left) [10]. Without major air leaks a pressure of 10^{-2} mbar is reached within a few days.

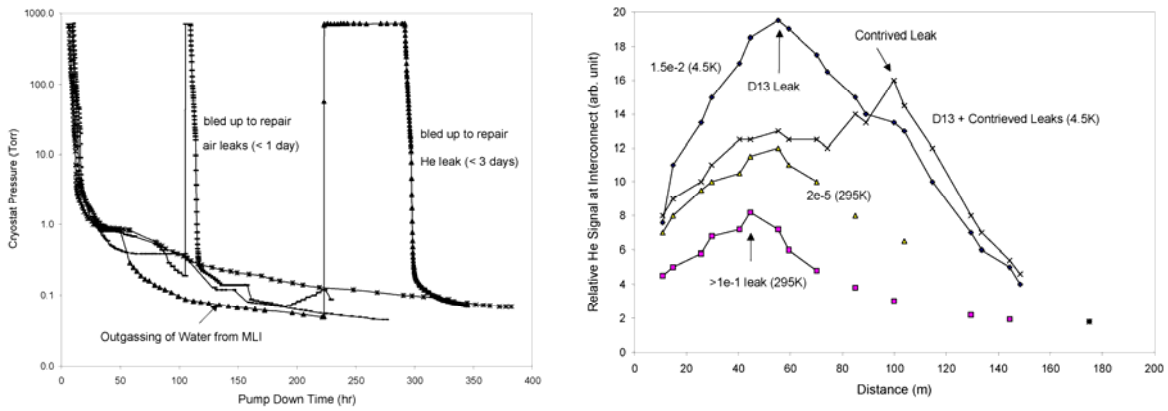


Fig. 5: Pump-down curve of the RHIC 494 m arc cryostats (left) and distribution of the helium signal at the various detectors in case of a leak in a process line interconnection (right) [10]

One should be aware that both the dead time of the leak detection system and the helium background within the tunnel may be significant, the latter due to leaks of the cryogenic helium supply lines within the tunnel. This could reduce the sensitivity of the leak detectors by helium penetrating through the exhaust of the roughing pumps or giving false signals when permanently entering through leaks. In extreme cases different tracer gases should be used. Owing to the moderate pressure requirements for the insulating vacuum, O-rings are often used as gaskets. In contrast to metal seals the permeation of especially light gases from the outside, e.g., helium sprayed during the leak check, is not negligible and finally determines the detection limit for leaks. As the permeation takes some time, sufficient experience is required to distinguish leaks from helium signals due to permeation.

Following the search for air leaks, the process lines need to be checked. Usually the same equipment is used as before, however, one should make sure that the pump system disposes of sufficient helium compression such that in case of leaks from the supply lines into the insulating vacuum the helium is efficiently removed. With the insulating vacuum well evacuated all process lines are pumped out. After checking the background level of the detectors, one process line after the other is pressurized with several bars of preferably helium gas. In this way an integral leak rate of each line within the insulating vacuum will be measured.

Once a leak in a process line has been detected, finding its location will be the more difficult task. With sufficient diagnostics available, the profile of the helium signal within the various detectors should clearly indicate the area of the leak as shown in the example of Fig. 5 (right) [10]. After venting the insulating vacuum, the corresponding sliding sleeve of the vacuum vessel needs to be opened and the leak check of the joints of the pressurized process line continued. Diagnostics could be done by a sniffer or a vacuum-jacketed fixture as described above. In any case, locating and repairing leaks in the process lines is a difficult and time-consuming task requiring lots of experience and sometimes innovative approaches.

The last step of the insulating vacuum systems commissioning is the cool-down. The pressure of the insulating vacuum will significantly drop once substantial surfaces are well below 0°C and thus act as very efficient pumps for water. Typical pressure values are 10^{-3} mbar before and 10^{-6} mbar or below after cool-down for routine operation. Once the pressure has stabilized, part of the mobile pump stations from the pump-down can be removed.

During the cool-down one should check for spots with condensation and/or freezing of water due to thermal bridges. The helium signal should be continuously monitored as additional leaks in the process lines might open up during this phase. Usually one has to live with such leaks at least till the next warm-up of the system, when the leak will eventually be localized and repaired. However, not all leaks can be localized or repaired, some cold leaks even disappear once the system gets warmed-up. Depending on the size of the leak, additional pumping equipment might be added in order to keep the pressure within an acceptable level. An example from HERA is shown in Fig. 6 [12]. Permanent air leaks at the flanges of the insulating tank might be sealed off by using glue or rubber mastic.

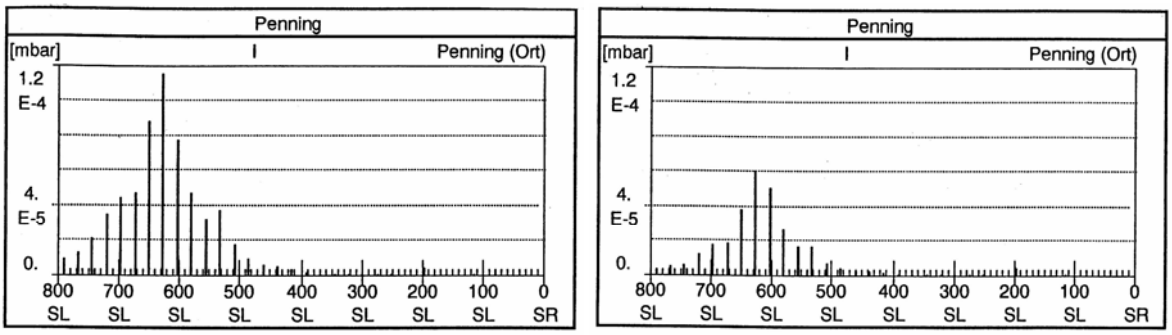


Fig. 6: HERA SL section having a large leak from the helium supply lines into the insulating vacuum: pressure distribution with normal distribution of pumping stations (left) and an additional pumping station (right) [12]

The beam vacuum should be observed frequently during commissioning of the insulating vacuum, e.g., during leak check of the process lines and during cool-down. Direct leaks from the (liquid) helium at the superconducting magnets or cavities or even combined leaks from the process lines through the insulating vacuum into the beam vacuum could be observed. Therefore it is advantageous to supply sufficient pumping speed for helium at the beam vacuum by sputter ion pumps with enhanced helium pumping speed or charcoal, for example. One should be aware that in the case of a cold beam tube, helium leaks might not been seen immediately as the helium will be partially pumped by the cold surfaces and thus a helium front will propagate through the beam tube.

7 First beam operation

Having successfully finished the previous steps the vacuum system should be well prepared for the first passage of beam particles. There are many possible reasons why the operators are not able to fill particles into the accelerator. Prominent ones are of course closed vacuum valves or obstacles in the path of the beam, e.g., RF fingers. But there are many more reasons like malfunctioning magnets, etc.

Once a significant amount of beam is filled, the behaviour of the vacuum pressure should be carefully monitored. Noticeable changes in the pressure should be well understood. In the presence of substantial synchrotron radiation a significant pressure increase might be expected initially due to the photon-induced desorption of gas molecules from the surfaces. The rise of the dynamic pressure depends on the accumulated photon dose. Thus operating the accelerator with beam will condition the vacuum pressure by cleaning the surfaces by photon bombardment. An example for the decrease of

the normalized dynamic pressure with beam dose is shown in Fig. 7 for the conditioning of the LEP vacuum system [13]. However, a pressure rise could also be due to other reasons like heating of components due to insufficient cooling or RF shielding, or beam–gas effects.

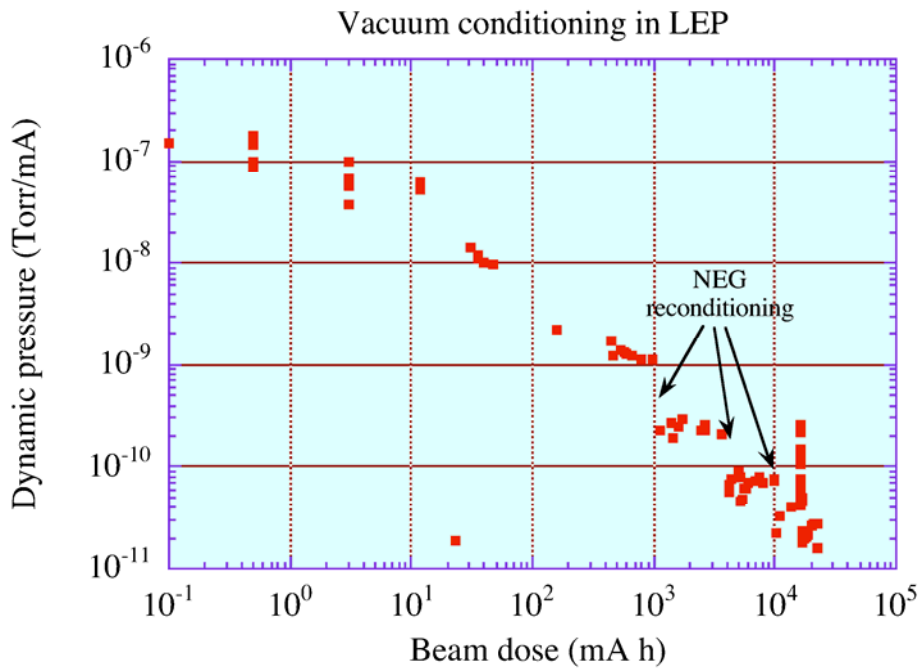


Fig. 7: Conditioning of the LEP vacuum system resulting in a decrease of the normalized dynamic pressure with accumulated beam dose [13]

8 Concluding remarks

The commissioning of accelerator vacuum systems should be the (final) proof of the correct layout, design, manufacture, and installation of the system. Owing to the complexity of the systems and the care required it should be clear that, in addition to well-trained and experienced personnel, sufficient time for the various tests of the systems is also necessary. As the installation of the vacuum system is one of the final steps of an accelerator installation, the available time slot is often reduced owing to delays in preceding steps. Since one is under time pressure, one should keep in mind that the commissioning phase is an essential part of the overall work package of delivering a subsystem of an accelerator. The foreseen checks and tests are an important basis for a reliable performance of the vacuum system during beam operation.

A proper planning of the commissioning phase is thus an integral part of the vacuum system planning. In the case of cryogenic applications, in particular one should be aware that locating and repairing helium leaks in the insulating vacuum systems will be a difficult and time-consuming task. Commissioning parts of the systems as early as possible will be advantageous in order to apply modifications to the remaining parts of the system when necessary.

Acknowledgements

I would like to thank my colleagues of the vacuum group MVP at DESY for their support and many fruitful discussions.

References

[1] D. Trines *et al*., XVth Int. Conf. on High Energy Accelerators, Int. J. Mod. Phys. A (Proc. Suppl.) **2A** (1993) 344.

[2] J.G.M. Römer and D. Trines, Vacuum **38** (1988) 613.

[3] R. Ballion *et al*., Vacuum **41** (1990) 1887.

[4] LEP Vacuum Group, Vacuum **41** (1990) 1882.

[5] K. Zapfe, Vacuum **73** (2004) 213.

[6] K. Zapfe, Leak detection, these proceedings.

[7] O. Gröbner, CERN, private communication (2006).

[8] J.G.M. Römer *et al*., Vacuum **39** (1989) 835.

[9] H. Weise, Proc. of the 2003 Particle Accelerator Conference, Portland (2003), p. 673.

[10] H.C. Hseuh *et al*., Proc. of the 1999 Particle Accelerator Conference, New York (1999), p. 557.

[11] M. Böhnert *et al*, The Cold Bore Beam Pipe Vacuum System of the HERA Proton Storage Ring, DESY HERA 92-02 (Hamburg, 1992).

[12] M. Böhnert *et al*, The Insulating Vacuum Systems of the HERA Proton Storage Ring, DESY HERA 91-22 (Hamburg, 1991).

[13] O. Gröbner, Vacuum **43** (1992) 27.

Analytical and numerical tools for vacuum systems

R. Kersevan
ESRF, Grenoble, France

Abstract
Modern particle accelerators have reached a level of sophistication which require a thorough analysis of all their sub-systems. Among the latter, the vacuum system is often a major contributor to the operating performance of a particle accelerator. The vacuum engineer has nowadays a large choice of computational schemes and tools for the correct analysis, design, and engineering of the vacuum system. This paper is a review of the different type of algorithms and methodologies which have been developed and employed in the field since the birth of vacuum technology. The different level of detail between simple back-of-the-envelope calculations and more complex numerical analysis is discussed by means of comparisons. The domain of applicability of each method is discussed, together with its pros and cons.

1 Introduction

The vast majority of particle accelerators need low pressures in order to work efficiently and reduce beam losses and radiation background. Therefore in the following only the case of molecular flow conditions will be considered, i.e., when the interaction of residual gas molecules with each other is negligible, and the molecules are supposed to be hitting only the walls of the vacuum chamber. We will refer often in the following to this situation as 'UHV conditions'.

In general, for the calculation of pressure profiles and conductances, two different approaches are possible: analytical methods and numerical methods. The former are appealing, as they allow the calculation using formulae, a solution which eases the exploration of the space of different configurations of the vacuum system under study. For example, one could implement one or more formulae into a spreadsheet and then change some parameters such as the spacing of the pumps, their size, or the geometry of the chamber, until the specified design performance is reached. Unfortunately, there are no universal formulae which could be used. Only very simple geometries can be tackled analytically in a correct way, but a real life accelerator often departs from this case.

The other approach, employing numerical methods, is more widespread, although this often makes use of existing formulae, derived analytically. One variant of the numerical approach, the Monte Carlo (MC) method, directly implements the formulae of the kinetic theory of gases, with no simplifications. Its major drawback is that it is CPU-intensive.

With few exceptions, the vacuum systems of particle accelerators fall under the category of conductance limited systems. In order to clarify what this means, the basic concept of conductance is revisited, using both analytic tools and numerical ones.

We then move on to review the different ways of solving the basic equations, mass-flow balance equations, which describe the behaviour of gases inside a vacuum system. Different ways of solving such equations are given, with particular emphasis on their applicability.

A rather large body of literature has been used to cover all possible aspects of the subject. Several references are given throughout the text: they should be consulted carefully and critically and used as a guidance for the implementation of vacuum calculations.

2 Conductance calculations

Much has been written concerning the concept of conductance [1–3]. Alongside with the basic equation of UHV technology

$$P = \frac{Q}{S} \tag{1}$$

it is one of the primary concepts through which many other interesting quantities can be calculated, such as the effective pumping speed and the average pressure.

Conductance is defined as the ratio of throughput to pressure differential, under steady conditions, where the pressures refer to two isobaric sections inside the pumping system, as indicated in Fig. 1.

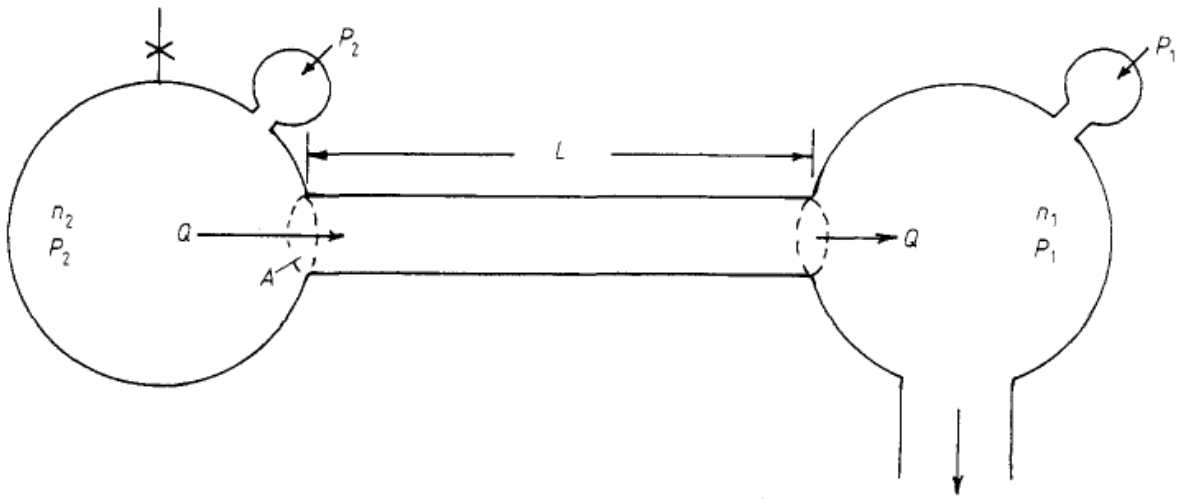


Fig. 1: Experimental set-up for the measurement of the flow resistance (conductance) of a tube [4]

In a less rigorous way, the conductance of a vacuum component is its capacity to let a volume of gas pass from one end to the other, regardless of the pressure differential. It is, basically, a function of the geometry of the system.

2.1 Analytic calculations

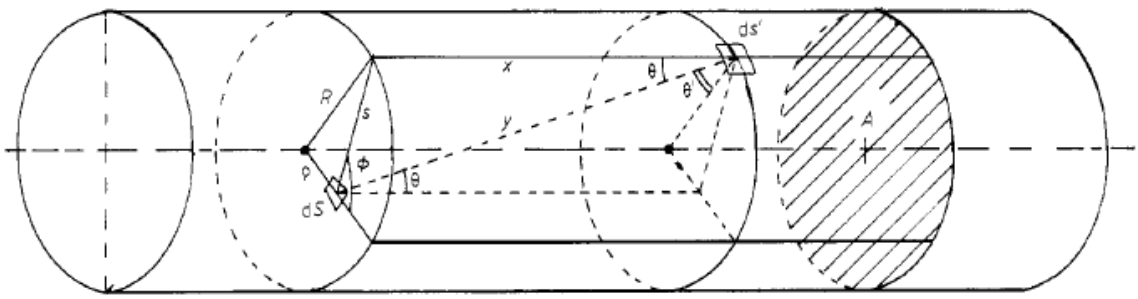
In one of the fundamental papers of vacuum technology [1], Knudsen wrote that the 'flow resistance' Z (1/conductance) of a tube of length L, cross-sectional area A, and perimeter S is given by

$$Z = \frac{3}{8}\sqrt{\pi/2}\int_0^L \frac{S}{A^2}\,\mathrm{d}l \quad . \tag{2}$$

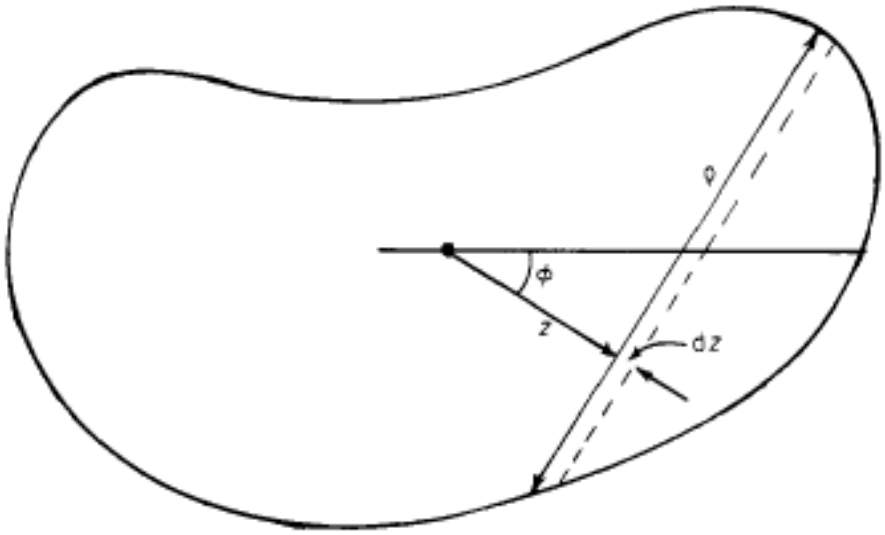
Smoluchovski and Clausing [2,3] realized that Knudsen's formula was correct only in the particular case of a circular cross-section. In all other cases a correction factor had to be introduced, and the simple integral in Eq. (2) changed to the double integral

$$I = \int_S \int_{-\pi/2}^{\pi/2} \frac{1}{2}\rho^2 \cos\theta \,\mathrm{d}\theta\,\mathrm{d}s \quad . \tag{3}$$

Figure 2 (a) and 2 (b) show the definition of the variables.



(a)



(b)

Fig. 2: (a), (b) Definition of variables for the calculation of Smoluchovski's integral [4]

It was shown later that the double integral above can be replaced by [4]

$$I = \pi \cdot \mathrm{avg}\left(\int \rho^2 \, \mathrm{d}z\right) \ . \tag{4}$$

As already pointed out by Clausing [2], for a tube of arbitrary shape it is better to use the concept of transmission probability $P_{\mathrm{i}\to\mathrm{j}}$, i.e., the ratio of the molecules which reach the 'exit' j of the tube once they are injected through its 'entrance' i (whether or not i and j are the real entrance and exit of the tube). The relationship between the conductance C in litre · s^{-1} and $P_{\mathrm{i}\to\mathrm{j}}$ is given by (Ref. [5], p. 86)

$$C = \frac{\bar{v}}{4} \cdot A \cdot P_{\mathrm{i}\to\mathrm{j}} \ . \tag{5}$$

where A is the entrance area in cm^2, and $\bar{v}$ is the average molecular speed in m · s^{-1} given by

$$\bar{v} = \sqrt{8kT/\pi m} = \sqrt{8RT/\pi M} \tag{6}$$

with k = Boltzmann's constant ($k = 1.381 \cdot 10^{-23}$ J · K^{-1}); m = molecular mass (kg); $R = 8.314$ J · mole^{-1}; M = molar weight (kg · mole^{-1}).

The resolution of Eq. (3) becomes difficult as soon as the cross-section of the tube is not as simple as a circle, ellipse or rectangle. A typical vacuum system of an accelerator is a sequence of tubes with different cross-sections, connected together. In this case it would be tempting to think of the total conductance of the system as the one obtained using an electrical circuit analogy of the flow resistance mentioned above Z_{tot}, where the resistance of each section of the tube having a given cross-section is calculated separately assuming that it connects two separate volumes as shown in Fig. 1, and then the reciprocal of these are summed up in order to obtain $1/Z_{tot}$ (Ref. [5], p. 96). This procedure would generate the wrong result, as it would not take into account the so-called beaming effect [2–4], which describes the fact that while the molecules move along the tube from the entrance to the exit, their average velocity vectors become more and more parallel to the axis of the tube. This had been recognized early on by Clausing [2], and re-visited later by several authors [4], as shown in Fig. 3.

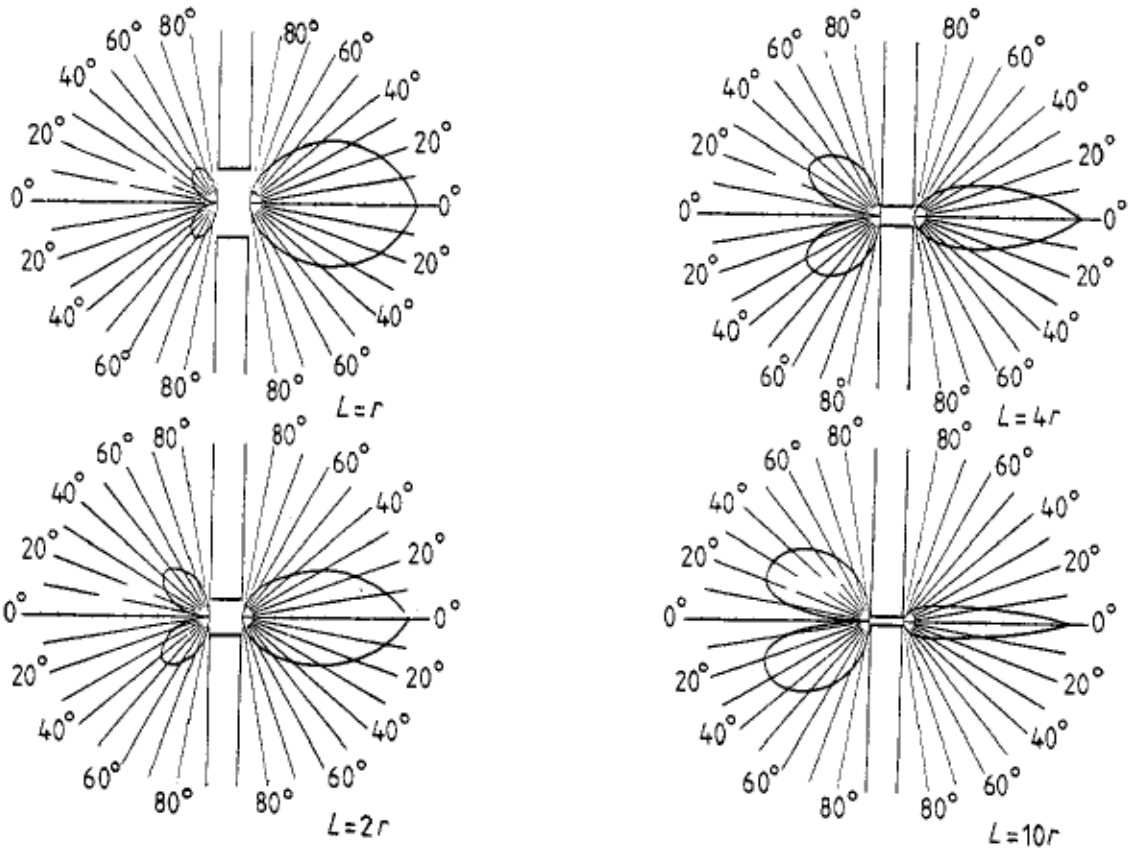


Fig. 3: Beaming effect for circular tubes of different length-to-radius ratios [4,6]

As soon as computers became available for this kind of calculation, they were used to calculate the transmission probability of simple and complex geometries, referring to real vacuum components, such as tubes, elbows, baffles, pumps [7,8]. All that needs to be done is to simulate the microscopic behaviour of the molecules under UHV conditions, with the assumption that their direction of emission follows the cosine law, frequently referred to as Lambert's law, i.e., that a molecule resting on a surface, has a probability $\mathrm{d}n/N$ of being desorbed within the solid angle $\mathrm{d}\omega = \sin\theta \cdot \mathrm{d}\theta \cdot \mathrm{d}\phi$ as shown in Fig. (4), and given by

$$\frac{1}{N}\frac{\mathrm{d}n}{\mathrm{d}\omega} = \frac{\cos\theta}{\pi} \quad . \tag{7}$$

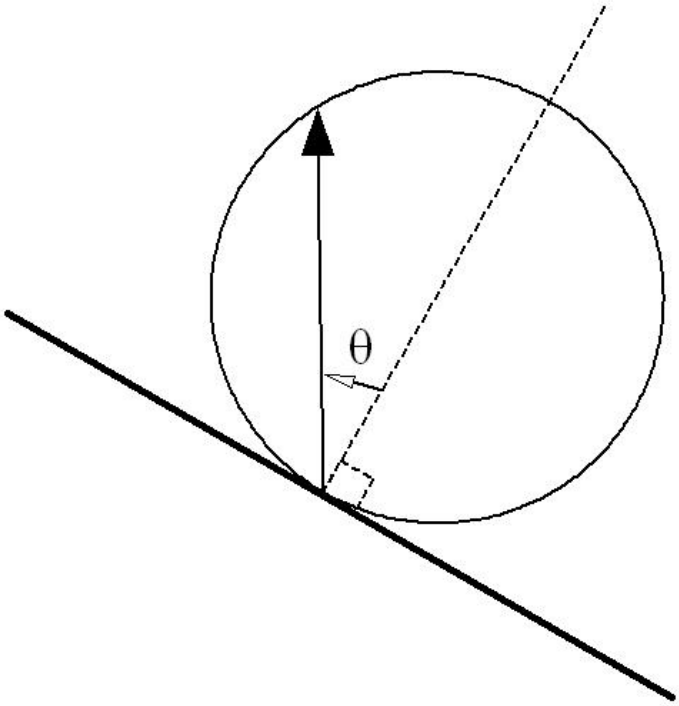
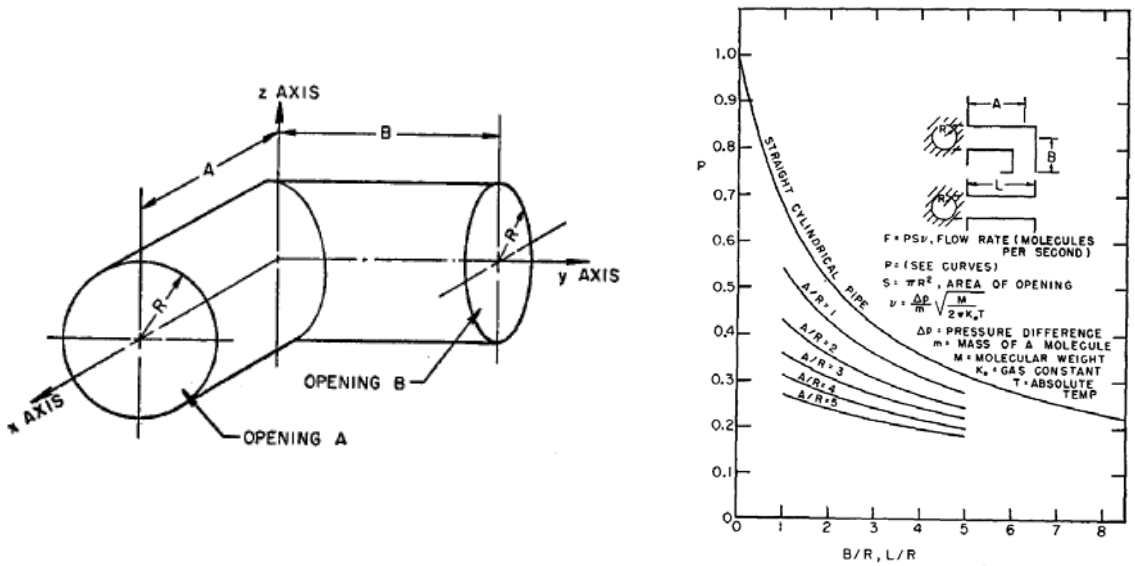


Fig. 4: Definition of the zenith angle of emission θ following Lambert's law. The azimuthal angle φ is a rotation about the axis perpendicular to the plane.

2.2 The Monte Carlo method

The Monte Carlo (MC) method has been used extensively for this purpose, as shown in Ref. [9] and Ref. [10] and references therein. The MC algorithm under UHV conditions, frequently called the Direct Simulation Monte Carlo method (DSMC) generates molecules one by one and tracks them from 'source' to 'sink' [11].

Figures 5 and 6 show examples of such calculations.



(a) **(b)**

Fig. 5: (a) Geometry for the MC simulation of an elbow; (b) its transmission probability [7]

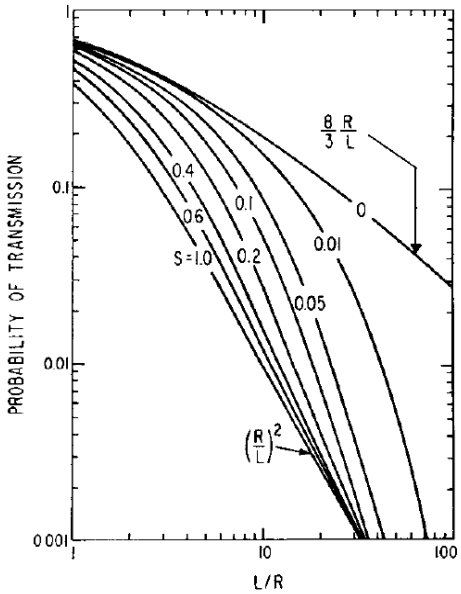
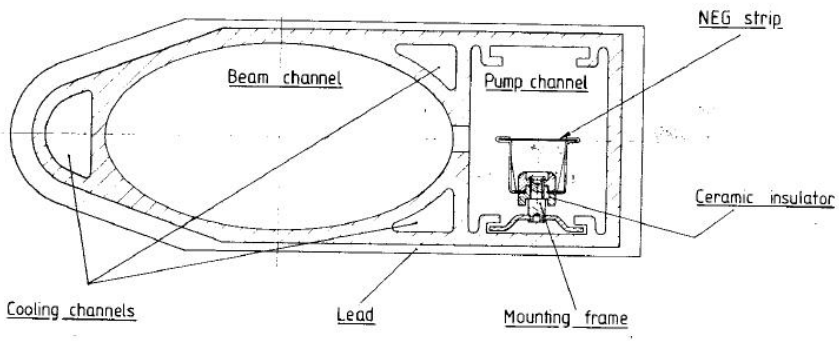
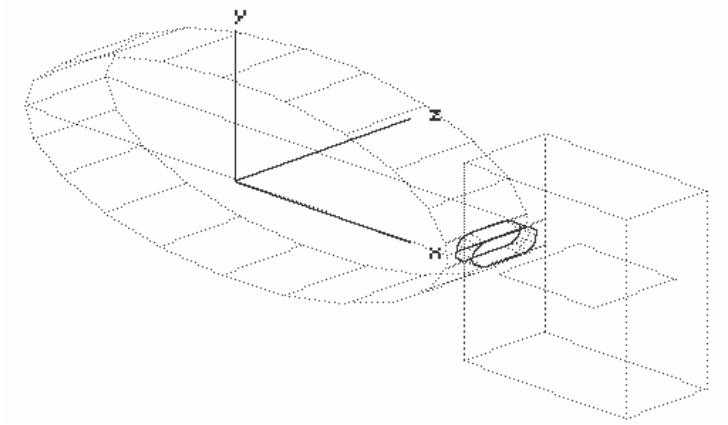


Fig. 6: Transmission probability for circular tubes of length *L* and radius *R* for different values of the sticking coefficient [8]

This can be generalized to any cross-section, such as the one for LEP, as shown in Fig. 7 [12] and modelled in Fig. 8 [10].



(a)



(b)

Fig. 7: (a) Cross-section of the dipole LEP chamber with NEG strip [12]; (b) Twenty-five mm long slice of the LEP chamber, with 20×9 mm^2 racetrack-shaped slot between chamber and antechamber [10,12].

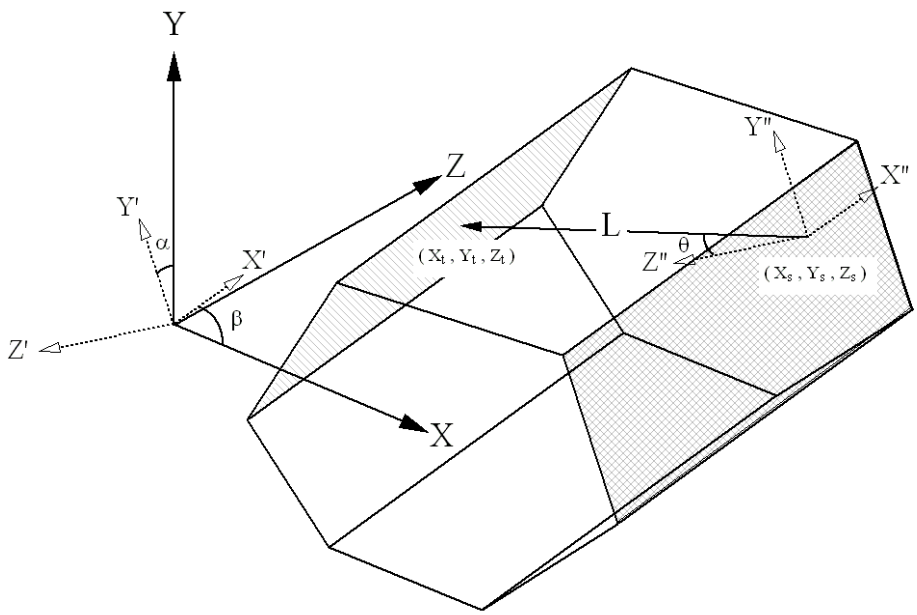


Fig. 8: Definition of some geometrical variables used for making a 3D MC simulation [10]

A three-dimensional (3D) description of the vacuum component is done. The easiest way is to make it up by using planar facets (polygons) as done in Refs. [10,11].

In this case, the ray-tracing calculation goes as follows, see Fig. 8:

- let *XYZ* be an arbitrary frame of reference;
- let *X''Y''Z''* be the frame of reference whose origin corresponds to the the position of the molecules on a facet, with *Z''* perpendicular to the facet (normal vector);
- let *X'Y'Z'* be a frame of reference parallel to *X''Y''Z''*, with the same origin as *XYZ*;
- let β be the rotation angle about the *Y* axis which takes the *X* axis onto *X'*, and α the one about *X'* which takes *Z* onto *Z'*.

The following transformation equations can be written

$$\begin{pmatrix} X'' \\ Y'' \\ Z'' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \cdot \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} , \tag{7a}$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \cdot \begin{pmatrix} X'' \\ Y'' \\ Z'' \end{pmatrix} . \tag{7b}$$

A molecule at position (X_s, Y_s, Z_s) ('source') is generated with a desorption (zenith) angle θ following a cosine distribution with respect to Z'' (φ is the azimuthal angle of rotation about *Z''*, not shown on Fig. 8) according to the equations

$$\begin{cases} X'' = L\sin\theta\cos\varphi \\ Y'' = L\sin\theta\sin\varphi \\ Z'' = L\cos\theta \end{cases} . \tag{8}$$

The angle θ is obtained by generating a random number r with uniform distribution in the interval [0,1], and then using the relation [13]

$$\theta = \sin^{-1}(\sqrt{r}) \quad . \tag{9}$$

The molecule will hit another facet at point (X_t, Y_t, Z_t) ('target')

$$\begin{cases} X_t = X_s + L_t(\sin\theta\cos\varphi\cos\beta_s - \sin\theta\sin\varphi\sin\alpha_s\sin\beta_s - \cos\theta\cos\alpha_s\sin\beta_s) \\ Y_t = Y_s + L_t(\sin\theta\sin\varphi\cos\alpha_s - \cos\theta\sin\alpha_s) \\ Z_t = Z_s + L_t(\sin\theta\cos\varphi\sin\beta_s + \sin\theta\sin\varphi\sin\alpha_s\cos\beta_s + \cos\theta\cos\alpha_s\cos\beta_s) \end{cases} \tag{10}$$

with L_t given by

$$\begin{cases} L_t = -(A_t X_s + B_t Y_s + C_t Z_s + D_t)/U \\ U = A_t(\sin\theta\cos\varphi\cos\beta_s - \sin\theta\sin\varphi\sin\alpha_s\sin\beta_s - \cos\theta\cos\alpha_s\sin\beta_s) + \\ \quad + B_t(\sin\theta\sin\varphi\cos\alpha_s - \cos\theta\sin\alpha_s) + \\ \quad + C_t(\sin\theta\cos\varphi\sin\beta_s + \sin\theta\sin\varphi\sin\alpha_s\cos\beta_s + \cos\theta\cos\alpha_s\cos\beta_s) \end{cases} \quad . \tag{11}$$

With A_t, B_t, C_t and D_t the coefficients defining the plane $A_tX + B_tY + C_tZ + D_t = 0$ which contains the polygon defining facet t. The MC code is such that it calculates L_t for each facet defining the model of the vacuum system except the source one, and then takes as the good L_t the one corresponding to the shortest positive L_t. It then repeats the procedure until a molecule is adsorbed by a surface: this is accomplished by assigning to each facet a sticking coefficient corresponding to the pumping speed of that facet. The sticking coefficient is the probability that a molecule hitting a facet will be permanently pumped. The relationship between sticking coefficient s and pumping speed S (l/s) of a facet is given by Eq. (12)

$$S = s \cdot A \cdot \frac{\overline{v}}{4} \tag{12}$$

which has the same form as Eq. (5), with the sticking coefficient s taking the place of the transmission probability $P_{i\to j}$ and the pumping speed S the place of the conductance C. A 1 cm^2 surface with $s = 1$, i.e., a perfect pump, has a pumping speed S of 11.77 l/s/cm^2 for N_2 (or CO) at a temperature of 20°C.

The transmission probability w follows a binomial distribution with a standard deviation (N = number of generated molecules) [10, 14,15]

$$\sigma = \sqrt{\frac{w \cdot (1-w)}{N}} \quad . \tag{13}$$

The conductance C of a tube of any cross-section is obtained by calculating w using the DSMC method, and then inserting it in place of $P_{i\to j}$ into Eq. (5). The standard deviation of C follows immediately from that of w.

For a MC simulation to be effective, it is of paramount importance to use an appropriate random number generator (RNG). It must have a suitably large period, in order to avoid that the generated number be correlated, thus skewing the results. For some high-precision simulations, where billions of molecules need to be generated in order to reduce the statistical error, the RNGs can have periods as large as 2^{19937}~10^{6000} [14,15].

2.3 Comparisons

References [4,5,9] can be used in order to check the relative accuracy of the different methods discussed so far. In general, for tubes of constant circular cross-section, most methods agree to a few per cent, with notable deviations when the length-to-radius ratio becomes small. For chambers with complicated cross-sections (chamber–antechamber, coaxial tubes, triangular or rectangular profiles) the deviations become more and more pronounced. See Section 3.9 for further comments on this.

3 Pressure profile calculations

After having reviewed the concept of conductance and some of the methods which can be used to calculate it for a system of given geometry, we can now move on to look at the different ways of calculating pressure profiles.

Since most accelerators or accelerator components (e.g., beamlines) have one dimension which is much bigger than the two others (circumference or beamline length vs cross-section of the chamber), a good starting point is to consider a uni-dimensional mass-balance equation of the type

$$V\frac{\mathrm{d}P(x,t)}{\mathrm{d}t} = Q(x,t) - S(x,t)\cdot P(x,t) + c(x,t)\cdot\frac{\mathrm{d}^2P(x,t)}{\mathrm{d}x^2} \tag{14}$$

where V is the volume, P the pressure, Q the outgassing yield, S the pumping speed, and c the specific conductance in litres·m/s.

Most particle accelerators run under conditions which can be thought of as a steady state: a given beam is injected into the machine, and then circulates in it for times that are normally much longer than the time taken by the molecules to move along the chamber and be pumped. As we have seen in Eq. (6), these average speeds are of the order of several hundred metres per second, and therefore any changes in the conditions of the machine (beam current and energy, beam position) which happen on a time scale longer than, say, a few seconds, can be considered as stationary from the point of view of pressure. The time evolution of the pressure can therefore be neglected, thus simplifying the analysis. Equation (14) then becomes

$$c(x)\cdot\frac{\mathrm{d}^2P(x)}{\mathrm{d}x^2} - S(x)\cdot P(x) = -Q(x)\ . \tag{15}$$

In the literature, it is solutions to this equation that are found most often. Some of them are obtained analytically, others numerically.

Looking at Eq. (15) it becomes clear now why we have discussed the conductance before coming to this point. Together with the pumping speed, it is the conductance which determines the shape of $P(x)$. We will see that S could in principle be made very large (at least locally), while c has an upper bound due to the physical dimension of the accelerator, mainly the physical aperture of the magnets. This fact is often referred to by saying that accelerators are conductance-limited systems.

3.1 Analytic solutions

When c and S are constants, a solution can easily be obtained by integrating the second-order differential equation (15).

The simplest model to start with is the one represented by a vacuum chamber of constant cross section, with a uniform outgassing yield q (in mbar · l/s/cm^2), a number of lumped pumps equally spaced L metres apart, and no distributed pumping. It can be thought of as a periodic system, with

period equal to the spacing of the pumps. Let A be the specific surface of the vacuum chamber A, in cm^2/m, as indicated by the example shown in Fig. 9.

$$c\frac{d^2P(x)}{dx^2} = -Aq \quad . \tag{16}$$

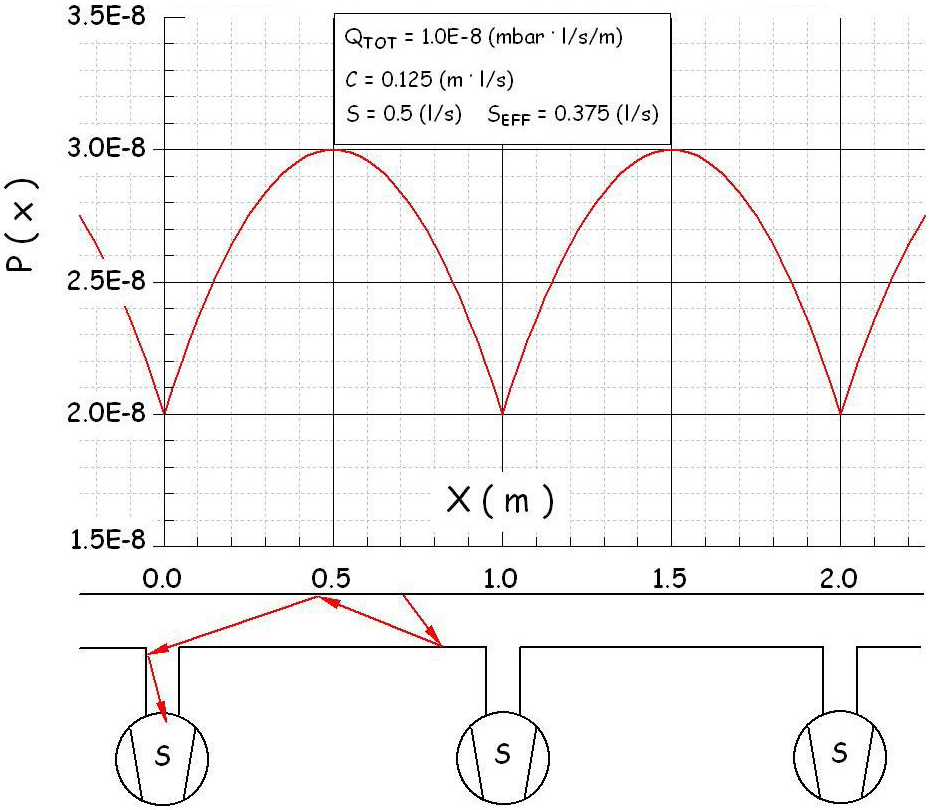


Fig. 9: Pressure profile for a periodic system

By symmetry, the boundary conditions are

$$\begin{cases} \dfrac{dP}{dx}(x = L/2) = 0 \\ P(x = 0) = AqL/S \end{cases} \quad . \tag{17}$$

The solution is

$$P(x) = \frac{Aq}{2c}(Lx - x^2) + \frac{AqL}{S} \quad . \tag{18}$$

The average and the maximum pressures, which can be relevant for the operation of the accelerator, are given by

$$\begin{cases} P_{avg} = \dfrac{1}{L}\displaystyle\int_0^L P(x)dx = AqL(\frac{L}{12c} + \frac{1}{S}) = AqL(\frac{1}{S_{eff}}) \\ P_{max} = AqL(\dfrac{L}{8c} + \dfrac{1}{S}) \end{cases} , \tag{19}$$

with S_{eff}, the effective pumping speed, given by

$$S_{\text{eff}} = \left(\frac{L}{12c} + \frac{1}{S}\right)^{-1} \quad . \tag{20}$$

It now becomes evident why accelerators are said to be conductance-limited systems. There is no point in increasing the pumping speed S beyond few times c/L, since even for a very large S the effective pumping speed would be limited to $12 \cdot c/L$, as indicated in Fig. 10.

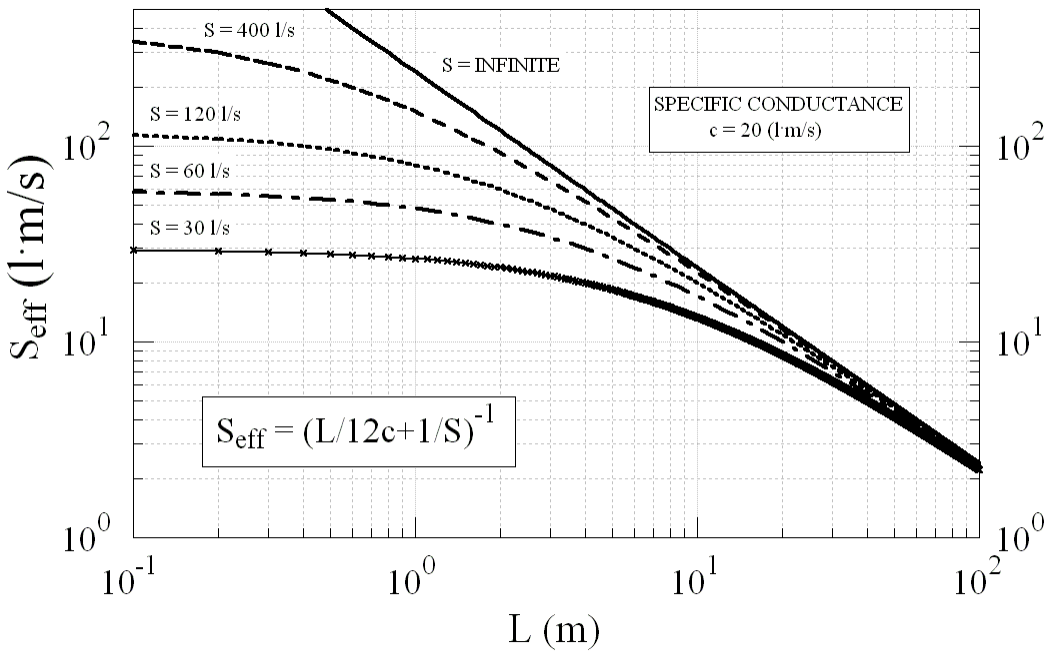


Fig. 10: Effective pumping speed vs pump spacing, for selected values of the specific conductance

As an example, let us consider the case of the ESRF: at a nominal beam current $I = 200$ mA, and energy $E = 6$ GeV, the total photon flux F is

$$F = 8.08 \cdot 10^{17} \cdot E \cdot I \qquad \text{(ph/s)} \ . \tag{21}$$

For a well-conditioned system, a value of the photodesorption yield η of 10^{-6} mol/photon can be reasonably assumed [16], which gives a total outgassing rate of $Q = F \cdot \eta \cdot k$, where $k = 4.047 \times \times 10^{-20}$ mbar · l/molecule, or $Q = 3.53 \cdot 10^{-5}$ mbar · l/s. Using Eq. (1), we obtain S as a function of Q and P, the measured average pressure. For the ESRF, a typical average pressure during operation is $1.5 \cdot 10^{-9}$ mbar, which yields $S = 3.53 \cdot 10^{-5}/1.5 \cdot 10^{-9} \approx 23\ 500$ l/s. This would be the total pumping speed to install along the storage ring were not the system conductance limited. The real installed pumping speed, on the other hand, is ~ 4700 l/s per standard cell. Since there are 32 cells in the ESRF, the total goes to ~ 150 000 l/s, or ~ 6.5 times the value calculated by using Eq. (1).

If either c, Q, S, or L are not constant, then 'piece-wise' solutions can be found [16]. In this case the pressure profile looks like a series of parabolae. It should be kept in mind that this solution is only approximate, since it does not take into account the effect of the changes of the cross sections.

If the distributed pumping speed is not zero throughout the system, then Eq. (15) becomes ($Aq = Q$, specific gas load in mbar · l/s/m)

$$c\frac{\mathrm{d}^2P(x)}{\mathrm{d}x^2}+Q-SP=0 \quad . \tag{22}$$

Its solution is given by

$$P(x)=\frac{Q}{S}+Ae^{\sqrt{\alpha}}+Be^{-\sqrt{\alpha}} \tag{23}$$

where $\alpha = S/c$. The solution is of an hyperbolic sine/cosine type. The boundary conditions between areas with distributed pumping and areas with lumped pumping, or no pumping at all, are given by the continuity of $P(x)$ and its first derivative $\mathrm{d}P/\mathrm{d}x$ at the boundaries. If the vacuum chamber is 'sliced' into N sections, and each section has its own c_i, Q_i, S_i and P_i, then the boundary conditions give $2N$ equations with $2N$ unknowns, which can be solved.

