TOTAL AND PARTIAL PRESSURE MEASUREMENT IN VACUUM SYSTEMS

J H Leck

Blackie

TOTAL AND PARTIAL PRESSURE MEASUREMENT IN VACUUM SYSTEMS

J.H. LECK

Emeritus Professor
Department of Electrical Engineering
and Electronics
University of Liverpool

Blackie
Glasgow and London

Blackie & Son Limited
Bishopbriggs, Glasgow G64 2NZ

7 Leicester Place London WC2H 7BP

First published 1989
Softcover reprint of the hardcover 1st edition 1989

British Library Cataloguing in Publication Data

Leck, J.H.

Total and partial pressure measurement in vacuum systems
1. Vacuum technology
I. Title
621.5'5

ISBN-13:978-1-4612-8224-2 e-ISBN-13:978-1-4613-0877-5
DOI:10.1007/978-1-4613-0877-5

To Esma

Phototypesetting by Thomson Press (India) Limited, New Delhi

Preface

This book deals with the underlying theory and practical aspects of pressure gauges that are at present in general use. Because of the ever-increasing demands to provide a wider range of sophisticated and reliable vacuum equipment a good understanding of these instruments is of vital importance to all workers in the research and industrial sectors.

Of the gauges considered only the mechanical types are absolute, in the sense that they measure pressure directly as a force upon a liquid column or a solid surface. Under ideal conditions it is possible to calculate their sensitivities, which are the same for all gases and vapours. The recent developments in the viscous or molecular damping gauges indicate that these may also be considered absolute. Other gauges are indirect in that they involve the measurement of some secondary phenomenon which is pressure-dependent and therefore these gauges can only be used for measurement after calibration against an absolute standard. The radiometer or Knudsen type gauge has been excluded from the text since these are now only of historic interest. Also no mention is made of the integration techniques involving surface changes (such as work function) although these could have application under very special circumstances. The McLeod gauge is dealt with in some detail, for even though this gauge has few practical applications, it is the most sensitive absolute gauge available and has value as a reference standard.

Throughout, emphasis is on the precision and reproducibility of measurement. A short but important chapter is devoted to the subject of calibration, as this is important in many laboratories where the reproducibility of measurement must be guaranteed over long periods.

Two chapters are devoted to the subject of the mass spectrographic analysis of gases at low pressure, and, in particular, to the theory, design and operation of the residual gas analyser. One chapter deals with the quadrupole mass filter, which is universally used as a monitor of gas composition in all types of vacuum system. This is a good example of an instrument which, provided its characteristics are understood and it is operated correctly, can be an almost invaluable tool to the vacuum engineer.

At the present time the units for pressure measurement are not standardized. Because of the move towards SI units, the traditional "torr" has been replaced in many laboratories by either the Pascal (Pa) or the millibar (mbar). However, the millibar has been used throughout the book because it is presently favoured by the majority of equipment manufacturers and system operators. In sections where published results are quoted, original units have been retained.

The author gratefully acknowledges the help and encouragement he has received both from his colleagues in the Department of Engineering & Electronics at the University of Liverpool and from other laboratories. Most of all his thanks are to his wife Esma without whose assistance in typing the whole of the manuscript and correcting spelling and grammatical errors, the book would not have been completed.

JHL

Acknowledgments

Permission to reproduce the following material is gratefully acknowledged:
Figs. 1.8, 1.9 (Berman); 3.4 (Bartmess and Georgiadis), 3.22, 3.23 (Poulter and Sutton); 3.25, 3.26 (Kudzia and Słówko); 3.27 (Kuo); 5.4, 5.6 (Peggs); 6.8 (Craig and Harden); 7.16 (Reid and James); 7.18, 7.19, 7.20 (Mao *et al.*, Mao and Leck) from *Vacuum* with the permission of the authors and of Pergamon Journals Ltd, Oxford.
Figs. 1.13 (Sullivan); 1.14 (Hyland and Tilford); 1.15 (Ono *et al.*); 1.17 (Fremerey); 3.9, 3.10 (Watanabe); 3.11 (Blechschmidt); 3.19 (Pittaway); 3.21 (Poulter *et al.*); 5.7 (McCulloh *et al.*); 7.13 (Blanchard *et al.*); 7.22, 7.23 (Reagan *et al.*) from *J. Vac. Sci. Technol.* with the permission of the authors and of the American Institute of Physics, New York.
Figs. 1.16 (Fremerey and Boden); 2.14, 2.15 (English *et al.*); 2.17, 2.18 (Steckelmacher and Fletcher); 5.1 (Poulter); 7.9 (Holme *et al.*) from *J. Phys. D, J. Phys. E, J. Sci. Instrum.* with the permission of the authors and of the Institute of Physics, Bristol.
Fig. 3.13 (Pittaway) from Phillips Research Rept. **29**, with the permission of Phillips Int. BV, Eindhoven.
Fig. 7.25 (Holme *et al.*) from *Int. J. Mass Spectrom. Ion Phys.* with the permission of the authors and Elsevier Scientific Publications BV, Amsterdam.

Units of Pressure

1 bar	=	10^5 Pascal (Pa)
1 millibar (mbar)	=	100 Pascal (Pa)
760 torr	=	1 standard atmosphere
1 torr	=	1 mm of mercury to within 1 part in 7×10^4
1 millibar (mbar)	=	0.750062 torr
	=	0.75 torr to within 1 part in 10
1 torr	=	1.333231 millibar (mbar)
	=	1.33 millibar (mbar) to within 0.25%

Contents

1 Mechanical manometers

1.1 Liquid manometers

The simplest vacuum pressure gauge is the glass U-tube containing mercury or some other low vapour-pressure liquid. One limb of the tube is connected to the vacuum chamber, the pressure above the other being held at some fixed value by means of an auxiliary vacuum system. Usually the pressure in the auxiliary vacuum is maintained at a very low level compared with that in the main chamber, so that the unknown pressure is given directly by the difference in levels of the two liquid columns. Differences in level of 0.1 mm can just be detected by eye. This sets a limit of minimum pressure change detectable to 0.1 mbar for a mercury filling. The sensitivity