As an example, taken from Ref. [16], the pressure profile along one cell of a particle accelerator, calculated using this method, is shown in Fig. 11.

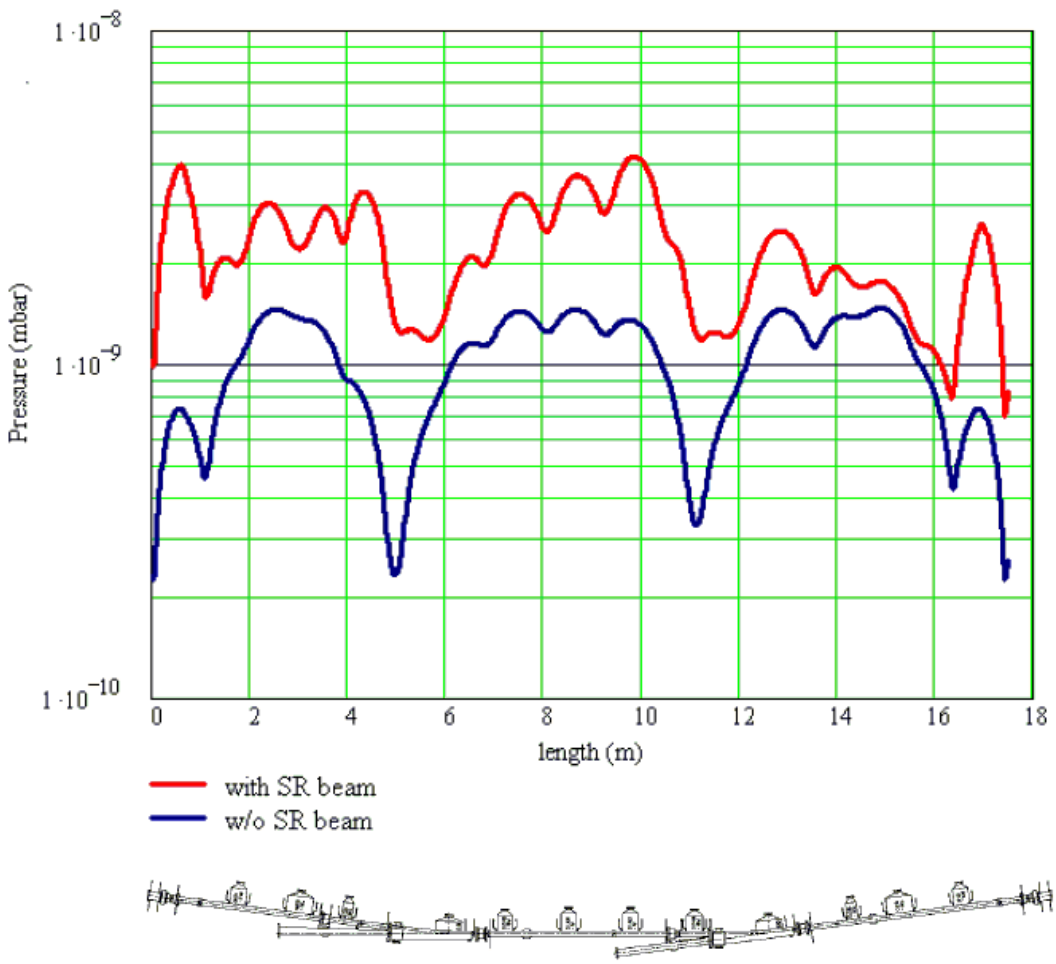


Fig. 11: Pressure profile along one cell of Diamond [16]

3.2 Transfer-matrix method

Under the assumption that c, S and Q are piece-wise constants, Eq. (22) and its solution Eq. (23) can be solved by using algorithms which are implemented in machine lattice codes such as

TRANSPORT [17], where the particle beam is tracked through the magnetic lattice. In this case, a corresponding 'vacuum lattice' is thought of [18,19], which obeys equations of the type

$$\begin{pmatrix} P(L) \\ \mathrm{d}P(L)/\mathrm{d}z \\ 1 \end{pmatrix} = \begin{pmatrix} \cosh(\alpha L) & \dfrac{\sinh(\alpha L)}{\alpha} & -\dfrac{q}{c}\dfrac{\cosh(\alpha L)-1}{\alpha^2} \\ \alpha\sinh(\alpha L) & \cosh(\alpha L) & -\dfrac{q}{c}\dfrac{\sinh(\alpha L)}{\alpha} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} P_0 \\ \mathrm{d}P_0/\mathrm{d}z \\ 1 \end{pmatrix} \tag{24}$$

with c, s and α as defined above. This equation maps the initial pressure and gradient, P_0 and $\mathrm{d}P_0/\mathrm{d}x$, to their values after a length $z = L$.

An example is shown in Fig. 12, taken from Ref. [18].

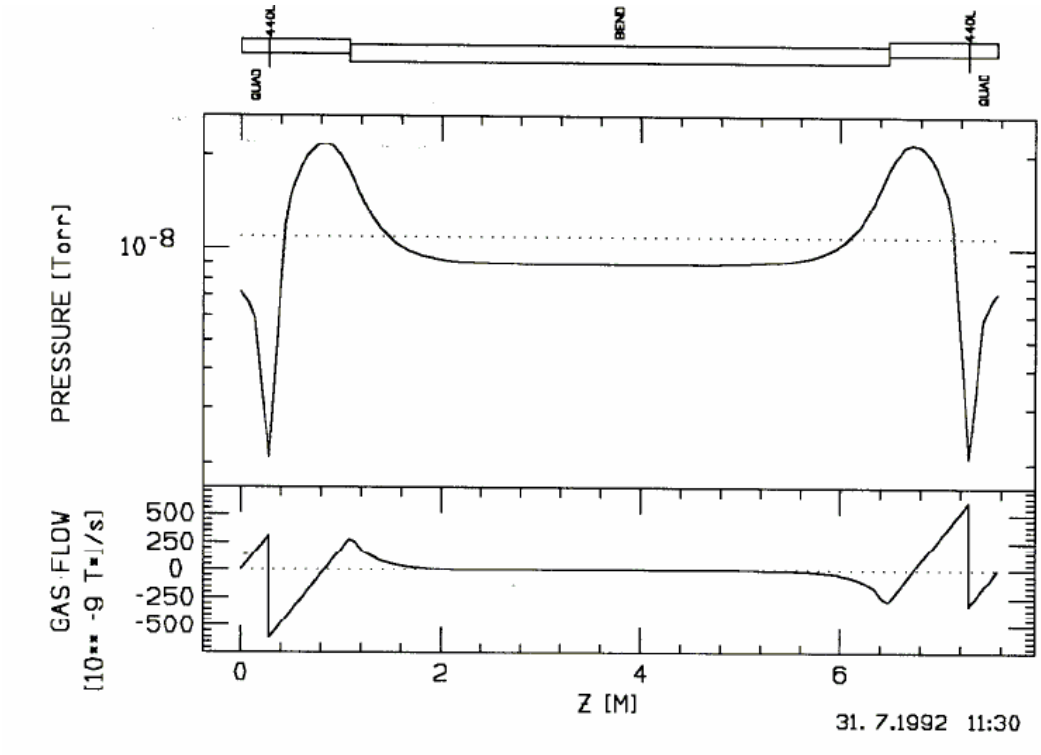


Fig. 12: Pressure profile and its gradient calculated with VAKTRAK [18]

A similar approach has been used in Ref. [20], where Eq. (22) is re-written as

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(c\frac{\mathrm{d}P}{\mathrm{d}x}\right) - SP + Q = 0 \ . \tag{25}$$

In this case c, S and Q can vary. A solving technique based on the finite-difference method is used, after discretization of the system (slicing) into N parts

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(c_{\mathrm{i}}\frac{\mathrm{d}P}{\mathrm{d}x}\right) = \frac{(c_{\mathrm{i+1}} - c_{\mathrm{i-1}})P_{\mathrm{i+1}} + (c_{\mathrm{i}} - c_{\mathrm{i-1}})P_{\mathrm{i-1}}}{2\Delta x^2} - \frac{(c_{\mathrm{i+1}} - c_{\mathrm{i-1}} + 2c_{\mathrm{i}})P_{\mathrm{i}}}{2\Delta x^2} \ . \tag{26}$$

This, together with the boundary conditions

$$Q_i = -c_i \frac{P_{i+1} - P_{i-1}}{2\Delta x^2} \tag{27}$$

becomes a tri-diagonal matrix which is multiplied by the pressure vector Pi. A solution can be found by employing, for instance, a Gaussian elimination and back-substitution technique, an example of which is shown in Fig. 13.

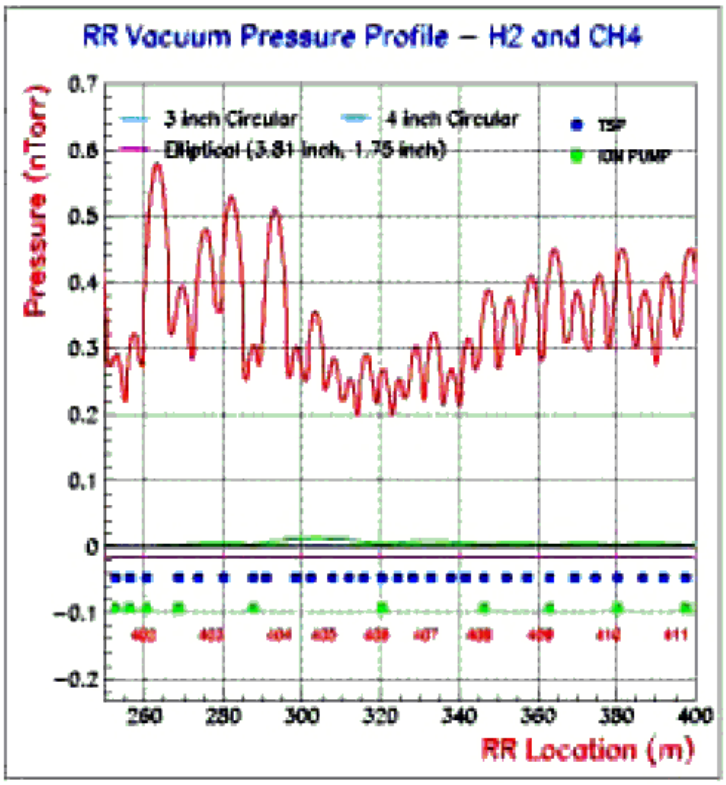


Fig. 13: Pressure profile along FNAL's Recycler Ring [21]

3.3 The continuity principle of gas flow

Another way of solving the mass-flow balance equation is the so-called Continuity Principle of Gas Flow (CpoGF), which can be stated after discretization of the vacuum system as per Fig. 14.

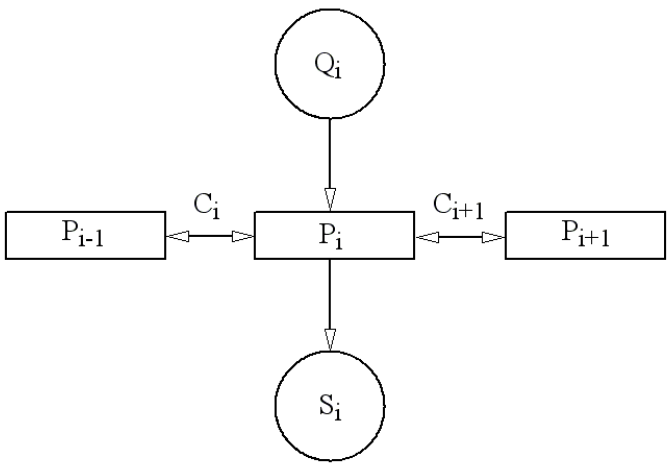


Fig. 14: The continuity principle of gas flow [22]

Each segment of the vacuum system is assigned its S_i, Q_i and C_i, and then its pressure P_i is obtained by solving the set of equations

$$c_i(P_{i-1} - P_i) + c_{i+1}(P_{i+1} - P_i) + Q_i = S_i P_i \quad . \tag{28}$$

Several types of boundary conditions (*BC*) can be thought of [22]:

– Periodic *BC*:

$$\begin{cases} P_0 = P_n \\ P_{n+1} = P_0 \end{cases} \tag{29}$$

which introduced into Eq. (28) gives

$$\begin{cases} C_1(P_n - P_1) + C_2(P_2 - P_1) + Q_1 = S_1 P_1 \\ C_i(P_{i-1} - P_i) + C_{i+1}(P_{i+1} - P_i) + Q_i = S_i P_i \\ C_n(P_{n-1} - P_{n1}) + C_1(P_1 - P_n) + Q_n = S_n P_n \end{cases} \tag{30}$$

where $i = 2, 3, \ldots (n - 1)$.

– Smooth *BC*:

$$\begin{cases} P_0 = P_1 \\ P_{n+1} = P_n \end{cases} \tag{31}$$

which introduced into Eq. (28) gives

$$\begin{cases} C_2(P_2 - P_1) + Q_1 = S_1 P_1 \\ C_i(P_{i-1} - P_i) + C_{i+1}(P_{i+1} - P_i) + Q_i = S_i P_i \\ C_n(P_{n-1} - P_{n1}) + Q_n = S_n P_n \end{cases} \tag{32}$$

where $i = 2, 3, \ldots (n - 1)$.

– Fixed-pressure *BC*:

When P_0 and P_n are known (or set to a given value), Eq. (28) becomes

$$\begin{cases} -C_1 P_1 + C_2(P_2 - P_1) + (Q_1 + C_1 P_0) = S_1 P_1 \\ C_i(P_{i-1} - P_i) + C_{i+1}(P_{i+1} - P_i) + Q_i = S_i P_i \\ C_n(P_{n-1} - P_{n1}) - C_1 P_n + (C_{n+1} P_{n+1} + Q_n) = S_n P_n \end{cases} \tag{33}$$

where $i = 2, 3, \ldots (n - 1)$.

The 3 *BC* cases can be solved by a forward substitution and chase-back technique [22], to obtain the pressure profile of Fig. 15.

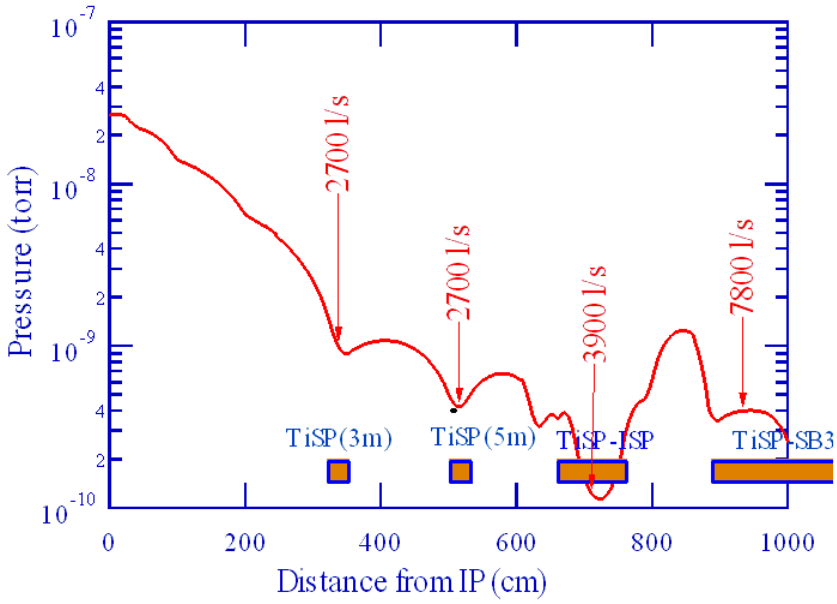


Fig. 15: The CPoGF algorithm applied to the interaction region of CESR-III – CLEO-III. IP = Interaction Point; TiSP = Titanium Sublimation Pump [22]

3.3.1 Commercial programs

Commercial programs have been used by some authors to calculate vacuum pressure distributions: examples are the use of MathCad, for analysing the pressure inside a RF waveguide [23–25].

3.4 Time-dependent analytic calculations

It may be interesting to calculate time-dependent pressure profiles, such as when the effect of gas leaks or a sudden air in-rush need to be calculated. One possible way of solving the problem is to start with the equation [26]

$$c\frac{\partial^2 P(x,t)}{\partial x^2} = -q(x,t) + v\frac{\partial P(x,t)}{\partial t} \tag{34}$$

where v is the specific volume (volume per unit length of the system), and q the specific outgassing yield.

As an example let us consider the case of Ref. [26], depicted in Fig. 16:

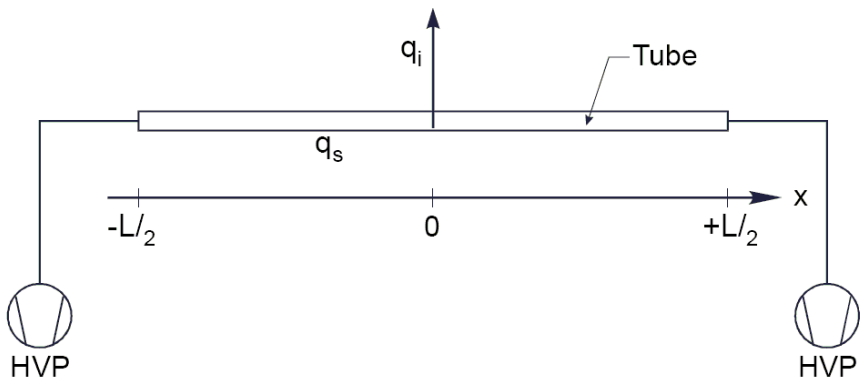


Fig. 16: Tube of length L pumped at the extremities, with a pulsed localized leak q_i in the middle of it, and one high-vacuum pump (HVP) at each end [26]

If the localized pulsed leak q_i is described by the equation $q_i(x,t) = q'\delta(x)\delta(t)$, then Eq. (34) can be solved analytically as

$$P(x,t) = -\frac{q}{2c}x^2 + \frac{Q_T}{2}(\frac{1}{S} + \frac{L}{4c}) + \frac{q'}{\sqrt{4\pi cvt}}\exp(-\frac{v}{4ct}x^2) \quad . \tag{35}$$

Where the first part of the equation is the parabolic pressure profile due to a spatially uniform outgassing yield q [similar to Eq. (18)], and the second part comes from the gas burst q'. With reference to the geometry of Fig. 16, the solution to this equation is shown in Fig. 17.

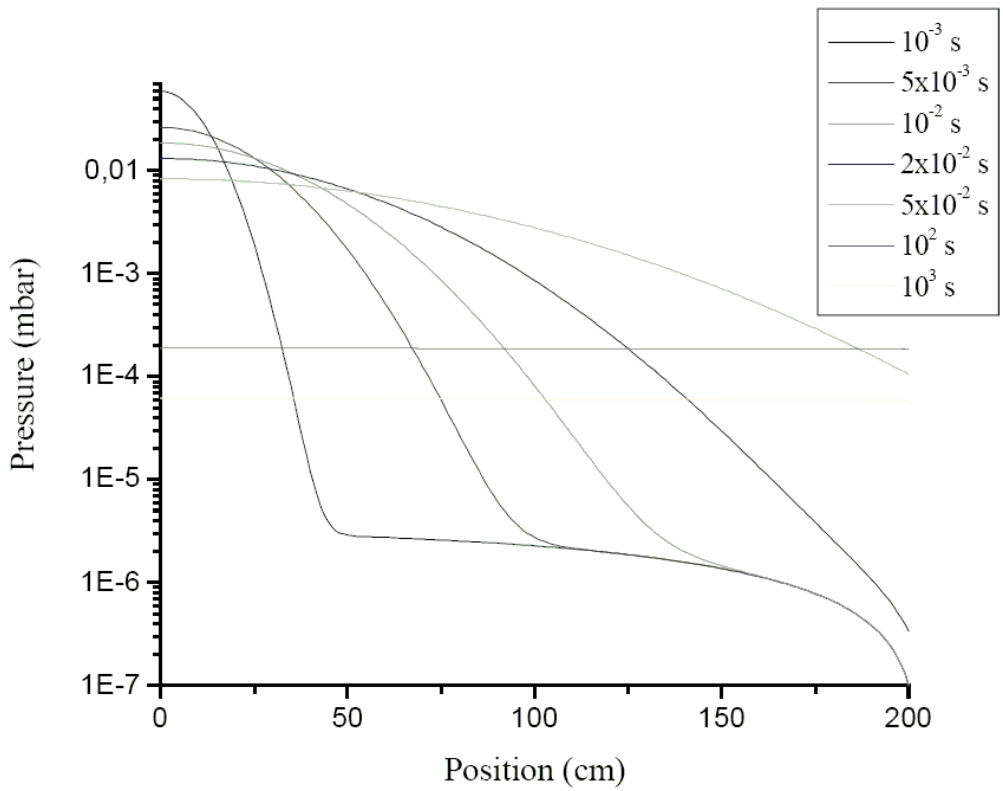


Fig. 17: Solution to Eq. (35) when L = 400 cm, q' = 0.01 mbar · l, q = 4.7 · 10^{-8} mbar · l/s/cm [26]

3.5 The Monte Carlo method

We have already introduced in Section 2 the main features of the MC method, for the calculation of transmission probabilities. There is not much else that needs to be done to use this method for calculating pressure profiles. A geometrical description (model) of the system is made, and then the MC simulation program is started. We have seen that for the calculation of transmission probabilities the molecules are made to desorb from one end of the system, usually a tubular geometry, and the program tracks all molecular trajectories and computes the fraction of the molecules which reach the other end. For the pressure profile calculation, for an accelerator the source of gas is, more often than not, distributed along the chamber in ways that depend on many factors, such as the type of accelerator (synchrotron radiation (SR) light source, high-energy hadron machine, linac, etc.), and the particular situation or mode of operation (fixed energy vs ramped, pulsed machine, gas leak, etc.).

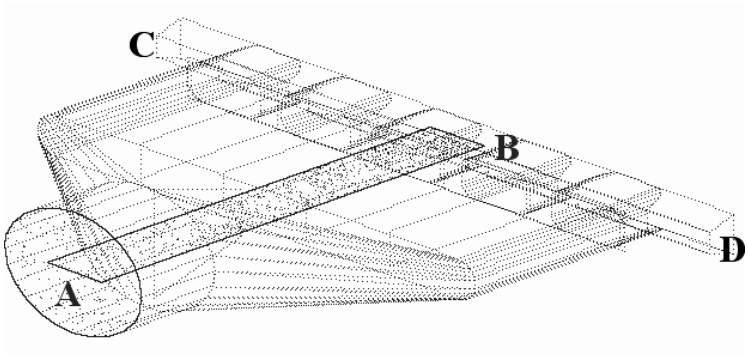
Therefore, it is a pre-requisite to the MC calculation to calculate the desorption profile along the chambers. It should be clear that the accuracy of the result will then depend not only on the statistical accuracy of the MC algorithm, but also on the accuracy of the estimation of the desorption profile. As

an example, for a SR light source one could be tempted, as a first approximation, to set the desorption profile proportional to the photon flux of Eq. (21). This would not take into account the fact that it is not only the instantaneous photon flux which matters, but the integrated photon flux D_γ as well, i.e., how many photons have hit a particular region of the chamber in the past. The instantaneous photon flux F at each point of the vacuum chamber needs to be multiplied by the correct photodesorption yield η, which is a function of D_γ [16]. The calculation of F can be accomplished by performing a ray-tracing analysis, or by using some analytic formulae [27]. Clearly, this needs to be done in close relationship with machine-physics experts, who can advise on the orbit trajectories and properties (emittances, beta functions, dispersion, etc.) [28]. The coefficient η can be obtained by measurement, or extrapolated from data in the literature [16].

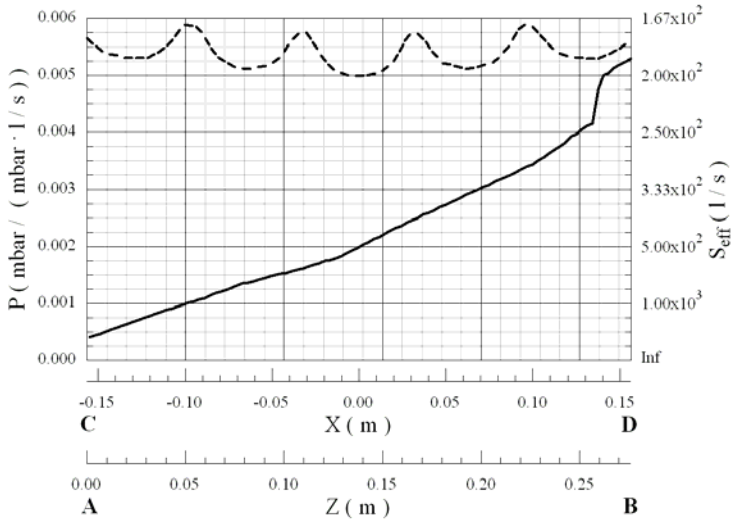
Let us now go back to the MC calculation of pressure profiles. All programs which have been written so far [7–12] work on the common principle of tracking a large number of molecules and keeping a record of all molecular hits on the surfaces making up the model of the vacuum system under study. The advantage of this method as compared to those discussed previously, is that it does not need the prior computation of the conductances or effective pumping speeds. In fact, it will calculate these quantities as a by-product of the calculation of the pressure profiles.

One example will clarify this point. Figure 18 (a) shows the 3D model made with Molflow of a side pumping-port which is located on the dipole chambers of the ESRF. Each chamber has two such pumping ports which stick out of the magnet yoke, between the magnet coils. There is first a narrow rectangular box which is welded on one side to a conical taper that goes into a CF63 flange, where one pump is connected. Both the rectangular box and the conical taper have some internal reinforcing ribs in order to prevent collapse due to external pressure. On the opposite side, near the dipole chamber, there are five rectangular slots machined onto a thick stainless steel bar which constitutes the backbone of the dipole chamber. One pumping port has a 60 l/s ion-pump installed on it, the other one a 200 l/s NEG pump. The MC calculation is done in order to find the conductance of each pumping port, which then yields the effective pumping speeds at the beam chamber. It also gives the pressure drop between the beam chamber and the pumps' inlets. Figure 18 (b) shows this pressure profile (solid line). The circular facet in A, the pump inlet, has a sticking coefficient of 1 (ideal pump), which corresponds to an installed pumping speed of 914 l/s [Eq. (12) for 20ºC CO gas]. The pressure profile is calculated for a unitary outgassing yield: the reciprocal of the vertical scale is therefore the effective pumping speed at the corresponding point on the horizontal axis, via the relation $S = Q/P$ derived from Eq. (1). This means that the effective pumping speed at point B is 1/0.0053 = 188.7 l/s. The pressure at the pump inlet A derived from the graph is 1/0.0004 = 2500 l/s, while it should be 1/914 = 0.00109: the discrepancy is due to the fact that the pressure under molecular flow conditions is not isotropic [9], it depends on the direction of measurement. Figure 19 shows the angular distribution of incident angles at the circular section between the taper and the pump inlet, 53 mm in front of the pump inlet circle. The large departure from the cosine-like distribution is evident. The transparent facet A–B is perpendicular to the average flow of molecules from C–D. The circular facet in A, the one that simulates the 914 l/s pump inlet, gives the correct value for the pumping speed, since it has a pressure of 0.001094 after 1 990 000 molecules have been generated.

One advantage of the MC method is the possibility of changing the microscopic properties of, for instance, the vacuum surfaces [29], and seeing what impact this has on the pressure profile or transmission probability of the system. It allows the exploration of 'what if' scenarios, such as finding out what changes would take place if the theoretical cosine-like desorption from vacuum surfaces were replaced by another angular distribution. Or, it can be used to simulate the desorption from a facet which takes the place of the exit of a tubular structure (like a gas injection manifold, for instance): in this case, the desorption from the facet takes a $\cos^n \theta$ power law, in order to simulate the beaming profile, with the power exponent n extrapolated from data such as those in Fig. 3 or Fig. 19.



(a)



(b)

Fig. 18: (a) Molflow model of one ESRF dipole pumping port; Point A = pump's inlet; B = beam chamber. Molecules are generated along the beam chamber C–D. The rectangular facet A–B is a transparent facet used to keep track of the molecular hits (small dots). (b) Solid line: pressure profile along the pumping port, from A to B, *Z* scale; Dashed line: pressure profile along the beam chamber, from C to D, *X* scale.

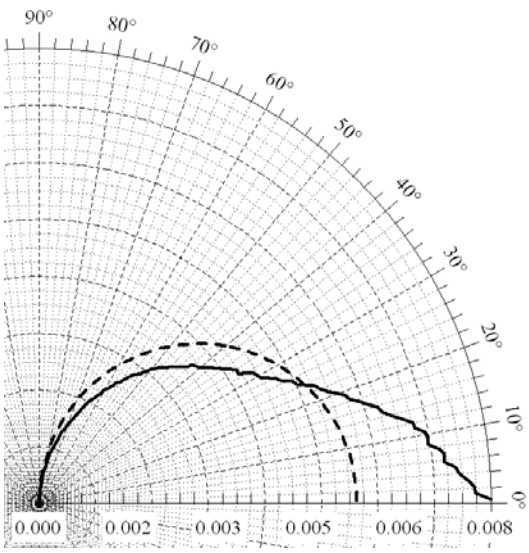


Fig. 19: Angle of incidence profile (solid line) compared to a cosine-like (dashed line). The 0° line is the normal vector to the pump inlet. The beaming effect is evident.

3.5.1 Time-dependent DSMC

There may be interest in simulating the time-dependent behaviour of gas species under UHV conditions. One way to do this is to follow the example set in Ref. [30]. The lengths of the molecular trajectories are translated into time intervals by means of the average molecular velocity given in Eq. (6). A suitable trajectory length step L_0 is chosen, corresponding to a time-step of $\tau_0 = L_0/\bar{v}$. A large number of molecules are generated, to get statistical accuracy, and the position of the molecules at each time-step interval is recorded. Figures 20 (a) and 20 (b) exemplify this with clarity.

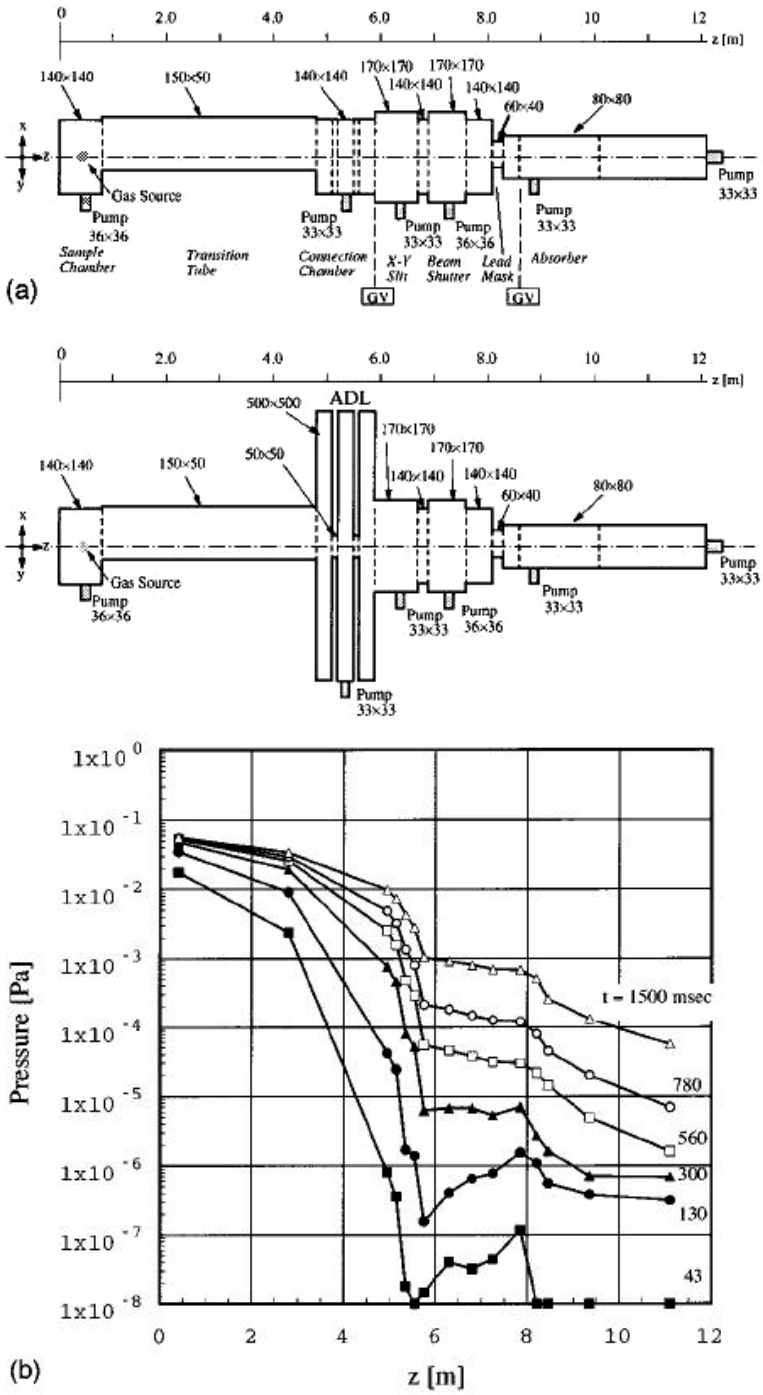


Fig. 20: Time-dependent MC simulation of an acoustic delay line (ADL) [30]; (a): geometry and dimensions of the ADL; (b): time-dependent pressure profiles

3.5.2 Commercial products

Instead of coding the MC ray-tracing algorithm using some high-level languages (Fortran, C, Pascal) [10–12], some authors have left the burden of making the 3D model and the ray-tracing calculations to some commercial programs [25]. In this case, one or more script files have to be created, and then the symbolic processor computational engine takes charge of tracking the molecular trajectories, storing the data, and post-processing them in order to visualize the results. The obvious advantage is that the user does not need to write hundreds of lines of code, the disadvantage is that the software licence can be expensive and the training time to become proficient in using the software can be long.

3.6 Diffusion equation

Several authors [31–33] have used the diffusion equation method in order to calculate both steady-state and time-dependent pressure profiles. The uni-dimensional diffusion equation is of the type [5,9]

$$\frac{\mathrm{d}n}{\mathrm{d}t} = D\frac{\partial^2 n}{\partial x^2} \tag{36}$$

where D is the diffusion coefficient of the gas specie under study, where n, the molecular density has been used instead of the pressure P. The equation $P = nkT$ gives the relationship between these two quantities. The diffusion coefficient D is related to the average molecular velocity and to the mean free path λ through the relationship $D = 1/3 \cdot \bar{v} \cdot \lambda$. Some interesting papers and analyses of vacuum systems of particle accelerators have been made using this method [31–33]. It is useful when the system under study can be thought of as being uni-dimensional. This means that the details of the pressure or density distributions at points where the local geometry of the system is important, cannot be obtained by this method. An example of this could be the geometry of a pumping port, or the side connection of a vacuum gauge to a vacuum chamber. Once solved, Eq. (36) could give the value of P or n at any given point in space and time (x,t), but the relationship between this and the pressure value read on a gauge attached to the chamber via a 'tee' or an elbow would need to be calculated separately, for instance using a MC algorithm.

3.7 Gas dynamics models

It is beyond the scope of this paper, which deals mainly with the case of molecular flows, to discuss the application of gas dynamics methods, which are more suited to describing the flow of gases at small Knudsen numbers. Since a dedicated lecture at this school has covered the subject (see F. Sharipov's lecture), it is just worth while to say that the integro-differential kinetic equation method (Ref. [34] and references therein) could certainly model a vacuum system under UHV conditions.

3.8 The view-factor (VF) or angular coefficients method

It has been realized that there is a mathematical analogy between the laws of molecular flow in systems with surfaces which desorb diffusively (cosine-like) and the radiative exchange of heat between walls at given temperatures [9,16].

For two sufficiently small surfaces, as depicted in Fig. 21 the following relation can be written for the molecular flux incident on any surface i

$$\nu_{\mathrm{inc}} = \sum_{\mathrm{j}=1}^{\mathrm{n}} (\varphi_{\mathrm{i}\to\mathrm{j}} \cdot \nu_{\mathrm{j}}) \tag{37}$$

which leads to the system of algebraic equations

$$\nu_i = \nu_{0i} + (1-\varepsilon_i)\cdot\sum_{j=1}^{n}(\varphi_{i\to j}\cdot\nu_j) = \nu_{0i} + (1-\varepsilon_i)\cdot\nu_{inc} \quad . \tag{38}$$

The coefficients ε_i are taken as sticking probabilities. The $\varphi_{i\to j}$ are the angular coefficients and

$$\nu_{ads} = \varepsilon_i \cdot \sum_{j=1}^{n}(\varphi_{i\to j}\cdot\nu_j) \tag{39}$$

is the adsorbed flux.

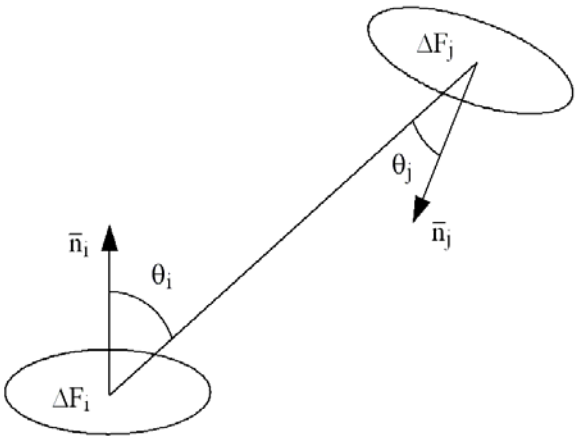


Fig. 21: Basis of the view-factor (VF) method

As an example, Fig. 22 shows the comparison of the calculation of the ratio of the pressure at the inlet and outlet of a tube of length L = 500 mm and diameter 35 mm, obtained by applying two different methods: the CpoGF method and the VF method.

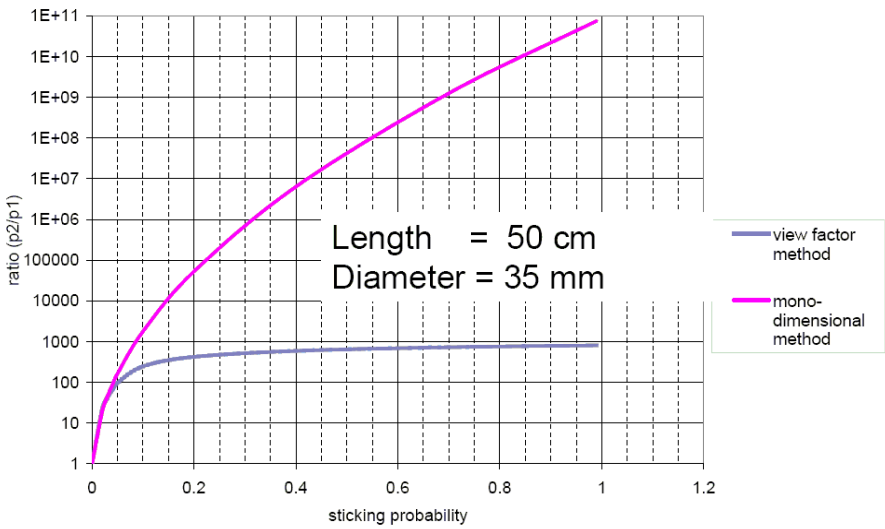


Fig. 22: Comparison between a uni-dimensional analytic method and the VF method [35]

The VF method gives the correct answer, the one that matches the measurements, while the CPoGF method greatly overestimates the pressure ratio at the two ends. This is due to the fact that the latter misses the beaming effect of the molecular flow inside the tube. The two methods agree to some extent only when very small values of the sticking coefficient of the wall are valid, of the order of a few per cent [35].

3.8.1 Commercial programs

Figure 22 has been obtained by applying a proprietary code which implements Eqs. (37–39). Other authors have used some commercial programs which implement the finite-element method (FEM), such as ANSYS, in order to integrate the vacuum and the mechanical calculations all within the same working space [35–37]. In this case, the analogy between radiative heat exchange and molecular flow is such that pressure P is equivalent to the radiation temperature T^4, and the gas flow Q to the radiation heat flow q, through the transformations [36]

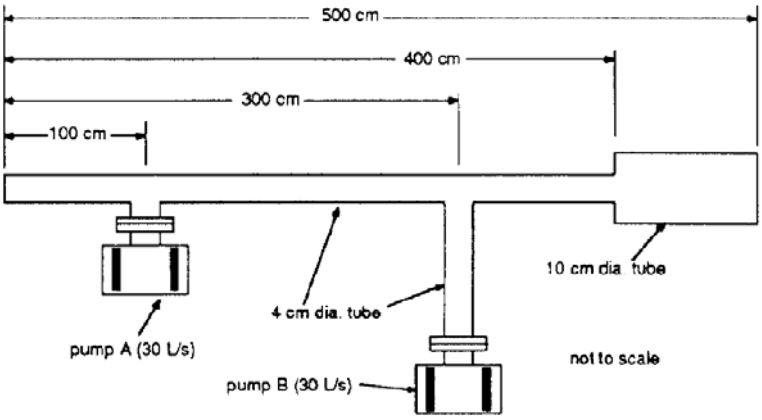
$$\begin{cases} \dfrac{P}{k}\sqrt{\dfrac{R}{2\pi MT}} \Leftarrow \sigma T^4 \\ Q \Leftarrow q \end{cases} . \tag{40}$$

A pump is simulated by transforming the pumping speed S of a pump of area A_p into the 'emissivity' ε_p of the surface simulating the pump

$$\frac{S}{A_p}\sqrt{\frac{R}{2\pi MT}} \Leftarrow \varepsilon_p \quad . \tag{41}$$

The gas conductance in l/s has its counterpart in the thermal conductance in W/°C.

(a)



(b)

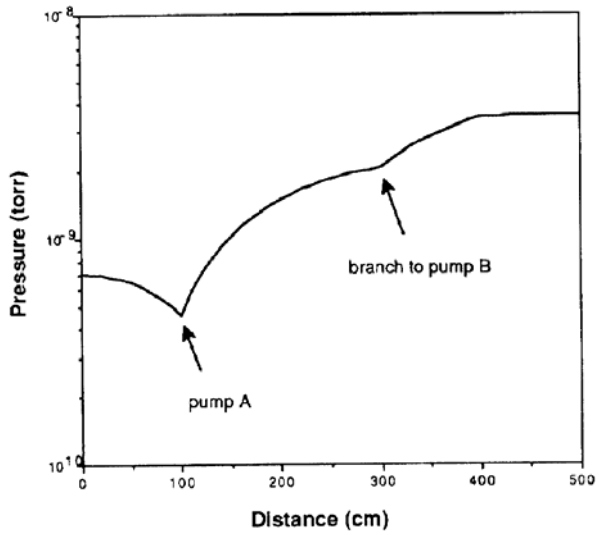


Fig. 23: (a) Geometry of the vacuum system; (b) Pressure profile calculated with ANSYS [37]

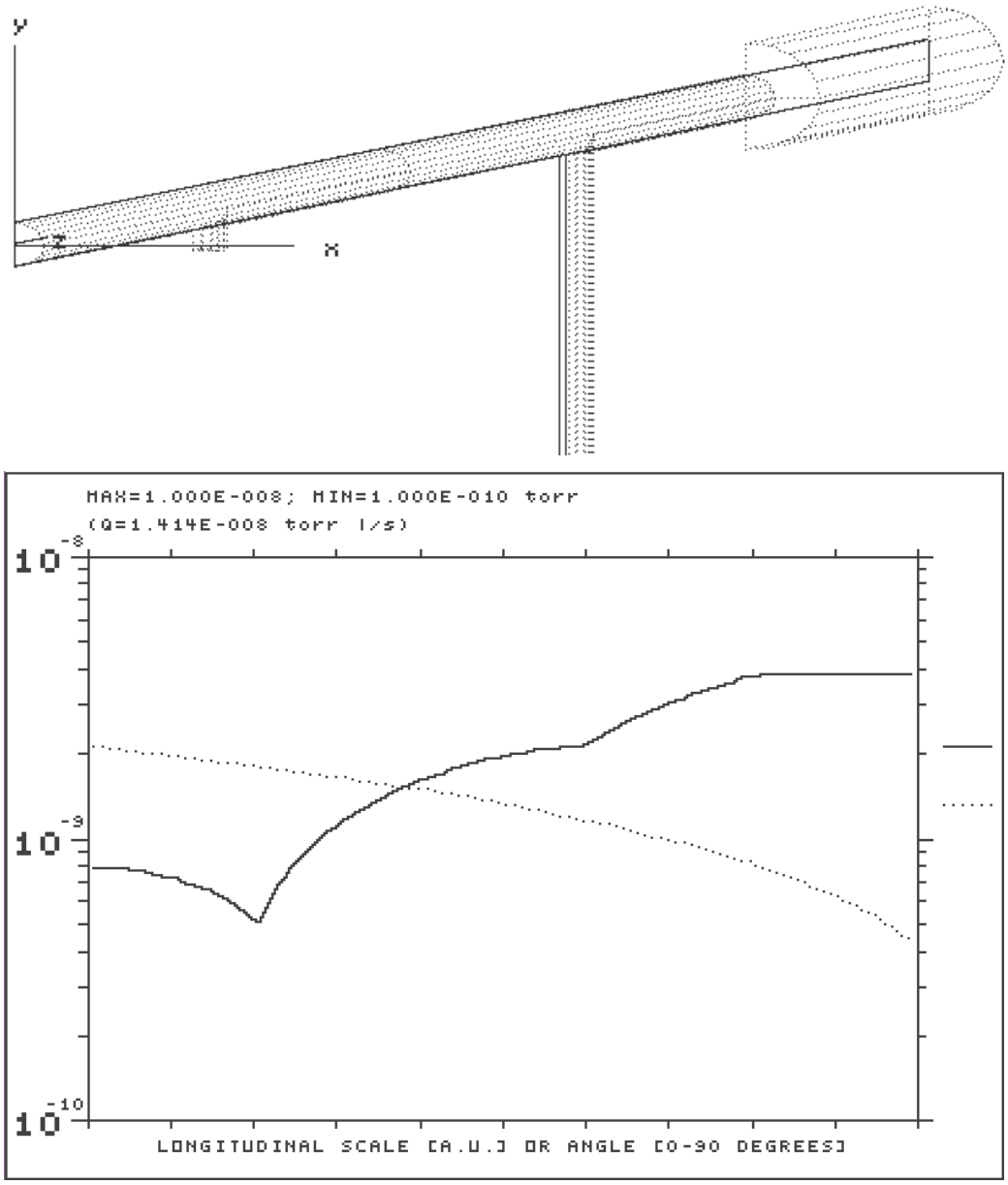


Fig. 24: Molflow model of the vacuum system of Fig. 23. Only ½ of the system has been modelled, using the mirror symmetry along the *YZ* plane. The molecules hitting this plane are mirror reflected rather than being reflected diffusively [10]; Pressure profiles calculated with Molflow: solid line = along the 500 cm long tubes; dotted line = along the 100 cm long vertical branch.

Figure 23 shows an example taken from Ref. [37] where the pressure profile along two circular tubes of different diameter connected in series, with a 'tee' along one of the two and two ion-pumps has been calculated using ANSYS. For comparison, Fig. 24 shows a DSMC calculation done with Molflow on the same system.

3.9 Specific conductance calculation

After reviewing many of the different ways to calculate vacuum pressures and other relevant quantities, we come back to the subject of the calculation of the specific conductance of the vacuum chamber of particle accelerators. As mentioned before, the choice of the cross-section of the vacuum chamber is a very important parameter, which could affect the dimension and field quality of the

magnets, the physical aperture of the beam, the beam lifetime, and so on. We have seen that analytic formulae have been developed, together with numerical methods. We propose here to bridge the two fields, by proposing to determine the specific conductance c by using a DSMC calculation and relating it to one analytical model of the pressure profile, the parabolic pressure profile. Let us consider a tube of constant cross-section, with uniform outgassing rate from the walls, pumped at the two ends by two equal pumps. The DSMC model will transform the pumping speeds at the ends into sticking coefficients, as per Eq. (12). The pressure profile obtained by the DSMC method will look like a parabolic one, with a maximum in the middle, and minima at either end, as indicated by Eq. (18). A second-order polynomial fit of the pressure profile will be of the type

$$P'(x) = A_1 x^2 + A_2 x + A_3 \quad . \tag{42}$$

Therefore,

$$\begin{cases} A_1 = -\dfrac{Aq}{2c} \\ A_2 = \dfrac{AqL}{2c} \\ A_3 = \dfrac{AqL}{S} \end{cases} \rightarrow \begin{cases} c' = \left|\dfrac{Aq}{2A_1}\right| \\ c'' = \dfrac{AqL}{2A_2} \\ S' = \dfrac{AqL}{A_3} \end{cases} , \tag{43}$$

where c', c'', and S' are fitted estimates of c and S.

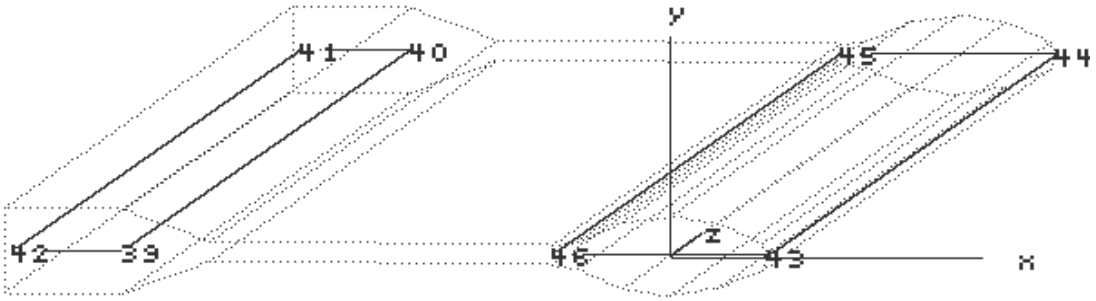
The following example will show how, for the same cross-sections, estimates of c varying by more than 25% can be obtained, depending on the choice of the method. Figure 25 (a) shows the cross-section of a chamber (pseudo-elliptical part) with an antechamber, similar to what has been used for the Advanced Photon Source synchrotron radiation light source [10]. It is approximately 28 cm in width, and 100 cm long, with a narrow slot of 1 cm height and 12.5 cm width. Molflow has been used to perform two DSMC calculations of the system.

- The first one to calculate the transmission probability, i.e., molecules are generated at the entrance on the plane $Z = 0$ and tracked till the exit at $Z = 100$ cm. Both the entrance and exit facets have a unitary sticking coefficient.
- The second one to calculate the pressure profile when all the side-walls are uniformly desorbing. In this case the pressure profile turns out to be parabolic, and it can be fitted to a second-order polynomial. In this case the 'test facet' which keeps record of all molecular hits used to make up the pressure profile can be chosen in different ways. In this case there are four different types: 1) along the beam chamber (points 43–46); 2) along the antechamber (points 39–42); 3) along the slot (not shown, it lies between the first two); 4) across, i.e., over the whole width of the chamber (i.e., the rectangle identified by points 43, 44, 41, 42). Case no. 3 is a sort of average across the chamber.

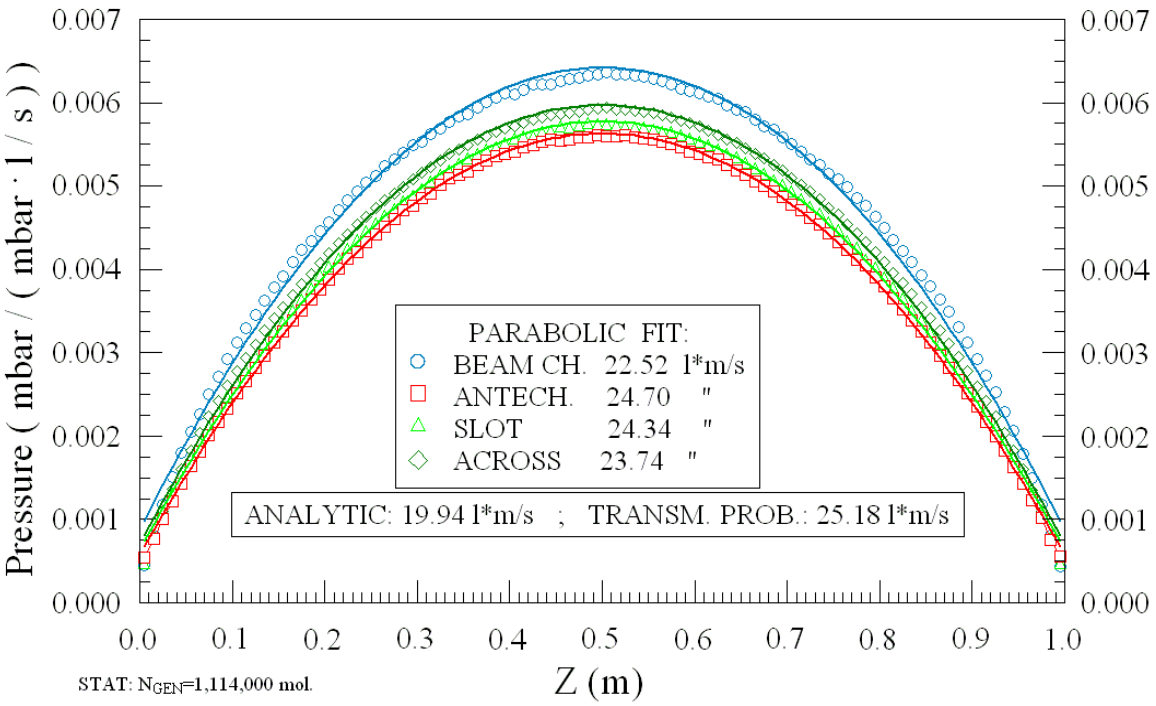
The calculation has also been performed using Knudsen's Eq. (2).

Note the large discrepancy between the value of c obtained by the transmission probability and parabolic fit methods, although they are both based on the same DSMC code. This should come as no surprise, as the former assumes molecular effusion (i.e., $\cos\theta$-like with respect to the axis parallel to Z) and therefore there is a non-negligible probability for molecules of travelling straight through the 1 m long chamber, while the latter assumed $\cos\theta$ desorption from the walls, i.e., perpendicular to the longitudinal Z axis of the chamber.

In light of this result, it is suggested to use the DSMC method with a parabolic fit to the pressure profile curve.



(a)



(b)

Fig. 25: (a) Molflow model of the chamber; (b) Parabolic pressure profiles calculated with Molflow. Note the large discrepancy between the value of *c* obtained by the transmission probability and parabolic fit methods, although they are both based on the same DSMC code.

4 Realistic gas compositions

In the previous sections, we have always made the implicit assumption that there was only one gas species inside the vacuum system. As is well known, this is never the case. Under UHV conditions, it is possible to assume, to a good degree of accuracy, that the total pressure is the sum of the partial pressures [19, 32, 33], which can be calculated as shown before. It is basically Eq. (6) which tells us how the various quantities of interest—pressure, conductance, molecular density and so on—will

change. Lighter gas species will have a higher conductance, all other parameters being the same, as will hotter gas species, through the $\sqrt{T/M}$ relationship.

5 Conclusions

In the previous sections, it has been shown that the recent literature is full of examples of pressure calculations performed by using a number of codes and methods, both analytical and numerical. It is left to the reader to make the appropriate choice for the calculation of pressure profiles and related vacuum quantities.

Many examples have been given, trying to point out the strengths and the weaknesses of each method.

Suggestions have been given about the rational choice for the calculation of the specific conductance of the vacuum chamber of particle accelerators.

Acknowledgements

It is a pleasure for me to acknowledge fruitful discussions with and suggestions by E. Al-Dmour, A. Rossi, B. Mc Donald, L. Rumiz, O. Malyshev and F. Sharipov.

References

[1] M. Knudsen, Ann. Phys. (Leipzig) [4] **28** (1909) 999–1016.

[2] P. Clausing, Ann. Phys. (Leipzig) [5] **12** (1932) 961, reprinted in J. Vac Sci. Technol. 8 (1971) 636–646.

[3] M. Smoluchovski, Ann. Phys. (Leipzig) [4] **33** (1910) 1559.

[4] W. Steckelmacher, Rep. Prog. Phys. **49** (1986) 1083–1107.

[5] J.M. Lafferty (ed.), *Foundations of Vacuum Science and Technology* (John Wiley & Sons, New York, 1998).

[6] B. Dayton, J. Vac. Sci. Technol. **9** (1972) 243–245.

[7] D.H. Davis, J. Appl. Phys. **31** (1960) 1169.

[8] C. G. Smith, G. Lewin, J. Vac. Sci. Technol. **3** (1965) 92.

[9] G. L. Saksaganski, *Molecular Flow in Complex Vacuum Systems* (Gordon and Breach, New York, 1988).

[10] R. Kersevan, Molflow User's Guide, Sincrotrone Trieste Tech. Note ST/M-91/17, 1991.

[11] A. Pace and A. Poncet, Monte Carlo simulations of molecular gas flow: some applications in accelerator vacuum technology using a versatile personal computer program, Proc. IVC-11, Cologne, 1989.

[12] T. Xu, J. M. Laurent and O. Groebner, Monte Carlo simulation of the pressure and the effective pumping speed in the LEP collider, CERN Note CERN-LEP-VA/86-02.

[13] J. Greenwood, Vacuum **67** (2002) 217–222.

[14] J. Gomez-Goni *et al.*, Vacuum **76** (2004) 83–88.

[15] J. Gomez-Goni *et al.*, J. Vac. Sci. Technol. A **21** (2003) 1452–1457.

[16] O. Malyshev, Calculation of Gas Flow Through Complex Vacuum Systems, IOP Seminar, London, October 2005.

[17] K. Brown *et al.*, TRANSPORT Manual, SLAC-91, 1977.

[18] V. Ziemann, SLAC-PUB-5962, October 1992.

[19] J. Kurdal, Proc. 45th IUVSTA Workshop on NEG Coatings, Catania, Italy, April 2006.

[20] M. K. Sullivan, A method for calculating pressure profiles in vacuum pipes, SLAC Note, 1993.

[21] K. Gounder *et al*., Residual gas pressure profile in the recycler ring, Proc. PAC 2003, Portland, USA.

[22] Y. Li *et al*., Calculation of pressure profiles in the CESR hardbend and IR regions, Proc. Int. Workshop on Performance and Improvement of e^-e^+ Collider Particle Factories, Tsukuba, 1999.

[23] P. Ladd, SNS, private communication.

[24] R. Kersevan, presentation slides, downloadable at CAS web site: http://cas.web.cern.ch/cas/Spain-2006/Spain-lectures.htm

[25] G. B. Bowden, RF accelerator pressure profile, Linear Collider Collaboration Tech. Note LCC-078, Stanford, 2002.

[26] F.T. Degasperi *et al*, Proc. PAC 2001, Chicago, USA.

[27] B. Plan *et al*., Blue Book, ESRF Note, 2003.

[28] G. K. Green, Spectra and optics of synchrotron radiation, Note BNL 50522, 1976.

[29] D. H. Davis *et al*., J. Appl. Phys. **35** (1964) 529–532.

[30] Y. Suetsugu, J. Vac. Sci. Technol. A **14** (1996) 245–250.

[31] W. C. Turner, Proc. PAC 1993, Washington D.C., USA.

[32] A. Rossi, VASCO: multi-gas code to calculate gas density profile in a UHV system, unpublished CERN note, private communication.

[33] C. Baumgarten *et al*., Eur. Phys. J. **D18** (2002) 37–49.

[34] F. Sharipov *et al*., J. Vac. Sci. Technol. A **20** (2002) 814–822.

[35] A. Bonucci, In-situ characterization of NEG pipe coatings: the transmission factor method, Proc. 45th IUVSTA Workshop on NEG Coatings, Catania, Italy, April 2006.

[36] M. Lafontaine *et al.*, Modeling photoresist outgassing pressure distribution using a finite-element method, ANSYS Conference, 2002.

[37] J. Howell *et al*., Proc. PAC 1991, p. 2297.

Getter pumping

C. Benvenuti
R&B Energy Research, 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 currently 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 πr_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 5×10^6 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

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

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

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

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

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

1.3 Surface adsorption capacity

The monolayer capacity of an atomically flat surface is of the order of 5×10^{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 7×10^{-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* ν is

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

where ν_0 is an attempt frequency, of the order of the vibrational energy of the adsorbed molecules $\left(\nu_0 \cong 10^{13}\ \text{s}^{-1}\right)$ 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 τ is much larger than the duration of the experiment. Typical values of τ are given in Table 1 [1]. Large τ 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 adsorption energy, assuming $\tau_0 = 10^{-13}$ s

Energy (kcal/mol)	**Typical cases**	τ_a **(s)**
0.1	He	1.2×10^{-13}
1.5	H_2 physisorbed	1.3×10^{-12}
3.5–4	Ar, CO, N_2, CO_2, (physisorbed)	1×10^{-11}
10–15	Weak chemisorption	3×10^{-6}
	Organics physisorbed	2×10^{-2}
20	H_2 chemisorbed	100
25		6×10^5 (1 week)
30	CO chemisorbed on Ni	4×10^9 (> 100 yr)
40		1×10^{17} ($\approx$ age of the Earth)
150	O chemisorbed on W	10^{1100} ($\approx 10^{1090}$ 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, characterized 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, characterized 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 oxidized 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 localized 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 Ref. [7].

The performance of a given NEG is characterized 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 (Table 2), which are the best suited for UHV applications.

Table 2

	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*)	$\log(p_{H_2}) = 4.28 + 2\log(q) - 7000/T$	$\log(p_{H_2}) = 5.14 + 2\log(q) - 6250/T$

*) Pressure p in mbar, H_2 quantity q in mbar ℓ g^{-1}, T in degrees 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: Owing 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 was selected to provide the main pumping for the Large Electron–Positron Collider (LEP) [13] (see Fig. 1).

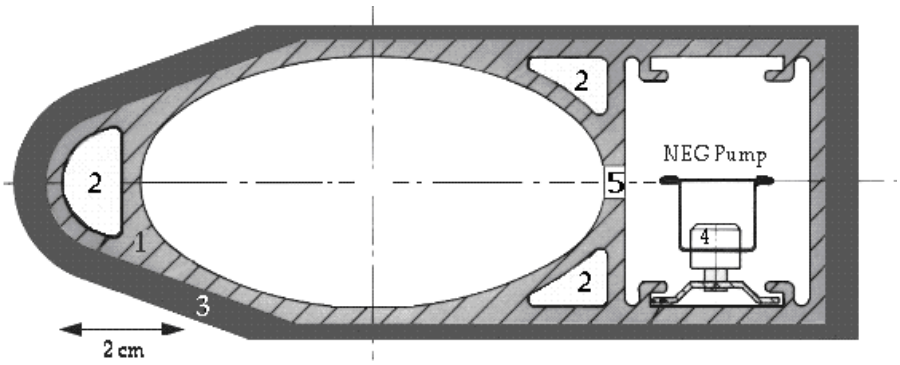


Fig. 1: Cross-section of the LEP dipole vacuum chamber

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 2×10^{-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 5×10^{-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].

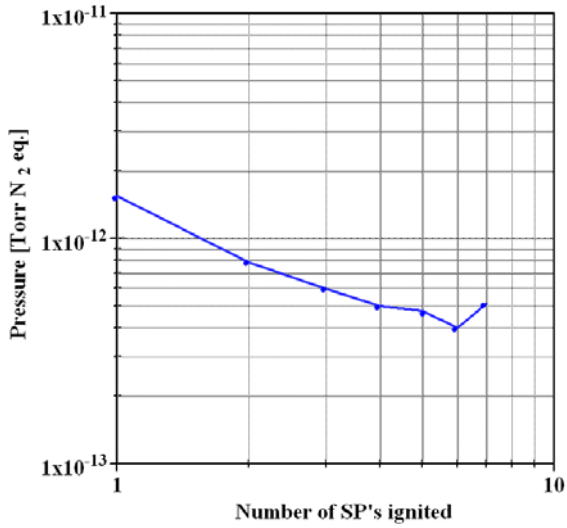


Fig. 2: Variation of the total pressure measured on a 12-m long LEP chamber equipped with a NEG pump and seven sputter-ion pumps, as a function of the number of sputter-ion pumps ignited

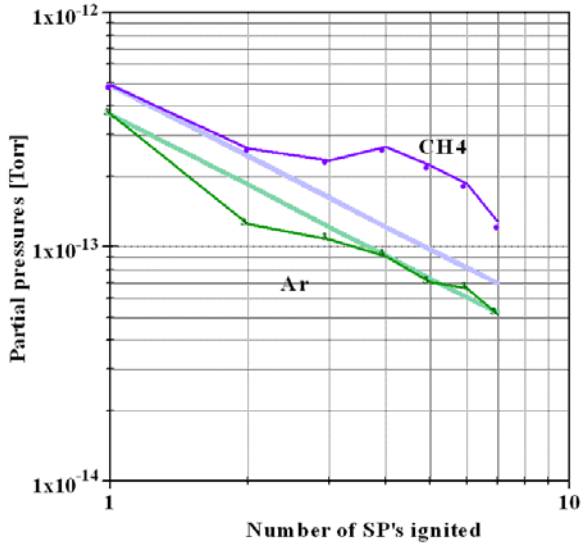


Fig. 3: Same variation as in Fig. 2, for the argon and methane partial pressures

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

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

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 one to estimate 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 (two-site 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.

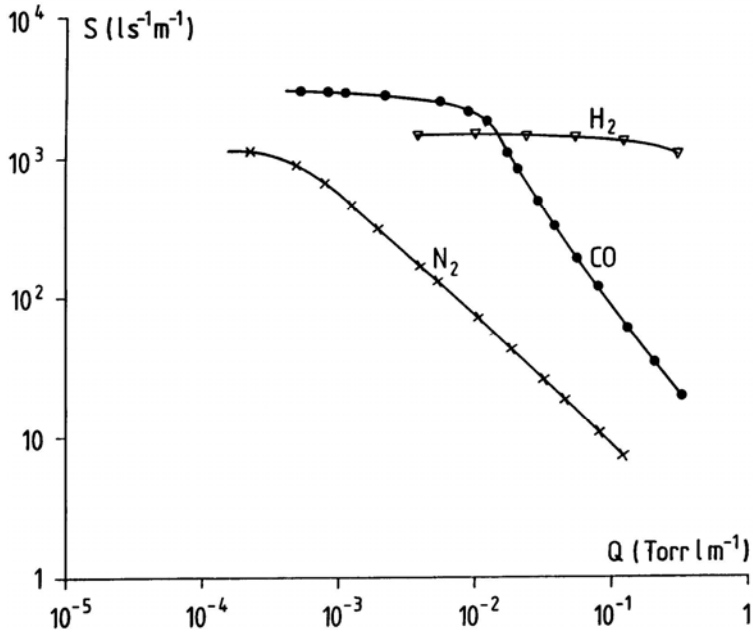


Fig. 5: Pumping speed variation of the St 101 NEG strip (30 mm wide) for H_2, CO, N_2 as a function of the amount of pumped gas

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 during 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 were studied at CERN, 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 was 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 stabilization provided by the much thicker substrate. The main results achieved using this coating are shown in Figs. 6 and 7. This study has been completed and the TiZrV coating has been applied to the vacuum chambers of many accelerators, and in particular to the warm sectors of the LHC at CERN.

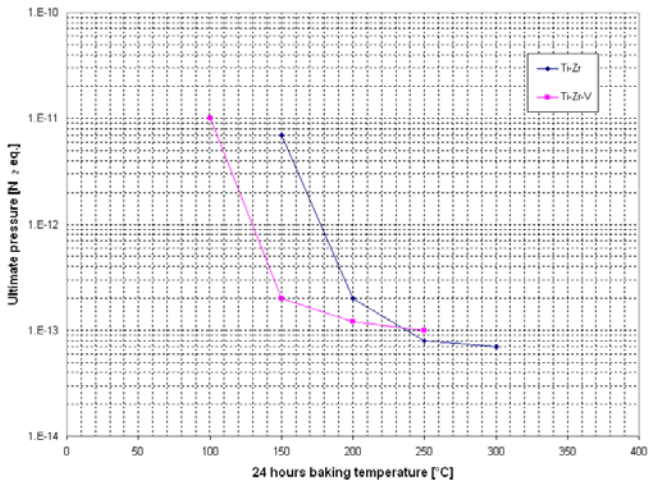


Fig. 6: Comparison of the ultimate pressures achieved in stainless-steel chambers coated with TiZr and TiZrV thin films. The chamber is 2 m long, its diameter is 100 mm, and it is connected at one extremity to a sputter-ion/titanium sublimation pumping station via an orifice of 25 $\ell\ s^{-1}$ conductance 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).

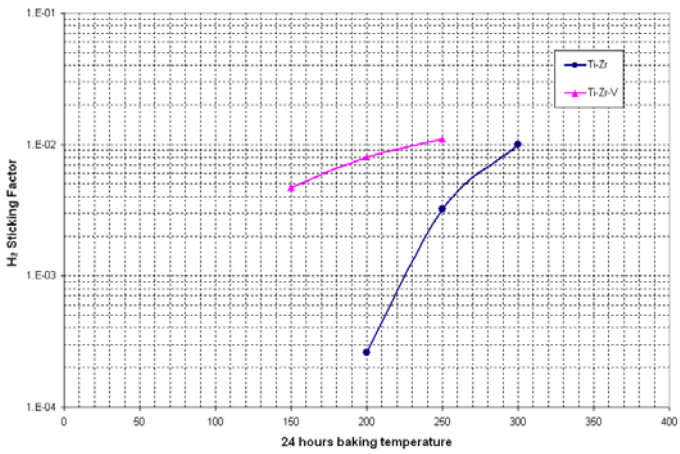


Fig. 7: Comparison of the H_2 sticking factors under the same conditions as described for Fig. 6

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 microns 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 (Wiley, New York, 1998), p. 551.

[2] Ibid., p. 275.

[3] Ibid., pp. 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] Ibid., 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 (Wiley, New York, 1998), p. 297.

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

[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** (1981) 1117.

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

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

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

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

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

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

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

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

[20] C. Benvenuti, J.M. Cazeneuve, P. Chiggiato, F. Cicoira, A. Escudeiro Santana, V. Johanek, V. Ruzinov and J. Fraxedas, IVC14, Birmingham, 1998, Eds. J.P. Coad *et al.*, Vacuum **53** (1999) 219–225.

Cleaning and surface properties

M. Taborelli
CERN, Geneva, Switzerland

Abstract
Principles of precision cleaning for ultra high vacuum applications are reviewed together with the techniques for the evaluation of surface cleanliness. Methods to verify the effectiveness of cleaning procedures are discussed. Examples are presented to illustrate the influence of packaging and storage on the recontamination of the surface after cleaning. Finally, the effect of contamination on some relevant surface properties, like secondary electron emission and wettability is presented.

1 Introduction

In an ultra high vacuum (UHV) system a low residual gas density can be obtained and preserved only by using constituent materials having a sufficiently low vapour pressure at the working temperature [1]. For the same reason, namely to reduce residual gas pressure, the surfaces facing vacuum of all the constituent parts must be free of organic additives, oils, greases, packaging residues, which were used for instance during the manufacturing process.

In an accelerator the adverse influence of the residual gas manifests itself in the interaction with the particle beam and the related degradation of the beam quality [2]. The strength of such an interaction depends on the type of beam particles, but for most of them it increases faster than linearly as a function of the atomic number of the residual gas atoms and molecules. In general, carbon-containing molecules are more harmful than, for instance, hydrogen molecules which might be outgassed from the constituent metallic parts.

There are further possible reasons to apply cleaning procedures to the constituents of the accelerator. For instance, chlorine-based lubricants should be either avoided during manufacturing or completely removed from internal and external surfaces by cleaning in order to preserve a stainless steel vacuum system from corrosion during the entire lifetime of the machine. As a further example, the surface of the metal pipes delivering the working gas of gas ionization detectors (drift tubes, wire chambers, etc.) must be free of silicones to guarantee a constant high efficiency and gas purity. Cracking of silicone molecules present in the working gas generates an insulating silicon oxide coating on the electrodes and deteriorates the detector response [3].

Thus, different contaminations are relevant at different sites and the definition of cleanliness and contaminant is related to the application of the parts. It is, in other words, a specification which must be based on a method of control to be performed before assembly and operation.

The next section illustrates the basic concepts of cleaning and cleaning methods. In the following sections the methods of cleanliness control and packaging after cleaning will also be reviewed together with the effects of contamination on some of the surface properties related to accelerator physics. Previous reviews on cleaning for accelerator technology can be found in Refs. [4,5].

2 Methods of precision cleaning for UHV applications

This section presents an overview of the general principles of cleaning methods suitable for UHV components. Cleaning at this level of accuracy is often called precision cleaning. No universal recipe will be given since commercial cleaning products evolve quickly and the most appropriate solution must be selected by considering the particular application.

2.1 Cleaning with solvents

In a practical description a solution is a system where a solute, in our case the contaminant, is uniformly distributed at the molecular level in a solvent without formation of aggregates or precipitates. The interactions between the molecules of the solute (contaminant) and those of the solvent will determine whether solution or aggregation will occur. If the solute–solute interaction is particularly strong compared to the solvent–solute interaction, aggregation will dominate and the system will separate in two phases. Otherwise solution will occur.

The formal definition of the solubility in thermodynamics is related to the chemical potential of the solvated and aggregated state of a substance. The solubility X_s (in mole fraction) of a contaminant in a solvent is expressed through the equilibrium condition:

$$\mu_0 = \mu_{\text{solution}} + k_B T \cdot \ln X_s \ .$$

μ_i is the chemical potential or Gibbs free energy per molecule, $G = \mu N$. μ_0 is the value for a molecule in an aggregate (or bulk) of contaminant immersed in the solvent and μ_{solution} the respective value for a molecule in solution. The difference in free energy between the two states must be as small as possible to favour solubility:

$$X_s = e^{-\frac{(\mu_{\text{solution}} - \mu_0)}{k_B T}} \ .$$

The free energy per molecule includes the energy due to the interaction of the molecule with its environment. A strong attractive solute–solute interaction has the tendency to lower the chemical potential μ_0 and induces aggregation. An attractive solute–solvent interaction has the tendency to lower μ_{solution}, lowers the difference in chemical potentials and solubility becomes significant when the difference ($\mu_{\text{solution}} - \mu_0$) becomes comparable to $k_B T$. In other words, solution is favoured if the interaction strength for solute–solute is similar to that for solute–solvent.

Molecular interactions depend on the chemical species of the solute and solvent and as a result a given solvent will be able to dissolve and remove only a certain type of contamination. A combination of various solvents can be applied in sequence, where each substance has a specific contaminant as target. A simplified view of the principles and interactions governing solubility is given in the following. The types of interactions are either ionic, Van der Waals, hydrogen bonds or based on the hydrophobic effect [6].

In a solvent the molecules held by ionic bonds, as salts, are dissociated when dissolved. The strength of the Coulomb interaction holding together cations and anions in an ionic solid decreases by a factor corresponding to the dielectric constant ε of the surrounding solvent. Therefore in a solvent the anion–cation attraction is weaker than in air, μ_0 is higher than in air, the difference ($\mu_{\text{solution}} - \mu_0$) is low and the dissociated state of the molecule is favoured. For instance a salt like NaCl exhibiting a strong bond between anion and cation in the crystalline lattice can dissociate and therefore be dissolved in water ($\varepsilon = 78.5$). This occurs less effectively in a solvent having a much lower dielectric constant as propanol ($\varepsilon = 20.2$), where indeed the solubility is 100 times lower than in water [6]. Solvents with high dielectric constant are often polar solvents [7] (by definition a polar molecule is one carrying a permanent electric dipole moment). Thus, ionic species dissolve better in polar solvents than in non-polar ones and perfectly in water. In a more refined picture, hydrogen bonds, described below, should also be considered to show the entire mechanism governing solubility.

Molecules held by covalent bonds interact through Van der Waals forces. This interaction has three main components: orientation between permanent dipoles (polar molecules), polarization between permanent and induced dipoles (polar with non-polar molecules), and dispersion between instantaneous dipoles (in non-polar molecules). The dispersion interaction does not need the presence of any permanent dipoles. For sufficiently large molecules the dispersion term dominates and the polar groups on the molecules, are less important. The dispersion term always provokes attraction between the solute molecules, and its strength (in the approximation of a continuous medium) is governed by the difference in refractive indices $(n_{solvent}^2 - n_{solute}^2)^2$ [6]. Therefore solution is favoured with respect to aggregation when the refractive index of the contaminant is similar to that of the solvent and this attraction is weak. Instead, for small molecules carrying strong permanent dipoles, like water, dispersion is no longer dominating and other interactions occur. Water can dissolve effectively many substances thanks to the strong interaction with polar groups. In some cases it can also form hydrogen bonds with the solute molecules. Hydrogen bonds are particularly strong, directional dipole–dipole interactions and are formed between chemical groups carrying $O^{\delta-}$–$H^{\delta+}$, $N^{\delta-}$–$H^{\delta+}$ and $F^{\delta-}$–$H^{\delta+}$ bonds. These groups are strongly polarized and neighbouring molecules are oriented though Coulomb interaction between the dipoles.

The solubility in water of non-polar hydrocarbon molecules, such as typical oils and greases, is instead very low first because their Van de Waals interaction is weak with water: the orientation and polarization terms dominate in this case, but for molecules like alkanes and similar long aliphatic chains all three terms of the interaction are weak. Solution is hindered by the so-called hydrophobic effect [8]: non-polar molecules, which cannot participate in hydrogen bonds (hydrophobic), induce around them an arrangement of the water molecules which is unfavourable from the entropy point of view. In other words, the entropy in such a configuration is decreased. Since water does not interact strongly with such molecules the loss of entropy upon solution would not be balanced by any decrease in enthalpy and would result in an increase of $\mu_{solution}$. The final result is segregation in two phases rather than solution. Therefore water as a solvent cannot clean greases and oils or other non-polar molecules.

Examples of simple solvents are alcohols (methanol, ethanol, propanol, etc.) and halogenated hydrocarbons. Alcohols can be easily used in laboratory application in small amounts. For UHV applications it is recommended to use high purity grades to avoid residues left on the surface. They can have Van der Waals interaction with polar as well as with non-polar molecules, the strength of which depends on the respective aliphatic chain length. In addition they can form hydrogen bonds. They are able to dissolve various chemical species, but are rather ineffective on some grease, like for instance Apiezon® vacuum grease. Their main disadvantages are flammability and toxicity, so that large amounts cannot be conveniently handled.

Halogen-based solvents (trichloro-ethylene, trichloro-ethane, chloroform, freon, perchlorethylene) are only slightly polar or non-polar and very effective in dissolving many types of greases, but some are highly toxic. Some of them are nowadays banned by security and health regulations and allowed emissions are strictly regulated. Perchlorethylene is tolerated in working areas when the concentration is below 20 mg/m^3 [9]. More refined solvents and blends are nowadays available on the market, based also on non-polluting halogen molecules as hydrofluoroethers (HFE), which can have remarkably low surface energies (as 14.5 mJ/m^2) and wet most of the surfaces. The result of a recent series of tests on the performance of solvents for UHV application including HFE can be found in Ref. [10].

After selection of the best solvent for the actual application and contamination, cleaning can consist in complete immersion of the work-piece to be cleaned in the solvent bath and agitation of the piece or of the bath (see later). After extraction the work-piece is dried by letting the solvent evaporate. Clearly the solvent must not leave on the surface any residues of itself or of dissolved contaminants and rinsing with pure solvent can be necessary. Drying must be performed through a

controlled and reproducible procedure. Evaporation of the solvent often cools the part due to the absorbed heat of evaporation and recontamination through adsorption and condensation from air can be avoided only by keeping the part warmer than the surrounding atmosphere, or by drying in an oven.

The alternative procedure, vapour degreasing, does not leave dry residues on the surface and can be applied to solvents having a sufficiently high vapour pressure slightly above room temperature. The bath of solvent is warmed up—for instance 120ºC for perchlorethylene—to obtain vapour above it and the colder work-piece is suspended in the vapour region. The vapour condenses on the surface of the part to be cleaned. The condensed solvent with the dissolved contaminants will fall into the underlying bath again. The interesting aspect of this method is that the solvent is continuously distilled and only pure solvent condenses on the work-piece, whereas contamination accumulates into the bath which can be filtered.

The same principle can be applied with CO_2, which is interesting for its non-toxic properties (the process does not produce CO_2, it just recycles it). This solvent is non-polar and is especially effective to dissolve aliphatic chains shorter than 20 methylene units and even silicones [11]. It is less effective for polar contaminants and residues of C = O and COOH groups which are not eliminated from the surface and can be detected for instance by X-ray Photoemission Spectroscopy (XPS). In a simple procedure CO_2 is used in the form of snow, where a jet of solid CO_2 forms a liquid layer upon impact on the surface to be cleaned [12]. During treatment the work-piece should be kept above room temperature to avoid re-contamination. In a more refined method, supercritical CO_2 ($SCCO_2$) is used. The gas is compressed (Fig. 1) and heated above the triple point (see dotted region in Fig. 2).

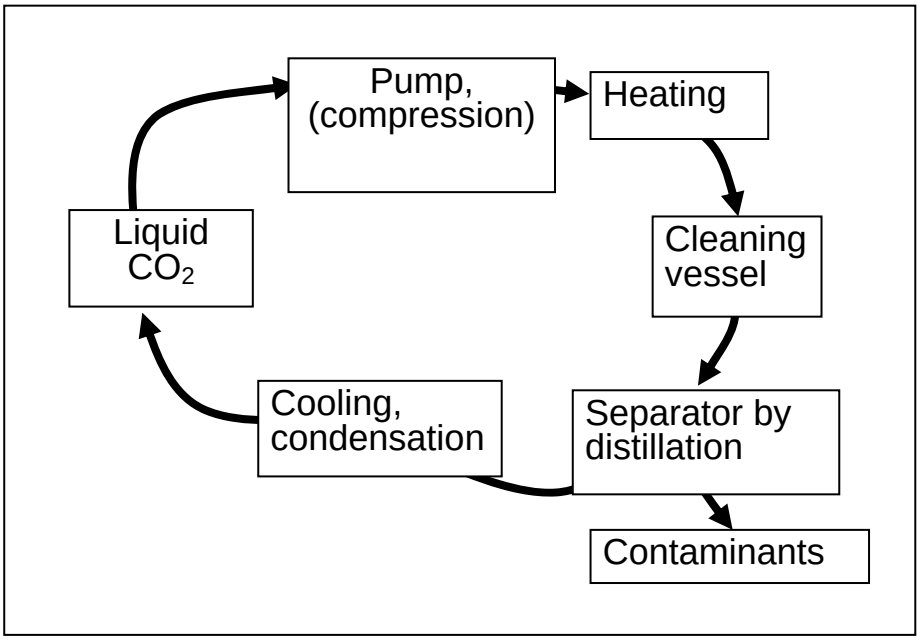


Fig. 1: Typical cycle for cleaning with $SCCO_2$. Note that the tank must be compatible with the high pressures used for the production of the supercritical phase.

The supercritical fluid can perfectly wet any surface, since it has an extremely low surface energy (1 mJ/m^2 for CO_2 to be compared to 32 mJ/m^2 for perchlorethylene and 72 mJ/m^2 for water) and a much lower viscosity than in its liquid phase. The solvation properties are similar to those for the liquid phase. To our knowledge this technique is not yet available on a commercial scale for the cleaning of parts for UHV and developments are still in progress. Suitable co-solvents and surfactants have been developed in order to improve its effectiveness in removing polar substances.

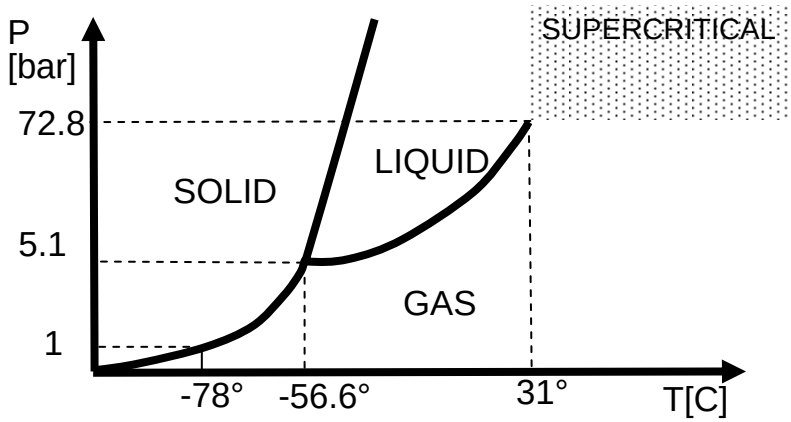


Fig. 2: Simplified phase diagram of CO_2. The supercritical region is marked with a dotted area above the critical point.

2.2 Cleaning with detergents

A detergent is a blend of substances designed for cleaning applications in combination with a solvent. In the following, only the most frequent case where the solvent is water will be considered. The main constituent is a surfactant (contraction of surface active agent), a substance which is able to wet virtually any surface. Such surfactant molecules are amphiphilic, i.e., are able to attract both hydrophilic (through H-bonds) and hydrophobic chemical groups. They consist of a polar or ionic hydrophilic head group and a long hydrophobic tail, as an aliphatic chain. The molecule can have attractive interaction with virtually any molecules, hydrophilic or hydrophobic, polar and non-polar with the most appropriate of its ends. Therefore such molecules like to sit at the interface between hydrophilic and hydrophobic media as water–air, water–oil. Above a critical concentration in the solvent—critical micelle concentration or CMC—a surfactant can build micelles (Fig. 3). Micelles are ordered aggregates of molecules which expose their heads to water. The hydrophobic effect is the driving force to form micelles and order the surfactant molecules. Inverse micelles can form in a hydrophobic liquid.

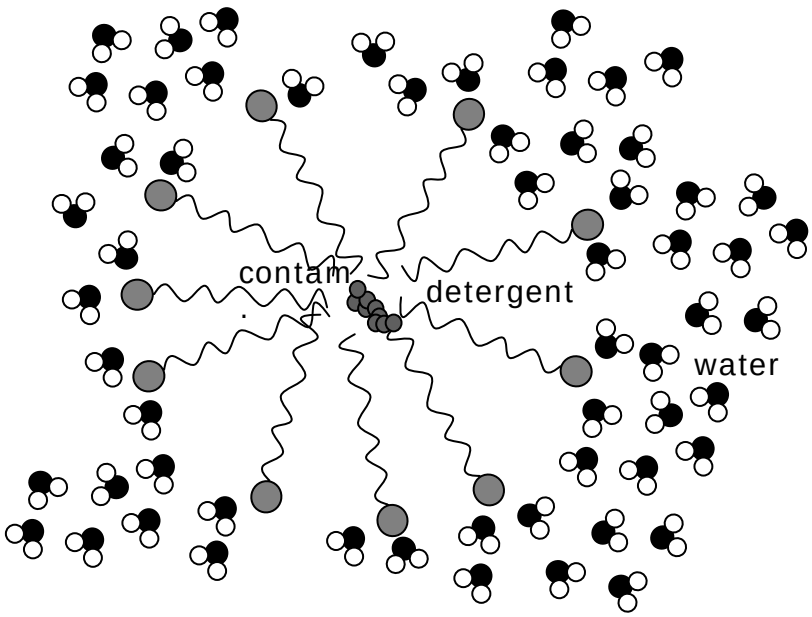


Fig. 3: Arrangement of surfactant molecules forming a micelle in water and encapsulating a hydrophobic contamination particle

The cleaning mechanism is twofold. First, the surfactant removes contamination from surfaces, since it can attract many types of chemical groups and wet both the surface and the contaminant. In this way it lowers the surface energy of the substrate avoiding re-adsorption of the contamination. The contact area of the contaminant with the surface decreases and removal through agitation is made easy. Second, the surfactant surrounds the contaminant, which could not otherwise be dissolved, and encapsulates it in a micelle. Recontamination is avoided and the micelle is then dispersed in the solvent. Continuous filtering of the bath removes the contamination. For instance, in water the surfactant molecules will surround an oil droplet, encapsulate it in a micelle (Fig. 3) exposing polar heads toward water.

The detergent can contain water 'softeners', like soluble silicates, which avoid calcium deposits. It often has a basic pH, an aspect which should be considered in view of the possible resulting surface damages. However, in general, slight etching eliminating the surface oxide is beneficial, since contaminants can be trapped inside. The performance of the detergent is generally optimized at a given concentration and temperature (50–60ºC), which should be compatible with the parts to be cleaned. At the end of the treatment in the detergent bath an extensive rinsing with tap water followed by demineralized water is necessary to eliminate residues of surfactant or additives, as the silicates, from the surface.

In order to guarantee a constant effectiveness of the cleaning bath over time, its quality must be periodically monitored, through pH, conductivity, concentration of detergent measurements, the latter being relevant for the formation of micelles. Another way of monitoring the cleaning effectiveness of the bath is to analyse the result on test specimen (see Section 3). If in a plant cleaning of one type of metal is frequent and some slight surface etching occurs during cleaning, it is safe to monitor the concentration of this metal in the bath to avoid transfer of such ions to other materials passing in the same cleaning station. Complicated cases in this respect are brazed joints, where one of the metals or one of the components of the brazing alloy can be spread on the entire surface of the work-piece.

As described above, the cleaning power of solvents is in principle improved by detergents, which operate always in combination with a solvent. A nice comparison between various solvents and detergents used for precision cleaning of UHV components is given in Ref. [13]. The main reason why solvents are still applied is the need to rinse the detergent cleaned parts with water in order to remove completely the traces of detergent from the surface. Such a rinsing can only be effective when the shape of the cleaned part does not trap water and residues through pockets, pores and meander-like shapes. Residues can provoke long-lasting outgassing or corrosion. For this reason vapour cleaning by solvents is necessary for bellows, porous materials such as ceramics, narrow curved pipes, valves and similar manifold components.

Moreover, since volatile solvents can be recycled by distillation during the cleaning process, they are particularly suitable for a gross cleaning phase of parts which are heavily contaminated by oils and greases after manufacturing. The gross cleaning can be used in combination with a subsequent treatment in a detergent bath. In this way the contamination of the generally expensive detergent bath can be limited and its lifetime prolonged.

In some cases detergents can provoke chemical deterioration of surfaces and alter their functional properties. A striking example is given by TiZrV non-evaporable getter (NEG) coatings used for pumping purposes in accelerator pipes. Figure 4 illustrates the effect on the surface composition of such a coating of cleaning by a detergent having a basic pH and commonly used in standard CERN cleaning procedures for vacuum chambers. The surface chemical composition is completely altered with a marked reduction of the vanadium content. As a consequence the functional behaviour of the getter is deteriorated and no activation occurs in the useful temperature range (Fig. 5).

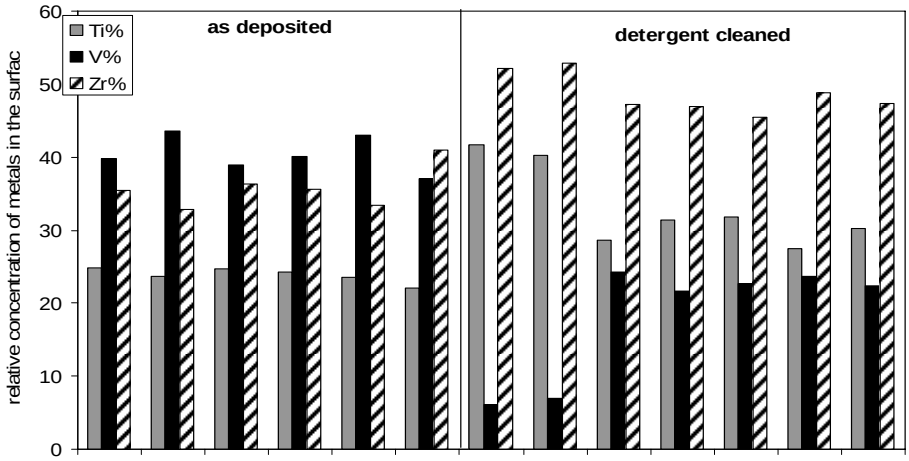


Fig. 4: Comparison of the surface chemical composition of several samples of TiZrV NEG coating in the as-deposited state (left side) and after treatment in a detergent bath (right side). The composition is measured by XPS.

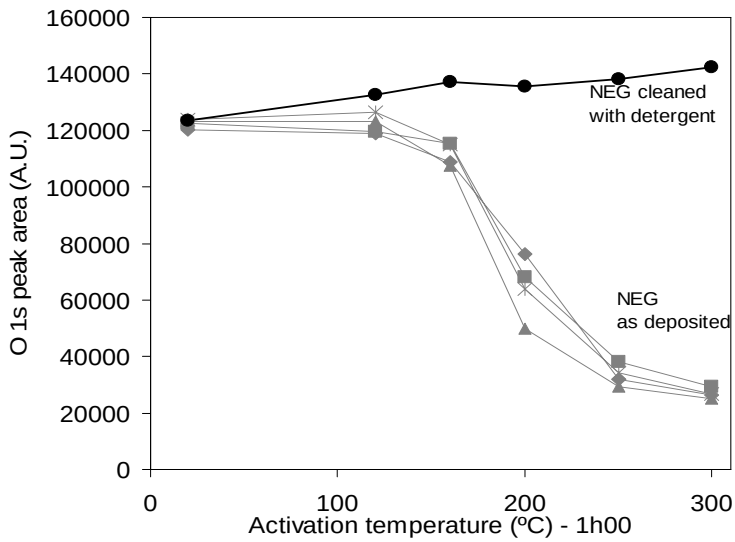


Fig. 5: Activation behaviour of TiZrV NEG coating illustrated by the decrease of the oxygen O 1s intensity in XPS. The sample cleaned in a detergent bath does not exhibit any activation.

2.3 Agitation

In the case of precision cleaning by immersion in a bath for both solvents and detergents, ultrasonic agitation [14] of the bath is applied. In this way removal of soils is much more effective, as illustrated in Fig. 6. Ultrasonic waves (20–120 KHz) are mechanical pressure waves generated by piezoelectric transducers placed in the cleaning tank. The waves create bubbles of some 10–100 microns in the liquid medium. Bubbles grow up to implosion and energy is then released. In such a way adsorbed contaminants and particles can be removed even from crevices and blind holes. Agitation is mandatory for all samples having a complex shape. The power of the ultrasonic actuators must be correctly dimensioned and depends on the bath volume, their position and orientation, and on the shape of the bath tank. For long pipes which cannot be easily immersed the ultrasonic agitation can be replaced by turbulent flow. The cleaning fluid is forced to stream through the pipe in a turbulent flow regime.

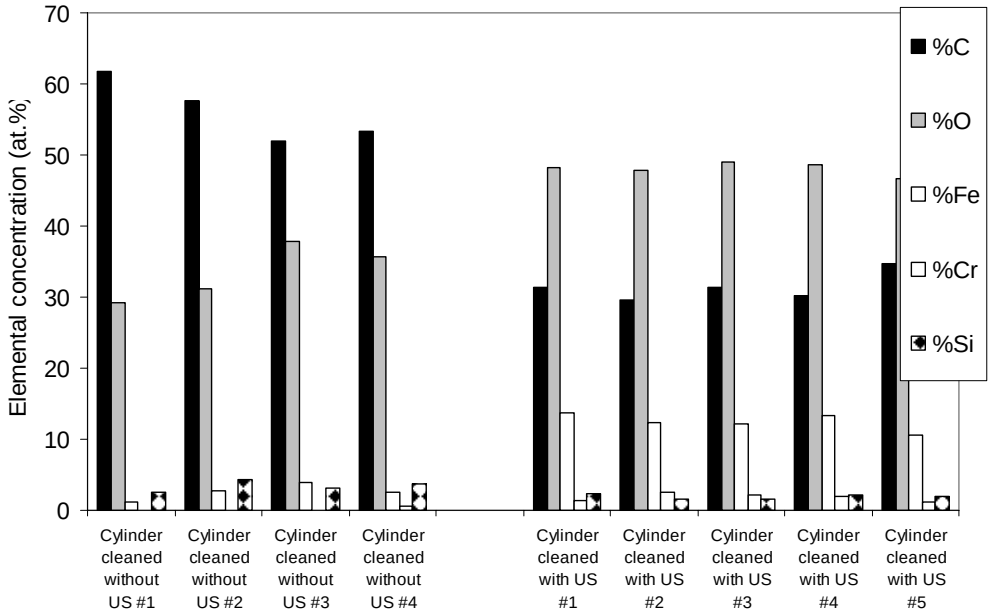


Fig. 6: Example of samples cleaned in the same detergent, with and without ultrasonic agitation, respectively. The XPS composition reflects the higher efficiency of cleaning with ultrasonic agitation.

Examples of cleaning procedures including some optional steps are illustrated in the schemata in Fig. 7. The application of the various optional steps depends on the level of contamination of the parts to be cleaned, on their shape, material, application, and so on. For instance, pre-cleaning in a solvent is applied on heavily contaminated parts. High-pressure water rinsing has been found successful for improving the performance of superconducting radiofrequency cavities [15].

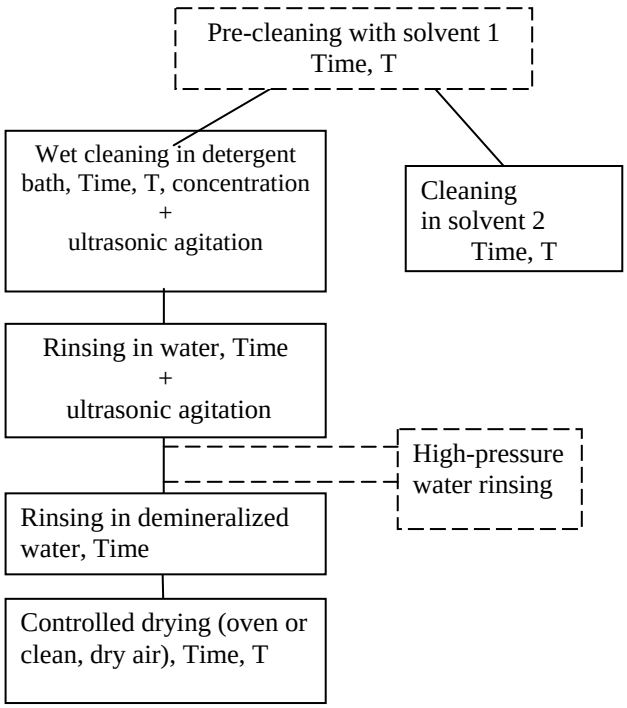


Fig. 7: Typical sequence for cleaning of UHV parts with solvents or detergents and typical parameters to be controlled. Dashed lines represent optional steps depending on the particular application.

2.4 Special cases

It is worth mentioning an example where cleaning seems to be impossible. Stainless steel such as 316LN is often used for UHV applications and is vacuum fired at 950ºC in vacuum for degassing purposes. If the amount of boron concentration in such steel is higher than some 9–10 ppm [16], boron nitride (hexagonal) can form at the surface during the vacuum firing treatment (Fig. 8). Boron nitride itself is not harmful for UHV, however, such a layer prevents the adhesion of a further coating made by evaporation, sputter deposition, or electroplating. Adhesion of the coating will be hindered by the fragile lamellar structure of boron nitride and its low surface energy. Water cannot wet such a surface and the usual detergents are not able to remove such a layer. Only etching or electropolishing of the surface is possible in order to enable further coating.

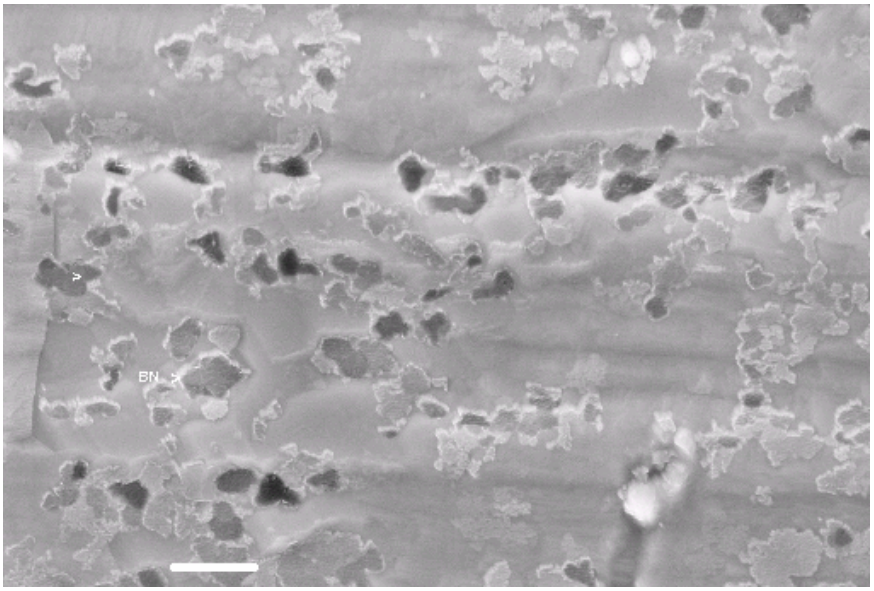


Fig. 8: Scanning electron microscope image of stainless steel 316LN after vacuum firing (scale bar is 2 microns). The leaflets visible on the surface are boron nitride crystals formed through surface segregation and reaction of boron (15 ppm in the volume) with the nitrogen of the steel.

Another interesting case is represented by silicones as contaminants. Vacuum greases often contain silicones which are difficult to eliminate from the surface. Indeed, silicones can float on a cleaning bath, and during extraction of the work-piece from the bath the contamination wets the surface again. As mentioned previously, some silicone species have non-negligible solubility in $SCCO_2$, but no commercial system exists yet to apply this technique for cleaning of large UHV parts. An excellent solvent for many silicone greases and compounds is hexane [17]. Its effectiveness has been verified for instance by analysing by XPS a copper surface previously contaminated with silicone grease and cleaned with hexane. The silicon signal was below detection limit. Unfortunately hexane is highly volatile. It must be used in low amounts in well vented rooms, so that the average concentration remains below 200 mg/m^3 [18]. Therefore this treatment can be applied to small parts or extraction for analytical purposes (see below).

This review considers only the cleaning methods. In addition, etching to eliminate damaged layers or treatments to induce surface passivation and avoid for instance re-oxidation can be performed if necessary. Some treatments are described in Ref. [4]. They are not suitable for parts which need to preserve accurate dimensions and precise surface finishing. Moreover, it is worth noting that the surface in operational conditions is often not the same as after cleaning since bake-out is frequently used for UHV systems.

2.5 Cleaning with ions: glow discharge

After cleaning with the methods described in Sections 2.2 and 2.3, a further processing step can be applied in order to remove oxides or contaminants adsorbed upon air exposure. Glow discharge treatment can be applied to vacuum pipes after evacuation and filling with a suitable gas at low pressure (10^{-1}–10^{-3} mbar). Parts with other shapes can be inserted in a vacuum chamber designed for this purpose and in this case the geometry of the electrodes should be studied carefully. Glow discharge can be applied with DC, AC or rf voltages and by using various gases. The principle consists in accelerating ions toward the surface so that the surface is sputter cleaned. In addition, especially in the cases of AC and rf discharges, also ultraviolet light is emitted by the plasma and can help contamination cracking through photochemical reactions. In the DC case the typical voltages used are 300–1000 V. For long pipes, such as accelerator beam chambers, the anode is a wire placed along the chamber axis and the chamber is kept at ground potential. In general, the wire is removed after the treatment by venting and opening the chamber. Such a technique was applied on the Intersecting Storage Rings (ISR) vacuum chambers at CERN. If the air exposure time to remove the electrode is minimized, the beneficial effect is partly preserved [19]. Venting must be carefully performed with clean (or at least dry) gases, like pure N_2. Depending on the dose of ions, the treatment can remove the topmost hydrocarbon contamination or even the native oxide layer on the surface. This configuration is exploited also if additional cleaning or oxide removal is necessary prior to magnetron sputter deposition of coatings on long pipes; the anode is already in place and acts then as a cathode during the deposition process.

Sputter cleaning with a noble gas such as argon does not induce chemical reactions with the surface to be cleaned, but some implantation occurs. In stainless steel implanted argon can be removed by baking at 350ºC [20]. A mixture of gases can be used to remove more effectively a specific contaminant so that the total ion dose can be reduced. For instance mixtures containing oxygen (Ar with 5–10% O_2) are very effective in the elimination of carbon from the surface [20, 21]. A 100 times lower dose of ions is sufficient to remove surface contamination of carbon from stainless steel when 10% O_2 is added to Ar, compared with the case where pure Ar is used. This is due to chemical reactions forming volatile species like CO and CO_2, which are evacuated together with the working gas. Hydrogen and helium are used as working gas for glow discharge treatment especially in fusion reactor walls [22] where such gases are not harmful and are less frequently used for UHV systems. A special use of helium is the so-called helium processing performed to condition radio-frequency niobium superconducting cavities. The process is again a sputter cleaning and helium is used just for convenience, since pure gas is available for the cryogenic circuit [23].

The main disadvantage of glow discharge treatment is the possible coating of insulating parts or windows with the sputtered material. The treatment can be applied on various metals such as stainless steel, aluminium, copper, titanium and beryllium. Glow discharge on beryllium is attractive, since this material cannot be easily handled in a wet cleaning facility owing to the toxicity of its oxide. In the case of beryllium [24] it is of great advantage that at sufficiently low energy (below ~300 eV) the sputtering coefficient of O_2^+ on C is higher than on beryllium thanks to a chemical reaction producing CO and CO_2. The effect is even more marked for beryllium oxide. Tests made at CERN with glow discharge in pure oxygen on small samples showed a strong decrease of carbon levels on the beryllium surface and no detectable (XPS) amount of beryllium on the mounting used to hold the sample.

In other domains glow discharge treatment is used to increase wettability and reactivity of polymer surfaces, since the plasma or sputtering breaks surface bonds.

3 Cleanliness and cleaning effectiveness

3.1 Methods for assessing cleanliness

Cleanliness must be evaluated with respect to the application for which the surface is intended. Often it is not possible to verify directly the functional performance without assembling an entire device and therefore a control technique is chosen providing sufficient sensitivity to the known crucial contaminants. The control procedure measures the amount of contaminant still present on the surface or a quantity related to it and compares it with the preset limit for acceptance. This limit defines cleanliness. More generally if contamination can be quantitatively monitored, one can define cleanliness classes, which are specifications for each application.

In accelerator technology various contaminants present on the vacuum chamber surface can deteriorate vacuum (hydrocarbons, intermediate vapour pressure compounds), propagate through the system (low vapour pressure metals like Cd and Zn), promote corrosion (halogens), transform into insulating layers upon irradiation (silicones). For the assessment of surface cleanliness many techniques and procedures have been used (Table 1) [25,26]. Most of them reveal the presence of the contaminants, without verifying directly the functional performance of the surface. Many of them require special sample size or shape and cannot be applied on the cleaned part itself, but only on a test specimen, which has followed the same treatment. The techniques can be divided into two types. Analytical techniques can identify and more or less quantify the contamination. Other methods measure a quantity, which is related in a complex and sometimes obscure way to surface cleanliness, but can indicate at least whether excessive contamination is present and enable one to reject the part before insertion in UHV.

The advantage of the analytical techniques is that by identifying the contamination they often enable one to understand its origin. Common surface analyses, such as XPS, Auger Electron Spectroscopy (AES) and Secondary Ion Mass Spectroscopy (SIMS) are well adapted to detect and identify a broad range of organic and inorganic contaminants and obtain a high surface sensitivity. XPS enables easy and fast identification of the elements present on the surface and has a detection limit close to 1 at% in the probed depth of 1–3 nm for most elements. With a monochromatized X-ray source, even distinction between organic species is possible. AES is as good as XPS from the point of view of sensitivity, but Auger lines are wider in the energy spectrum and overlap is frequent. Moreover, care should be taken by using AES to limit the current density of the electron beam, otherwise it can induce surface modifications through Electron Stimulated Desorption (ESD) and also influence the local carbon coverage through stimulated diffusion [27]. ESD measurements indicate that desorption yields on air-exposed surfaces decrease by about a factor 10 for 10^{16} electrons/cm^2. The typical primary beam currents used in Auger analysis (for instance 10^{-8}–10^{-7} A on an area of $10 \times 10\ \mu m^2$ up to $100 \times 100\ \mu m^2$) result in some 6×10^{15}–6×10^{18} electrons/cm^2 in 10 seconds, which can be a typical time to acquire a full spectrum. There is experimental evidence that the carbon concentration measured by AES depends on the impinging beam current density [27] and decreases with irradiation. Therefore, the current density and the dose should be well controlled, when comparing data of surface cleanliness. Using XPS, without high spatial resolution options, the amount of surface damage is reduced by a factor 10 to 100 [28]. SIMS, especially in its static version (SSIMS) and high mass resolution, is superior from the point of view of identification of the chemical species of contaminants and has a better sensitivity. The information is, in principle, more detailed; at the same time the interpretation and quantification, which is affected by matrix effects, is not straightforward.

Table 1: Techniques for evaluation of surface cleanliness. Sensitivities should be taken as estimated orders of magnitude.

Technique	On site, on cleaned pieces	Limitations (Analytical or not)	Cost	Time	Principle of measurement (detection limit for carbon species, when known)
XPS, ESCA (X-ray Photoemission Spectr.)	No	UHV, small sample size (analytical)	High	Slow	Electron photoemission (3% at. in the probed depth, corresponding to ~10^{14} atoms/cm^2)
AES (Auger Electron Spectr.)	No	UHV, small sample size (analytical)	High	Slow	Electron-induced electron emission (~10^{14} atoms/cm^2)
SSIMS (Static Secondary Ion Mass Spectr.)	No	UHV, small sample size, interpretation (analytical)	High	Slow	Ion erosion coupled to mass spectroscopy (10^{12} atoms/cm^2 [29])
ESD (Electron Stimulated Desorption)	No	Needs suitable sample shape and size (partly analytical)	High	Slow	Electron-induced desorption of adsorbates detected by mass spectroscopy (sensitivity depends on irradiated sample size)
Outgassing rate	Yes	Sample shape, (partly analytical)	High	Slow	Thermal desorption monitored by mass spectroscopy (depends on accumulation time)
FTIR (Fourier Transform Infrared)	No	Needs rather smooth surface to get high surface sensitivity, partly overlapping peaks (analytical)	High	Slow	Infrared absorption (10^{12} atoms/cm^2 in multiple-internal-reflection mode, MIR-FTIR on flat Si wafer [29]) or grazing incidence reflection or based on extraction
UV-vis	No	Needs extraction through solvent, overlap of absorption lines (analytical)	High	Slow	UV absorption, extraction method
TRXRF (Total Reflection X-ray Fluorescence)	No	Needs smooth flat surface (analytical)	High	Slow	X-ray fluorescence
Ellipsometry	No	Needs smooth flat surface	High	Slow	Rotation of light polarization upon reflection (10^{12} molec/cm^2) [29]
OSEE (Optically Stimulated Electron Emission)	Yes	UV light modifies (cleans) the surface. Does not distinguish between oxides and contaminants	Low	Fast	Photoelectron emission in air (10^{15} molec/cm^2 [26])
Wettability by water	Yes	Depends on experience of the operator	Low	Fast	Wettability of the surface by water
Water contact angle	Yes, for portable instruments	One model tested and found unreliable	Medium	Fast for portable system	Wettability of the surface by water
SPD (Surface Potential Difference)	Yes	Instability of reference surface	Low	Fast	Measure changes of work function due to oxidation and adsorbates (~ 10^{14}–10^{15} molec/cm^2)
Gravimetry	No	Insufficient sensitivity	Low	Slow	Measure weight loss of the sample due to cleaning
Surface tension markers	Yes	Insufficient sensitivity, pieces must be re-cleaned	Low	Fast	Wetting by inks of various surface energy (surf. en. < 44 mJ/m^2)
Radiactive tracer	No	Only on test sample contaminated on purpose	High	Slow	Measure decrease after cleaning of radioactivity of a sample contaminated with a tracer [13]

In the following the application of XPS is discussed more in detail. In general the result of an XPS analysis is not the absolute concentration, as molecules per cm^2 for instance. A so-called surface concentration or relative atomic concentration (at%) in the probed depth is obtained. This quantity is calculated from the measured intensities for each element, as peak areas, and the corresponding elemental calibration factors. In fact this quantity corresponds to the true relative atomic concentration only for the case where the distribution of the considered elements is uniform in the probed depth. For contamination this is rarely the case, since it is by definition on top of the surface. However, such a quantity can be used safely for the quantitative characterization of cleanliness and the comparison of results obtained with the same excitation source, the same analyser parameters and geometry (source-sample-analyser) [30]. The case of carbon, which is one of the most common contaminants, is considered more in detail as an example. It is present as hydrocarbons left by lubricants, cutting oils, rotary vane pump oils, residues of packaging materials, fingerprints, and finally airborne hydrocarbons from storage in unprotected environment. An upper limit for the amount of carbon on the surface is therefore often adopted as criterion for surface cleanliness, possibly accompanied by an upper limit for the total amount of other minor impurities. The validity of the so-called surface concentration, at%C, to assess cleanliness has been discussed in detail in Ref. [30]. At CERN for UHV applications a level of 40% at of C on stainless steel is defined as upper acceptable limit (non-monochromatic MgKα, PHI-ESCA 5400 analyser, slit 4, 45° escape angle, 35eV pass energy). This value is chosen based on the accumulated experience on UHV applications, showing that such a surface will have an acceptable degassing rate and will enable one to achieve UHV conditions for static vacuum. Moreover, recontamination kinetics justifies such a limit (see Section 3.5). The thickness of such a layer is estimated around 0.5 nm, assuming the common electron attenuation length values [31] and assuming a homogeneous layer of pure carbon to simulate the contaminant. Care should be taken when comparing cleanliness of different materials. The attenuation of the XPS signal from deeper layers provoked by the attenuation length of the photoelectrons is energy dependent. As a result the same absolute amount of hydrocarbons on two different metals will give a different relative atomic concentration as measured by XPS. Conversion factors can be established, based on the experimental data, to compare the values measured on different materials, as described in Ref. [30]. For instance the same amount of carbon contamination on stainless steel and copper will result in 40% at C and 44% at C, respectively.

A weakness of XPS, at least when used with a non-monochromatized X-ray source is its incapacity to distinguish between silicones and silicates. The chemical shift of the silicon line Si $2p^{3/2}$ is similar in both cases and distinctions based on the detected amount of oxygen are unreliable due to variations in silicone species and adsorbates. The problem is not only academic, since silicates are not so harmful as silicones and are sometimes left on the surface after detergent cleaning (silicates are often included in the cleaning agent). Two techniques can help, SIMS and FTIR. Even in low-sensitivity SIMS fragments like $Si(CH_3)_3^+$ at 73 m/e can be detected and do not overlap in the spectrum with other intense hydrocarbon fragments. For FTIR the sample can be rinsed with hexane, which is a good solvent for silicone oils and greases as mentioned before. The resulting solution is deposited directly on the window used for the IR reflection or transmission measurements and the data acquisition is carried out after evaporation of hexane. The typical absorption features are in the 800–1300 cm^{-1} region. The sensitivity of this elution method is potentially high and depends on the size of the rinsed surface.

Thermal outgassing is another method of testing surface cleanliness which probes the functional performance in static vacuum conditions. The sensitivity of the method depends on the surface which is heated compared to the surrounding surface of the vacuum vessel, and generally samples of the suitable shape must be prepared. Comparison of cleaning methods by this technique has been carried for instance in Refs. [10, 32].

3.2 Tests for dynamic vacuum performance

In the particular case of surfaces, which are designed to be exposed to a particle beam, as in the case of an accelerator beam-pipe, the actual pressure or dynamic vacuum is determined by particle-induced desorption. For such cases it is safe to verify the cleanliness levels also by ESD. For the most common cleaning procedures used at CERN this has been done, as described in Ref. [13]. Often ESD is applied to tube-shaped samples measured after baking at 150°C to reduce the level of water in the residual gas, which would mask other fragments. The typical irradiation dose where the ESD yield decreases by a factor of 10 is about 10^{16} electrons/cm^2 (10^{-3} Clb/cm^2) due to a progressive cleaning of the surface. This means that the dose must be limited in order to acquire relevant data for the characterization of the surface in an unconditioned state.

A similar technique is based on ion-stimulated desorption [33]. The ion source in the keV range can be obtained from usual SIMS or sputtering ion guns, where defocusing or scanning should be applied, since one should keep in mind that the sensitivity for constant dose is proportional to the irradiated surface. The main tendencies of ion-stimulated desorption, as desorbed species and intensities, correlate well with ESD [20]. A special case of ion-stimulated desorption is represented by high-energy (MeV/nucleon), highly ionized heavy ions [34]. This phenomenon is particularly important for ion storage rings where the impact of lost ions can induce pressure bursts. The lifetime of such particles is extremely sensitive to residual gas pressure and has a positive feedback mechanism, since ions with a modified charge will diverge from the beam, impinge on the chamber wall, and desorb more gas. It has been shown that desorption coefficients are some orders of magnitude higher for such ions than for instance for electrons. The phenomenon is not completely understood yet, but the usual relation between surface contamination level and desorption yield is confirmed. For instance coatings of the surface as getters or noble metals, which can be easily cleaned by baking in situ, exhibit lower yields than bare stainless steel-surfaces.

To measure Photon Stimulated Desorption (PSD), which is relevant in all the cases where synchrotron radiation impinges on the beam chamber walls, the only suitable source is generally obtained from synchrotron radiation itself at the necessary critical energy. Experiments along this line have been performed [20] by operating a large machine rather than a small set-up for laboratory-size experiments.

3.3 On-line and off-line quality control

The ideal quality control for a cleaning plant consists in the real-time monitoring of the surface cleanliness immediately after processing. This would enable one to provide the necessary bath maintenance in time and avoid delivery of parts which are not perfectly satisfactory. Some of the methods considered in the literature for the characterization of surface cleanliness are listed in Table 1. Unfortunately, none of these techniques can be applied as a fast selection test on-line on a series of cleaned objects having different base material, shape, size as in the case of parts cleaned in a facility for an accelerator. It is worth noting that the simplest fast monitoring for the cleanliness of treated parts is the observation of the wetting behaviour of the piece immediately after rinsing. The piece is considered clean if it remains covered by a uniform layer of water when it is lifted out from the rinsing bath or after spraying some water on it. This type of control relies on the experience of the operator of the cleaning station. Available commercial portable instruments for water contact angle measurement in situ have shown low reliability.

3.4 Evaluation of the effectiveness of a cleaning procedure

The evaluation of the effectiveness of a cleaning method is verified by contaminating a sample with a well-known blend, cleaning and analysing the surface for instance by XPS. A sufficiently large area and number of samples (4–5) should be measured, in order to average over statistical variations within the same cleaning run. A similar rationale is recommended also by international standards [35]. The

mixture should be chosen to contain chemicals which are representative of a real contamination supposed to occur on the parts arriving at the cleaning plant. For instance, a mixture of oils and greases used during machining at the local workshop or vacuum pump oil which might be present on previously used parts can be a reasonable choice [32]. At CERN such a method has been adopted also to assess the quality of cleaning procedures used by external manufacturers of parts to be inserted in UHV.

An especially elegant way of measuring the amount of residues left on the surface from a previous well-defined contamination is the method of the radioactive tracer [13]. The contamination molecules carry radioactive isotopes and the level of radioactivity after cleaning measures directly how effectively such a contamination has been removed over the whole sample surface.

3.5 Packaging and storing cleaned parts

Recontamination occurs through adsorption of contaminants from air or through incorrect packaging methods. Upon air exposure a clean metallic surface, for instance a sputter-cleaned copper surface, forms a layer of oxide, then possibly part of it converts to hydroxide or can be covered by adsorbed water. This occurs because the surface energy of an atomically clean metal or an oxide is some 10–100 times higher than that of water or hydrocarbons (1850 mJ/m^2 for clean metallic copper, 72 mJ/m^2 for water and 25 mJ/m^2 for alkanes). Hydrocarbons have the lowest surface energy and can cover such a surface in a dynamic process, which possibly results in a contamination layer including water and hydrocarbon molecules, the latter with the non-polar regions pointing toward air. Such a process results in the growth of carbon contamination illustrated in Fig. 9. Similar curves were measured by ellipsometry [29] on precision-cleaned silicon wafers. Two further conclusions should be drawn from Fig. 9. First, large parts which will inevitably be exposed to air before packaging or tight closure will always exhibit non-vanishing hydrocarbon coverage on the surface. Second, comparison of effectiveness of cleaning procedures is meaningful only when air exposure time and the storage method (see below) have been correctly defined in advance.

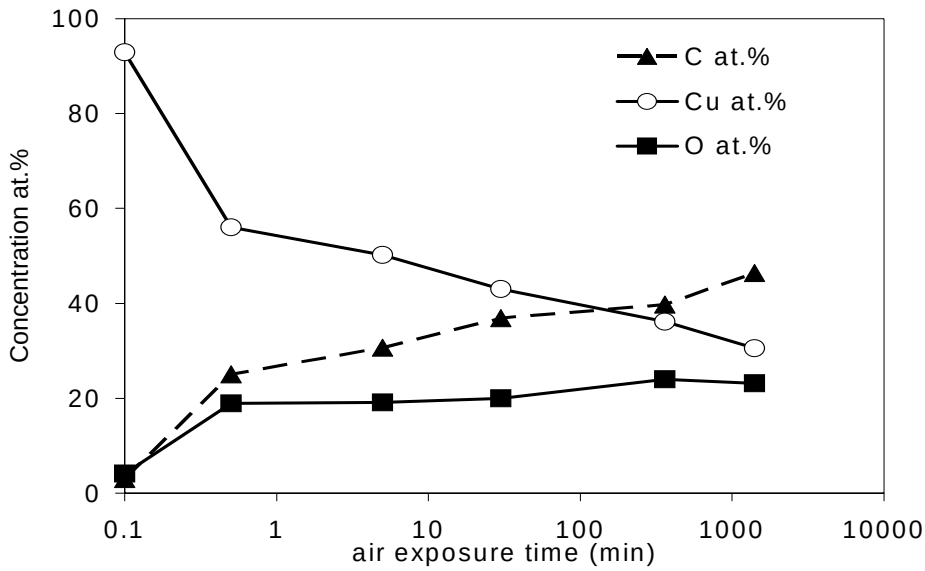


Fig. 9: Evolution for the composition, measured by XPS, of a sputter-cleaned copper surface as a function of the air exposure time in the laboratory

In order to avoid recontamination, parts should be used as soon as possible after cleaning. This is not always possible in case of construction of large plants where large series are cleaned, transported, and installed. A proper packaging to protect the part from contamination during storage is suitable. In Fig. 10 [30] a comparison of simple storage and packaging methods is shown. All the

samples were cleaned in the same run in a detergent bath, measured in XPS, and stored wrapped in aluminium foil, in air (protected from dust in a Petri dish), in a polyethylene bag, in a pure-polyethylene bag, and inserted in a polyethylene bag after wrapping in aluminium foil, respectively. It is clear that inserting the samples in a polyethylene foil can have dramatic effects and moreover the result depends on the polyethylene quality. However, packaging in a polymer bag has obvious advantages of protection from macroscopic contamination during transport. Wrapping the parts in clean aluminium foil (in this case common grade used for food packaging) before packaging them in the polyethylene bag removes completely the effect of the surrounding polymer bag. The sample remains as clean when as wrapped in aluminium foil alone. Such a method is obviously not suitable in cases where aluminium traces can provoke adverse effects on the parts.

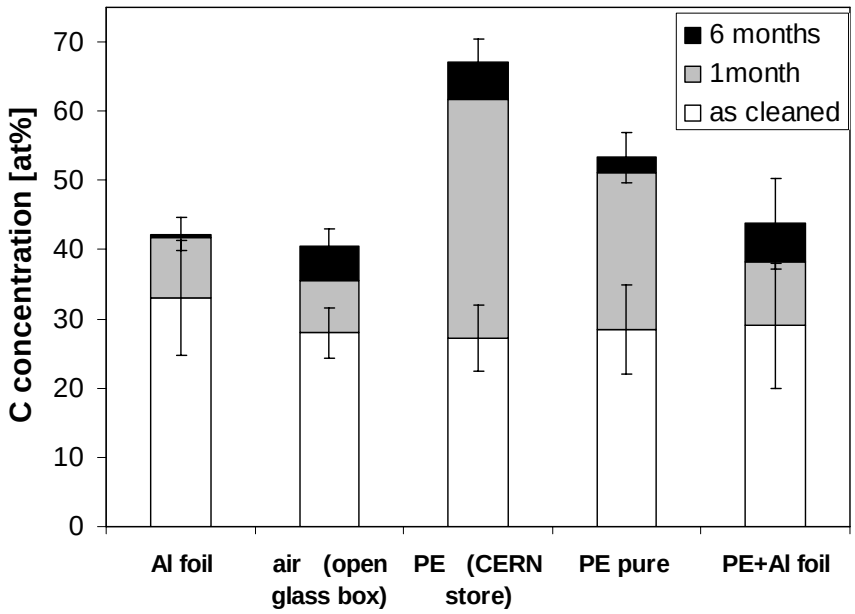


Fig. 10: Effect of recontamination of copper samples cleaned in a detergent bath and stored in different ways (average over four samples in each case)

It is worth noting that contamination increases during the first month of storage, but its amount saturates and the further increase measured after six months exposure is moderate. This translates the fact that initial adsorption on the high-energy surface is much more favoured than on the contaminated low-energy surface.

4 Surface properties and contamination

4.1 Wetting and surface energy

In addition to static and dynamic vacuum, other properties which are relevant for UHV application can be influenced by surface cleanliness. During air exposure, adsorbates like hydrocarbon and water cover easily any clean surface and lower its surface energy. This prevents the adhesion of further coatings, for instance deposited by magnetron sputtering or evaporation. In principle a single monolayer of tightly packed organic molecules covering the surface, like a deposited self-assembled monolayer or a Langmuir Blodgett film, is sufficient to reduce drastically the surface energy [36]. In the case of hydrocarbons adsorbed in a disordered layer during air exposure this occurs more gradually and a larger total thickness is needed. It has been shown [37] that the surface energy decreases down to the level of bulk hydrocarbons when the overlayer reaches a thickness of 2–3 nm. The behaviour

found as a function of thickness is summarized in Fig. 11. A similar set of data illustrates [29] the increase of water contact angle, proving the decrease in wettability, as a function of air exposure time.

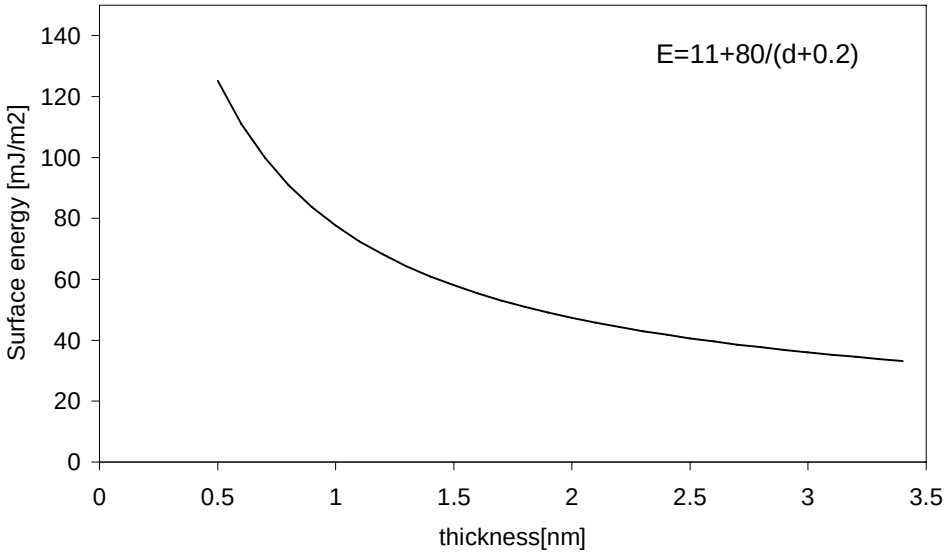


Fig. 11: Behaviour of the surface energy of stainless steel as a function of the thickness of the adsorbed contamination. The equation on the top gives the energy in mJ/m2 as a function of the thickness in nm [37]

4.2 Secondary electron yield

The secondary electron yield (SEY) of the surfaces exposed to the particle beam is a further quantity which is relevant for particle accelerators. This quantity influences the beam stability and in the worst case can maintain the so-called multipacting or resonant electron multiplication or electron cloud effect [38]. For common technological materials like stainless steel, aluminium, and copper the SEY of the air-exposed surface is much higher than for the clean metal or the corresponding oxide. As can be concluded from the shift of the maximum yield toward lower primary energies [39], the SEY increase is mainly due to the coverage with a layer of contaminant, which has itself a high SEY. Since thin layers of water have been shown to be ineffective to justify such an increase, the origin of the effect must be in the adsorbed airborne hydrocarbons.

For getter materials the activation process removes the surface contamination by transforming the hydrocarbons in carbides and by letting the oxygen of the oxide diffuse into the bulk; simultaneously the SEY is reduced [40]. For the more common metals used for accelerator vacuum chambers SEY can be progressively reduced by irradiation by electrons or photons (synchrotron radiation). This process is usually called conditioning and consists in surface cleaning through particle stimulated desorption and de-hydrogenation of the adsorbed hydrocarbons [27].

As already presented in Section 3.5, cleaned surfaces can be recontaminated due to improper storage materials or even by storage in air. The effect of such a recontamination on the SEY is shown in Fig. 12. SEY always increases with storage time for all methods used for packaging. Therefore parts which are cleaned and inserted in the accelerator after longer storage time will need a longer conditioning phase. In addition, the curves show that the adverse effect can be limited by selecting suitable storage conditions. For instance wrapping the parts in aluminium foil before inserting them in a polymer bag improves the situation in comparison with packaging in the polymer bag alone. The results of the SEY behaviour (Fig. 12) correlate well with those found for the carbon contamination increase as a function of storage time (Fig. 10). This fact proves that the high SEY of air-exposed surfaces is related to contamination from airborne hydrocarbons.

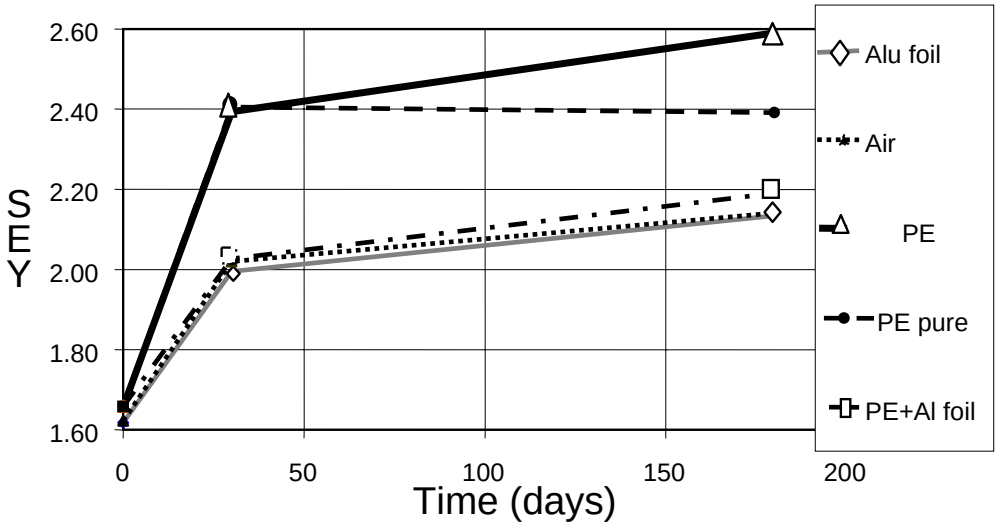


Fig. 12: Secondary electron yield of copper as a function of packaging method and storage time (B. Henrist, N. Hilleret, C. Scheuerlein, M. Taborelli, unpublished)

5 Conclusions

The various methods of cleaning presented above are not a complete and universal recipe, but give some guidelines in order to optimize a cleaning method for a particular application. Even the most accurately selected procedure should be tested on the real parts to be cleaned or on a specimen which is representative of them for shape, size, contamination, and surface composition. The possibility to clean the parts effectively should be considered, and implemented from the initial stage of development, design, manufacturing and assembly process, in order to build 'cleanable' parts. A good rule is always to remember that cleaning is only necessary because at some stage there is contamination added to the parts. Typically this occurs during the manufacturing process. Therefore, the safest way of proceeding is not to rely on cleaning, but to conceive a fabrication process which avoids the use of the most relevant contaminants.

Acknowledgements

I wish to thank J.P. Bacher, C. Benvenuti, S. Calatroni, P. Chiggiato, L. Marques Antunes Ferreira, N. Hilleret and Ch. Scheuerlein for all the stimulating and instructive discussions on this topic.

References

[1] S. Sgobba, Materials for high vacuum technology, an overview, *CAS, Vacuum in Accelerators*, Platja d'Aro, Spain, 2006.

[2] S. Pape-Møller, Beam residual gas interactions, *CAS, Vacuum in Accelerators*, Snekersten, Denmark, 1999.

[3] M. Capeans-Garrido, Aging and materials: Lessons for detectors and gas systems, *Nucl. Instrum. Methods Phys. Res.* **A515** (2003) 73.

[4] A.G. Mathewson, Making it work, *CAS, RF Engineering for Particle Accelerators*, Oxford, UK, 1991.

[5] R. Reid, Cleaning for vacuum service, *CAS, Vacuum in Accelerators*, Snekersten, Denmark, 1999.

[6] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London 1993).

[7] An important contribution to the polarizability, which is related to the dielectric constant, is the orientation term arising from permanent dipoles on the molecules, see reference 6.

[8] C. Tanford, *The Hydrophobic Effect* (Wiley, New York, 1980).

[9] Council Directive 1999/13/EC Annex IIA, Activity 4.

[10] K.J. Middleman, J.D. Herbert and R.J. Reid, Cleaning stainless steel for use in accelerators—Phase I, *Vacuum*, in press.

[11] R. Sherman and W. Whitlock, The removal of hydrocarbons and silicone grease stains from silicone wafers, *J. Vac. Sci. Technol*. **B8** (1990) 563.

[12] R. Sherman, J. Grob and W. Whitlock, Dry surface cleaning using CO_2 snow, *J. Vac. Sci. Technol*. **B9** (1991) 1970.

[13] C. Benvenuti, G. Canil, P. Chiggiato, P. Colin, R. Cosso, J. Guerin, S. Ilie, D. Latorre and K.S. Neil, Surface cleaning efficiency measurements for UHV applications, *Vacuum* **53** (1999) 317.

[14] S. Awad, Ultrasonic cavitations and precision cleaning, *Precision Cleaning* **IV** (1996) 12.

[15] C. Benvenuti, S. Calatroni, P. Darriulat, M.A. Peck, A.-M. Valente, C.A. Van't Hof, Study of the residual resistance of superconducting Nb films at 1.5 GHz, LANL Report LA-13782-C (tha004), *Proceedings of the 9th Workshop of RF Superconductivity*, National laboratory, Los Alamos, 2000, B. Rusnak ed. (Los Alamos, Santa Fe, 1999).

[16] M. Fassone, S. Sgobba, J.P. Bacher, Evaluation of the formation of BN film on stainless steels performed by different industries, CERN EST/SM/SB report Nr.00-11-25, 2000, unpublished.

[17] C. Dias-Soares, S. Ilie, C. Petitjean, Identification de produits silicones présent comme contaminants sur différents supports, CERN, 2003, EDMS Nr.410 981.

[18] SUVA Schweiz, Threshold limit values (TLV) of classical compounds at working places, 1997, N 1903.

[19] A.G. Mathewson, The surface cleanliness of 316 LN stainless steel studied by SIMS and AES, *Vacuum* **24** (1974) 505.

[20] A.G. Mathewson, The effect of cleaning and other treatments on the vacuum properties of technological materials used in ultra-high vacuum, *Il Vuoto* **XVII** (1987) 102.

[21] E. Hoyt, M. Hoyt, R. Kirby. C. Perkins, D. Wright and A. Farvid, Processing of OFE copper beam chambers for PEP-II high energy ring, *Proceedings of the 1995 Particle Accelerator Conference*, Vol. 3 (1995), p. 2054.

[22] H.F. Dylla, Glow discharge techniques for conditioning high-vacuum systems, *J. Vac. Sci. Technol*. **A6** (1988) 1276.

[23] J. Knobloch and H. Padamsee, Microscopic investigation of field emitters located by thermometry in 1.5 GHz superconducting niobium cavities, *Seventh Workshop on RF Superconductivity*, Gif sur Yvette, France, 1995, *Part. Accel.* **54** (1996) no. 1 p. 53.

[24] S. Calatroni, private communication.

[25] Instrumental methods of determining surface cleanliness, SAE International, standard ARP4252 (1989) Warrendale, PA, USA.

[26] M. Farella, G.E. Mitchell and L.A. Bonadies, *Proceedings of the International CFC and Halon Alternatives Conference*, Baltimore, Maryland, 1991, p. 84.

[27] N. Hilleret, C. Scheuerlein, M. Taborelli, A. Brown, M. Baker, An AES study of the room temperature surface conditioning of technological metal surfaces by electron irradiation, *Appl. Surf. Sci.* **202** (2002) 57.

[28] J. Cazeau, The influence of radiation damage (microscopic causes) on the sensitivity of Auger electron spectroscopy and X-ray photoemission spectroscopy, *Appl. Surf. Sci.* **20** (1985) 457.

[29] A. Roche, S. Marthon, J.F. Pie, N. Rochat, A. Danel, M. Olivier, M. Juhel and F. Tardif, Detection and identification of organic contamination on silicon substrates, *Solid State Phenom.* **76-77** (2001) 111.

[30] C. Scheuerlein and M. Taborelli, The assessment of metal surface cleanliness by XPS, *Appl. Surf. Sci.* **252** (2006) 4279.

[31] P.J. Cumpson and M.P. Seah, Elastic scattering corrections in AES and XPS. II. Estimating attenuation lengths and conditions required for their valid use in overlayer/substrate experiments, *Surf. Interf. Anal.* **25** (1997) 430.

[32] J.D. Herbert, A.E. Groome and R.J. Reid, Study of cleaning agents for stainless steel for ultra high vacuum use, *J. Vac. Sci. Technol*. **A12** (1994) 1767.

[33] M.-H. Achard, R. Calder and A. Mathewson, The effect of bakeout temperature on the electron and ion induced gas desorption coefficients of some technological materials, *Vacuum* **29** (1979) 53.

[34] E. Mahner, J. Hansen, D. Küchler, M. Malabaila and M. Taborelli, Ion stimulated gas desorption yields of electropolished, chemically etched and coated (Au, Ag, Pd, TiZrV) stainless steel vacuum chambers and St707 getter strips irradiated with 4.2 MeV/u lead ions, *Phys. Rev. Special Topics, Accel. Beams* **8** (2005) 053201.

[35] ASTM G121 (1998) Preparation of contaminated test coupons for the evaluation of cleaning agents standard on the contamination procedure.

[36] A. Uhlmann, *An Introduction to Ultrathin Organic Films* (Academic Press, New York, 1991).

[37] M. Mantel and J.P. Wightman, Influence of the surface chemistry on the wettability of stainless steel, *Surf. Interf. Anal*. **21** (1994) 595.

[38] O. Groebner, Lifetime, ion-electron trapping, dust, electron cloud, *CAS, Vacuum in Accelerators*, Platja d'Aro, Spain, 2006.

[39] N. Hilleret, C. Scheuerlein, M. Taborelli, The secondary-electron yield of air-exposed metal surfaces, *Appl. Phys*. **A 76** (2003) 1085.

[40] B. Henrist, N. Hilleret and C. Scheuerlein, M. Taborelli, The secondary electron yield of TiZr and TiZrV non-evaporable getter thin film coatings, *Appl. Surf. Sci*. **172** (2001) 95.

The use of NEG pumps and coatings in large vacuum systems: experience and limitations

F. Mazzolini
Elettra - Sincrotrone Trieste S.C.p.A., Trieste, Italy

Abstract
Today Non Evaporable Getter (NEG) materials are commonly used for vacuum production in large vacuum systems, in particular in particle accelerators. Recently, a new technology based on NEG sputtering techniques was successfully developed by CERN and utilized at various synchrotron light sources and ion collider facilities. The aim was to improve the vacuum performance in some conductance limited vacuum systems in order to minimize the gas bremsstrahlung emission and to reduce the impact of the electron-cloud instability discovered in some machines. The most recent results obtained are presented.

1 Introduction

In the last thirty years Non Evaporable Getters (NEGs) have been successfully used for Ultra High Vacuum (UHV) applications. However, the adoption of NEGs in a vacuum system requires an *in situ* bake-out. Moreover, NEGs are not suitable for pumping every gas (e.g. noble gases, methane) and for vacuum systems routinely open to air.

The need for a distributed pumping system in large particle accelerators and storage rings has been met by the adoption of linear Sputter-Ion Pumps (SIPs) or long NEG strips integrated in the vacuum vessels. But the presence of SIP magnetic fields close to the particle orbit may not be acceptable. In addition, the design of the chambers has to be modified with the integration of the longitudinal antechamber for these pumps. A new solution involves the use of traditional 'localized' pumps and a thin film coating of NEG material directly on the vacuum chamber walls [1], [2].

By doing this, three aims are achieved: a distributed pumping system, the reduction of the coated chamber degassing [3], and the simplification of the vacuum vessel design. The NEG coating proves to be very suitable in reducing both the static and dynamic pressure inside the narrow and long vacuum chambers of the insertion devices of a synchrotron radiation storage ring, where the photon-induced degassing is strong [4], [5]. NEGs activated at a lower temperature (180–200ºC for Ti-Zr-V getters) can be used on materials like aluminium and copper [6]. Therefore the conditioning time of extruded aluminium chambers, and the bremsstrahlung radiation caused by outgassing, can be significantly reduced by means of NEG coating [4].

2 NEG pumps

NEG pumps were originally developed and industrialized by CERN and SAES Getters Company. Originally they were produced in strip form, in order to be installed in some of the LEP accelerator vacuum chambers at CERN. In this case, the NEG powder is 'pressed' and stuck on a thin metallic strip. A good choice is Constantan that can act both as support and heater. It should be electrically insulated from the vacuum chamber and it should conduct an adequate current, so electrical feedthroughs are required. Its typical installation is in a longitudinal antechamber connected to a straight, narrow and long vacuum chamber.

The metallic strip can be folded many times in order to obtain a more compact pumping unit, having the same active surface. They are commonly referred to as wafer modules. Both types are illustrated in Fig. 1.

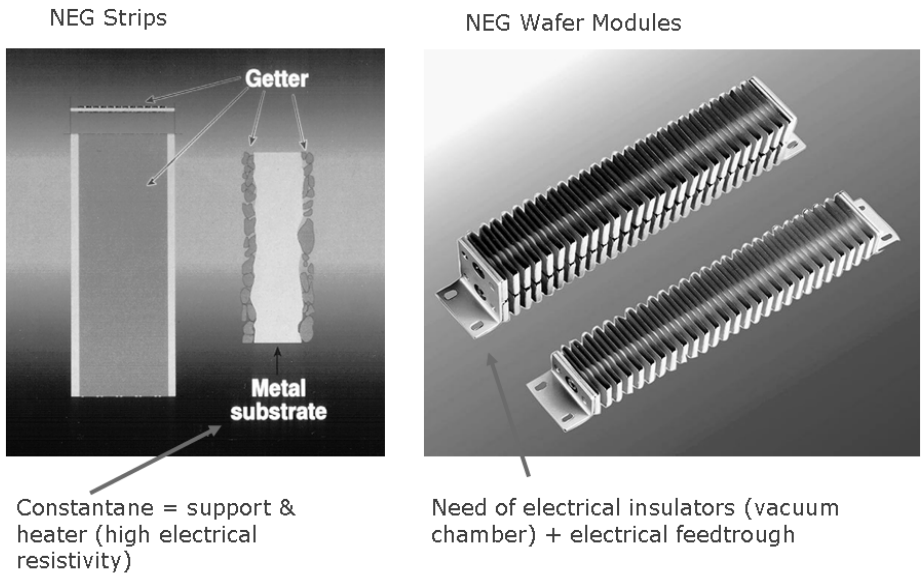


Fig. 1: NEG strip and wafer module

A third type of NEG pump is the NEG cartridge pump. It is based on a wafer module, bent in a cylindrical shape and installed on a Conflat flange. Heaters and a temperature gauge, together with their feedthroughs, are also installed on the same flange, as shown in Fig. 2.

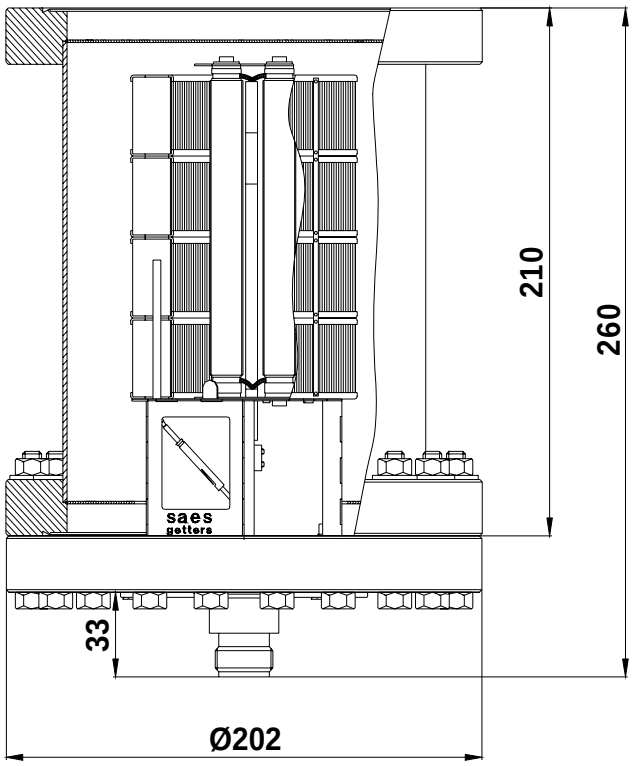


Fig. 2: NEG cartridge pump (SAES GP500)

Usually NEG pumps can be easily integrated in a vacuum system. In particular the cartridge version can be simply added to an already existing vacuum system, because of its quite compact size and light weight. Their main limitation is the activation procedure that implies a heating process up to 350–400°C (in the case of SAES St 707 alloy) and the relatively high initial cost.

Their use offers many advantages:

- a high pumping speed for all active gases (especially for hydrogen);
- a high capacity, that means a very long life. As an example, NEG cartridges installed in the front-ends at Elettra have never been changed after their initial installation more than 12 years ago;
- a very clean vacuum can be obtained, because of the virtual lack of oil, grease and similar contaminants. Theoretically, NEG powder detachment from the cartridge is always possible, but it is not evident in common use;
- their operation is totally vibration-free;
- after the initial activation, their operation does not require any cabling, controller and power consumption;
- the possibility to be operated in the presence of a high magnetic field;
- a reversible behaviour with respect to hydrogen pumping, if required;
- their inability to pump noble gases is essential during a leak test procedure: it can also be conveniently utilized for noble gas purification.

The activation procedure of a NEG pump installed in a vacuum system hosting other vacuum equipment implies many different steps. An example concerning a typical vacuum system of a particle accelerator (that includes ion pumps, NEG pumps, vacuum gauges, residual gas analyser, vacuum valves, etc.) is provided:

- pump the vacuum system down (e.g. by means of a turbomolecular pump connected to the vacuum chamber through a vacuum valve)
- bake the vacuum system; for example at 150°C for 24 hours
- cool the system down to 100°C and hold this temperature
- ‘degas’ all the ion pumps (switch them on–off 3–4 times, until the pressure decreases)
- switch all the pressure gauges off
- activate all the installed NEG pumps at 430°C for 1 hour, as shown in Fig. 3.
- ‘degas’ all the filaments (RGAs, etc.)
- switch all the ion pumps and all the pressure gauges on
- cool the vacuum system down to room temperature
- close the valves connecting the turbomolecular pumps and switch the turbomolecular pump off.

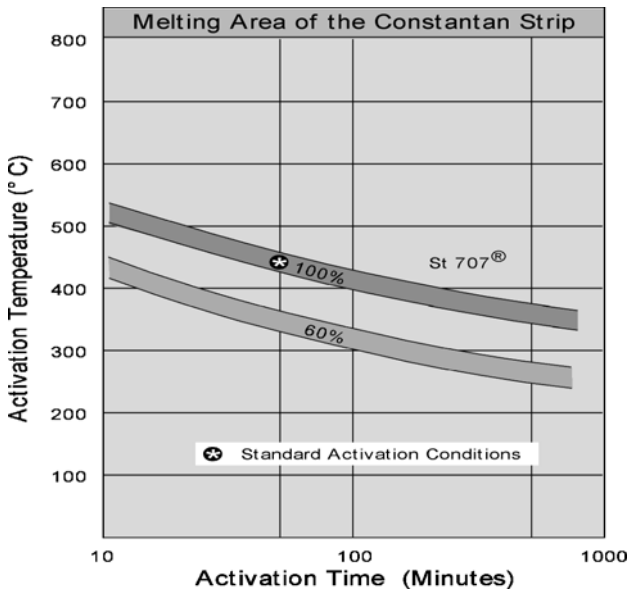


Fig. 3: NEG activation: time vs. temperature for SAES GP series (St707 alloy)

The previous example should be modified according to the complexity of a specific vacuum system, but in principle is always applicable [7].

The two main limitations of the NEG activation procedure for a given vacuum system are the large amount of gas initially released, which must be evacuated by means of appropriate auxiliary pumping units, and the high temperature required for the activation.

3 Towards the NEG coating

Owing to their specific geometry, the vacuum systems of particle accelerators typically require a distributed vacuum pumping. Actually, the vacuum chambers are usually very long, with a reduced cross-section. This is due to the fact that, for technical and economic reasons, the poles of magnets steering the particle beam should be as close as possible to the beam itself, limiting the transverse dimensions of the pipe. In this case, vacuum performance is significantly limited by the poor vacuum conductance of the chamber. Therefore, a practical way to optimize the pressure profile is to distribute the pumping. This can be obtained by replacing big pumps with more, smaller pumps, and reducing the relative distance between them (Fig. 4, distance *L*).

In some specific cases this is not possible because of the lack of space needed to add additional flanges for the added pumps and because of the already existing components (magnets, diagnostic, ancillary subsystems, etc.) that greatly reduce the available space around the vacuum vessel for subsequent new installations.

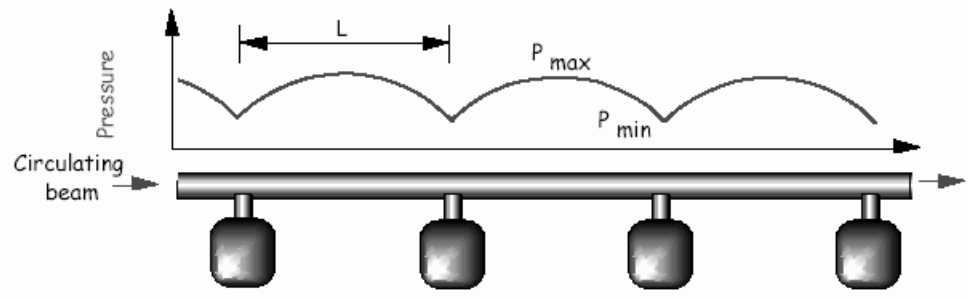


Fig. 4: Pressure profile (discrete pumping)

Another common way to obtain a distributed pumping requires the use of a so-called antechamber, parallel to the main one and connected to it by means of a longitudinal slot. A typical cross-section is represented in Fig. 5. This is a compromise that allows the magnetic pole to stay close to the beam while increasing the cross-section and, consequently, the vacuum conductance of the pipe. But the 'chamber plus antechamber' design, besides the increase of the vacuum conductance, offers the further advantage of facilitating the installation of additional pumps inside the antechamber: linear sputter ion pumps or NEG strips. The first solution is more complex to realize, expensive, and not optimized like a commercially available sputter ion pump. The second offers the advantage of merging the benefits of NEGs with properties of other vacuum pumps. For example, in particle accelerators ultra high vacuum is typically maintained by means of sputter ion pumps because they guarantee an air-tight, closed system even in the case of electrical power failure, they are clean, vibration-free, UHV compatible.

Drawbacks of these solutions are the increase in the dimensions of vacuum vessels; the increase in their complexity; their production cost; the increase in the inner surface and in the consequent thermal outgassing; the difficulty of maintenance in a working particle accelerator.

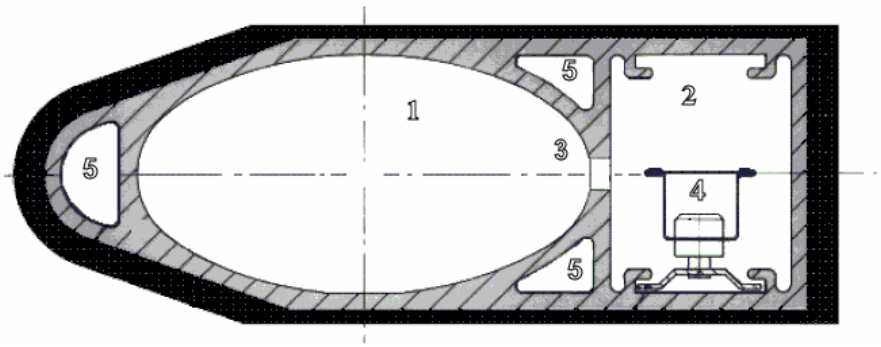


Fig. 5: Cross-section of the LEP dipole vacuum chamber: 1) chamber, 2) antechamber, 3) slot, 4) NEG strip and its support, 5) cooling channels

Almost fifteen years ago, a new NEG technology was developed at CERN. It was based on the deposition of a thin film of NEG material directly on the inner surface of the vacuum vessel by means of sputtering techniques and on the discovery of the new NEG alloy (TiZrV), with a remarkably low activation temperature: 180–200°C instead of 400–450°C of the previous generation of NEG (SAES St707).

The TiZrV coating has the same properties in pumping gases as a traditional NEG material, but it does not require an additional space to be installed. It is also characterized by a very low desorption, both thermal and photon-induced [3]. Moreover it does not require a dedicated heating system, but it can be 'passively' activated during the normal UHV bake-out procedure, owing to its relatively low activation temperature.

This gave the opportunity to develop new types of vacuum chambers, especially suitable for particle accelerators: it was possible to obtain a narrow pipe with a very low, flat pressure profile. Figure 6 shows a qualitative comparison between a pipe pumped by means of two pumps placed at both ends (top curve), the same pipe with two additional pumps in between (middle curve) and by two pumps at both ends and a NEG coating (bottom curve).

The first application of the NEG coating technology to a vacuum vessel installed in a working particle accelerator as a permanent component was the realization of an insertion device vacuum chamber for the European Synchrotron Radiation Facility (ESRF) in 2000 [8], followed by a similar chamber installed at Elettra, the Italian Synchrotron Light Source, at the beginning of 2002 [9].

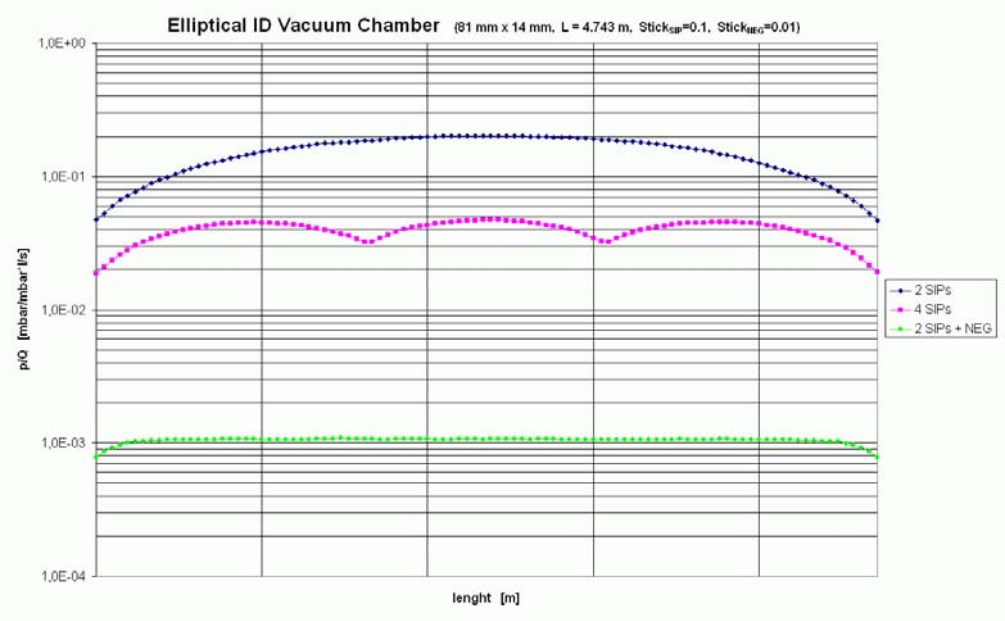


Fig. 6: Pressure profile comparison (see text)

4 Six years' experience in working particle accelerators

It was immediately possible to observe a reduction of the conditioning time of the new chamber and a decrease of the gas bremsstrahlung emission, a clear indication of a very good vacuum performance [10].

In a particle accelerator, the measurement of the gas bremsstrahlung provides indirect information about the total pressure inside the vacuum chamber. Figure 7 shows the fast bremsstrahlung reduction obtained after the installation of a new NEG-coated vacuum chamber at Elettra. It clearly shows the recovery of the original bremsstrahlung level after less then 10 Ah [11].

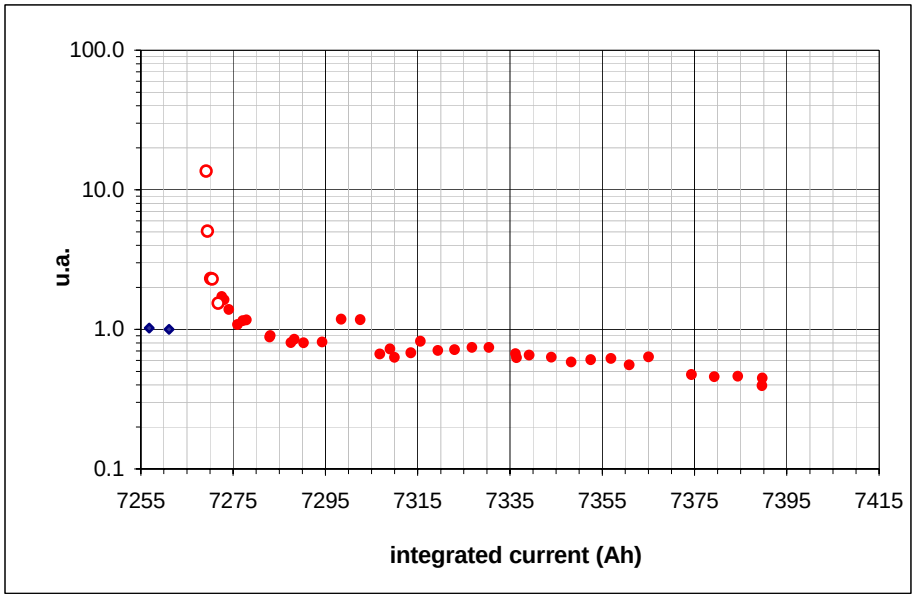


Fig. 7: Gas bremsstrahlung reduction during the commissioning of a NEG-coated vacuum chamber installed at Elettra (K. Casarin, Health Physics Group, Elettra)

"The machine division of the ESRF, in conjunction with Elettra and Soleil, has carried out throughout several years a rather detailed study of the effect of NEG-coated chambers, both made of SS and Al, as far as the resistive-wall impedance budget is concerned, especially on its effect on the single-bunch threshold. The rather poor resistivity of Ti, Zr and V should enhance the imaginary part of the resistive wall impedance, affecting the single-bunch threshold: coating thicknesses have been reduced, from 1μm to 0.5μm (nominal)." [12]. This decrease of thickness did not reduce the vacuum performance, as predicted by P. Chiggiato [13].

In these six years it has also been demonstrated that ageing (the maximum number of venting and reactivations that a NEG coating can safely withstand without noticeable reductions of this vacuum performance) is sufficiently high for normal operations: at least 30 times, but probably more.

In addition, in ion colliders "the most promising remedies to e-cloud seems to use thin film coatings: TiZrV NEG will also bring linear pumping upon activation" [14].

5 Conclusion

While NEG materials have been widely used in vacuum technology since their discovery thirty years ago and their properties have been thoroughly investigated, TiZrV NEG coating techniques and their applications are relatively new. However, there is a huge increase in the adoption of this technology, especially in particle accelerators. In synchrotron light sources an improved vacuum performance, a reduction of commissioning time, and a suitably long lifetime of coatings have been demonstrated over the last six years in different laboratories. Moreover the pressure rise due to the negative impact of electron-cloud instability in ion colliders has been reduced by means of NEG coatings.

Acknowledgements

I wish to thank all the colleagues who enthusiastically shared their experience gained from the use of NEG technology in their Laboratories. In particular I wish to express my gratitude to C. Benvenuti (CERN, Switzerland), who made the NEG coating technology possible, P. Chiggiato (CERN, Switzerland) and R. Kersevan (ESRF, France) who installed a NEG-coated chamber in a working machine for the first time, collecting and sharing the results of their pioneering measurements, J. Setina (IMT, Slovenia) who helped me in organizing the 41st IUVSTA Workshop partially dedicated to the use of NEG coating in particle accelerators, O. Malishev (Astec, UK) and P. Manini (SAES Getters, Italy), who organized the 45th IUVSTA Workshop specifically devoted to the use of NEGs in particle accelerators and vacuum systems.

References

[1] C. Benvenuti, Non-evaporable getters: from pumping strips to thin film coatings, Proceedings of EPAC 1998, Stockholm, June 1998.

[2] C. Benvenuti *et al.*, A novel route to extreme vacua: the non-evaporable getters thin film coatings, Vacuum **53** (1999) 219–225.

[3] C. Benvenuti, P. Chiggiato, F. Cicoria and V. Ruzinov, Decreasing surface outgassing by thin film getter coatings, Vacuum **50** (1998) 57–65.

[4] R. Kersevan, Vacuum system of the ESRF: operational experience and status report, Vacuum **60** (2001) 95–99.

[5] P. Chiggiato and R. Kersevan, Synchrotron radiation-induced desorption from NEG-coated vacuum chambers, Vacuum **60** (2001) 67–72.

[6] C. Benvenuti *et al*., Vacuum properties of TiZrV non-evaporable getter films, Vacuum **60** (2001) 57–65.

[7] A. Rossi, 45th IUVSTA Workshop on NEG coatings for Particle Accelerators and Vacuum Systems, http://www.aiv.it/neg_documenti/session_7/Adriana_Rossi.pdf .

[8] R. Kersevan, Performance of a narrow-gap, NEG-coated, extruded aluminium vacuum chamber at ESRF, http://accelconf.web.cern.ch/accelconf/e00/PAPERS/THP5B11.pdf .

[9] F. Mazzolini *et al*., Performance of the insertion device NEG-coated aluminium vacuum chamber at Elettra, Synchrotron Radiation News, Vol. 15, No. 3, May/June 2002.

[10] F. Mazzolini *et al*., Performance of insertion device vacuum chambers at Elettra, http://accelconf.web.cern.ch/AccelConf/e02/PAPERS/WEPDO026.pdf .

[11] F. Mazzolini, 45th IUVSTA Workshop on NEG Coatings for Particle Accelerators and Vacuum Systems, http://www.aiv.it/neg_documenti/session_5/Fabio_Mazzolini.pdf .

[12] R. Kersevan, 45th IUVSTA Workshop on NEG Coatings for Particle Accelerators and Vacuum Systems, http://www.aiv.it/neg_documenti/session_1/Roberto_kersevan.ppt .

[13] P. Chiggiato, 41st IUVSTA Workshop on Vacuum System Design for Particle Accelerators: A Multidisciplinary Approach, http://www.imt.si/iuvsta/index.html .

[14] F. Le Pimpec, 45th IUVSTA Workshop on NEG Coatings for Particle Accelerators and Vacuum Systems, http://www.aiv.it/neg_documenti/session_4/Frederic_Le_Pimpec.pdf .

Bibliography

For theory, principles and history behind NEGs and NEG coatings see C.Benvenuti's lecture (these proceedings).

For applications of NEGs and NEG coatings to particle accelerators and large vacuum systems, see the: *45th IUVSTA Workshop on NEG Coatings for Particle Accelerators and Vacuum Systems* documents.

In April 2006 the Accelerator Science and Technology Centre (CCLRC Daresbury Laboratory, UK) and the Italian Vacuum Association (AIV), under the sponsorship of the International Union for Vacuum Science Technique and Applications (IUVSTA), organized the 45th IUVSTA Workshop on NEG Coatings for Particle Accelerators and Vacuum Systems. The aim of this workshop was to promote an exchange of information among researchers investigating NEG properties, people involved in the design of vacuum chambers and components, as well as UHV pump manufacturers.

The workshop was organized in seven sessions:

Session 1. NEG coating technology

Session 2. Underlying physics, properties, preparation technology, and future directions.

Session 3. Laboratory measurements: methodologies and experimental results.

Session 4. Electronic properties, surface chemistry, film morphology, film mechanical properties.

Session 5. Experience of NEG-coated vessels installed in working machines.

Session 6. Electron, photon and ion stimulated desorption from NEG films.

Session 7. Use of NEG coating in future projects and further areas of research.

Contributions presented at this workshop, summary, and final report can be found at the following link:
http://www.aiv.it/ita/scuole/neg_aiv.asp

Additional contributions on the use of NEG coatings from the *41st IUVSTA Workshop on Vacuum System Design for Particle Accelerators: A Multidisciplinary Approach* can be found at the following link:
http://www.imt.si/iuvsta/index.html

Cold/sticky systems

V. Baglin
CERN, Geneva, Switzerland

Abstract
The understanding of complex and/or large vacuum systems operating at cryogenic temperatures requires a specific knowledge of vacuum science at such temperatures. At room temperature, molecules with a low binding energy to a surface are not pumped. However, at cryogenic temperatures, their sojourn time is significantly increased, thanks to the temperature reduction, which allows a 'cryopumping'. This pumping mechanism is described by different regimes. Sticking probabilities, capture factor and thermal transpiration concepts are also used to characterize the pumping mechanism. At cryogenic temperature, a gas load into a vacuum system turns into an increase of the surface coverage and of its associated vapour pressure. Some adsorption isotherms of H_2 and He which differ with key parameters such as surface nature and temperature are also presented. As an application of this field of vacuum technology, the vacuum system of the CERN Large Hadron Collider is introduced. The implementation of cryosorbers and the consequences of He leaks in the accelerator beam tube are reported.

1 Introduction

At sufficiently low temperature, the vacuum in vessels can be achieved by the adsorption of the molecules due to the attractive van der Waals forces. One of the first applications of the cryopumping mechanism was the realization of storage containers by Sir J. Dewar after the production of liquid hydrogen in the early 1900s [1]. At the end of the 1950s, the early space projects and the requirement for producing large pumping speed in space simulating vacuum chambers at 80 K gave rise to the building of the first large cryopump cooled by helium at $T < 20$ K [2,3]. In the meantime, with the invention of the Gifford–McMahon process [4] used to refrigerate low-noise amplifiers, the modern cryopump was born.

The Large Hadron Collider (LHC) [5], currently under construction at CERN, is one of the world's largest instrument which uses and contributes to the development of the field of vacuum technology at cryogenic temperature. This storage ring will collide protons at 14 TeV in the centre of mass. By means of 8.6 T superconducting dipole magnets operating at 1.9 K, the energetic protons are kept on their circular orbit 26.7 km long. To allow a proper operation of the machine, the beam life time, dominated by the proton scattering on the nucleus of the residual gas, must be greater than 100 h. To fulfil this requirement, the LHC vacuum system has been under careful study and design for more than a decade [6,7].

In Section 2, the main concepts of the cryo-vacuum are described. Sojourn time, regimes of cryopumping, sticking probability, capture factor, and thermal transpiration are discussed and illustrated by examples taken from the literature.

In Section 3, the adsorption isotherms are described in detail. Some models of adsorption isotherms are introduced. The saturated vapour pressure is introduced. Examples of H_2 and He adsorption isotherms are given. The relevance of some parameters such as gas species, temperature, and the porosity of the substrate are supported by data in the literature data.

How the knowledge of vacuum at cryogenic temperature could be applied to the construction of the LHC, particular, the cryosorbers and the helium leaks in the beam vacuum tube are discussed.

Section 4 outlines the principle of the LHC vacuum system with emphasis on the behaviour of the dynamic vacuum in the LHC cryogenic beam pipe. The requirements, the performances and the operation of the cryosorbers to be installed in some of the superconducting magnets of the machine are presented.

Finally, Section 5 discusses the case of the helium leaks in the LHC vacuum tube. A model of the leak propagation is compared with the experimental data. The consequences of the helium leaks, the possible diagnostic means and the remedies are addressed in the context of LHC operation.

2 Cryopumping

2.1 Desorption from a surface

When a molecule is desorbed from a surface at a given activation energy of desorption (or binding energy) E, its rate of desorption is given by the Frenkel equation

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\,\theta\,\nu_0\,e^{-\frac{E}{kT}} \tag{1}$$

where ν_0 is the frequency of vibration of the molecule, k the Boltzmann constant ($86.17 \cdot 10^{-6}$ eV/K), and T the temperature of the surface. The frequency ν_0 is expected to be ~ 10^{13} Hz at room temperature but can increase to ~ 10^{16} Hz for a localized adsorption of a rather immobile molecule. The equation is written for the first-order desorption in which the desorption rate depends linearly on the surface coverage θ. The first-order desorption applies for physisorbed molecules and for non-dissociated chemisorbed molecules (H_2, N_2 and O_2 are dissociatively chemisorbed on metals) [8].

The desorption process is characterized by the sojourn time τ given in Eq. (2). This relation shows that the higher the temperature, the shorter the sojourn time. This fact is routinely observed during a bake-out of a vacuum system when mainly water molecules are desorbed from the surface above 100ºC. Moreover, the equation shows that molecules with low binding energy, i.e., physisorbed molecules, have significant sojourn time at low temperature:

$$\tau = \frac{e^{\frac{E}{k\,T}}}{\nu_0}\ . \tag{2}$$

Table 1 gives some activation energies of desorption of physisorbed (for low surface coverage) and chemisorbed molecules on technical surfaces. The physisorbed molecules have activation energy of desorption in the range of ~0.1 eV and the chemisorbed molecules are in the range of ~1 eV. The activation energy is measured by recording the pressure while the sample's temperature T is increased at a constant rate β, i.e., $T = T_0 + \beta t$ [9]. At the maximum of the desorption rate, i.e., at the pressure peak of the desorption spectrum, the temperature T_P is recorded and the activation energy of the desorption is computed from Eq. (3). This equation is derived from the derivative of (1) which equals zero at the maximum desorption rate:

$$\frac{E}{k\,T_\mathrm{P}^2} = \frac{\nu_0}{\beta} e^{-\frac{E}{kT_\mathrm{P}}}\ . \tag{3}$$

In the case of physisorbed molecules, a significant sojourn time, above 100 h, is obtained for a temperature of the substrate in the range of 5 to 65 K. So, the pumping of most of the gases by

physisorption is achieved below 65 K. A simple cooling of a metallic surface to the liquid nitrogen temperature (77 K) will only pump water molecules.

In an unbaked vacuum system, the residual gas composition is dominated by water which is chemisorbed at room temperature. Indeed, at room temperature, the sojourn time is 100 h for water while it is 3 minutes for hydrogen and more than 6 months for carbon monoxide. Increasing the temperature of the vacuum vessel above 100ºC reduces the sojourn time of water to a minute, and hence strongly reduces the surface coverage and also the outgassing rate of water at room temperature after bake-out. However, molecules which are strongly bounded to the material's surface can be desorbed by stimulated desorption (in an accelerator for instance). Therefore, a bake-out up to at least 300ºC is required to reduce the stimulated desorption of carbon monoxide molecules with 1.7 eV binding energy.

Table 1: Activation energies of desorption *E* of some physisorbed (with low surface coverage) and chemisorbed molecules

Molecule	***E* [eV]**	**Surface**
Physisorbed		
H_2	0.017	Stainless steel 316 L [10]
H_2	0.065	Stainless steel 316 L [10]
H_2	0.081	Saran charcoal [8]
H_2	0.23	Carbon fibre [11]
Chemisorbed		
H_2	0.9	Stainless steel 304 L [12,13]
H_2O	1.1	Aluminium [14]
CO	1.2	Stainless steel 316 L [15]
CO	1.7	Stainless steel 316 L [15]

2.2 Cryopumping regimes

At cryogenic temperature, a molecule interacting with a surface is physisorbed, owing to the Van der Waals force, if the surface is 'cold enough'. Several regimes of cryopumping are defined.

- Physisorption is the regime of the sub-monolayer coverage. The van der Waals force acts between the adsorbed molecule and the material surface. The binding energies are those given in Table 1. For hydrogen, the binding energy varies from 20 to 65 meV for smooth and porous material. One hour sojourn time is obtained at 5 K and 20 K, respectively. Since the binding energy for physisorption is larger than the heat of vaporization (10 meV for hydrogen), sub-monolayer quantities of all gas can be physisorbed in sub-saturated conditions at their boiling temperature. Pumps relying on physisorption are called cryosorption pumps [16].
- As far as the surface coverage increase, the van der Waals force acts between the molecules themselves. This is the cryocondensation regime. While increasing the surface coverage, a saturation equilibrium between gas adsorption and desorption is reached. This is the saturated vapour pressure. The activation energy of desorption equals the energy of vaporization. It ranges from 10 to 175 meV for hydrogen and carbon dioxide, respectively. Cryocondensation is the pumping of gas due to a phase change into a solid or a liquid (i.e., an ice or frost of the gas). Pumps relying on condensation are called cryocondensation pumps [17].

– The possibility to use a condensable gas to trap a non-condensable gas with a high vapour pressure is called cryotrapping. For instance, argon is used to trap helium and hydrogen in cryopumps. The trapped molecules are incorporated in the condensable carrier so that the equilibrium pressure is significantly lower than in pure physisorption.

2.3 Sticking probability, pumping speed, and capture factor

The sticking coefficient for cryosorption and cryocondensation is the ratio of the average number of molecules which stick when impinging on a cold surface divided by the total number of impinging molecules. The sticking coefficient ignores the effects of the vapour pressure. It varies between 0 and 1. The value is a function of the gas species, the surface nature, the surface coverage, the temperature of the gas, and the surface temperature. Figure 1 shows the sticking coefficient of hydrogen at 300 K incident onto a surface at 3.1 K as a function of surface coverage [18]. The sticking coefficient increases with the surface coverage. For thick surface coverage above 10^{17} H_2/cm^2 (~ 100 monolayers), the sticking coefficient approaches unity. It is then named condensation coefficient since the pumping is in the cryocondensation regime. Figure 2 shows the variation of the hydrogen condensation coefficient onto a surface at ~ 4 K for different temperatures (i.e., energies) of the incoming hydrogen [19]. For low energies, the condensation coefficient is close to one. It is still above 0.8 for hydrogen held at room temperature.

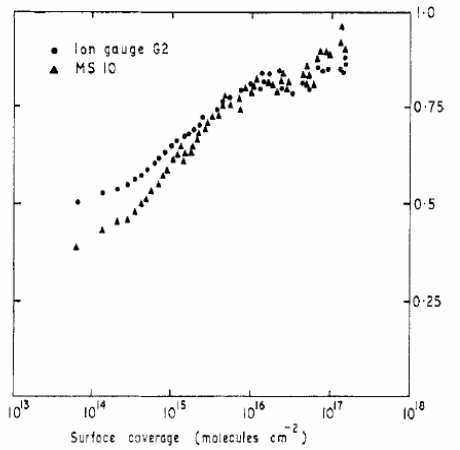


Fig. 1: Sticking coefficient of hydrogen as a function of the surface coverage. The hydrogen is at 300 K incident onto a surface at 3.1 K [18].

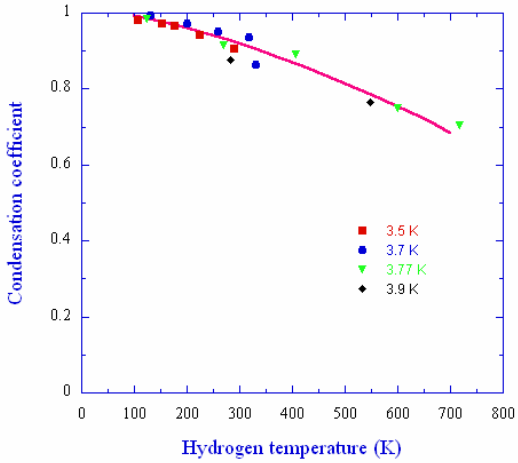


Fig. 2: Hydrogen condensation coefficient onto a surface at ~ 4 K as a function of the hydrogen temperature [19]

Since a cold surface acts as a pump, its pumping speed must be evaluated. If the temperature at the surface of the condensate is so high that the vapour pressure P_{vap} of the solid gas becomes comparable with the pressure of the incident gas P, the pumping speed S of the cryopump is given by Eq. (4) and the effective pumping speed reduces to zero when $P = P_{vap}$. However, most cryopumps do not operate in such conditions so that the pumping speed is simplified to the product of the sticking probability σ times the aperture conductance of the cold surface.

$$S = \frac{1}{4}\sigma\left(1 - \frac{P}{P_{vap}}\right)A\overline{V} \approx \frac{1}{4}\sigma\, A\overline{V} \ , \tag{4}$$

where A is the geometrical area of the surface and $\overline{V}$ the mean molecular velocity. A practical formula of the pumping speed in $l \cdot s^{-1} \cdot cm^{-2}$ is given by

$$S = 3.64\sigma\sqrt{\frac{T}{M}} \tag{5}$$

where T is the temperature of the surface (i.e., when the temperature of the gas is accommodated to the surface temperature) and M the molecular weight of the incident gas. At 4.2 K, the maximum pumping speed of hydrogen and carbon monoxide equals 5.3 and 1.4 $l \cdot s^{-1}cm^{-2}$.

In a cryopump, on account of heat load, the cold surface is shielded from the part of the vacuum vessel held at room temperature which must be evacuated. So the molecules must find their way towards the cold surface to be pumped. This geometrical effect is taken into account by the capture probability. The computing of the capture probability is required to optimize the design of the vacuum system. Monte Carlo methods or angular coefficients methods can be used to compute the capture probability. For instance, the cryopump head is shielded from the radiation heat load by chevron-type baffles. In this case, the pumping speed of the cryopump is reduced compared to the pumping speed of the cold surface. The capture probability is estimated to be 0.3 [20,21]. In another example, the angular coefficients method has been used to optimize the position of the holes in the electron shield with respect to the position of the pumping holes in the LHC beam screen [22].

2.4 Thermal transpiration

To reduce the heat load on the cryogenic system, the measurement of a vacuum system held at cryogenic temperature is usually made by vacuum gauges located in a room-temperature vacuum vessel. In this case, a thermal transpiration correction must be applied (the Knudsen relationship) [23]. When two vessels at two different temperatures T_1 and T_2 communicate by a small aperture, the collision rate is conserved when the steady state is established. By equating the fluxes, and since the mean velocity scales like $\sqrt{T}$, the ratios of the pressure P_1/P_2 and the gas density n_1/n_2 in the two vessels are given by

$$\frac{P_1}{P_2} = \sqrt{\frac{T_1}{T_2}} \quad \text{and} \quad \frac{n_1}{n_2} = \sqrt{\frac{T_2}{T_1}} \ . \tag{6}$$

In practice, when measured at room temperature, the pressure inside a vacuum vessel cooled with liquid helium (4.2 K) is the measured pressure divided by 8. The measured pressure must be divided by two for a vacuum vessel cooled with liquid nitrogen (77 K).

Figure 3 demonstrates the validity of the thermal transpiration correction. The figure shows the adsorption isotherms of hydrogen at 4.2 K measured by a gauge located at room temperature and a gauge immersed in liquid helium at 4.2 K [24]. For ease of comparison, the thermal transpiration correction is applied to the gauge located at room temperature. The pressure is therefore given in both cases for a 4.2 K temperature. As expected from the thermal transpiration theory, the values are similar for the two gauges' readings at least over six decades!

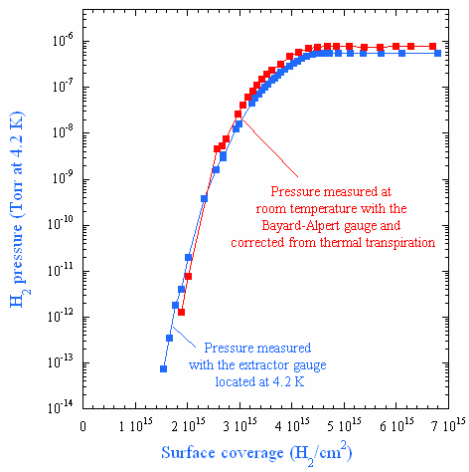


Fig. 3: Adsorption isotherms of H_2 at 4.2 K measured by an ionization gauge located at room temperature and an ionization gauge immersed in liquid helium at 4.2 K [24]

3 Adsorption isotherms

3.1 Some models

The measurement at constant temperature of the equilibrium pressure (vapour pressure) as a function of the surface coverage is the adsorption isotherm. The adsorption isotherm is a function of the molecular species, the temperature of the surface, the nature of the surface, and the gas composition. In the following, some measurements of the isotherms are extracted from the literature to illustrate the effects of the various parameters. The examples are limited to hydrogen and helium since they are the main gases present in a vacuum system at liquid-helium temperature.

There are several semi-empirical adsorption isotherms. At low surface coverage, when there are no lateral interactions between the adsorbed gas molecules, the vapour pressure follows Henry's law. It predicts that the surface coverage θ will vary linearly with pressure P:

$$\theta = cP \ . \tag{7}$$

In the sub-monolayer range, for metallic, glass and porous substrate, the isotherm is better described by the DRK (Dubinin, Raduskevic and Kanager) model. The model is valid at low pressure and offers a good prediction of the isotherm as a function of temperature [25]. The DRK Eq. (8) connects the surface coverage, θ, the monolayer surface coverage, θ_m, the temperature, T, the saturated vapour pressure, P_{sat}, and the pressure, P, via a constant D. A plot of the measured isotherm in the DRK coordinates, i.e., $\ln(\theta)$ vs. ε^2, yields a straight line from which the constant D and the monolayer capacity can be computed:

$$\ln(\theta) = \ln(\theta_m) - D\varepsilon^2 = \ln(\theta_m) - D\left[kT\ln\left(\frac{P_{sat}}{P}\right)\right]^2 \ . \tag{8}$$

The BET (Brunauer, Emmet and Teller) is a multimonolayer description of the isotherm. It links the surface coverage, θ, the monolayer surface coverage, θ_m, the saturated vapour pressure, P_{sat}, and the pressure, P, with a constant α. The constant α is much larger than one so the BET Eq. (9) is simplified. Again, from the measured isotherm, the monolayer capacity can be computed from a linear plot in the BET coordinates $P/[\theta(P - P_{sat})]$ vs. P/P_{sat}. It should be noted that the monolayer surface coverage derived by the BET model is different from the one derived by the DRK model. Despite this, the model covers the entire pressure range from Henry's law to the saturated vapour pressure of the adsorbate; it is only an adequate description of the experimental data in the range $0.01 < P/P_{sat} < 0.3$ [26]:

$$\frac{P}{\theta\,(P - P_{sat})} = \frac{1}{\alpha\theta_m} + \frac{(\alpha - 1)}{\alpha\theta_m}\frac{P}{P_{sat}} \approx \frac{1}{\theta_m}\frac{P}{P_{sat}} \ . \tag{9}$$

3.2 Saturated vapour pressure

The saturated vapour pressure of the gases is the pressure of this gas over its liquid or its gas phase, i.e., when many monolayers of gas are condensed. The saturated vapour pressure follows the Clausius–Clapeyron equation (A and B are constant):

$$\mathrm{Log}(P_{sat}) = A - \frac{B}{T} \ . \tag{10}$$

Figure 4 shows the saturated vapour pressure of the most common gases as a function of temperature [8]. The pressures, corrected from the thermal transpiration, are at room temperature. The figure indicates that below 20 K, the saturated vapour pressures of most of the gases are below 10^{-13} Torr. Therefore, large quantities of such gases can be pumped below 20 K. At liquid-helium

temperature, 4.2 K, only helium and hydrogen are not pumped. At 4.2 K, the saturated vapour pressure of hydrogen is ~ 10^{-6} Torr as already shown in Fig. 3.

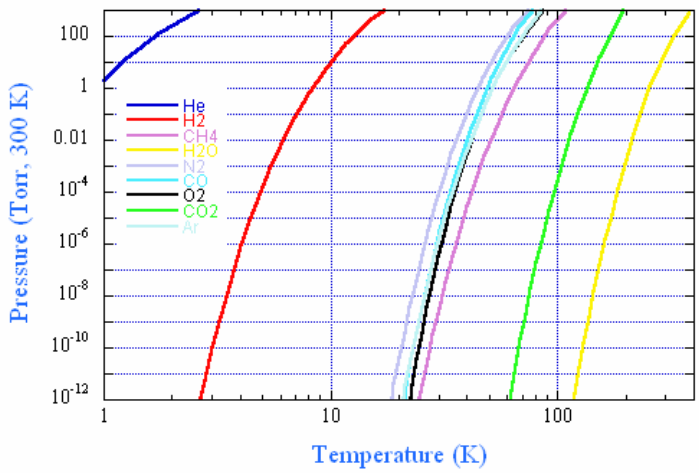


Fig. 4: Saturated vapour pressure curves as a function of temperature [8]

3.3 Hydrogen adsorption isotherms

By pumping on a liquid helium bath, the temperature of the liquid helium can be reduced to allow the pumping of large quantities of hydrogen with a condensation cryopump [27]. The condensation cryopump which operated from 2.3 K (50 Torr on the helium bath) to 4.2 K (atmospheric pressure) can be used to measure the adsorption isotherm of hydrogen as a function of temperature [28].

Figure 5 shows the adsorption isotherm of hydrogen onto a stainless steel surface as a function of temperature. The monolayer capacity is estimated to be $3 \cdot 10^{15}$ H_2/cm^2. Below 3 K, the saturated vapour pressure of hydrogen is negligible. Above 10^{16} H_2/cm^2 the saturated vapour pressure is reached. The results follow the Clausius–Clapeyron law illustrated in Fig. 4 and expand over five decades.

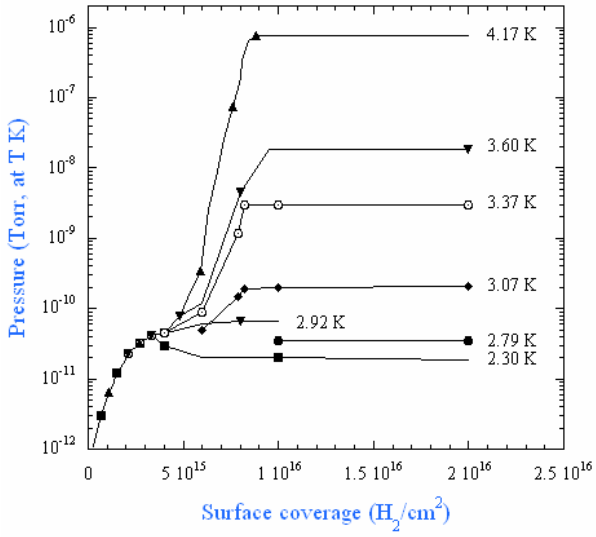


Fig. 5: Hydrogen adsorption isotherm onto a stainless steel surface as a function of the temperature [28]

The addition of a gas can drastically change the shape of the hydrogen isotherm [29]. For example, a frost of carbon dioxide, which is condensed onto the surface prior to the hydrogen injection, forms a porous layer. As shown in Fig. 6, the adsorption capacity of the electroplated Cu

surface is increased. The steep pressure rise is shifted from $2 \cdot 10^{15}$ to $6 \cdot 10^{15}$ H_2/cm_2 due to the carbon dioxide frost. The estimated DRK adsorption capacity of the condensate is 0.3 H_2/CO_2. Similarly, the admission of gas in the vacuum chamber might change the appearance of the adsorption isotherm. This is again the case of carbon dioxide mixed with hydrogen. As shown in Fig. 7, injecting a mixture of 45% H_2 and 55% CO_2 decreases the level of the saturated vapour pressure by 2 orders of magnitude. This phenomenon is applied in cryopumps when carbon dioxide is injected to enhance the pumping of hydrogen and helium by cryotrapping.

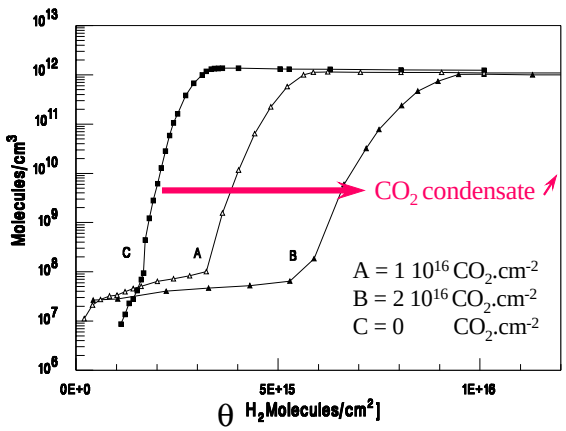


Fig. 6: H_2 adsorption isotherm at 4.2 K on a precondensed layer of CO_2 [29]

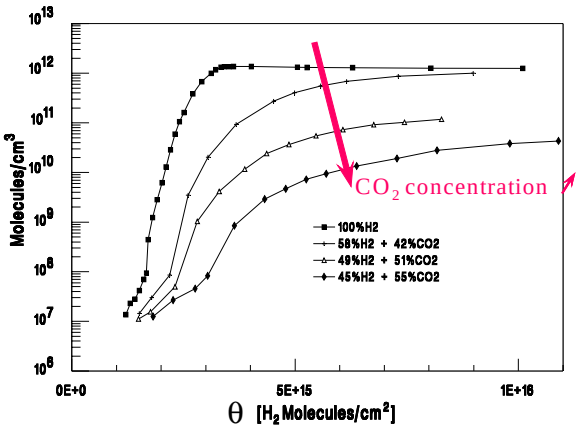


Fig. 7: H_2 co-adsorption isotherm at 4.2 K with CO_2 [29]

3.4 Helium adsorption isotherm

Another gas of interest at cryogenic temperature is helium. As shown by the saturated vapour pressure of helium in Fig. 4, one must operate below 4 K to obtain a low vapour pressure of helium. So, if one wants to perform measurements in UHV condition, one must work in the sub-monolayer range. Thus this is a good test of the validity of Henry's law. Figure 8 shows the helium adsorption isotherm on a stainless-steel tube as a function of temperature [30]. As expected, the isotherms are well described by the DRK model with $\theta_m = 1.3 \cdot 10^{15}$ H_2/cm^2 and $D = 2.92 \cdot 10^4$ eV^{-2}. Below 10^{-9} Torr, the isotherms start to approach the behaviour of Henry's law.

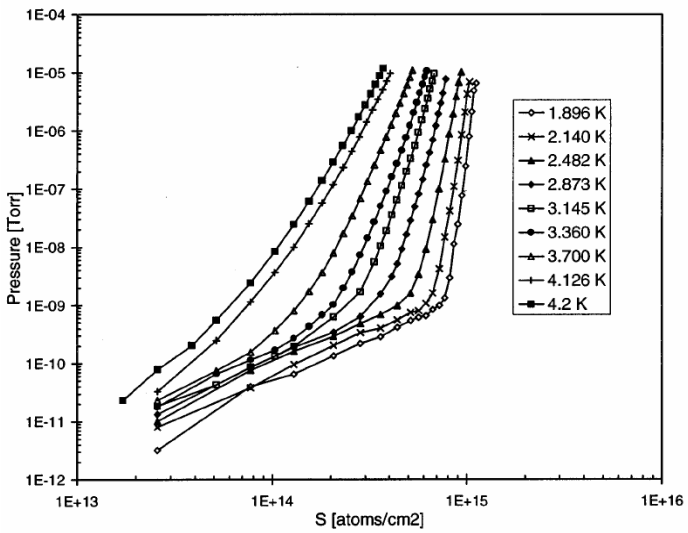


Fig. 8: Helium adsorption isotherm measured on a stainless-steel tube from 1.9 to 4.2 K [30]

3.5 Roughness factor

With the exception of the hydrogen condensation at low temperature, the hydrogen and helium capacities on technical metallic surface are very small, i.e., 10^{14} to 10^{15} molecules/cm^2. Cryosorbing materials are used to increase the capacity, the pumping speed, and the operating range of temperature of vacuum pumps. For instance: activated charcoal is used in cryosorption pumps. Its capacity is ~ 10^{22} H_2/g, i.e., 10^{21} monolayers [8]. Its sticking probability equals 0.3 at 30 K [31]. Thanks to these properties, activated charcoal is mounted onto cryopanels installed at the second stage of a cryopump to allow the pumping of hydrogen and helium at 20 K [16].

The capacity of cryopanels is usually measured with the BET method. From the xenon isotherm at 77 K, the BET monolayer θ_m is obtained. Since the area of the xenon molecules is about 25 Å^2, the roughness factor, i.e., the ratio of the real surface over the geometrical surface is computed. Table **2** shows the roughness factor of several technical surfaces. Metallic surfaces have roughness factors close to unity, i.e., a low surface capacity whereas porous surfaces such as anodized aluminium and Non Evaporable Getter (NEG) St 707 have large roughness factors [32].

Table 2: Roughness factor of some technical surfaces

Technical surface	Unbaked	Baked at 150ºC
Copper Cu-DHP acid etched	1.4	1.9
Stainless steel 304 L vacuum fired	1.3	1.5 (at 300ºC)
Aluminium degreased	3.5	3.5
Sealed anodized aluminium 12 V	24.9	Not measured
Unsealed anodized aluminium 12 V	537.5	556.0
NEG St 707	70.3	156.3

3.6 Carbon fibre cryosorber

Usually cryopanels of cryopumps are made of activated coconut charcoal. In some circumstances and for technological reasons, this material cannot be used. This is the case of the cryosorbers of the LHC which are made of woven carbon fibre due to installation and radiation issues [33]. This cryosorber has the appearance of a fabric with woven carbon fibre. A carbon fibre wire is ~ 1 mm diameter. The wires are woven. The fibres that compose the wire are ~ 10 μm diameter. Each fibre has pores which range from 50 to 500 nm. The pumping speed and the capacity are provided by the trapping of the molecules within these pores. The performance of the cryosorber is a function of the pore size distribution [10,16,34]. Figure 9 shows the hydrogen adsorption isotherms on a woven carbon fibre [11]. The capacity, which is defined here by the amount of gas which can be pumped with an equilibrium pressure not exceeding the LHC operating pressure, ranges from 10^{18} to 10^{17} H_2/cm^2 at 6 K and 30 K, respectively. The roughness factor of such a cryosorber is about one thousand times larger than a metallic surface. Figure 10 shows the sticking probability of the woven carbon fibre. At 10^{18} H_2/cm^2, the sticking probability decreases from 0.3 to 0.1 at 6 K and 30 K, respectively. The values are comparable with those of the activated charcoal.

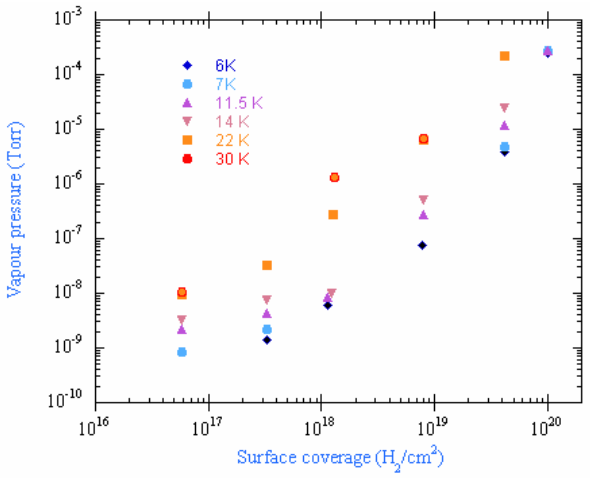


Fig. 9: Hydrogen adsorption isotherms on a woven carbon fibre from 6 K to 30 K [11]

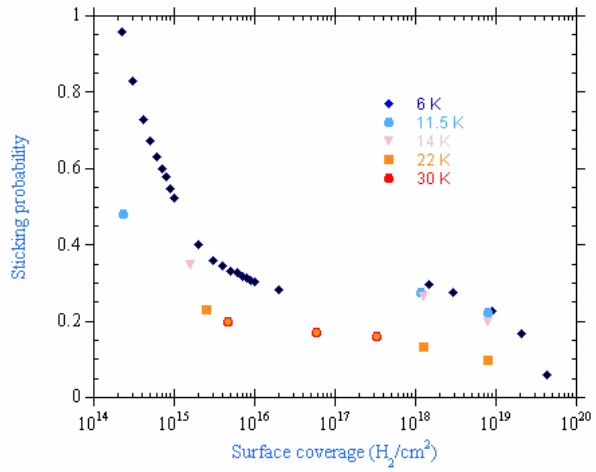


Fig. 10: Sticking probability of the carbon fibre cryosorber, in the 6 K to 30 K temperature range [11]

4 Cryosorbers in cold systems: the case of the LHC superconducting magnets operating at 4.5 K

4.1 LHC arc vacuum system principle

The LHC is a ring of 26.7 km circumference. The ring is divided into eight octants of 2.8 km each. An octant is made of a repetition of cells of superconducting dipoles and quadrupoles. The superconducting magnets operate at 1.9 K. There are eight insertions of 575 m long. The two proton beams of 530 mA and 7 TeV will collide in four of those insertions: the experimental areas.

In the arcs, the vacuum system is made of a cold bore held at 1.9 K surrounded by the superconducting coils assembled in a cold mass. A perforated beam screen operating at 5 K to 20 K is inserted into the cold bore to intercept the thermal loads induced by the circulating beam.

During LHC operation, the proton beams will stimulate molecular desorption by photon, electron and ion bombardment [6,35]. Consequently, the desorbed molecules must be pumped to ensure a beam lifetime of 100 h which is equivalent to a hydrogen pressure of 10^{-8} Torr when measured at room temperature. Figure 11 shows a cross-section of the actively cooled beam screen in the cold bore. The positions of the source of gases are sketched with green arrows. Chemisorbed molecules are desorbed from the beam screen surface by photons, electrons, or ions. These molecules are either physisorbed onto the beam screen surface or pumped through the holes where they are condensed onto the cold bore at 1.9 K. Hydrogen, which has a high vapour pressure (see Fig. 5) cannot be physisorbed in large quantities onto the beam screen and is therefore mainly pumped through the holes. Other molecules, which have low vapour pressure, are physisorbed onto the beam screen. However, scattered photons and electrons can recycle these molecules into the gas phase. When the amount of desorbed molecules is balanced by the amount of physisorbed molecules, an equilibrium pressure and an equilibrium surface coverage are reached [36]. The equilibrium density n_{eq} is driven by the pumping speed of the holes, C, the photon, electron and ion fluxes, $\dot{\Gamma}$, and by the primary desorption yield, η. As a consequence, the LHC vacuum system designer had only to play with the number of holes and the quality of the technical surface to fulfil the design requirements.

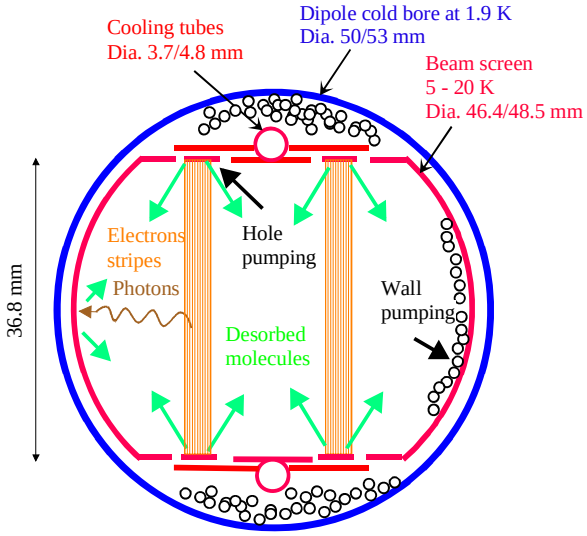


Fig. 11: Cross-section of the LHC vacuum system

$$n_{\mathrm{eq}} = \frac{\eta \dot{\Gamma}}{C} \tag{11}$$

The equilibrium surface coverage on the beam screen θ_{eq} is a function of the monolayer coverage θ_{m} driven by the pumping speed of the beam screen, σS, the holes' pumping speed, C, the primary desorption yield, η, and the recycling desorption η_0' at one monolayer. The equilibrium surface coverage is below a monolayer for low sticking probability and large recycling desorption yield:

$$\theta_{\mathrm{eq}} = \left(\frac{\sigma S}{C} \frac{\eta}{\eta_0'} \right) \theta_{\mathrm{m}} \ . \tag{12}$$

Figure 12 shows the hydrogen dynamic pressure inside a beam screen which is irradiated with synchrotron radiation simulating one third of the LHC design current [37]. The vapour pressure (~ 10^{-10} Torr) is subtracted from the data. Without the pumping holes, the pressure increases due to the recycling desorption by the scattered photons of the physisorbed hydrogen. Since the experimental apparatus has a limited length of 2 m, the hydrogen pressure levels off at 10^{-9} Torr due to the external pumping speed at the extremity. However, if the length had been longer, i.e., in the case of the LHC arcs, the pressure would have continued to increase above the design value. Therefore the beam screen must be perforated to allow the pumping of the recycled hydrogen towards the cold bore. When perforated with 2% holes, the hydrogen pressure levels off at $4 \cdot 10^{-10}$ Torr a value under the LHC design pressure.

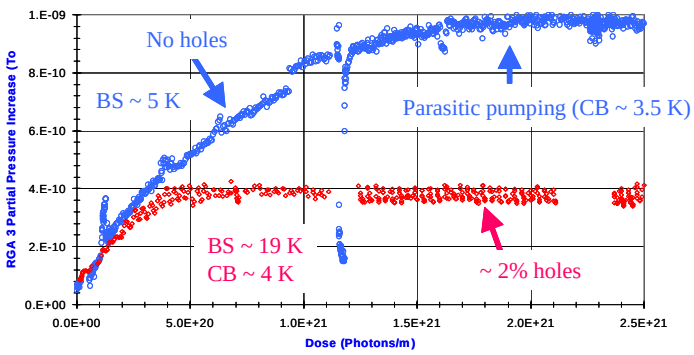


Fig. 12: Comparison of the H_2 photodesorption between a beam screen with and without holes [37]

4.2 LHC long straight sections

In the LHC insertions, the long straight sections, the perforated beam screen technology is still required for cryogenic reasons and for controlling the dynamic vacuum. But, some superconducting magnets operate with a cold bore at 1.9 K and others with a cold bore at 4.5 K. The cumulated length represents 660 m at 1.9 K and 740 m at 4.5 K. The vacuum system of the magnets operating at 1.9 K will behave as described in Section 4.1. However, the long-term operation of the LHC desorbs several monolayers of hydrogen. So, as shown in Figs. 4 and 5, the saturated vapour pressure will be negligible at 1.9 K but will be several orders of magnitude above the design pressure at 4.5 K. Therefore the vacuum system of the magnets operating at 4.5 K requires the insertion of a cryosorber to fulfil the vacuum requirements.

Figure 13 shows a picture of such a beam screen with a cryosorber. In the coaxial space between the beam screen and the cold bore an electron shield is clamped onto the cooling capillary. This electron shield is installed to protect the cold bore from the heat load dissipated by an electron cloud. The cryosorber is a black ribbon installed between the beam screen and the electron shield. By this means, 200 cm^2/m of cryosorber is installed in the narrow space between the beam screen and the cold bore. The chosen cryosorber for the LHC is woven carbon fibre [33].

The vacuum requirements for the LHC cryosorber are the following:

- Installed outside the beam screen in the shadow of the proton beam
- Operation from 5 K to 20 K
- Capacity larger than 10^{18} H_2/cm^2
- Capture coefficient larger than 15%

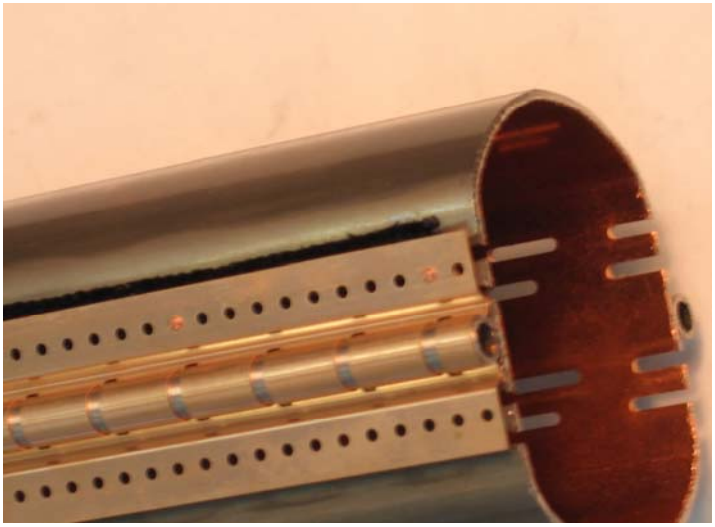


Fig. 13: Photograph of a beam screen equipped with a cryosorber inserted in superconducting magnets operating at 4.5 K

The principle of operation with the cryosorber was checked in a dedicated set-up [38]. In this experiment, the cryosorber was activated charcoal. To study the dynamic vacuum, a beam screen was irradiated with synchrotron radiation while the cold bore was held at 70 K. An equivalent of ~ 100 monolayers of H_2 desorbed from the beam screen was condensed onto the cryosorber prior to irradiation. Under stimulated photodesorption, the hydrogen pressure remains below the LHC design pressure while the beam screen temperature was in the range 15 K to 20 K. So, for a gas load twenty times larger than the design capacity, the cryosorber could maintain the required vacuum performances.

In the LHC, the cryosorbers are installed on the back of the beam screen to provide the required capacity and pumping speed for H_2. These cryosorbers are located in the superconducting magnets operating with a cold bore at 4.5 K. During the annual shutdown, a regeneration of the cryosorber is required. The cryosorber is regenerated at about 80 K. During the regeneration, the beams must not circulate and the beam screen must be warmed up to more than 80 K while the cold bore is held at more than 20 K. This action is required to empty the cold masses. While the hydrogen is desorbed from the cryosorber, it will be evacuated from the vacuum system by the mobile turbomolecular group. This manipulation is expected to last no more than a few days [39].

5 Helium leaks in cold systems: the LHC beam tube case

The LHC operates with superfluid helium at 1.9 K. Several thousand welds are made to assemble the different components of the LHC for its construction. During the design phase, the materials were carefully selected and special care was taken to eliminate welds between UHV vessels and liquid helium vessels. By design, full penetration welds are forbidden along the beam screen cooling capillaries. All the beam screens are leak checked before their insertion into the superconducting magnets. Thus the occurrence of a helium leak in the LHC beam vacuum system is minimized.

If a helium leaks appears in the beam tube held at 1.9 K, the gas is physisorbed locally before any possible detection. As a consequence, pressure bumps, radiation dose and even magnet quench (a transition from the superconducting to the resistive state of the magnet) might appear during machine operation. So, the estimation of the pressure evolution due to helium leaks is of great importance [40,41].

When a helium leak appears, a pressure wave develops with time along the beam vacuum chamber [42]. The He condenses onto the 1.9 K surface up to a monolayer. For larger surface coverage, the He pressure increases to the saturated vapour pressure (2261 Pa). With time, the helium accumulates and the helium wave can span over several tens of metres without being detected. Consequently, the local pressure bump yields to a local proton loss and a risk of quench.

Figure 14 shows a typical evolution with time of the helium pressure wave along the cold bore axis. The helium leak is located at $x = 0$. At $t = 0$, the pressure at the level of the leak, P_{XF}, is simply defined by the ratio of the leak rate Q to the pumping speed of the cold bore aperture. As time goes on, the cold bore is saturated with helium and the effective pumping speed at the level of the leak is decreased. As a result, the pressure at the level of the leak increases. The pressure profile along the cold bore is linear. At $t1$, the pressure at the level of the leak equals $P_{0,t1}$ and the pressure at the level of the front wave [$x = X_F(t1)$] is P_{XF}. At a larger time $t2$ the pressure at the level of the leak has increased and the helium wave arrives at $X_F(t2)$.

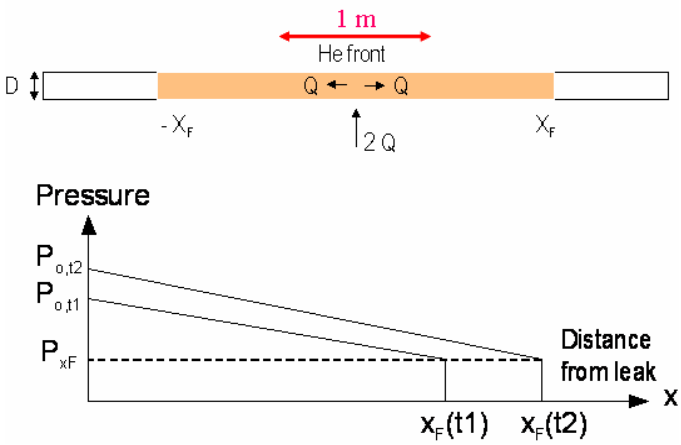


Fig. 14: Evolution of helium pressure with time along the cold bore axis

A model of the helium propagation wave was developed during the design of the Relativistic Heavy Ion Collider (RHIC) and is used for the LHC case [42]. This model of the He propagation wave was validated in a dedicated experiment performed in the LHC test string in a LHC type cold bore at 1.9 K [43]. Figure 15 shows the evolution of the pressure measured at the level of the leak (almost flat curve) and the pressure measured 73.5 m downstream from the leak. The predictions of the model are superimposed on the original data. The green curve is the predicted pressure at the leak and the orange curve is the predicted pressure 73.5 m downstream from the leak. The pressure is given at 1.9 K. At $t = 0$, a leak of $6 \cdot 10^{-5}$ Torr · l/s at room temperature is admitted into the system. It is interesting to observe that for such a leak rate the pressure at the level of the leak is 10^{-6} Torr corresponding to 10^{-5} Torr at room temperature. Therefore, after less than a few minutes, if a proton beam at 7 TeV had interacted with such a gas density, the superconducting magnets would have quenched. However, despite this large leak rate, the helium signal is observed 73.5 m downstream to the leak only 20 h after the opening of the leak. The model predicts the appearance of the leak signal after 17 h which is in good agreement with the observations.

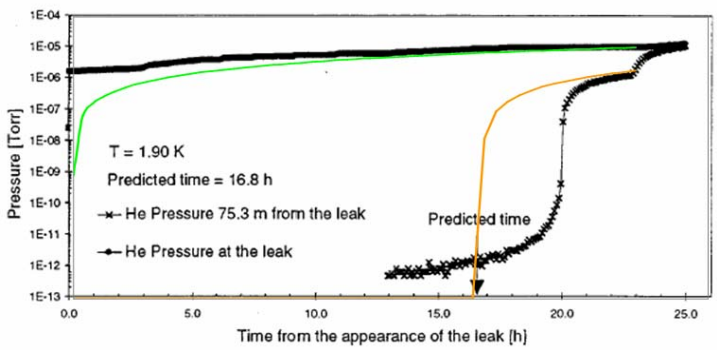


Fig. 15: Example of the evolution with time of a He pressure front observed in the LHC string experiment [43]. The predictions of the model are superimposed on the original data.

At 7 TeV, $8 \cdot 10^{6}$ protons/m/s provoke a magnet quench. So, when the amount of proton losses, due to the nuclear scattering on the helium gas in the vicinity of the leak, approaches this number, the quench of the superconducting magnet is triggered. It can be shown that a helium leak rate above $5 \cdot 10^{-7}$ Torr · l/s must be detected to avoid the risk of a magnet quench. Lower leak rates can be handled by pumping the beam tube each year but, for instance, a leak rate larger than 10^{-5} Torr · l/s triggers a quench within a day. So, the detection of a potential helium leak is of great importance.

In the LHC arc, the minimum distance between two consecutive vacuum gauges is 150 m. This distance is too large to allow a diagnostic of a leak before a quench occurs. So, other means of diagnostics are required. The beam-loss monitors might be used to detect leak rates below 10^{-6} Torr · l/s. The measurement of the dissipated power in the cold masses sounds a realistic means of diagnostic. If 1 W/m can be measured, this cryogenic diagnostic would allow one to detect leak rates up to 10^{-5} Torr · l/s.

In all cases, if a leak is suspected in a given area, an intervention in the tunnel will allow the diagnostic to be improved. When possible, the cold mass temperature must be increased to 4 K to shift the helium front towards the closest short straight section where a vacuum port is located. Afterwards, a residual gas analyser must be installed to perform the leak detection. If the leak needs to be monitored during the beam operation, mobile radiation monitors can be installed to follow the radiation wave associated with the pressure wave [44].

When a magnet has already quenched, it is identified by the triggered diode. During the quench, the cold bore of the magnet is warmed up to more than 30–40 K. So, the helium in the beam tube, if responsible for the quench, is flushed to the nearest unquenched magnet and condensed over a few metres. Again, helium leak detection must be done with a residual gas analyser.

The repair of a leaking magnet is considered in the following way. For a small helium leak rate, a regular warm-up of the cold bore above 4 K and a pump-out of the helium every month allows the machine to operate with a reduced beam current. The time estimate to perform such a job is one day. For larger leak rates, an exchange of the magnet must be foreseen. However, before doing such work, more time must be invested to be sure that the observed quench is indeed due to a helium leak. For this reason, it is imperative that helium has been clearly identified with a residual gas analyser. Moreover, the position of the leak and its rate must be known.

6 Summary

The understanding of the vacuum at cryogenic temperature is required to design and operate large vacuum systems held at cryogenic temperature. Vacuum at cryogenic temperature is encountered in many domains. The areas of fusion technology, space technology, cryogenic technology, and accelerator technology have driven many developments. A few basic concepts and the knowledge required to enter the field of vacuum at cryogenic temperature are recalled in the text. The interaction of a gas molecule with a surface at cryogenic temperature is mainly described by the pumping speed and the adsorption isotherm. These quantities are a function of several parameters: temperature, gas species, etc. In this paper, the characteristics of a cryosorber and of helium have been chosen to illustrate the consequences on the design phase and the operation phase of the LHC vacuum system. First, the insertion of the carbon fibre cryosorber ensures the vacuum performance of the superconducting magnets operating at 4.5 K. Second, in the case of helium leaks, the understanding of the behaviour of the helium gas in the 1.9 K beam tube is mandatory to optimize the exploitation of the LHC.

Acknowledgements

I am indebted to O. Gröbner and the late A.G. Mathewson who led me to the field of vacuum at cryogenic temperature. It is my pleasure also to thank my colleagues of the CERN vacuum group for their support and expertise, with a special credit to B. Jenninger for his technical skills and enthusiasm. The organizers of the school are gratefully acknowledged for having given me the opportunity to present this lecture at the CERN Accelerator School.

Books

The Physical Basis of Ultrahigh Vacuum, P.A. Redhead, J.P. Hobson and E.V. Kornelsen (Chapman and Hall, London, 1968).

Capture Pumping Technology, K.M. Welch (North-Holland, Amsterdam, 2001).

Cryopumping: Theory and Practice, R.A. Haeffer (Oxford Science Publications, Oxford, 1989).

Physical Adsorption of Gases, D.M. Young and A.D. Crowell (Butterworths, London, 1962).

Vacuum Technology, A. Roth (North-Holland, Amsterdam, 1990).

Scientific Foundations of Vacuum Techniques, A. Dushman and J.M. Laferty (John Wiley & Sons, New York, 1962).

References

[1] J. Dewar, *Collected Papers* (Cambridge University Press, 1927).

[2] B.M. Bailey and R.L. Chuang, *Trans. Nat. Vac. Symp.* **5** (1958) 262.

[3] B.G. Lazarev, Je.S. Borovik, M.F. Fedorova and N.M. Zin, *Ukr. Phys. J.* (1957) 176.

[4] W.E. Gifford and H.O. McMahon, *Adv. Cryog. Eng*. **5** (1960) 354.

[5] LHC Design Report, Edited by O. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole and P. Proudlock, CERN-2004-003, 2004, CERN, Geneva, Switzerland.

[6] A.G. Mathewson, Vacuum technology for superconducting colliders, *Proc. of the 15th Particle Accelerator Conf.* (PAC'93), May 1993, Washington, DC, USA.

[7] O. Gröbner, Overview of the LHC vacuum system, *Vacuum* **60** (2001) 25.

[8] P.A. Redhead, J.P. Hobson and E.V. Kornelsen, *The Physical Basis of Ultrahigh Vacuum* (Chapman and Hall, London, 1968).

[9] P.A. Redhead, Thermal desorption of gases, *Vacuum* **12** (1962) 203.

[10] G. Moulard, B. Jenninger and Y. Saito, Industrial surfaces behaviour related to the adsorption and desorption of hydrogen at cryogenic temperature in the LHC vacuum system, *Vacuum* **60** (2001) 43.

[11] V. Baglin, H. Dupont and T. Garcin, Vacuum characterisation of a woven carbon fiber cryosorber in presence of H_2, *Proc. of the 9th European Particle Accelerator Conf.* (EPAC'04), 2004, Luzern, Switzerland.

[12] S. Rezaie-Serej and R.A. Outlaw, Thermal desorption of CO and H_2 from degassed 304 and 347 stainless steel, *J. Vac. Sci. Technol., A* **12** (1994) 2814.

[13] J-P. Bacher, C. Benvenuti, P. Chiggiato, M-P. Reinert, S. Sgobba and A.-M. Brass, Thermal desorption study of selected austenitic stainless steels, *J. Vac. Sci. Technol., A* 167 (2003) **21**.

[14] M. Mohri, S. Maeda, H. Odagiri, M. Hashiba, T. Yamashina and H. Ishimaru, Surface study of type 6063 aluminium alloys for vacuum chamber materials, *Vacuum* **34** (1984) 643.

[15] A. Mathewson, R. Calder, A. Grillot and P. Verbeek, Thermal desorption of CO from some stainless steels, *Proc. 7th Intern. Vac. Congr. & 3rd Intern. Conf. Solid Surfaces*, Vienna, 1977.

[16] Ch. Day, Cryopumping – basics and applications, these proceedings.

[17] C. Benvenuti, Characteristics, advantages, and possible applications of condensation cryopumping. *J. Vac. Sci. Technol.*, **11** (1974) 591.

[18] J.N. Chubb, L. Gowland and I.E. Pollard, Condensation pumping of hydrogen and deuterium on to liquid-helium-cooled surfaces, *J. Phys., D* **1** (1968) 361.

[19] J.N. Chubb and I.E. Pollard, Experimental studies of hydrogen condensation on to liquid helium cooled surfaces, *Vacuum* **15** (1965) 491.

[20] C. Benvenuti and D. Blechschmidt, The optimization of a tubular condensation cryopump for pressures below 10^{-13} Torr, *Proc. Of the 6th Int. Vacuum Congr.*, Kyoto, 1974, *Jpn. J. Appl. Phys.* **2** (1974) 77.

[21] R.A. Haefer, Cryogenic vacuum techniques, *J. Phys., E* **14** (1981) 273.

[22] A.A. Krasnov, Molecular pumping properties of the LHC arc beam pipe and effective secondary electron emission from Cu surface with artificial roughness, *Vacuum* **73** (2004) 195.

[23] A. Dushman and J.M. Laferty, *Scientific Foundations of Vacuum Techniques* (John Wiley & sons, New York, 1962).

[24] V. Baglin. A. Grillot and A.G. Mathewson, Fonctionnement d'une jauge à ionisation à la température de l'helium liquide, Vacuum Technical note 95-01, January 1995, CERN, Geneva, Switzerland.

[25] J.P. Hobson, Physical adsorption isotherms extending from ultrahigh vacuum to vapour pressure, *J. Phys. Chem.* **73** (1969) 2720.

[26] J.P. Hobson, Analysis of ultrahigh vacuum isotherm data with the Brunauer–Emmett–Teller Equation, *J. Vac. Sci. Technol., A* **14** (1996) 1277.

[27] C. Benvenuti, Characteristics, advantages, and possible applications of condensation cryopumping, *J. Vac. Sci. Technol.* **11** (1974) 591.

[28] C. Benvenuti, R.S. Calder and G. Passardi, Influence of thermal radiation on the vapour pressure of condensed hydrogen (and isotopes) between 2 and 4.5 K, *J. Vac. Sci. Technol.* **13** (1976) 1172.

[29] E. Wallén, Adsorption isotherms of H_2 and mixtures of H_2, CH_4, CO, and CO_2 on copper plated stainless steel at 4.2 K, *J. Vac. Sci. Technol., A* **14** (1996) 2916.

[30] E. Wallén, Adsorption isotherms of He and H_2 at liquid He temperatures, *J. Vac. Sci. Technol. A* **15** (1997) 265.

[31] T. Satake, M. Hashiba, M. Mohri and Y. Yamashina, A fundamental study of cryopumping systems with charcoal sorption panel, *Fus. Tech.* **6** (1984) 511.

[32] V. Baglin, Mesure de la rugosité de surfaces techniques à l'aide de la méthode BET, Vacuum Technical note 97-03, January 1997, CERN, Geneva, Switzerland.

[33] V.V. Anashin, I.R. Collins, R.V. Dostovalov, Z.A. Korotaeva, A.A. Krasnov, O.B. Malyshev and V.A. Poluboyarov, Vacuum performance of a carbon fibre cryosorber for the LHC LSS beam screen, *Vacuum* **75** (2004) 293.

[34] D.W. Sedgley, A.G. Tobin, T.H. Batzer and W.R. Call, Characterisation of charcoals for helium cryopumping in fusion devices, *J. Vac. Sci. Technol., A* **5** (1987) 2572.

[35] N. Hilleret, Non-thermal outgassing, these proceedings.

[36] V.V Anashin, G. Derevyankin, V.G. Dudnikov, O.B. Malyshev, V.N. Osipov, C.L. Foerster, F.M. Jacobsen, M.W. Ruckman, M. Strongin, R. Kersevan, I.L. Maslennikov, W.C. Turner and W.A. Lanford, Cold beam tube photodesorption and related experiments for the Superconducting Super Collider Laboratory 20 TeV proton collider, *J. Vac. Sci. Technol., A* **12** (1994) 1663.

[37] V. Baglin, I.R. Collins, O. Gröbner, C. Grünhagel and B. Jenninger, First results from COLDEX applicable to the LHC cryogenic vacuum system, *Proc. of the 7th European Particle Accelerator Conf.* (EPAC'00), 2000, Vienna, Austria.

[38] V. Baglin, I.R. Collins, O. Gröbner, C. Grünhagel and B. Jenninger, Crysorber studies for the LHC long straight section beam screens with COLDEX, *Proc. of the 8th European Particle Accelerator Conf.* (EPAC'02), 2002, Paris, France.

[39] V. Baglin, Vacuum transients during LHC operation, *Proc. of the LHC Project Workshop* – Chamonix XIII, 2004, CERN-AB-2004-014 ADM, CERN, Geneva, Switzerland.

[40] V. Baglin, How to deal with leaks in the LHC beam vacuum. *Proc. of the LHC Project Workshop* – Chamonix XIV, 2005, CERN-AB-2005-014, CERN, Geneva, Switzerland.

[41] V. Baglin, He leaks in the CERN LHC beam chambers operating at cryogenic temperatures, *Vacuum* **81** (2007) 803.

[42] J.P. Hobson and K.M. Welch, Time-dependent helium and hydrogen pressure profiles in a long, cryogenically cooled tube, pumped at periodic intervals, *J. Vac. Sci. Technol., A* **11** 1566 (1993).

[43] E. Wallén, Experimental test of the propagation of a He pressure front in a long, cryogenically cooled tube, *J. Vac. Sci. Technol., A* **15** (1997) 4949.

[44] V. Talanov, V. Baglin and T. Wijnands, Radiation monitors as a vacuum diagnostic in the room temperature parts of the LHC straight sections, *Proc. of the 10th European Particle Accelerator Conf.* (EPAC'06), 2006, Edinburgh, Scotland.

Vacuum controls and interlocks

P. Strubin
CERN, Geneva, Switzerland

Abstract
The vacuum control system is, in most cases, a subset of the general control system of an accelerator. As such, it shares the architecture and communication infrastructure of the main control system. Considered as a 'slow process' to control in the frame of accelerators, the vacuum control system can be built using commercial industrial controllers (PLCs). A data driven approach allows for changes in configuration without changing the software code but at the expense of a solid database. Modelling the equipment allows for easy adaptation of a variety of control units with the same functionality but different physical interfaces. It also allows for a uniform display of the available data and status values. Interlocks are required to protect the vacuum equipment itself against abnormal conditions, but also to protect other systems, like RF, which need a good vacuum to operate. They are an integral part of any vacuum control system.

1 Introduction

The vacuum control system is, in most cases, a subset of the general control system of an accelerator. As such, it shares the architecture and the communication infrastructure of the main control system, but often requires specific solutions at the process level to accommodate a large variety of equipment (pumps, gauges, valves, interlocks, bakeout controllers, etc.). In an environment like CERN, running many accelerators (Fig. 1) over many years, an additional challenge is to follow the evolution of the control techniques while minimizing the modifications to the hardware.

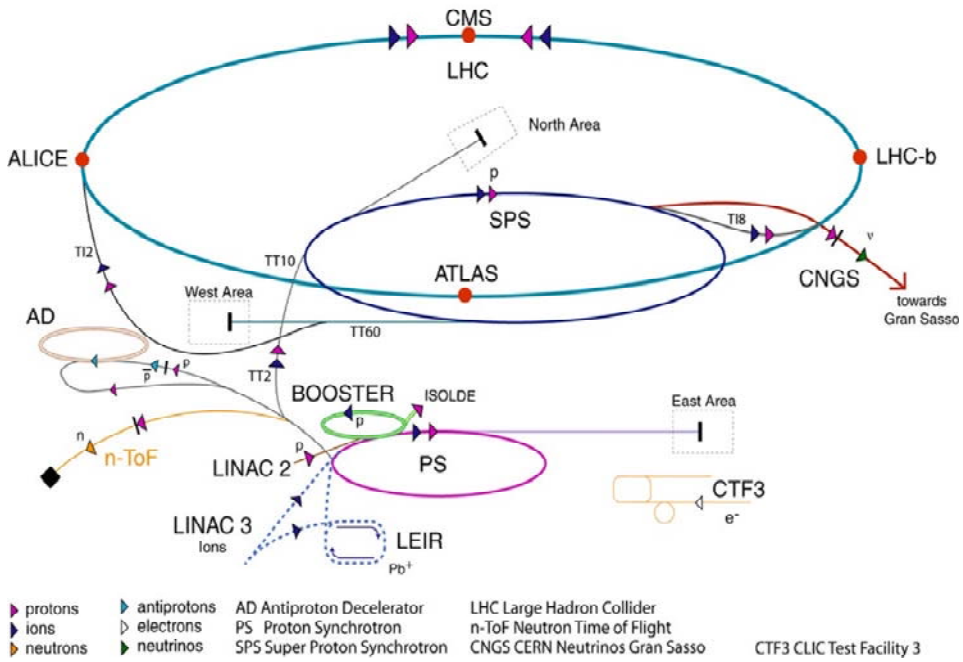


Fig. 1: The CERN accelerator chain

The format of presentation of the acquired data varies, depending on the final end-user and must allow for easy operation by machine operators as well as detailed diagnostic and fault hunting for the vacuum specialist.

Interlock strategy and implementation is also an important part of the vacuum control system to ensure the safe operation of a vacuum system.

Finally, unlike the control systems for most accelerator sub-systems, the vacuum control system must be up and running at all times, including shutdown periods.

2 Physical data to process and control

There is a large variety of equipment to control in a typical vacuum system, ranging from pumping and measuring devices (mechanical or static) to valves (separation or gate valves, roughing valves, venting valves, etc.). The equipment is most often defined by its functionality (e.g., a pumping station) and consists of an assembly of a number of individual components or sub-assemblies, including a control unit or power supply. From a vacuum user's point of view, it is enough to know the global status of the equipment during normal operation (e.g., off, pumping, etc.). Figure 2 gives an example of the possible sequence of the states of a pumping station [1]. But it may also be necessary to obtain more detailed information about the status of one component of an assembly for troubleshooting. The designer of the controls equipment must have access to every available signal and the control system, in particular the software, must allow for both cases.

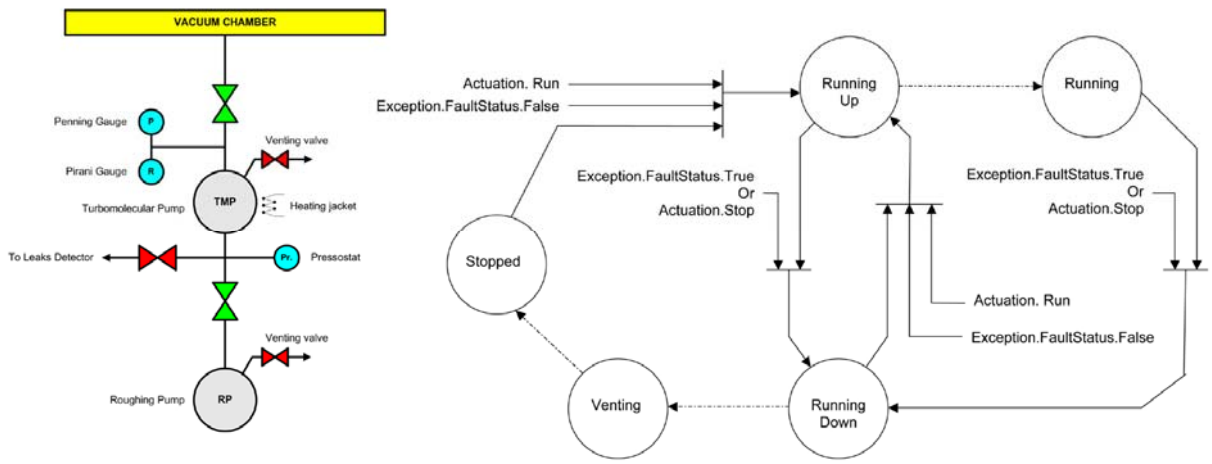


Fig. 2: Possible states of a pumping station

What is illustrated here is that the state of this particular equipment depends on different stimuli: an external action like a command (here referred to as 'actuation'), the status of subcomponents or faults (here referred to as 'exception'), but also external conditions and time. The latter is the case of the transition from the 'Running Up' state to the 'Running' state, illustrated by a turbo-molecular pump reaching its nominal speed.

On the measurement front, the main datum to be acquired is obviously pressure and, for more demanding applications, the composition of the residual gas. In many cases, the pressure is evaluated by the measurement of another signal (ion current, temperature, etc.), via a control unit. In this case also, the vacuum user must receive a meaningful end value, whereas the controls specialist must be able to read all signals. Table 1 shows an example of the required variables.

Table 1: Example of the required variables

Variables	Type	Values
Actuation	Discrete	Run, Stop
Status	Discrete	Stopped, RunningUp, Running, RunningDown, Venting
Exception	Structure	
FaultStatus	Boolean	False, True
FaultList	Set	TMOverheating, RPOverheating, ValveStuck, etc.
WarningStatus	Boolean	False, True
WarningList	Set	LowTMSpeed, etc.

3 Architecture

3.1 Definitions

The most common architecture in modern control systems is the so-called 'three tiers' architecture. The main feature of this architecture is a layered approach, allowing for a more generic approach for common services, like the communication layers or the application program interface. Figure 3 shows the way CERN implements this architecture for cryogenics controls [2], and Fig. 4 shows an example taken from DESY [3].

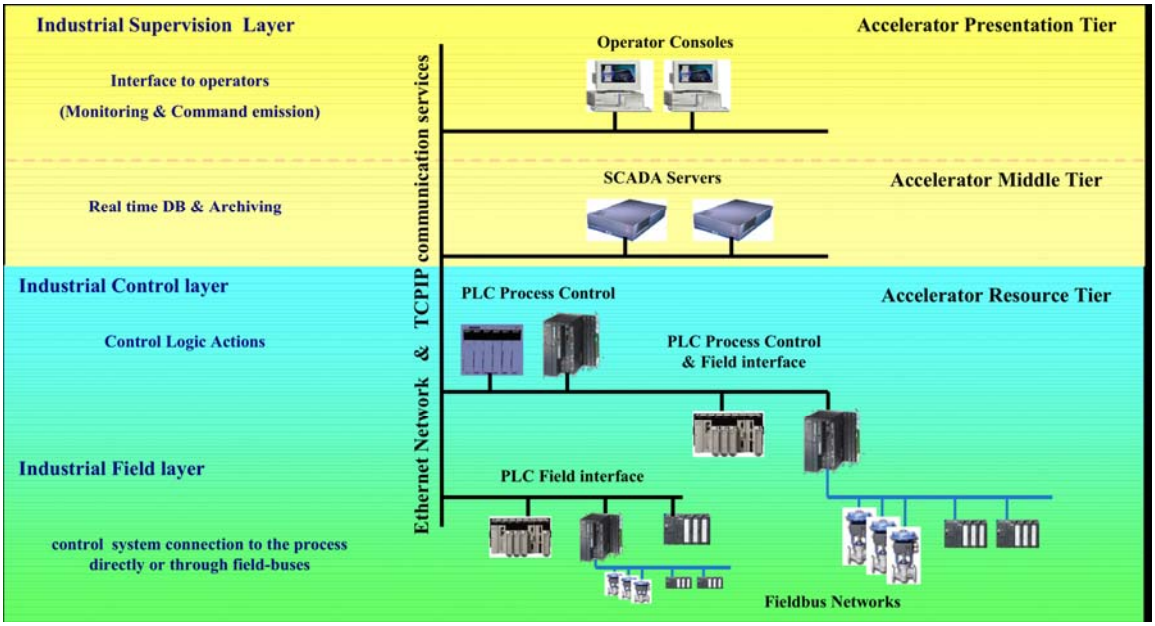


Fig. 3: Cryogenics controls architecture

The first example emphasizes the hardware layers, whereas the second one is more software oriented. It is not always evident to map the various software layers to physical ones. In general, however, it is more important to have clearly defined software layers and interfaces. The actual implementation of these layers matters less and, in any case, should not be the concern of someone developing vacuum controls (or any other process control, like cryogenics) who will concentrate on the lowest layer, close to the equipment.

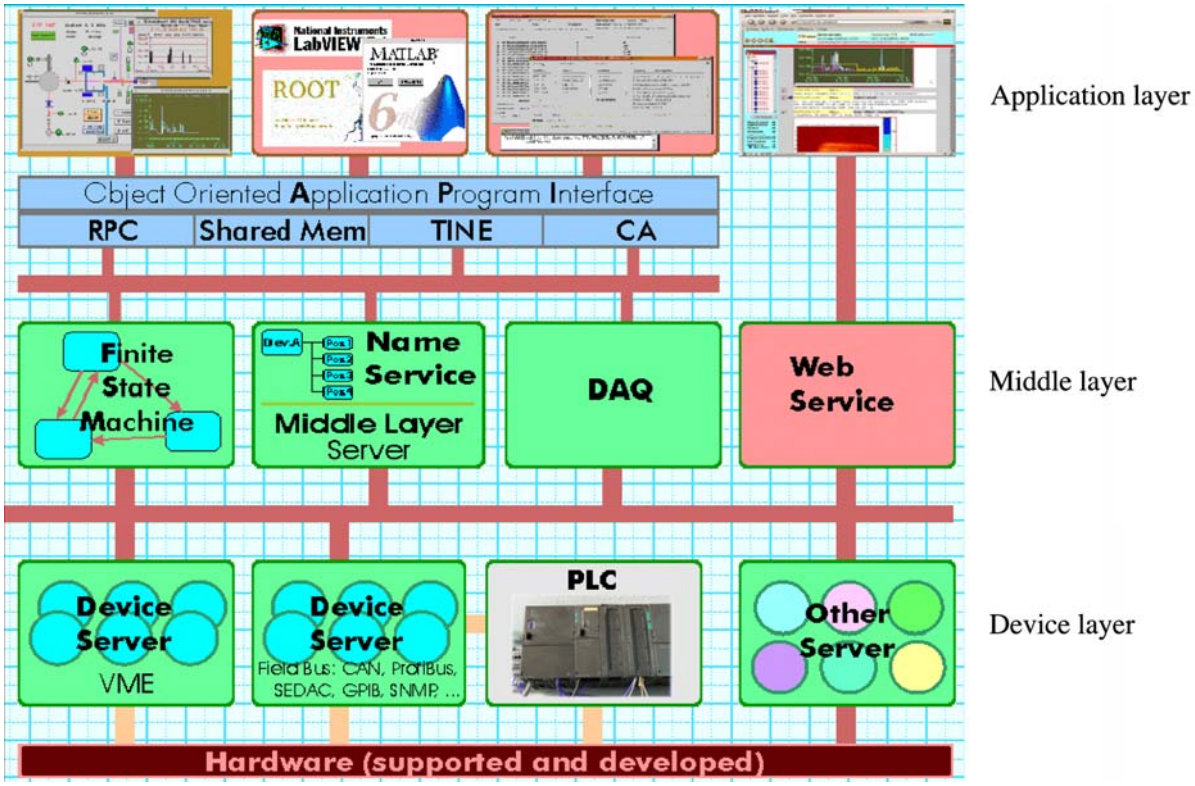


Fig. 4: Example taken from DESY

3.2 Standards

The driving force in modern control systems is the use of industrial standards (e.g. TCP/IP for communication protocols or Web-based servers for graphical interfaces) and industrial equipment (e.g., programmable logic controllers or PLCs for process control and data acquisition). The advantage is of course less expense, multi-vendor solutions as compared to custom-made development. The maintainability can also be easily outsourced. These solutions may not, however, always cover the needs, in particular for fast controls (e.g. beam diagnostic and beam steering), but they are suitable for vacuum controls. On the negative side, it is important to realize that using widely used and documented communication standards makes the control system more vulnerable to hackers. It already happened at CERN that the PLC programs controlling the cryogenic plant used to test the super-conducting magnets for the LHC were blocked by an outside attack. Similarly, it was found that malicious persons were remotely using the computing resources of a digital oscilloscope (essentially an industrial PC). In both cases, the equipment was connected to the main CERN backbone Ethernet network, using TCP/IP communication protocols. It therefore becomes very important to protect the networks. Using proprietary protocols like Profibus™ at the equipment level also helps.

3.3 Models

Another common feature of modern control systems is to develop an ‘object-oriented’ or ‘model-based’ approach, where the actual hardware is hidden behind a virtual representation which describes the data provided or requested by the equipment as well as the dynamic behaviour of the equipment, independently from the hardware. There exist many different implementations of the same concept, often motivated by the necessity to stay compatible with previous implementations of the control system and allow for a smooth transition. The layered approach allows for this and also takes into account that the lifetime of the lower layer (typically at least 10 to 15 years) by far exceeds the lifetime of the upper layer (often less than 5 years).

3.4 Architecture for LHC vacuum controls

CERN vacuum controls have been modernized many times during the 50 years of existence of the laboratory. Since 2000 and following the general trend, there is a strong push to get away from custom-made solutions towards industrial controllers (PLCs). The first vacuum control system to be converted to PLC was for the SPS, which then became the prototype for the future LHC [4]. The architecture follows the standards described above. Figure 5 shows the upper two layers of the 'three tiers' approach. The main part of the middle layer is the central data server (here a PVSS server). It allows for a common data representation to all users, from the vacuum specialist to the database and logging programs. This server communicates with so-called 'master PLCs' over Ethernet.

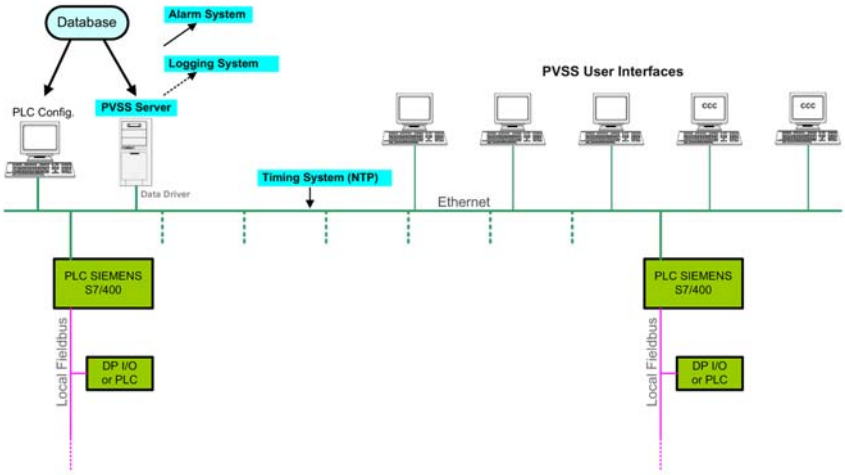


Fig. 5: Upper layers of the LHC vacuum control system

Figure 6 shows the way the lower layer, or equipment layer, has been implemented. The S7 PLC family from Siemens has been chosen in this case, together with Profibus™ and the Siemens proprietary communication protocol MPI. Not only is there a large diversity of equipment to connect, but some equipment with identical functionality, like Pirani or cold cathode gauges, comes with different hardware configurations. This is required in the LHC to minimize costs by using active gauges (sensitive to radiation) where the radiation level is low and passive gauges with a remote controller elsewhere. Slave PLCs are used to connect to the equipment which require a local control process (like a pumping station).

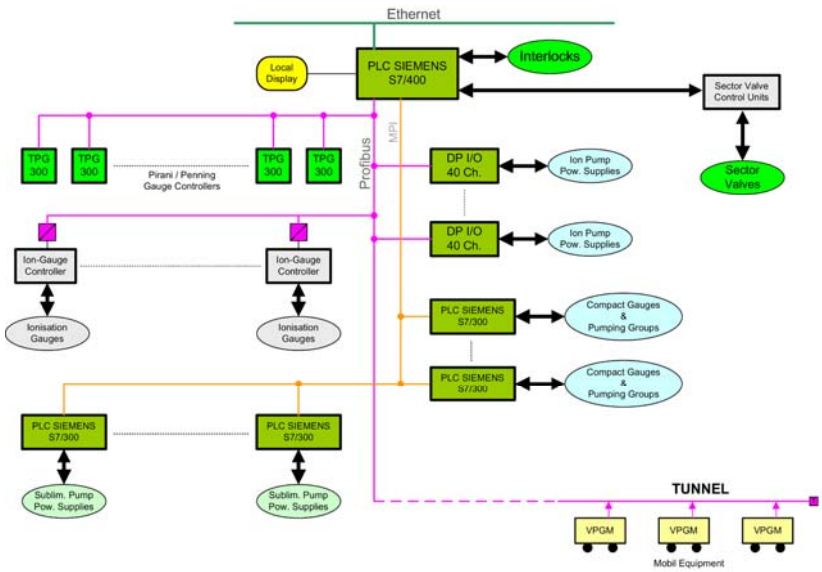


Fig. 6: Lower or equipment layer of the LHC vacuum system

Other equipment, in general self-protected, is connected to the master PLC using deported input–output modules. Finally, some equipment controllers come with a compatible interface and can be connected to the master using Profibus™.

One specificity of large modern accelerators is the use of mobile pumping and diagnostic means to reduce the cost. This requires that the control system be dynamically configured when mobile equipment is connected to it.

4 Describing the equipment

In order to minimize the variants of the applications driving or reading the vacuum equipment, it is necessary to have a uniform description of the properties of the equipment, independent of the actual hardware, as suggested in the 'object-oriented' approach. This technique was successfully implemented in LEP, where most of the equipment was described in generic control models. These models describe the data type, the data flow and the dynamic behaviour resulting from either actuation or an internal or external fault or protection. The models can be translated into 'classes' in the controls environment, allowing for inheritance of attributes and methods. The attributes describe the data flowing in or out of the model (commands, set values, raw and processed measurements, etc.). For example, an ion gauge is a sub-class of a generic class 'vacuum gauge' (a device which has at least the attribute 'pressure') [5]. Figure 7 shows the example of an ion gauge. The important feature of this approach is that the object to read or to control is the gauge, but this happens via the power supply or control unit. This is transparent to the user.

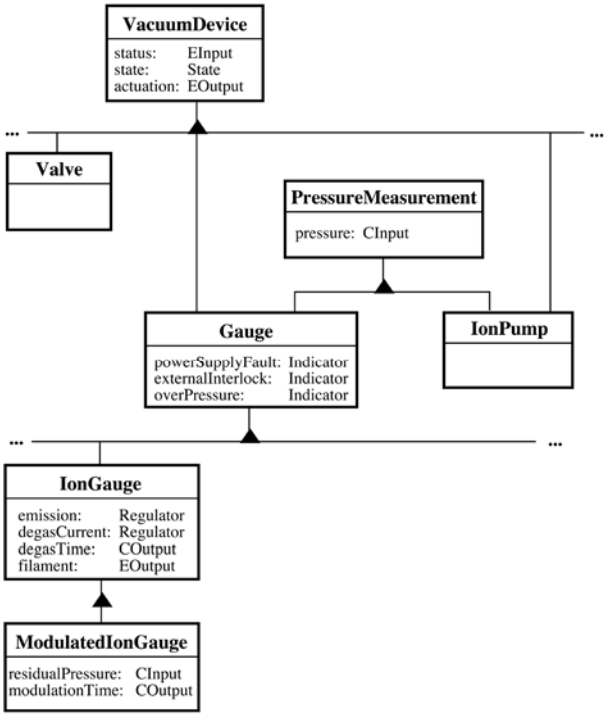


Fig. 7: Control model for an ion gauge

This approach is independent of the control system. Analysis of the control models reveals the common functionality and capabilities between vacuum devices. This leads to the definition of a class hierarchy in which the common features are placed at the top. Specialization of these generic classes provides the interface to more specific vacuum equipment, like gauges. Common functionality, like pressure measurement, can be added as another abstract class, from which 'gauges' would inherit. Finally, further specialization will lead to the classes representing the real equipment.

5 Mapping the model to the PLC

Once the attributes and the behaviour of an equipment have been defined in an abstract way, one needs to map these models to the real hardware. In the case of PLCs, the attributes are mapped to so-called 'data blocks' which contain the data at the level of the memory of the PLC. The way this has been implemented for the future LHC (and is already in use in the SPS and LEIR) is completely data driven, that is, the mapping between defined values and the location to access them is described in a database and can be automatically downloaded to the target PLC. The behaviour of the equipment, or how the equipment reacts to commands or faults, is controlled by small program loops. Figure 8 describes the global logical structure used. The structure is organized by the direction of flow of the data. Commands (actuations or changes of internal parameters, like the sensitivity of a gauge) are grouped in one type of data block ('Write Data Block'), acquired or computed values, as well as internal settings and timestamps are grouped in a second type of data block ('Read Data Block'). The 'Function Block' controls the behaviour of the equipment and how the physical values of interest (e.g. pressure) are computed. The differences between different physical devices of the same type (e.g., a cold cathode gauges) are described in a type of additional data block ('DEV_DB').

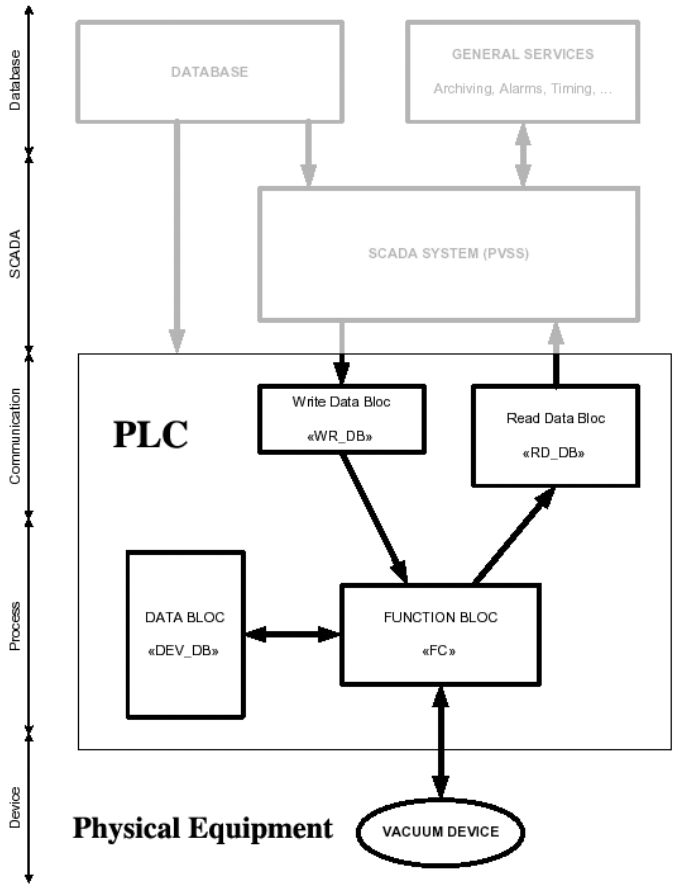


Fig. 8: Logical structure of the vacuum controls at CERN

The above description covers 'classes' of equipment. The PLC controller will manage a number of instances of every class. Hence, for each instance of equipment, a set of data blocks has to be defined and created in the memory of the PLC. Therefore additional information is required to define the physical location of the data blocks for every single instance of a piece of equipment. This is also done via the 'DEV_DB' data block. Figure 9 shows the structure and mapping of the attributes the 'DEV_DB' data blocks for some typical devices.

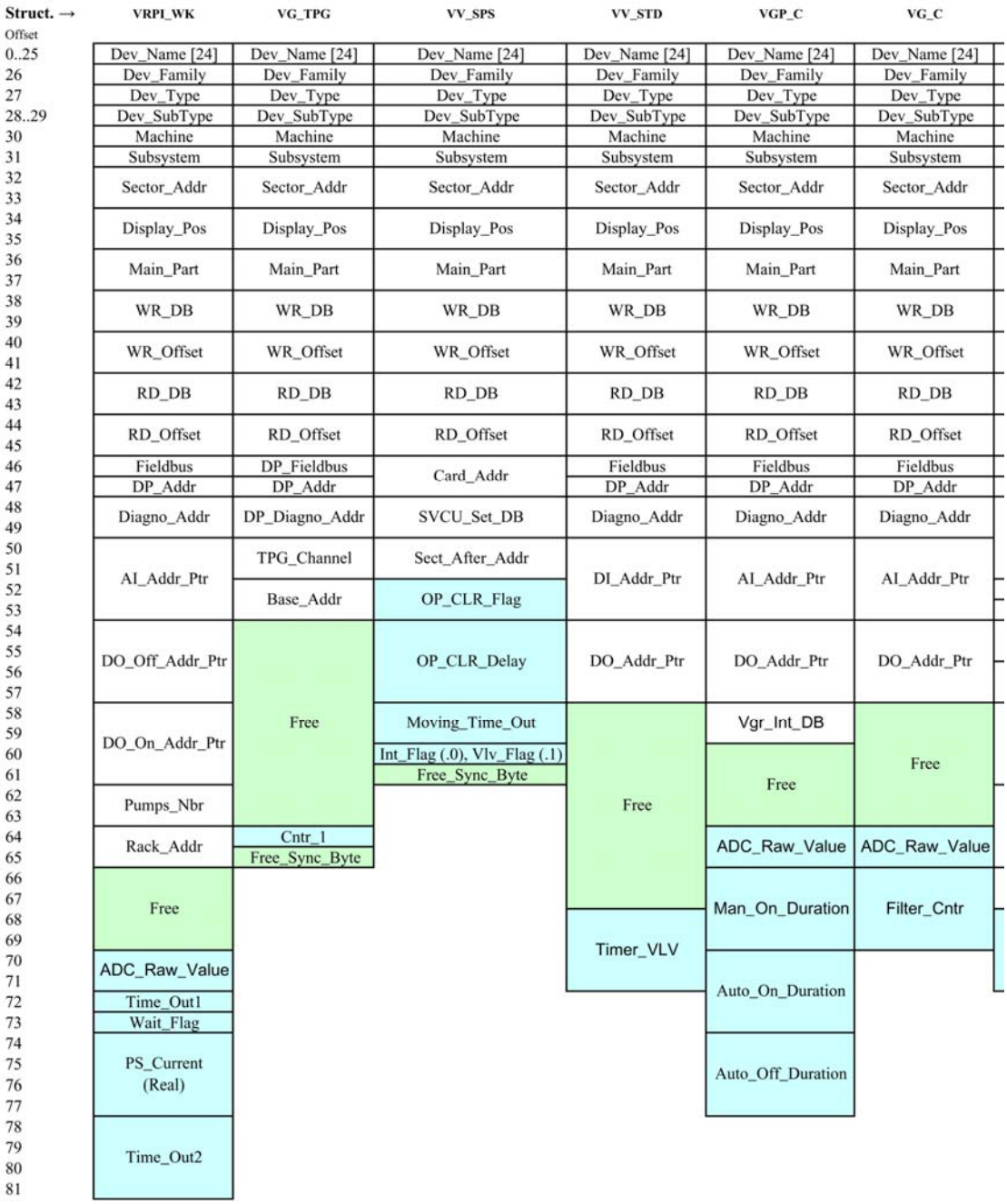


Fig. 9: Mapping of device attributes to physical memory blocks

6 The SCADA server

6.1 Functionality

The previous section describes one way to implement the lower or equipment layer of the 'three tiers' approach. At this point, the representation of the data is still quite close to the physical equipment, calling for an additional layer to make a homogeneous interface to the applications (graphical interface, logging, alarm services, etc.). At the same time, as the processing power of the PLC controllers is limited and best used to control the process, it makes sense to have a single access point

which is materialized by one or more servers, sometimes called a SCADA server (Supervisory Control And Data Acquisition). The standard SCADA functionality can be summarized as follows:

- data acquisition;
- data logging and archiving;
- alarm handling;
- access control mechanism;
- human–machine interface including many standard features e.g., alarm display or trending.

As such, the SCADA server can replace the two upper layers of the 'three tiers' architecture. In general, however, a SCADA server also includes a flexible Application Program Interface (API) allowing access to all its features from an external application that would typically be on the upper layer of the system.

A specific SCADA product has been chosen by CERN, PVSS from the Austrian company ETM [6]. The system, illustrated in Fig. 10 is built around an 'Event Manager' which dispatches the data to and from the possible users. Towards the lower part of the 'three tiers' architecture, PVSS connects to the hardware (typically PLCs) via a set of drivers. Towards the upper layer, PVSS provides an Application Programming Interface (API) and a possibility for processing data in the background (this is where the control over a distributed process would happen). PVSS also provides the tools to build so-called 'User Interface Managers', overlapping in this case with the 'Presentation Layer' of the 'three tiers' architecture. Finally, PVSS maintains a database which contains all data from all connected devices, be they physical or virtual (e.g., a software process).

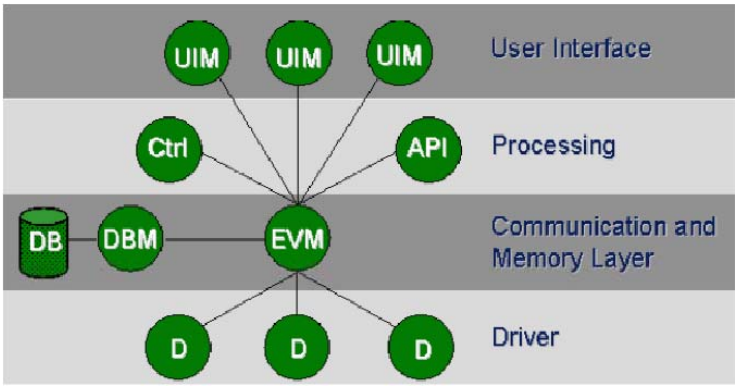


Fig. 10: Architecture of a PVSS system

As for the PLCs, the data to and from the equipment is collected into a data structure, in this case labelled 'data points' (roughly the equivalent of the data blocks in the PLC). The description of these data points is done via a 'data point type' structure, which is similar to a 'class' in object-oriented terminology. A data point contains the information related to a particular instance of an equipment defined by a data point type and a 'data point element' represent one particular value in a data point.

7 Databases

An important aspect of a modern control system is to avoid hard-coding device-specific features in the control programs as much as possible. This goal has been achieved by using the hardware and software structure described earlier. The data blocks describing the equipment in the PLCs are prepared in a central ORACLE based database. The same database also stores the description of the PVSS data points. Figure 11 shows the global database scheme used. The 'survey database' provides

static information about the layout (typically the geographical locations of the equipment), the 'LHC equipment database' describes how the vacuum equipment is distributed (e.g., to which vacuum sector it belongs). The native export facilities of the database management system are used to move the data from one database to the other, via intermediate text files.

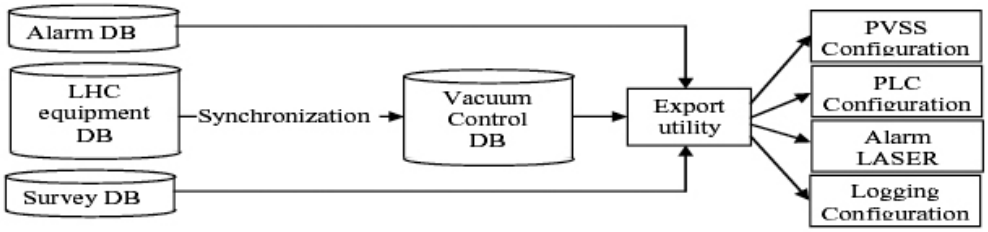


Fig. 11: Configuration database scheme

In an environment like CERN, it is desirable that a single database structure be used for all accelerators. This is achieved by having a single set of tables describing all families and types of equipment to control and a set of tables for every accelerator which is closely linked to the layout (includes, for instance, geographical information taken from the surveyors database). In this way it is possible to use the database not only to configure the controls equipment, but also to configure the way data is displayed on the user interface.

8 Data presentation and storage

Which data has to be displayed very much depends on the users. A control room operator is normally happy with a green light telling him or her that all pressures are below a given value and all sector valves are open to allow for the beam to circulate. As soon, however, as these conditions are no longer satisfied, it becomes important to have access to more detailed information.

8.1 State of the equipment

Although not strictly necessary, the states of the vacuum equipment are most often displayed on a synoptic showing the sequence of elements and using a colour coding for the state. Although it is recommended to use normalized coding for states, the colour scheme is a recurrent subject of discussion, as many operators are more responsive to, say, a red symbol than to its shape. Figure 12 shows the pumps and the valves of a part of the SPS. Some other accelerator equipment is also sketched, so as to allow for a better localization. A 'point and click' approach is chosen to select a piece of equipment and show more details or enable the control panel. Simply pointing with the cursor to the device will return its main data (e.g. pressure for a gauge, state for a valve) and its name. Figure 13 shows the examples of a valve control panel and an ion-gauge control panel. It is from these panels that actions on individual pieces of equipment will be launched.

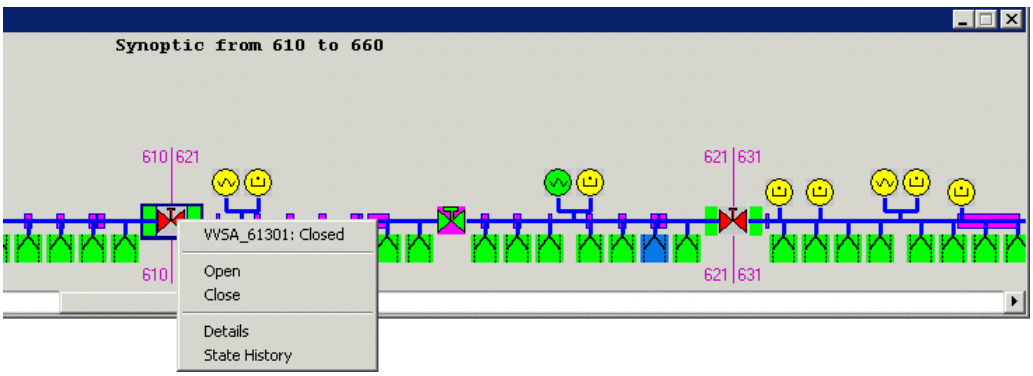


Fig. 12: Synoptic view of part of the SPS

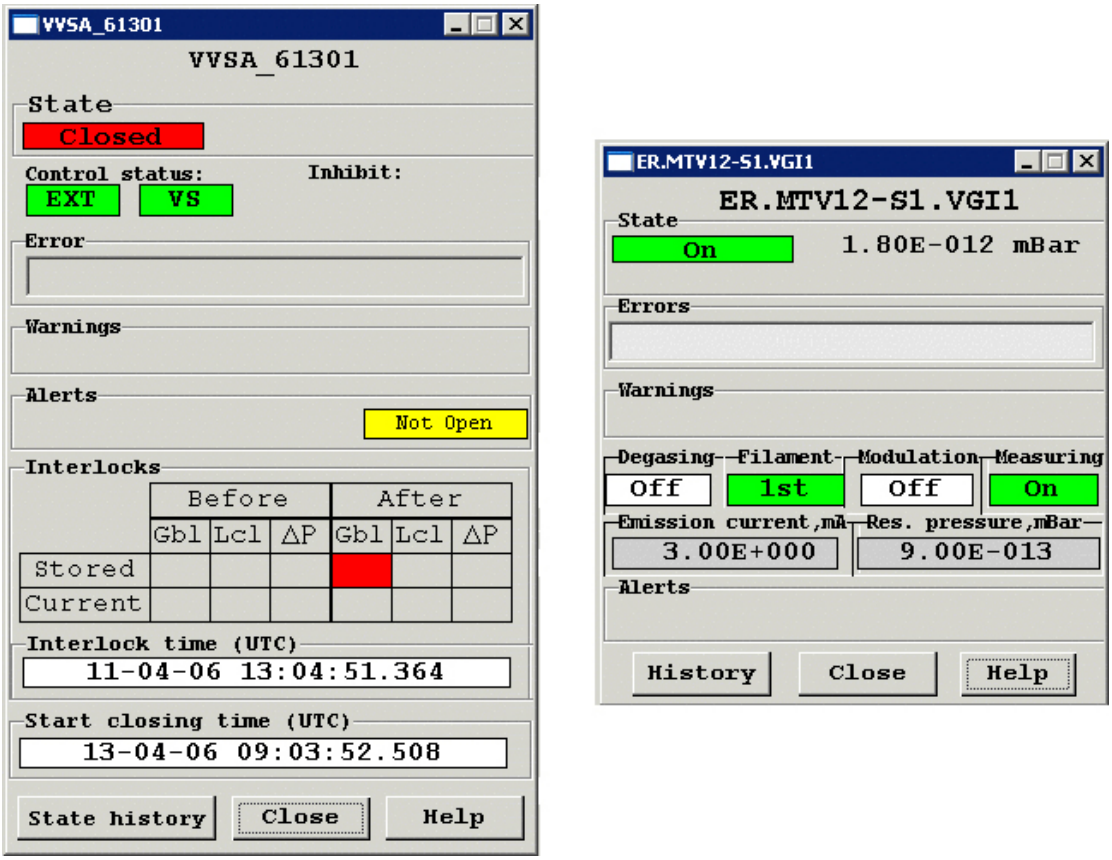


Fig. 13: Example of control panels for a valve and an ion-gauge

Grouped actions are very often required in a vacuum system. For example, switch on all pumps in one vacuum sector or close all sector valves. These are normally implemented as a menu item of the synoptic view (for sector wide commands) or as a menu item of the global schematic view.

8.2 Displaying pressures

Displaying the pressure values is most often done in two ways: a profile over a given part of the accelerator (e.g., one vacuum sector) or the evolution over time of one or more pressure readings.

8.3 Pressure profiles

There are many different vacuum devices able to return a pressure value (different type of gauges, and sputter-ion pumps), but their useful ranges do not overlap. This presents a difficulty for the display of the pressure profile, which can be illustrated in Fig. 14. On the right part of the figure, all devices returning a pressure are shown and one can see that the readings from the Pirani gauges, although valid, do not reflect the actual pressure. On the left part of the figure, the Pirani gauges have been removed from the items to display, leading to a more useful representation. At the same time, the scales of the graph have been adjusted. The choice of the active elements as well as the scales should be dynamically chosen, but the rules for the choice may not always be obvious. In the present implementation at CERN, these choices have to be done manually.

Every value is plotted as a variable length bar, with a colour coding to draw the attention of the operator in case of abnormal values. As for the synoptic, pointing and clicking on a bar will allow access to more details about the equipment returning the pressure.

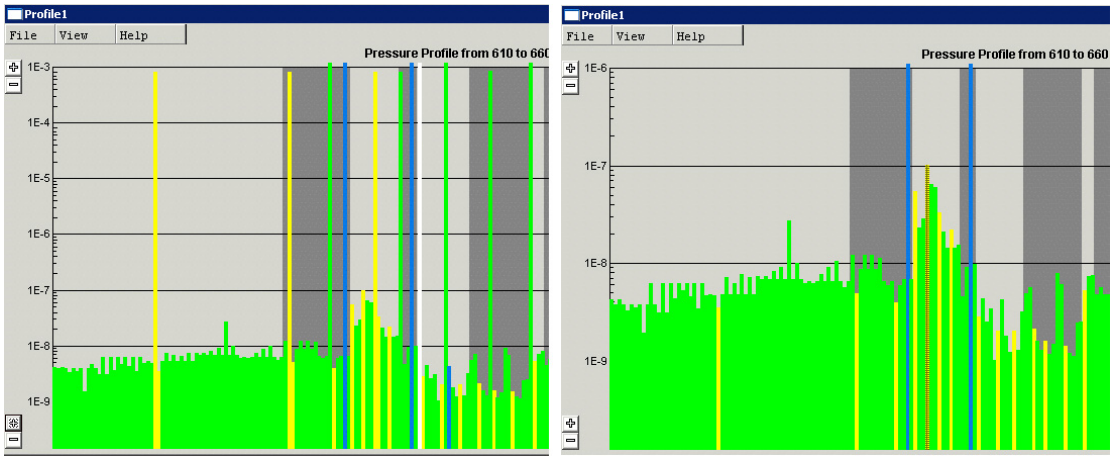


Fig. 14: Pressure profiles over on SPS sextant

8.4 Pressure history

Plotting one or more pressure values against time (histogram) has two different approaches: real-time data added to a graph or archived data over a given time span. The data available in the SCADA server is used for the first case. All data is stored in parallel in a long-term archive database from which it can also be retrieved and displayed. The user selects the equipment to be displayed (usually by selecting it on a pressure profile) and defines the time span of interest. Logarithmic scales are used by default for the pressure scale, linear and logarithmic scales must be available for the timescale, the logarithmic scale being most useful to follow up the pump-down of large volumes.

The histogram is a very powerful diagnostic tool. From the vacuum point of view it allows one to follow trends and give an early warning of a degradation of the vacuum conditions. Figure 15 shows the evolution of the pressure in the insulation vacuum when a part of the LHC helium distribution line was warmed up, showing that gases trapped onto the cold surfaces start to be released at higher temperatures. What is missing here, is the direct correlation between pressure and temperature. This must be done at the upper level of the architecture of the control system and emphasizes the need for a common presentation and access method to the data from all accelerator components.

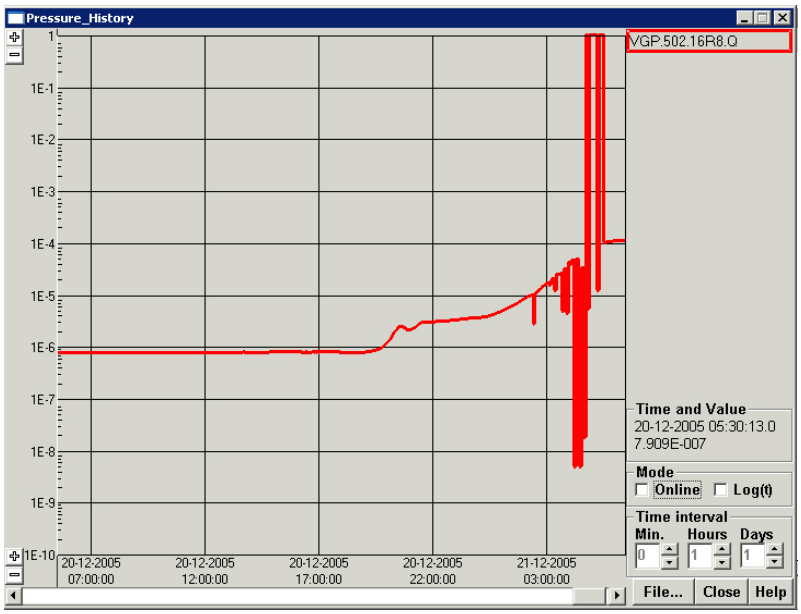


Fig. 15: Histogram of the pressure during the warm-up of the LHC helium distribution line

A histogram can also help in identifying beam-related effects (losses, electrically induced noise, etc.). Figure 16 represents the values returned by an ion-gauge in LEIR while an ion beam was injected into the accelerator. The equipment used is very new and still quite some debugging is required to understand the reason for these very unstable readings.

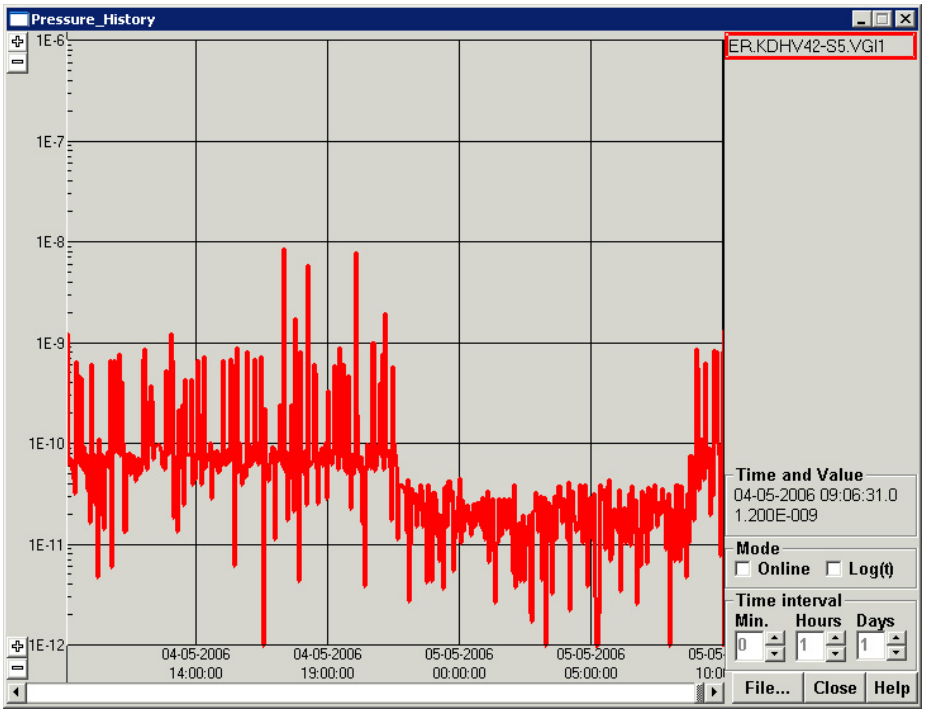


Fig. 16: Histogram of an ion-gauge reading in LEIR

8.5 Alarms

A typical vacuum system includes much controlled equipment, and hence may generate a considerable number of warnings and alarms. These range from pressure limits being reached to faulty power supplies or inappropriate valve states. The SCADA server is in charge of generating the alarms, usually based on a set of rules defined in a database. The SCADA server itself can display and log alarms, but they are most often also sent to a central alarm server which collates the alarms from all systems and presents them in an uniform way to the operators in the control room.

Reduction algorithms are used to minimize the number of messages an operator must cope with. They group similar alarms by vacuum sectors or sometime larger machine areas, like an octant. Figure 17 shows an example of an alarm screen, where some alarms are indeed reduced. It is possible though for the operator to display the details of the alarm sources.

Very often, the origin of the fault comes from an external event, like a power failure or a communication network problem. In this case, the reduction algorithm should make a global analysis on all systems and hierarchize the display of alarms. For instance, it does not make much sense to display a list of faulty power supplies if there is no power on the grid in the racks where the supplies are installed. So far, at least at CERN, the central alarm server does not allow for such overall analysis.

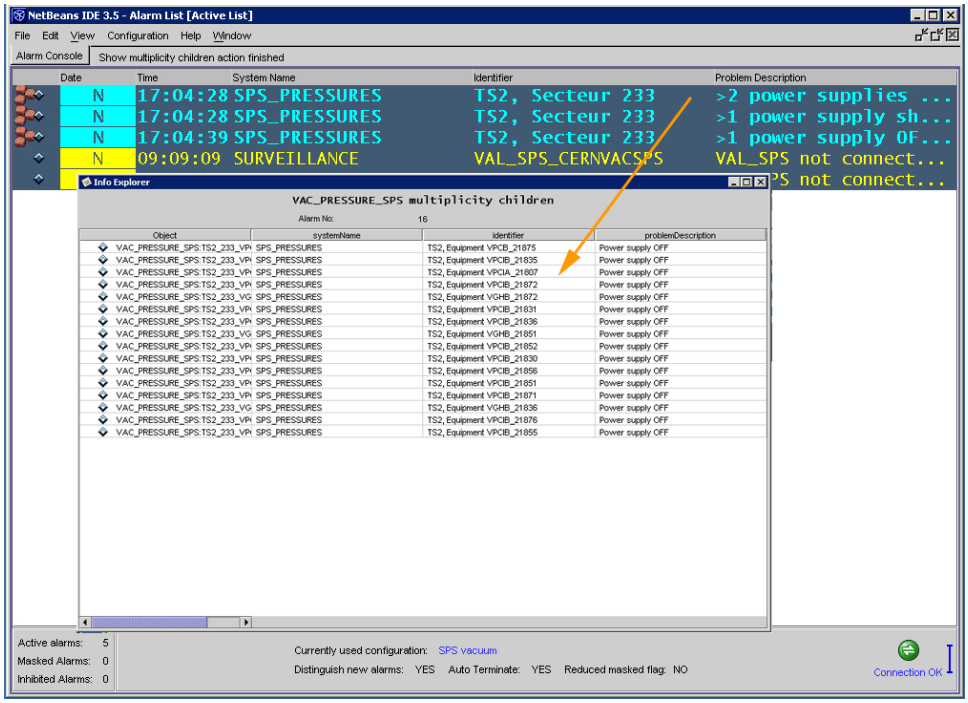


Fig. 17: Example of an alarm screen and details of a reduced alarm

9 Special cases

9.1 Residual gas analysers

If the most common types of vacuum equipment are connected to the main control system using the techniques described above, it sometimes also happens that a solution provided by a supplier has to be integrated. A typical example is the operation of sophisticated residual gas analysers (RGAs) which come from the manufacturer with a suite of programs to operate them. Figure 18 shows how it has been done at DESY [7], where an interface with an Ethernet connection has been used to make the RGA hardware controllable from a remote PC. In a way, it also follows the layered approach, but does not make use of the common vacuum SCADA server. With even more recent RGAs, the control unit has its own Ethernet interface and uses standard communication protocols (with the inherent vulnerability mentioned previously).

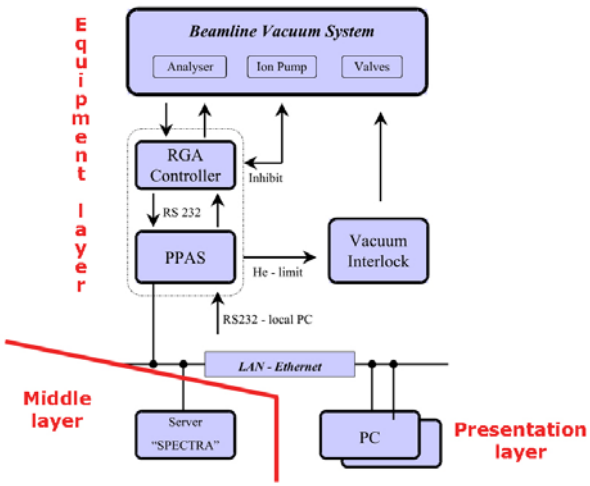


Fig. 18: Example of the remote control of a residual gas analyser

9.2 Control of bakeout process

Traditionally the bakeout process was controlled at CERN using one industrial PID controller per regulated channel. Easily available from industry, these controllers were, however, not very easy to integrate in a global system, at least if cost was a concern. With the development of PLC-based controls, it became attractive to use them for temperature regulation during the bakeout cycle. Among the advantages which were tested on prototypes at CERN, one can mention:

- built-in possibility to connect to the main control system;
- acquisition of external parameters, like power to the heating elements or pressure;
- selective recovery procedure from failures, based on external parameters;
- central alarm management and data logging.

Regulation and protection can be significantly improved using a PLC-based system. The program can automatically detect the time constants of the heated equipment. Measuring the power to the heating elements, allows for detection of faulty (or wrongly connected) thermocouples. Figure 19 shows how power supervision can prevent overheating a vacuum component.

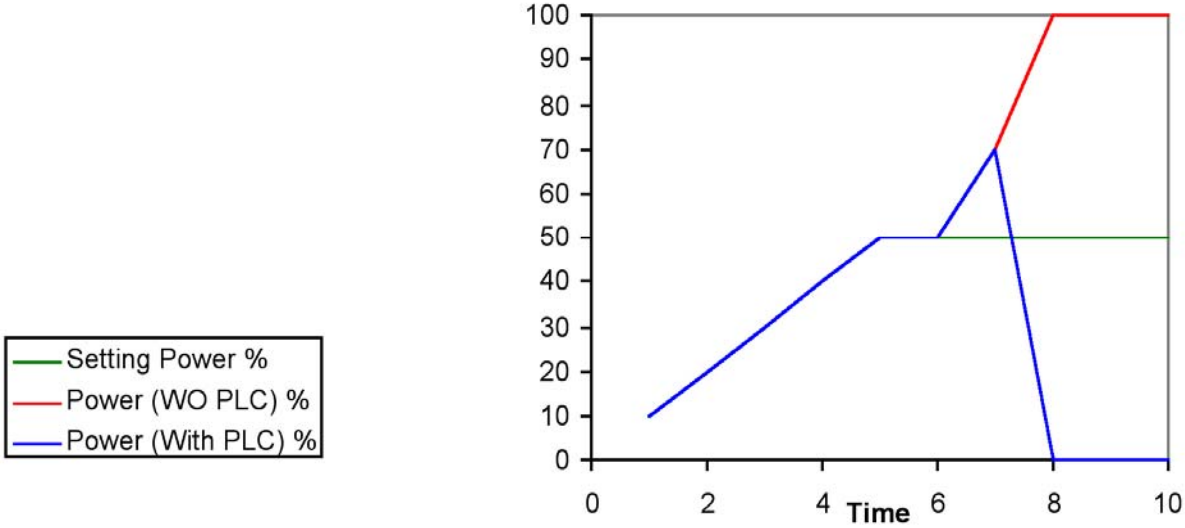


Fig. 19: Using power measurement for protection against overheating

Table 2 shows a set of possible scenarios for dealing with various faults, showing that the available processing power allows one to protect against or to recover from a number of failures using information from other equipment.

Using a PLC-based controller is also economically more attractive. A single PLC can control as many as 24 channels (the limitation is more the number of electrical connections to make, than the processing power) with a single interface to the control network.

Table 2: Failure recovery scenarios for a bakeout process

Events/ Actions	Non-NEG	NEG	Standard regulators	PLC BO rack
Short power cut (<5mn)	Y (1)	Y / N	Stop the bakeout	Either stop or continue the bakeout (setting) & generate an alarm
Long power cut (>5mn)	Y (2)	Y (2)	Stop the bakeout	Smoothly top the bakeout (with a decrease temp. slope) & generate an alarm
Too high temp. measurement	Y (1)	Y (1)	Nothing	Stop the channel, continue or smoothly stoop the other channels (with a decrease temp. slope) & generate an alarm
Too low temp. measurement	Y (1)	Y (1)	Nothing (100% of the power will be delivered)	Stop the channel, continue or smoothly stoop the other channels (with a decrease temp. slope) & generate an alarm
High pressure	Y (3)	Y (2)	Stop the bakeout (need hardware interlock)	Smoothly stop the bakeout (with a decrease temp. slope) & generate an alarm
Leak (RGA pick 40 leak detection)	Y (3)	Y (1)	Stop the bakeout (need hardware interlock)	Smoothly stop the bakeout (with a decrease temp. slope) & generate an alarm
Pumping group stopped or VVR closed	Y (3)	Y (3)	Nothing	Smoothly stop the bakeout (with a decrease temp. slope) & generate an alarm

10 Web-server-based applications

An alternative to central user interfaces (at the presentation layer) is to use embedded Web servers. This technique has been tested at ESRF for residual gas analysers and PLCs. Embedded Web servers are essentially interface modules that can access the equipment on one side and accept html commands from Ethernet on the other side. Web pages generated by the embedded server can be called from any browser.

Commercial Web servers come with graphical tools which are fine to develop simple control panels and display simple graphics, but their limits are quickly reached when approaching any scientific display. Also, the Web server is not designed to store any important amount of data, meaning that it does not replace the traditional control path. The main reason to use them is to have

access to local processes even when the main control system is not available (typically during some shutdown periods). An alternative reason is to allow for quick prototyping of new user interfaces.

Figure 20 shows a Web page produced to control a residual gas analyser. The Web server module is part of the control unit, itself located in the synchrotron ring and not accessible during operation. Therefore, the Web interface is exclusively used during shutdown periods, for diagnostic while the residual gas analyser is connected to the main control system during operation.

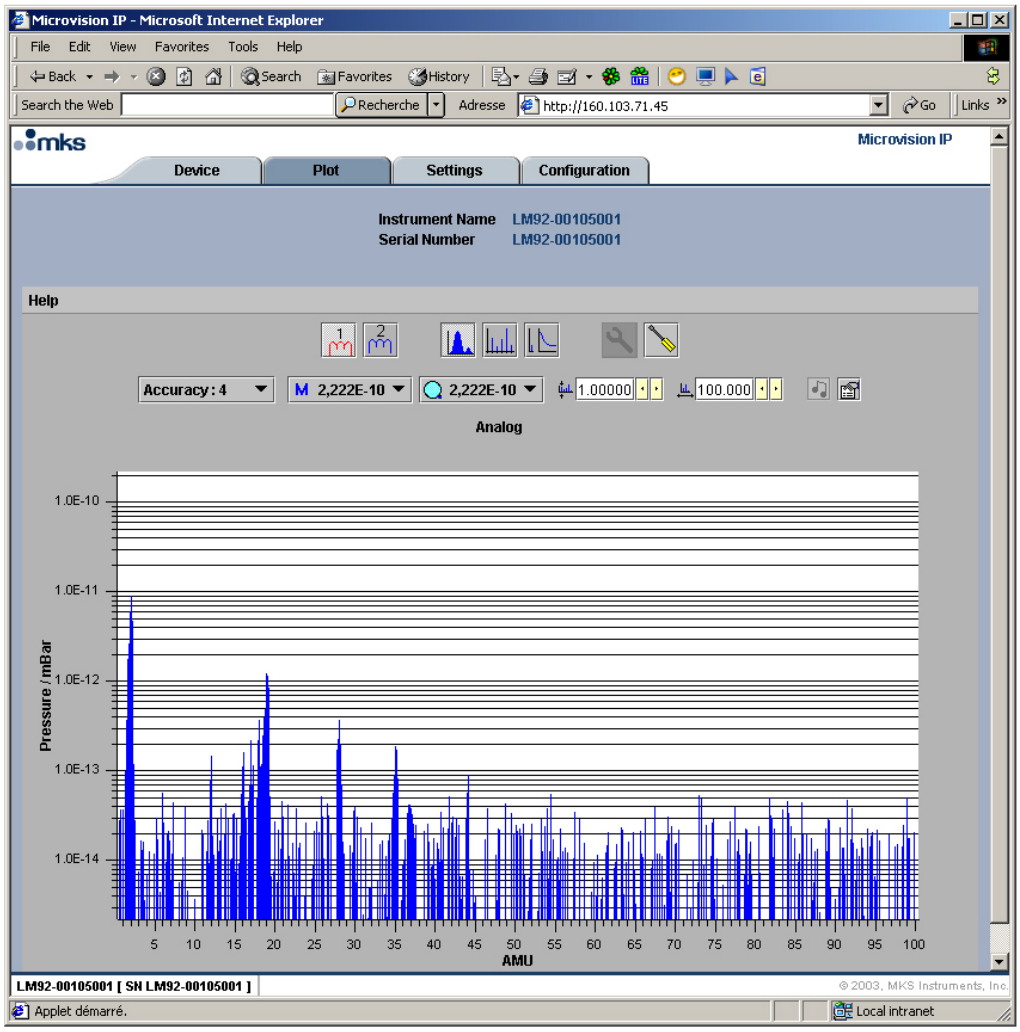


Fig. 20: Web page to control a residual gas analyser

A second example is shown in Fig. 21 with the use of an embedded Web server in a PLC dedicated to the control of valves, shutters, absorbers, and vacuum interlocks in beam lines. Using a Web server on beam lines appears really useful since the set-up changes very frequently, making changes and tests much quicker than from the central control system. In this case, the synoptic is displayed, a point and click approach allows one to open the specific control panel of each piece of equipment.

Other applications have been successfully tested with this approach at ESRF, like upgrades of bakeout controllers or even the controls for the recently installed NEG coating facility.

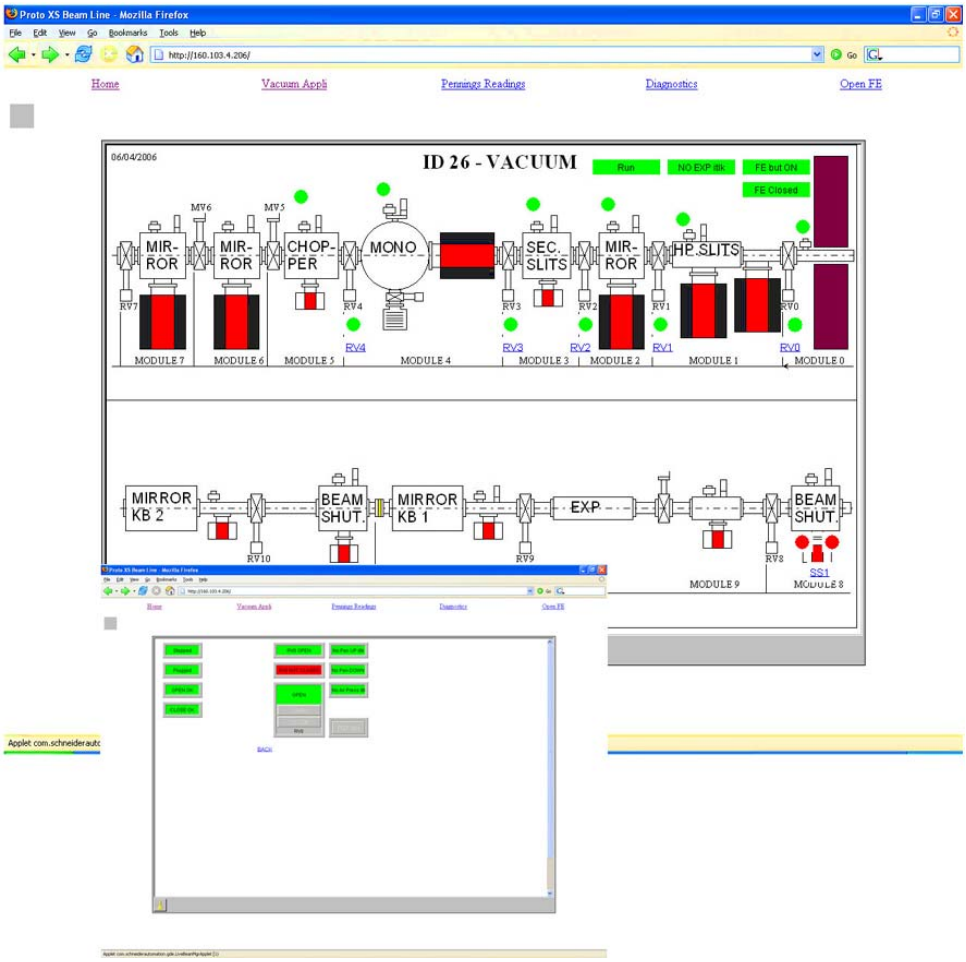


Fig. 21: Web page to control a beam line in ESRF

11 Interlocks

11.1 Protecting vacuum sectors

An accelerator vacuum system is normally divided into sectors to allow independent venting and pumping cycles. The choice of the length and location of the sectors can be driven by different reasons. For a room temperature accelerator, critical equipment like RF cavities or injection and ejection devices, are normally isolated by sector valves. For an accelerator including superconducting elements (like the magnets in the LHC), a good location to place sector valves is at the cold-to-warm transition. Large experimental facilities like in HERA and in the LHC are usually also isolated by valves. In all cases, the aim is to minimize the intervention time and to protect sensitive parts of the accelerator from major vacuum problems. The control of the sector valves needs to be interlocked with the vacuum conditions. The evaluation of the vacuum conditions is most often done using ion-pumps and cold-cathode gauges, as these devices are quite robust. A good practice is to use several instances in the same vacuum sector with a voting rule (e.g., 2 out of 3 alarms) to trigger the closure of the valve, but to require all devices to give a good indication to allow opening the valve. In most cases, the interlock chain for the sector valves is hardwired.

For high intensity machines (like the future LHC), closing a valve on a circulating beam would not only destroy the valve, but very likely severely damage adjacent equipment. An additional interlock is therefore required to trigger the beam abort in case of a vacuum problem, with a proper feedback to the valve controller to allow the valve to close only when the beam is ejected.

11.2 Protecting individual components

Many vacuum devices need to be protected against a too high pressure. As long as the equipment is on, monitoring a relevant parameter of the equipment (in most cases a current proportional to pressure) and switching the equipment off when a limit is reached is enough (self-protection). For more complex instruments, like hot-cathode ionization gauges, more than one parameter has to be monitored, for instance the value of the emission current. The situation is different when a piece of equipment is off. If some are quite robust (e.g., sputter ion pumps or cold cathode gauges) and in general will only suffer a degradation of performance if switched on at a too high pressure, some others can be destroyed (e.g., hot-cathode gauges or titanium sublimation pumps). One therefore has to establish a chain of interlocks based on various measurement devices, where at least one must be able to give a reliable indication at high or atmospheric pressure (e.g., a Pirani gauge).

11.3 Interlocks to other systems

In addition to the protection of the vacuum system proper, interlocks from the vacuum system to protect other systems are often required. As an example, RF cavities or electrostatic septa need a vacuum interlock, as they could suffer from sparking at high pressures. Although still often hardwired, the interlocks to protect the equipment are now also implemented as software signals. Figure 22 shows an example of the various interlocks which have been identified to run the LHC.

Interlocks should also be considered at a more abstract level and modelled, but little effort has been put into this so far in the field of vacuum. Basically, an interlock is just another process with a number of input signals (conditions) that trigger or filter actuations (commands to the equipment). With more and more distributed process control, it will become inevitable that the interlocks are separated from the equipment they protect. For example, during the bakeout and NEG activation of a vacuum sector in the LHC, the bakeout regulation has its own process controller, as do the pumping stations. The permanently installed pumps and gauges also have their controller. Every controller integrates its own protection mechanisms and can have a few external inputs or outputs (e.g., close the roughing valve if there is a fault in the pumping station). But when it comes to protecting the system as a whole, there should be a global monitoring process, which so far relies mostly on the operator.

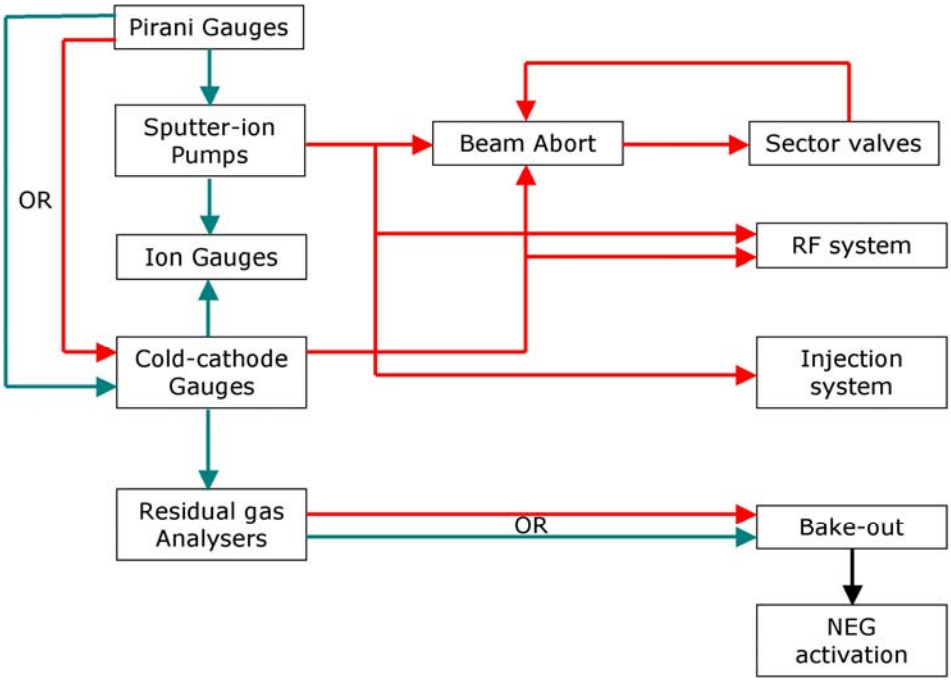


Fig. 22: Schematic of the interlocks of the LHC vacuum system

References

[1] R. Gavaggio, M. Steffensen and P. M. Strubin, Using models to allow for object oriented programming of the vacuum control systems, EPAC94, London, UK.

[2] P. Gayet and C. H. Sicard, Deployment and integration of industrial controls: the case of LHC cryogenics controls, Proceedings of ICALEPCS2003, Gyeongju, Korea.

[3] K. Rehlich, Status of TTF VUV-FEL control system, PCaPAC 2005, Hayama, Japan.

[4] R. Gavaggio and I. Laugier (CERN, Geneva, Switzerland), L. Kopylov, S. Merker and M. Mikheev (IHEP, Protvino, Russia), Development of the vacuum control system for the LHC.

[5] P.M. Strubin and N. Trofimov, Control and operational models for vacuum equipment, PAC97, Vancouver, Canada, Eds. M. Coryn *et al*. (IEEE, Piscataway, NJ, 1998), pp. 2440–42.

[6] http://itcobe.web.cern.ch/itcobe/Services/Pvss/

[7] U. Hahn, M. Hesse, T. Kracht, E. W. Weiner, K. Porges and U. Miße, The residual gas analyzer system for the HASYLAB beamline vacuum system, Nucl. Instrum. Methods Phys. Res. A **467/468** (2001), 805–808.

Vacuum aspects of synchrotron light sources

R.J. Reid
CCLRC Daresbury Laboratory, Warrington WA4 4AD, UK

1 Introduction to synchrotron radiation [1]

Synchrotron radiation is produced whenever a charged particle is accelerated. Classically, the radiated power P is described by

$$P \propto \frac{2e^2}{3c^3} \cdot a^2 , \tag{1}$$

where e is the electric charge of the particle, a is the acceleration, and c is the speed of light.

Where the particle is constrained to a circular orbit, the acceleration is the usual centripetal acceleration

$$a = \frac{v^2}{r} , \tag{2}$$

where v is the circumferential velocity of the particle and r is the radius of the orbit.

This radiation is emitted isotropically as shown in Fig. 1.

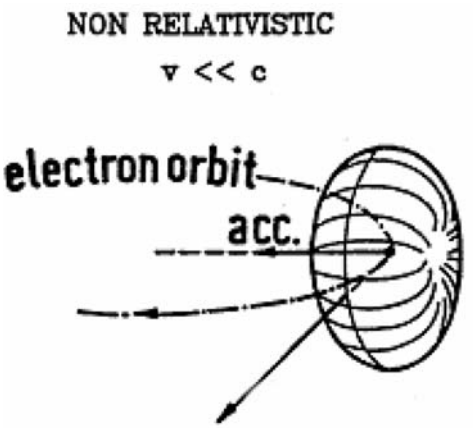


Fig.1: Schematic of classical synchrotron radiation

If the particle of mass m is moving with energy E at a velocity close to the velocity of light, the motion is relativistic. In this case, the power radiated becomes

$$P \propto \frac{2e^2}{3} \frac{\gamma^4 c}{r^2} \tag{3}$$

where

$$\gamma = \frac{E}{mc^2} \ . \tag{4}$$

An observer moving with the particle would still see the radiation emitted isotropically, but when transformed to the laboratory frame of reference the radiation appears 'beamed' forward like a lighthouse beam in the plane of the orbit with an opening angle of $1/\gamma$, as shown in Fig. 2.

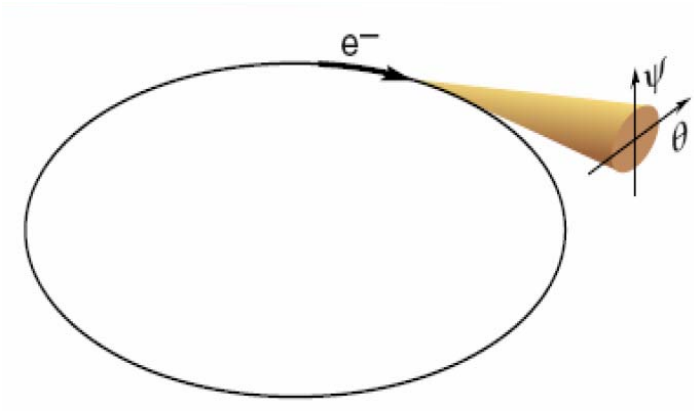


Fig. 2: Synchrotron radiation emitted by a relativistic charged particle, shown here as an electron

In a synchrotron light source, the particles are usually electrons, although positrons can also be used. In the following, it is assumed that electrons are the particles being considered.

The synchrotron radiation sweeps out a fan of radiation in the horizontal plane with a very small vertical dimension. This fan is pulsed and produces very bright pulses of radiation. The energy spectrum of the emitted radiation depends on the beam energy of the electrons and their bending radius and can be displayed as a universal curve as shown in Fig. 3.

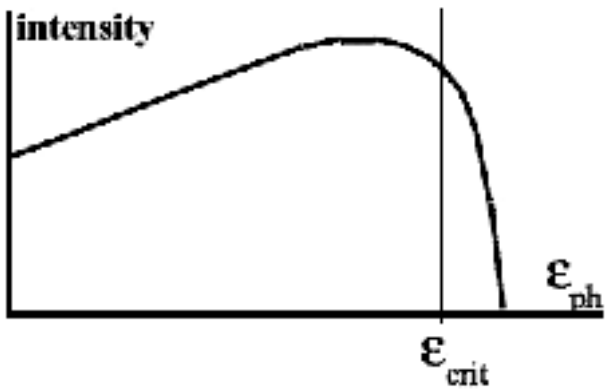


Fig. 3: Synchrotron radiation emitted by an electron constrained to a simple bending path

This is the sort of radiation emitted by an electron travelling around a simple arc. The emitted radiation spectrum is continuous. At low photon energies the intensity varies slowly with energy, but falls off rapidly at higher photon energies. The so-called critical energy $\varepsilon_{\text{crit}}$ is that at which the integrated power emitted above and below this value is equal and is given approximately by

$$\varepsilon_{\text{crit}} \approx \frac{B}{\gamma^2} = 0.665 E^2 B \,, \tag{5}$$

where B is the bending magnetic field in Tesla, E is the electron energy in GeV, and $\varepsilon_{\text{crit}}$ is in keV.

2 Overview of synchrotron light sources

In practice, synchrotron radiation is important because it provides what is currently the brightest tuneable source of light, especially in the X-ray region. Synchrotron radiation has been exploited for scientific research in many laboratories in many countries which have designed and built synchrotron light sources. These are generally based on electron storage rings.

The characteristics of synchrotron radiation are exactly calculable and can be matched to many different sorts of experiment. Sources provide multiuser facilities, thus making a huge range of science accessible.

Synchrotron radiation is characterized by

- High brightness and high intensity, many orders of magnitude more than with X-rays produced in conventional X-ray tubes
- High brilliance, exceeding other natural and artificial light sources by many orders of magnitude
- High collimation, i.e., small angular divergence of the beam
- Low emittance, i.e., the product of source cross-section and solid angle of emission is small
- Widely tunable in energy/wavelength by monochromatization (sub eV to tens of keV)
- High level of polarization (linear or elliptical)
- Pulsed emission (pulse durations below one nanosecond)
- Synchrotron light sources have developed over the past forty or so years and current-build devices are, in general, so-called third generation sources. Table 1 gives a brief selective overview of their development from first to third generations.

Table 1: Selective history of the development of synchrotron light sources

Evolution	Characteristics	Example
First generation sources	Parasitic on particle physics accelerators	NBS (1962) DESY (1964) Tokyo (1965) NINA (1966) Frascati (1966) etc.
Second generation sources	Dedicated, purpose-built electron storage rings – optimized for flux	SRS (1980) NSLS (1981) Aladdin (1981) Photon Factory (KEK) (1982) Bessy (1982) LURE (1984) etc.
Third generation sources	Dedicated electron storage rings with insertion devices – optimized for brightness Higher energies, large machines Low emittance (typically 3 nm rad horiz) Undulators/wigglers	ESRF (1994) APS (1996) Spring-8 (1997) Elettra (1994) etc.

Third generation light sources exploit the properties of special magnet arrays, so-called insertion devices, to tailor the emitted synchrotron radiation to meet the needs of particular experiments. There are basically two types of device – undulators which provide high spectral brilliance and wigglers which provide high flux. In their simplest forms both are based on periodic arrays of magnets. The principle is illustrated in Fig. 4.

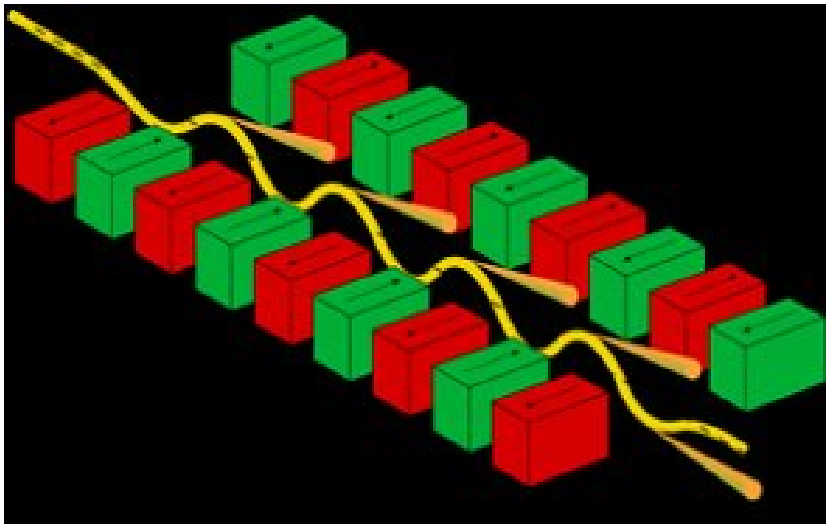


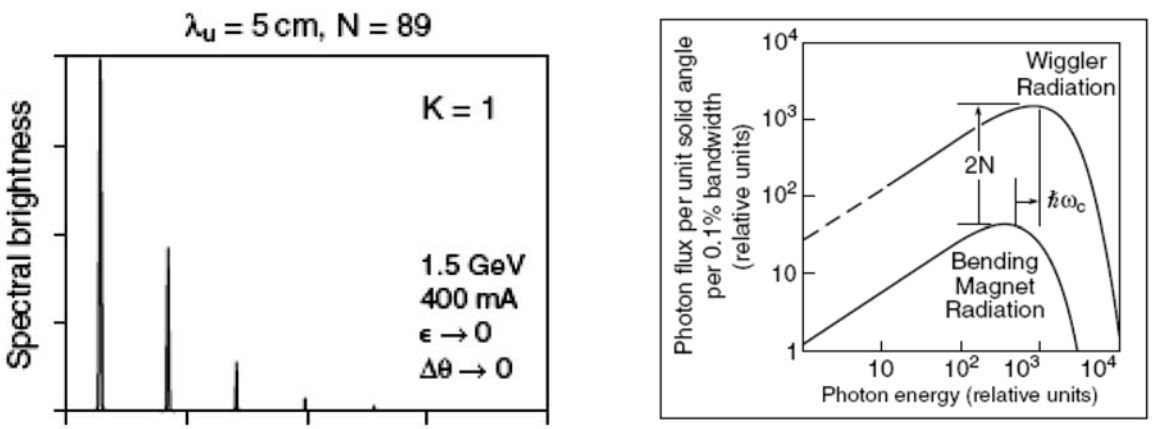
Fig. 4: Principle of an insertion device

For a particular insertion device there the so-called K parameter is given by

$$K = \frac{e \cdot B_0 \cdot \lambda_u}{2 \cdot \pi \cdot m \cdot c} \ , \tag{6}$$

where B_0 is the magnetic field and λ_μ is the magnetic period.

If $K < 1$, then the device is an undulator and if $K > 1$, the device is a wiggler.



(a) **(b)**

Fig. 5: Radiation from examples of an undulator (a) and a wiggler (b)

Figure 5 shows sample radiation outputs from such devices and illustrates the essential characteristics of each type. An undulator produces narrow spectral lines with high spectral brilliance and partial coherence [Fig 5(a)]. A given undulator is tunable over a defined range of photon energies and will exploit harmonics to extend the output range. The principle is that as the electron beam traverses the device it undulates around the mean orbit as shown in the figure. Radiation emitted at each undulation interferes constructively with that emitted at the other undulations to produce the defined spectral output. A wiggler, by contrast, moves the emitted spectrum to higher energies

[Fig. 5(b)], produces a continuous output spectrum and a high photon flux. In this case, the radiation emitted at each undulation (or wiggle) adds to that produced at the other wiggles without interference taking place.

The latest development in light sources, currently in the conceptual design phase at a number of facilities, is that of fourth generation devices. These are based on a different technology, namely free-electron lasers [2], although they can also include current technology photon sources such as wigglers and undulators. They will not utilize storage rings, but will be single-pass devices based on linear accelerators. This permits the use of energy recovery techniques which make them economically attractive since running costs are lower.

A conceptual design being worked on at the Daresbury Laboratory [3] is shown in Fig. 6.

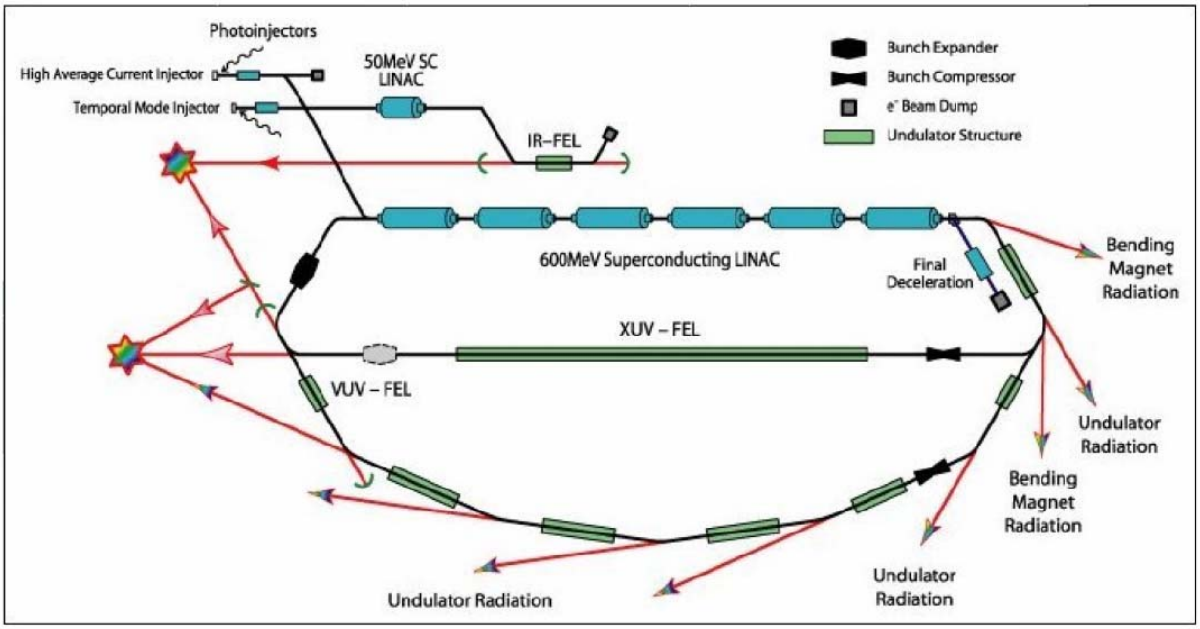


Fig. 6: A conceptual layout for the Daresbury 4GLS

In this design the main linac is a 600 MeV superconducting device. Electrons from a gun pass through the linac and are accelerated to up to 600 MeV. They are then guided along one of a number of return paths and re-enter the linac at a time when the rf phase is reversed, so that the electrons are decelerated, returning energy to the rf system. They are then dumped at low energy and a new pulse of electrons is injected from the gun. The advantage of such a system is its great flexibility, since it can be provided with a number of return paths, each of which can contain very different optical devices.

Such a source has a number of advantages: very short pulses of light are generated; combinations of sources are possible which facilitate pump-probe and two-colour dynamics experiments; the source is intense and tuneable; variable polarization FEL sources can be optimized for spectroscopy and imaging in the XUV, VUV and IR-THz frequency ranges. Energy recovery linac spontaneous light sources are available from soft X-ray to THz. This gives short-pulse, high-repetition-rate operation and an intense broadband source of coherent THz radiation.

There are a number of proposed fourth generation light sources, including IR-FEL (Jefferson Lab) (operational); LCLS (SLAC); XFEL (DESY); 4GLS (Daresbury).

3 Vacuum requirements of synchrotron light sources

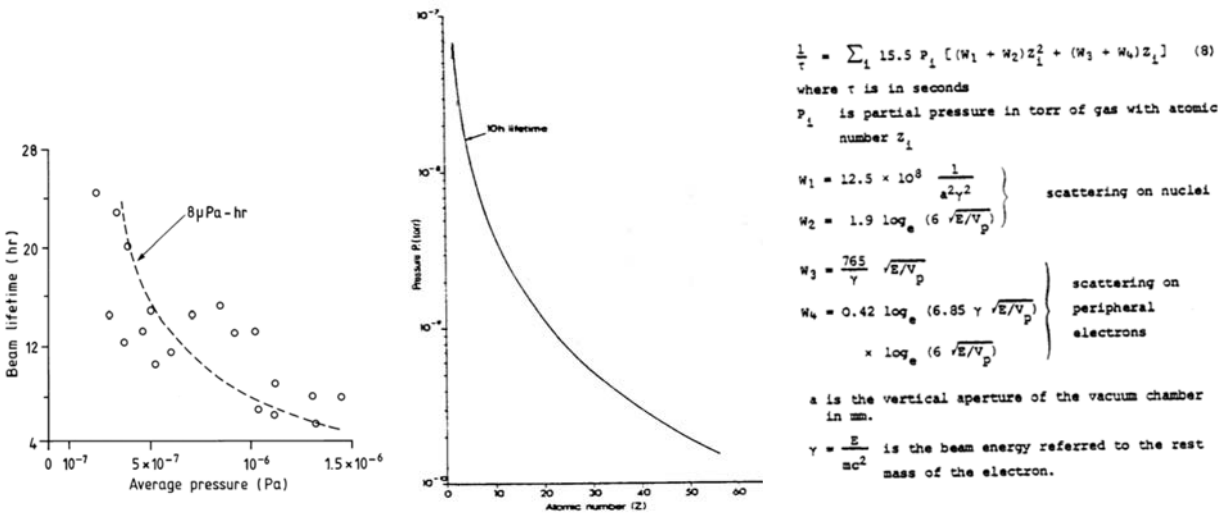
We shall confine our attention to the vacuum requirements of third and fourth generation light sources, since those of first generation sources were entirely dependent on the particle accelerators they were parasitic on and those of second generation light sources are similar to, but less stringent than, those of third generation sources.

For light sources, the dominant property is the beam lifetime, i.e., the time it takes a stored electron beam to decay to a value of $1/e$ of its initial intensity. Normally, this should be at least 12 hours. In practice many sources now operate with lifetimes greater than 24 hours. Lifetime also impinges on mechanical aspects of the machine, especially beam positional stability. A further consideration is personnel safety, essentially that stray radiation should be within safe limits.

The vacuum requirements of fourth generation sources are less clear, although they are likely to be dominated by ion scattering affecting the beam size and bunch length. Since these are single-pass machines, lifetime is not an issue. However, some areas will have to be at very good vacuum levels. In photocathode guns, the cathode is degraded by back bombardment of ions generated from the residual gas [4], so this has to be reduced to acceptable levels. In free-electron lasers, the mirrors can be degraded by the build up of carbon on the surfaces from cracking of hydrocarbons present in the residual gas or by ion bombardment if energetic ions are generated in the beam space and travel to the mirror surfaces. Where superconducting cavities are present, then particulates inside the vacuum envelope can lead to field emission and ultimately cavity quenches.

3.1 Beam lifetime

As noted, an important figure of merit for a source is the beam lifetime. Beam lifetime is essentially determined by scattering of the electrons from residual gas atoms. Both the number density and atomic number of scatterers are the important parameters. Thus it is the partial pressures of the various residual gas species in the beam space that are important for the vacuum scientist to measure (Fig. 7).



(a) **(b)** **(c)**

Fig. 7: (a) Beam lifetime as a function of average pressure measured for the Daresbury SRS; (b) beam lifetime as a function of atomic number of the scattering species; (c) an illustrative formalism for the various scattering mechanisms experienced by an electron

Beam lifetime is important because it determines the time between refills, i.e., the time available to an experimenter to carry out an experiment or set of measurements. A long lifetime promotes

stability. Since the power loading is relatively constant, thermal effects such as magnet and mirror movements are slow and minimized, so both the intensity and position of the photons on a sample at the end of what can be a long beam line, which can be many tens of metres, is stable. A long lifetime enhances efficiency statistics, i.e., the percentage of operational time that is useful beam time for the users to work. Therefore most third generation light sources aim for—and achieve—no more than one fill per day.

There is a mode of operation, 'top up', in which the stored electron beam current is kept relatively constant by the quasi-continuous injection of electrons to compensate for those lost from the beam by scattering, etc. Such a mode has to be approached with care, since it can lead to beam instability. However, it is expected to be widely deployed in future. In this case, the lifetime can be essentially infinite.

In most medium- and high-energy light sources, the lifetime is essentially determined by the gas scattering lifetime, with lifetimes by other effects, such as the Touschek effect, being much longer. A long lifetime is obtained by minimizing the pressure and a typical pressure specification will be $<10^{-9}$ mbar with beam. Reducing contamination, for example, high partial pressures of hydrocarbons and other contaminants also helps and this requires rigorous cleaning and preparation of vacuum vessels before installation. Originally, most storage rings were also baked out at temperatures of up to 250°C to reduce the pressure, especially of water. However, there is a tendency now not to have *in situ* bakeout and this requires the pre-installation preparation of vacuum vessels to be thorough. The driving force for this is to reduce costs, since no bakeout heaters and control systems are required. Magnet apertures can be reduced, although modern bakeout heaters can be installed which are less than 1 mm or so thick, including insulation. However, *in situ* bakeout is expensive, time-consuming and has to be carried out with care.

3.2 Photon stimulated desorption

The pressure inside a machine during operation is in practice determined by beam desorption, otherwise known as photon-stimulated desorption (psd). Here, adsorbed gas is desorbed from the surface layers of a material by energy transfer from an incident photon. This is not a well-understood process and desorption yields (in terms of molecules per incident photon) cannot be calculated for any given system. Therefore pragmatic values determined by experiment are usually used in design work and typical results are shown in Figs. 8 to 10. It will be seen that desorption yields are beam dose dependent.

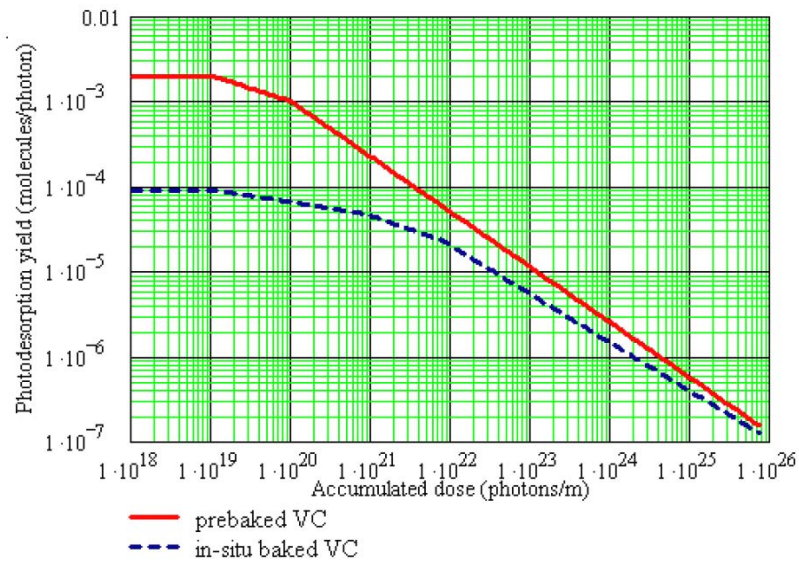
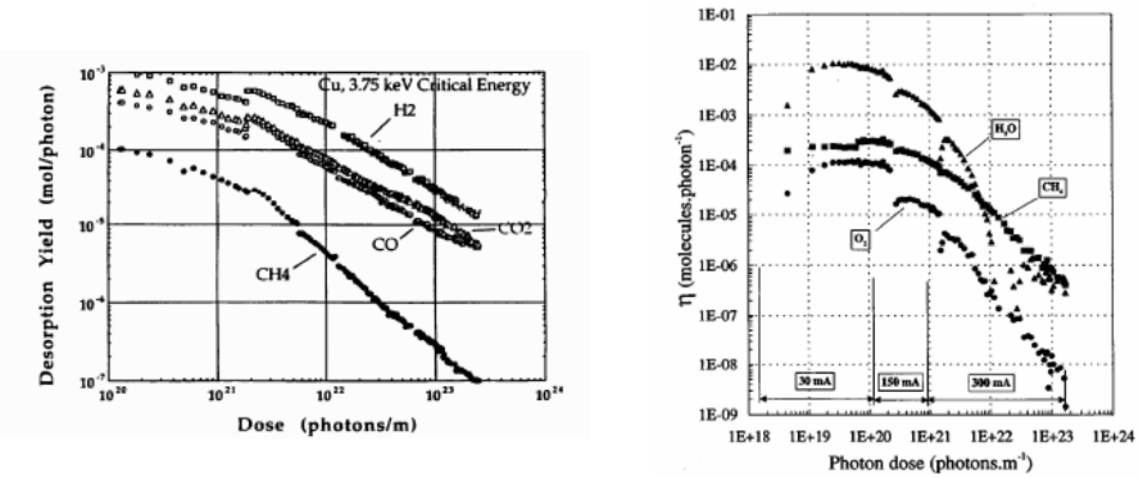


Fig. 8: 'Idealized' yield curves for baked and unbaked stainless steel [5]



(a) **(b)**

Fig. 9: Desorption yields as a function of beam dose for (a) copper and (b) stainless steel [6]

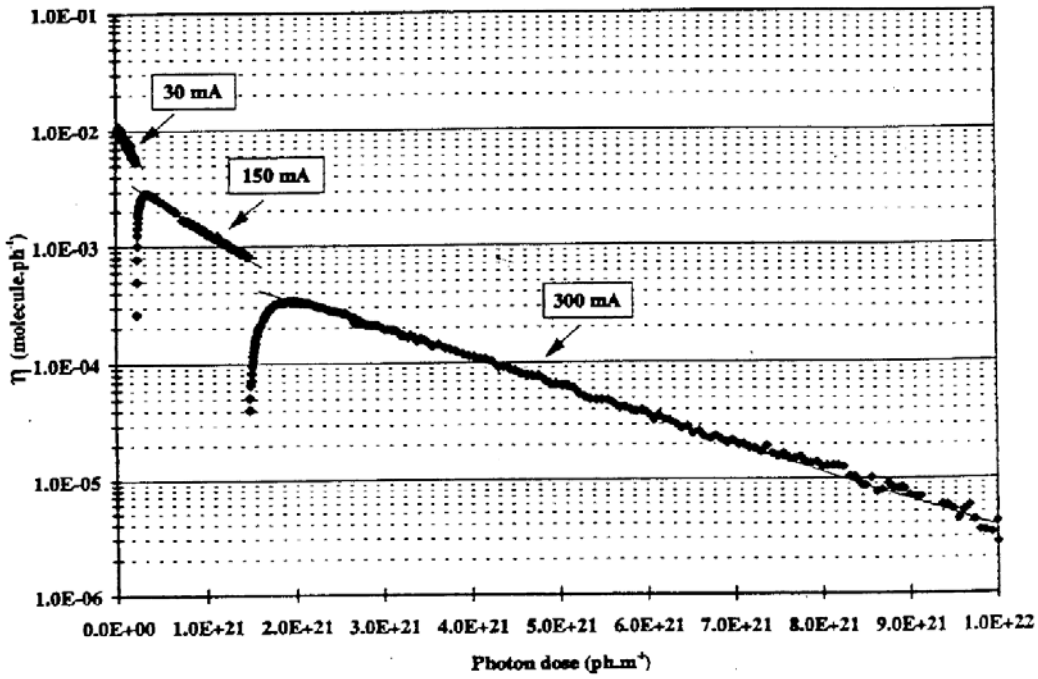


Fig. 10: Desorption of water as a function of beam dose [6]

These figures show clearly the phenomenon of beam 'clean-up'. Initially a substantial quantity of gas is desorbed, so that in a typical machine when beam is first introduced, the pressure rises to quite high levels. However, after moderate exposure to beam—typically an accumulated beam dose of 100 Ah—desorption is at quite low levels and good gas scattering lifetimes are achieved.

An important observation is that after this beam 'scrubbing', on subsequent exposure to atmosphere, the desorption increases, but not to the initial unscrubbed levels, and quickly recovers to where it was before the exposure. This recovery is assisted if the machine is let up to dry nitrogen, i.e., with a dew point < –70°C, rather than ambient air.

In an operating machine, it can be quite difficult to determine what the actual desorption yields are, although an approximate average value can sometimes be inferred. This is because of a number of

factors. Pumping speeds available at the beam space are not well determined – manufacturers' catalogue values for commercial pumps are only approximate and are history, pressure and species dependent. The calibrations of total and partial pressure gauges may not be known and the pressure may be measured at places far from where the desorption is taking place. Finally, because photon doses at different parts of the machine are different, clean-up rates will be different and there will therefore be a range of desorption yields present.

From a machine designer's point of view, the desorption properties of the main constructional materials, stainless steel, aluminium and copper, are rather similar and this is therefore not normally a strong driver in the choice of material used – other properties such as cost, mechanical properties and fabrication methods are usually more important.

Figure 11 [7] is interesting as it shows that desorption of water by photons is a rather efficient process. This is the justification for a number of third-generation machines now being designed without *in situ* bake facilities.

4 Vacuum system design

From a vacuum system design point of view, the main difficulty is getting enough pumping to cope with the gas load generated by desorption. This is because machines tend to comprise long, narrow-bore beam tubes, just large enough in diameter to accommodate the beam-stay-clear aperture. Minimizing vessel cross-sections minimizes magnet aperture and therefore reduces costs. Space for fitting fixed (lumped) vacuum pumps is strictly limited and is usually not where the maximum pumping speeds are needed. Distributed pumping, such as NEG coating [8], can greatly assist.

As noted earlier, the required average pressure in the machine will be determined essentially by beam lifetime, although another important consideration will be personnel safety. As electrons interact with the residual gas molecules, bremsstrahlung radiation, i.e., relatively high-energy X-rays, is produced. This radiation can beam directly down an experimental beamline, resulting in a high radiation environment at the experimental station. This has proved to be a problem [9] where narrow-gap, and therefore poorly pumped, insertion device vessels have been installed. Relatively high background pressures result with a corresponding increase in bremsstrahlung. This meant that long conditioning times were required before a beam line could be safely opened for use. The problem has, however, been almost completely tamed by the use of NEG coated chambers in such situations so that outgassing and desorption are greatly reduced.

The vacuum design process requires detailed calculation of pressure profiles for a number of different scenarios in order to optimize pump sizes and placement. Several iterations of vacuum and mechanical design may be needed to achieve a satisfactory average pressure around the beam orbit and therefore a satisfactory design lifetime. Fortunately, most machines being quasi one-dimensional in structure eases the problem of calculating average pressures. There are a number of computational calculation techniques available which can run on desk top computers. A popular method is test particle Monte Carlo simulation and a readily available package for this is Molflow [9], although many other codes are used. Running these programs is quite time consuming, especially if many iterations of design are required. Semi-analytical methods (e.g., the method of angular coefficients [7]) are much quicker, although if the one-dimensional approximation is not valid, these methods may not be applicable. Simple finite element methods and more complex matrix methods have also been used, and packages are available for this.

As examples, Fig. 11 [7] shows a typical design for a unit cell of a third-generation source, Diamond, where the pressure distribution has been calculated using the method of angular coefficients, and Fig. 12 shows that for ALBA, calculated using Molflow.

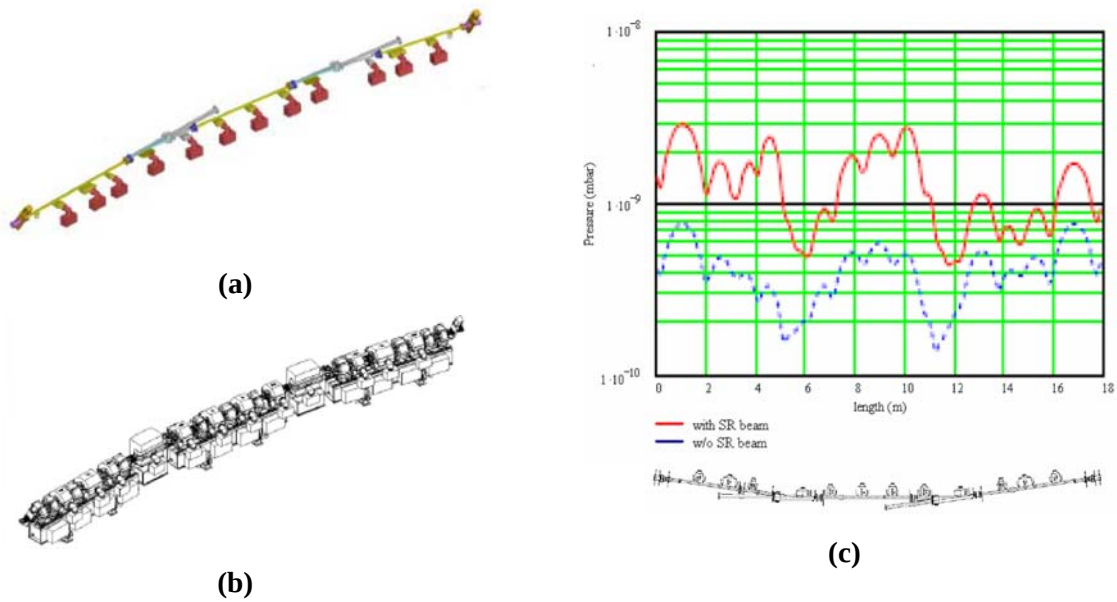


Fig. 11: The unit cell of a typical third-generation light source (Diamond); (a) the vacuum vessel string, showing lumped pumps on the inside; (b) the magnets and vacuum vessels mounted on a girder structure; (c) calculated pressure profiles through the unit cell with and without beam

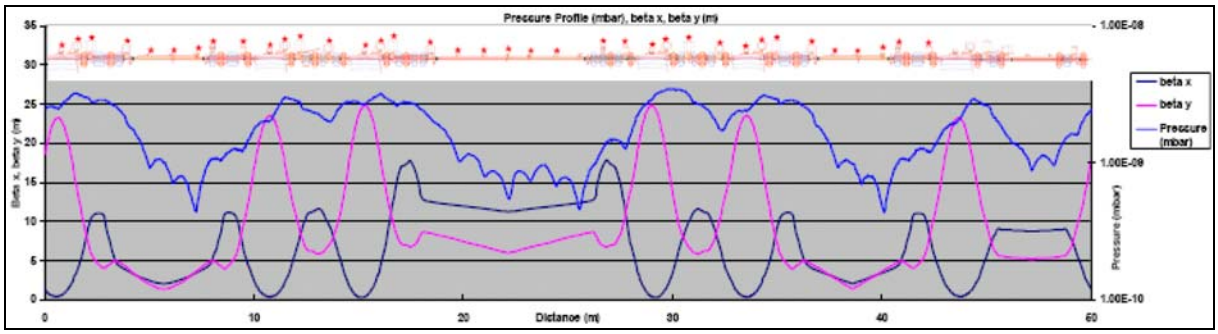


Fig. 12: Pressure distribution calculated for the ALBA source [10]

Most light sources now use relatively conventional vacuum technology, although much of this was pioneered on earlier generations of accelerators or light sources. Pumping will be by combinations of turbomolecular pumps, sometimes with magnetic bearings, and often of the 'wide range' type with integrated drag pumps, sputter ion pumps, titanium sublimation pumps, and NEG (non evaporable getter) cartridges. More recently, the use of sputtered NEG films, is becoming important, especially where conductances are very limited as in insertion device vessels. Rough pumping will almost always use clean pumps such as scroll pumps or diaphragm pumps to avoid oil contamination backstreaming into the main vacuum vessels.

Pressure measurement is now most often accomplished using inverted magnetron gauges, although hot filament Bayard–Alpert gauges have been extensively used in the past. The expense of very low pressure gauges such as extractor gauges is rarely justified in these applications. Ion pump currents are often used to indicate pressure distributions, although these have to be treated with caution. Partial pressure gauges are extremely beneficial in diagnosing pressure related problems and are excellent indicators of the 'health' of a vacuum system.

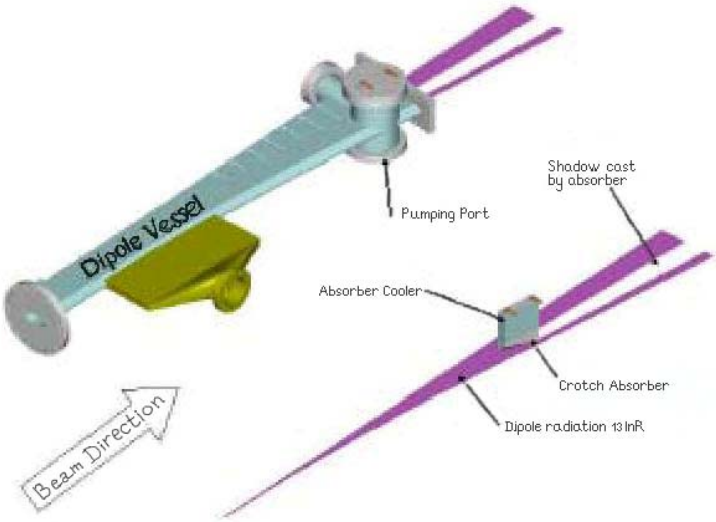
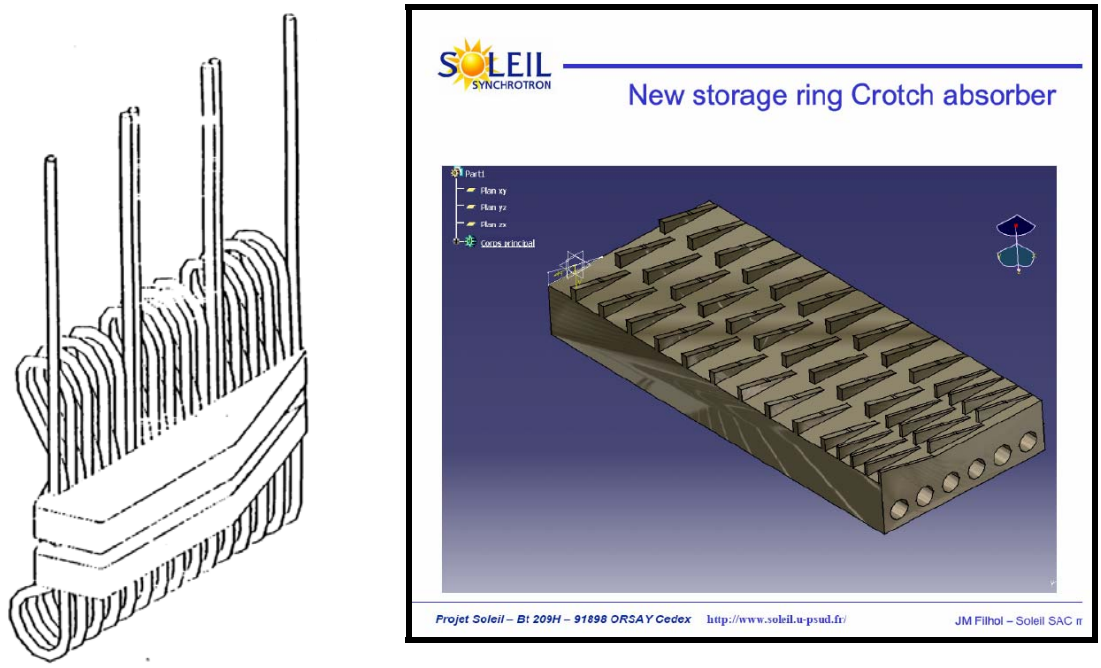


Fig. 13: A dipole magnet vessel showing the photons impinging largely on a discrete radiation absorber

The mechanical design of most vacuum vessels is relatively straightforward, although achieving the necessary high tolerances and distortions may not be easy. The usual good practice rules for UHV must be adhered to meticulously. One important aspect is thermal stability of the machine. Third-generation light sources require beam orbit stabilities of better than 1 micron, so magnets, vessels and mirror systems have to be stable enough to achieve this, even with the variable thermal loadings of different beam conditions. Modern design tends to concentrate the generated photons on discrete radiation absorbers which have to be designed to cope with power inputs comparable to those of electron beam welders. A typical example is shown in Fig. 13. This is the so-called crotch absorber configuration.



(a) **(b)**

Fig. 14: Two examples of high-power photon absorbers (a) from ESRF and (b) from Soleil

Figure 14 shows the detail of two such absorbers. That from ESRF illustrates the intricate cooling pipework required to handle the power while keeping the temperature of the absorber surface at reasonable levels. The complexity is caused by obeying the standard UHV rule of no water to vacuum joints being permitted. The example from Soleil shows the use of a profiled surface to spread the beam and therefore reduce power densities on the absorber surface, while still coping with the total power load.

Particular attention must be paid to the detailed design of bellows and transitions to ensure that they are capable of handling the power loadings of image currents induced by the beam current

5 Concluding remarks

Vacuum systems for synchrotron light sources are by now reasonably well understood and satisfactory designs can be produced with relatively low risk of untoward behaviour. The technology utilized in vacuum systems has evolved over the years but not dramatically, so is relatively mature. It can be said confidently that technological solutions are available for third-generation sources and probably for fourth-generation sources.

Acknowledgements

I have to acknowledge the many discussions with colleagues from the major light sources over the past three decades and the extensive support and assistance of my colleagues in the Vacuum Science Group of ASTeC.

References

[1] H. Wiedemann, *A Synchrotron Radiation Primer*, http://www-ssrl.slac.stanford.edu/primer.pdf .

[2] D.A.G. Deacon *et al.*, First operation of a free-electron laser, *Phys. Rev. Lett*. **38** (1977) 892–894.

[3] *4GLS Conceptual Design Report* (2006), CCLRC Daresbury Laboratory.

[4] T. Siggins *et al*., *Nucl. Instrum. Methods Phys. Res*. **A475** (2001) 549–553.

[5] C. Foerster, private communication.

[6] C. Herbeaux and P. Marin, *J. Vac. Sci. Technol*. **A17** (1999) 635.

[7] O.B. Malyshev *et al*., EPAC, 2002.

[8] P. Chiggiato, this School.

[9] R. Kersevan, this School.

[10] E. Al-Dmour, private communication.

Vacuum challenges for future machines

R.J. Reid
CCLRC Daresbury Laboratory, Warrington WA4 4AD, UK

1 The challenge

There are many new accelerators being designed or conceived around the world, and all of them bring their own challenges to the designer of their vacuum systems. Most are of conventional design and the ground rules are well established, but even so, care has to be taken so that design specifications are met. However, for some of the more adventurous accelerators there are a number of problems which remain to be solved and, doubtless, new hurdles will appear which will need to be overcome. Some of the challenges will be ones of scale, and others will be rather more fundamental. Other contributions to this School have given pointers to where many of these problems lie.

The challenge which the design team will effectively present to the vacuum designer remains as it has always been, viz., to produce a vacuum system where the base pressure is (almost) zero; this pressure is reached in a few minutes; there is no space for pumps or gauges and the cost is very low! Fortunately, there are more realistic challenges which have the possibility of being met.

In this summary, I shall simply outline some of the challenges on the current horizon.

2 Ion stimulated desorption and electron cloud effects

Heavy-ion machines, such as the FAIR project at GSI Darmstadt [1], and hadron machines, such as the LHC at CERN [2] may experience problems caused by ion desorption or electron cloud effects.

Both of these problems are being extensively studied, but it is by no means clear that they have been solved. In each case, the mediating mechanism is secondary electron production at the surface of the vacuum vessels. Reduction of secondary electron yields is difficult and there are active programmes of research in place. One possible means of reducing yields is to produce surfaces where there is a high probability of secondary electrons quickly striking another part of the surface rather than escaping into the beam space. This could possibly be achieved by producing microtextured surfaces using techniques which are standard in the semiconductor industry. A typical example may be seen in Fig. 1.

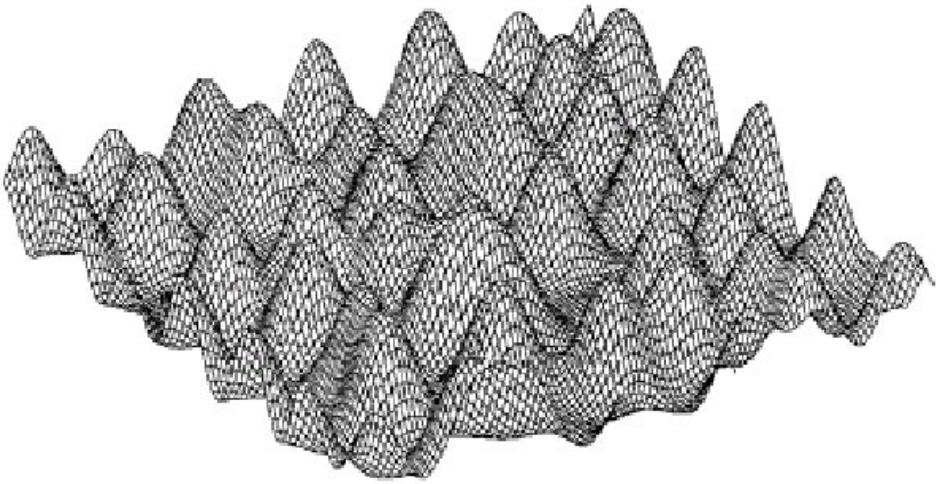


Fig. 1: An example of a microtextured surface

However, it is by no means clear that such coatings can be produced on a semi-industrial scale on the inner surfaces of vacuum vessels. Likewise it is not clear that they would actually work or that they would withstand the rigours of service life inside an operating machine. Nonetheless it is an area worth exploring.

3 The challenge of large size

Large machines such as the International Linear Collider (ILC) [3] are in the planning phase. This is a world collaboration project whose siting has yet to be determined. It is an electron–positron collider of energy 200–500 GeV, later 1 TeV. A possible design is shown in Figs. 2 and 3.

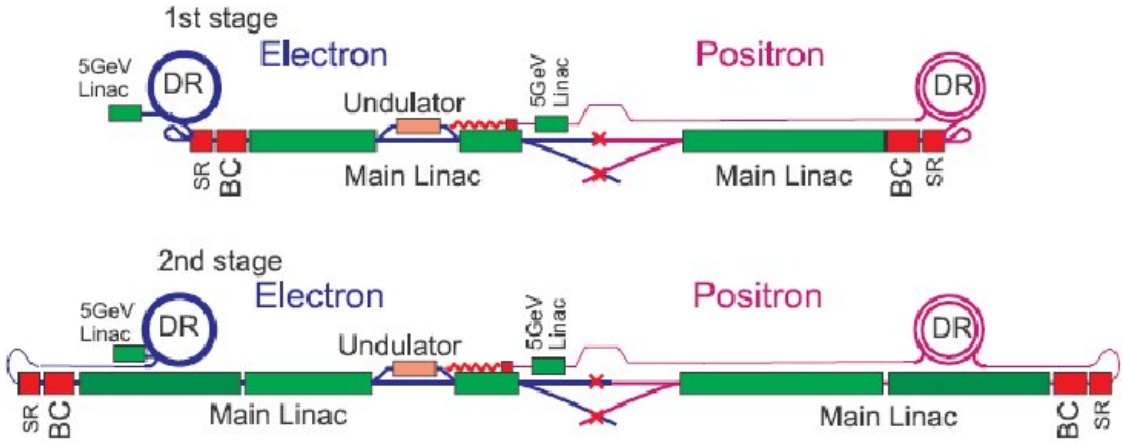


Fig. 2: Conceptual diagram of the ILC

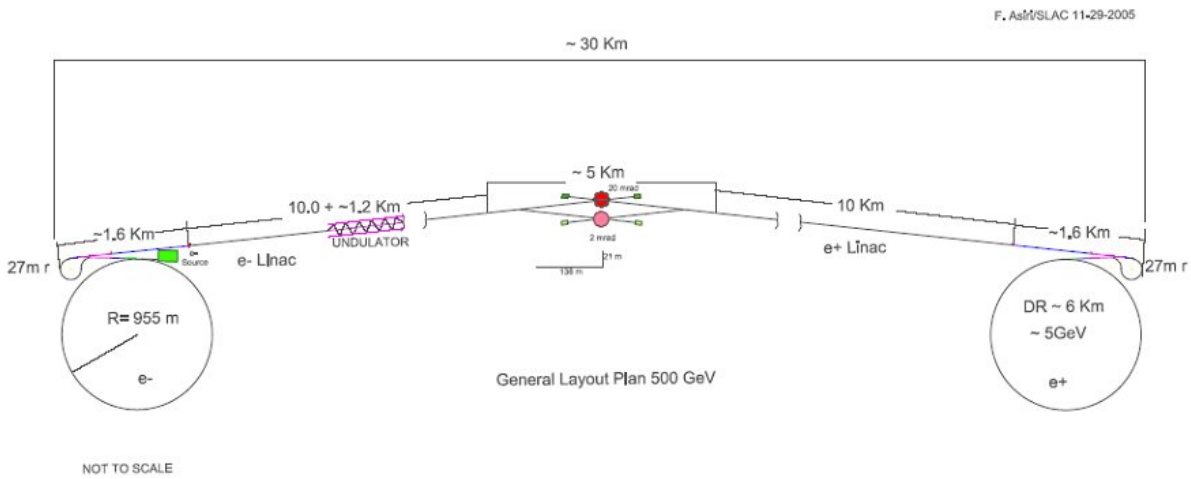


Fig. 3: Schematic of the ILC indicating its scale

Such machines present considerable challenges simply because of their scale. By definition, large systems are large and contain many components, so sophisticated tracking of items through production and installation will be required. Quality control and assurance testing encompass not just materials and manufacturing procedures, which are relatively well understood, but also vacuum processing and acceptance tests, which may be less well defined. They will include leak testing, where the requisite specifications may challenge the technical state of the art, and also cleanliness where it may not be at all obvious what is required and what can be measured, especially at manufacturers' premises.

In the vacuum design of such large systems, there will always be an economic argument to minimize numbers of pumps and vacuum diagnostics. The question to be addressed for diagnostics is: how much is enough to give the information we need? A careful analysis of what information is useful needs to be carried out at an early stage in the design.

In the ILC, the linear accelerators are superconducting devices and the industrial production of so many superconducting radiofrequency cavities in a relatively short period of time is uncertain. Such cavities are usually electropolished and techniques will have to be developed for surface characterization of the finished devices for quality control. This will require an understanding of the surface conditioning necessary for them to achieve their desired performance. At present, this is an area not well understood and several different pragmatic recipes exist. Vacuum and surface scientists will need to be involved in elucidating the physical and chemical mechanisms underlying the conditioning to know what actually needs to be done.

One interesting development here is the recent production of superconducting cavities from single crystals of niobium [4], which are felt to have superior properties. The question to be considered is that if this can be done for niobium, can it be done for other metals and if so, would they have superior, or at least more uniform, vacuum properties? This is a question for the future.

4 Leak-testing large systems

Leak-testing large installations is a time-consuming and difficult business, often because vacuum vessels are buried inside magnets and are relatively inaccessible. Here is a very speculative thought as to how a different technique may be used. Elsewhere in this School [5], H.F. Dylla discussed the work of J.P. Looney on optical laser interferometric techniques for measuring partial pressures of gases. Could such a technique be applied to leak detection by using a system tuned to a helium resonance to detect the plume of helium gas entering a vacuum system through a leak and using time domain reflectrometry techniques to locate the position of the plume?

5 The challenge of small size

Small vacuum systems present challenges of a different sort from large systems. There is pressure from accelerator designers to make beam tubes of smaller and smaller diameter and these of course require pumping to obtain good enough pressure levels for operation of the machine. The relatively new technique of NEG coating has been described in detail in this School [6] and the normal method of magnetron sputtering described. However, using this technique for very small diameter long tubes is technically very difficult.

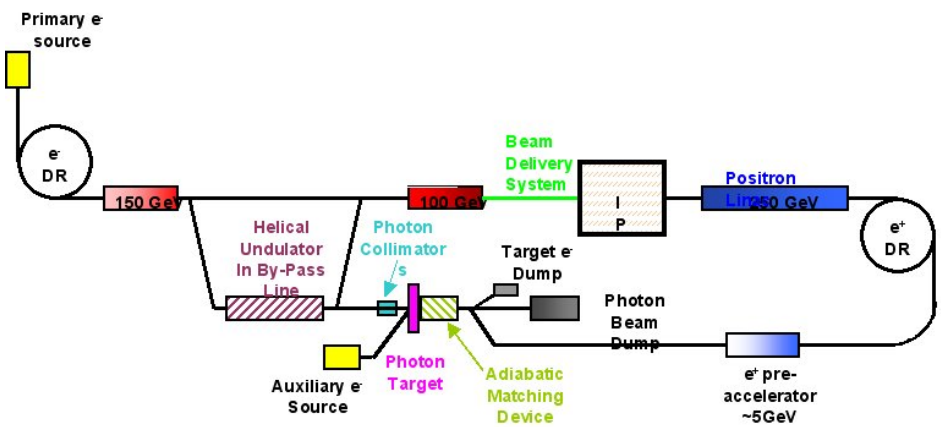


Fig. 4: A design for the positron source for ILC

In the ILC, one design of the positron source, shown in Fig. 4, uses a so-called helical undulator as source of polarized photons used to generate the positrons. The beam tube through this device is several hundred metres long and only 4 mm in diameter. Conventional pumping is totally inadequate to pump this to satisfactory pressure levels and it remains to be seen whether such a tube can be satisfactorily NEG coated by the conventional method of magnetron sputtering which will here be technically very difficult.

There is a completely different sort of accelerator which is of very small size, the so-called wakefield accelerator. One project looking at this is Alpha-X—Advanced Laser-Plasma High-energy Accelerators towards X-rays [7].

The basic principle is illustrated in Fig. 5 with a schematic in Fig. 6.

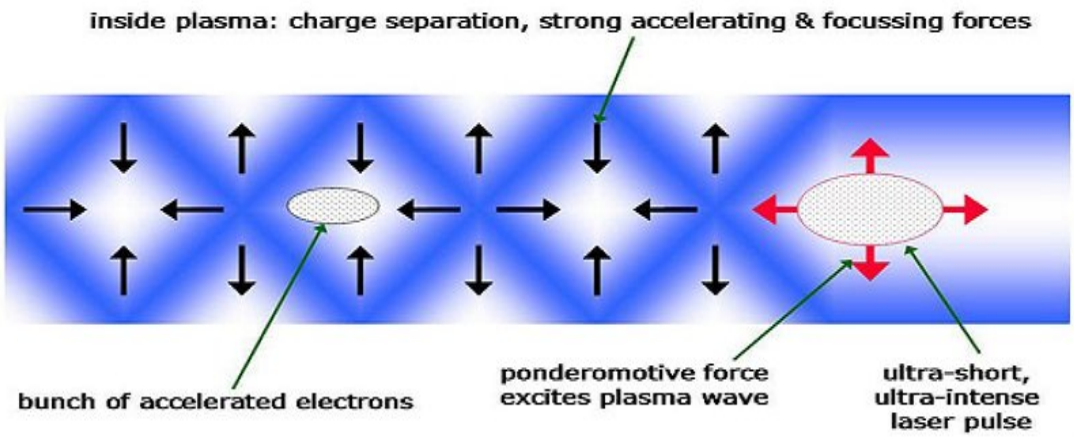


Fig. 5: The basic principle of a wakefield accelerator

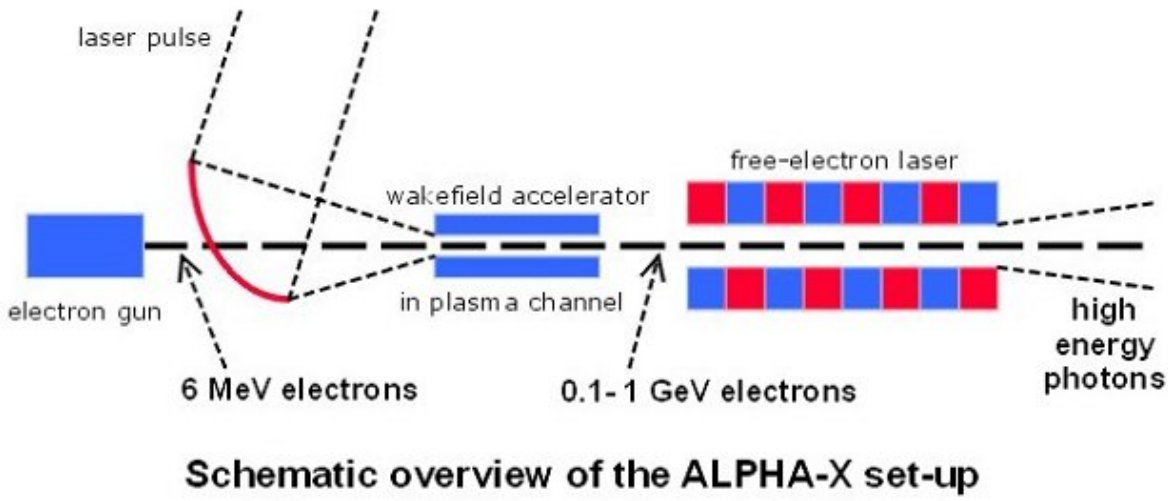


Fig. 6: A schematic of a wakefield accelerator

The heart of the accelerator is the plasma channel, where ionized gas at a relatively high pressure is confined in a capillary channel of the order of a millimetre in transverse dimensions and a few centimetres in length. Although the pressure is a few millibar, the gas flow can be molecular, because of the small dimensions. The pressure of this gas has to be maintained while the tube is coupled to much lower pressures at each end. The role of outgassing in such a device is unclear.

The properties of such a device from a vacuum point of view are rather similar to those found in some MEMS devices.

6 Concluding remarks

I have tried to illustrate some of the currently known challenges for the vacuum scientist and engineer, although of course I cannot predict what interesting and unknown challenges lie ahead in this field. What I do know is that they will be there and they will provide interest and excitement for the vacuum scientist and engineer for many years to come.

Acknowledgements

I gratefully acknowledge the stimulating discussions with many members of the accelerator vacuum community around the world which have maintained my interest in this field over the years.

References

[1] http://www.gsi.de/zukunftsprojekt/index_e.html

[2] http://public.web.cern.ch/Public/Content/Chapters/AboutCERN/CERNFuture/WhatLHC/WhatLHC-en.html

[3] http://www.linearcollider.org

[4] http://www.physorg.com/news4275.html

[5] H. F. Dylla, this School.

[6] P. Chiggiato, this School.

[7] http://phys.strath.ac.uk/alpha-x/index.html

Participants

ADER, C.	Fermilab, Batavia, US
ALONSO, J.	Tekniker Technological Centre, Eibar, ES
AMANN, C.	DESY, Hamburg, DE
ANACLERIO, G.	University of Padua, Padua, IT
ANDREOLETY, G.	CERN, Geneva, CH
ARNAU IZQUIERDO, G.	CERN, Geneva, CH
BACH, J.	ACCEL-Instruments, Bergisch Gladbach, DE
BELLACHIOMA, M.C.	GSI, Darmstadt, DE
BETZIOS, P.	University of Athens, Athens, GR
BIENGE, J.	DESY-Zeuthen, Zeuthen, DE
BIERI, T.	PSI, Villigen, CH
BONGIOVANNI, S.	Strumenti Scientifici Cinel, Vigonza, IT
BOURCEY, N.	CERN, Geneva, CH
BOUSSIER, B.	Diamond Light Source Ltd., Chilton, UK
BRAUN, E.	Leybold Vacuum GmbH, Cologne, DE
CANETTI, M.	Rial Vacuum Spa, Parma, IT
CASAGRANDE, F.	ORNL, SNS Project, Oak Ridge, US
CHEN, C.L.	NSRRC, Hsinchu, TW
CHEN, Z.	Fermilab, Batavia, US
CHOINSKI, J.	Warsaw University, Warsaw, PL
COLOMBET, T.	CERN, Geneva, CH
CORTES, M.	ELYTT Energy, Madrid, ES
CRUIKSHANK, P.	CERN, Geneva, CH
DE ZUBIAURRE WAGNER, A.	DESY, Hamburg, DE
DIAZ CONTRERAS, N.	Tecnovac S.L., Tres Cantos, ES
DIETEL, R.	Feinwerk und Messtechnik GmbH, Berlin, DE
DOLEGIEVIEZ, P.	GANIL, Caen, FR
DUPPICH, P.J.	PSI, Villigen, CH
ENGUITA, O.	Universidad Autónoma de Madrid, ES
FABRE, C.	CERN, Geneva, CH
FADIL, M.	GANIL, Caen, FR
GATTI, S.	Rial Vacuum S.p.A., Parma, IT
GIBSON, P.	ORNL, Oak Ridge, US
HEDLUND, E.	Uppsala University, Uppsala, SE
HERNANDO, J.	Iberdrola Ingeniería y Construcción, Madrid, ES
HESSE, M.	DESY, Hamburg, DE
HIRSEMANN, H.	DESY, Hamburg, DE
HSUEH, H.P.	NSRRC, Hsinchu, TW
IVERSEN, K.	Aarhus University, Aarhus, DK
KOS, N.	CERN, Geneva, CH
KRAEMER, A.	GSI, Darmstadt, DE
KÜHNEL, K-U.	Max-Planck-Institut für Kernphysik, Heidelberg, DE
KUMBARO, D.	MAX-lab, Lund, SE
LABRADOR, J.	Centro Nacional de Aceleradores, Seville, ES
LE PIMPEC, F.	PSI, Villigen, CH
LENNGREN, C.	MAX-lab, Lund, SE
LEPEULE, P.	CERN, Geneva, CH

LEVALLOIS, R.	GANIL, Caen, FR
LIU, X.	Cornell University, Ithaca, US
LÜDECKE, H.	DESY-Zeuthen, Zeuthen, DE
MAAN, W.	CERN, Geneva, CH
MACDONALD, B.	Diamond Light Source Ltd, Chilton, UK
MAESEN, P.	CERN, Geneva, CH
MAKAHLEH, F.	SESAME, Amman, JO
MANSSON, A.	MAX-lab, Lund, SE
MARKS, S.	LBNL, Berkeley, US
MARLETTA, R.	INFN-LNS, Catania, IT
MARTIN FERNANDEZ, R.	CELLS, Cerdanyola del Valles, ES
MARTINEZ, D.	Telstar Tecnologia Mecanica S.L., Barcelona, ES
MCGEE, M.	Fermilab, Batavia, US
MEUNIER, S.	CERN, Geneva, CH
MIDDLEMAN, K.	CCLRC Daresbury Laboratory, Warrington, UK
MILDNER, N.	DESY, Hamburg, DE
MOLENA SERAPHEM, R.	Brazilian Synchrotron Light Source, Campinas, BR
MORENO, D.	Telstar Tecnologia Mecanica S.L., Barcelona, ES
ORLOV, D.	Max-Planck-Institut für Kernphysik, Heidelberg, DE
PAGE, E.	CERN, Geneva, CH
PARDONS, A.	CERN, Geneva, CH
PARMA, V.	CERN, Geneva, CH
PEREIRA, A.M.	LIP, Coimbra, PT
PINILLA, M.	ELYTT Energy, Madrid, ES
POLTORATSKA, Y.	University of Dortmund, Dortmund, DE
RAMOS, D.	CERN, Geneva, CH
RATHJEN, C.	CERN, Geneva, CH
RECSEI, S.	KFKI RMKI, Budapest, HU
REMDE, H.	DESY, Hamburg, DE
RIBO, L.	CELLS, Barcelona, ES
ROSEBERRY, T.	ORNL, Oak Ridge, US
RUMIZ, L.	Sincrotrone Trieste S.C.p.A., Trieste, IT
SANCHEZ ALAMEDA, M.	Leybold Vacuum España S.A., Barcelona, ES
SARTA, G.	INFN-LNS, Catania, IT
SCHMIDT, J.	MAX-lab, Lund, SE
SCHROETER, C.D.	Max-Planck-Institut für Kernphysik, Heidelberg, DE
SEYVET, F.	CERN, Geneva, CH
SIEBER, T.A.	Max-Planck-Institut für Kernphysik, Heidelberg, DE
TOCK, J.P.	CERN, Geneva, CH
TOUSSAINT, P.	GANIL, Caen, FR
VOELKER, A.	Forschungszentrum Karlsruhe, Eggenstein, DE
WEISS, K.	CERN, Geneva, CH
WOHLENBERG, T.	DESY, Hamburg, DE
YADAV, D.	Raja Ramanna Centre for Advanced Technology, Indore, IN
YOSIFOV, D.	TRIUMF, Vancouver, CA
ZHUKOV, A.	Budker Institute of Nuclear Physics, Novosibirsk, RU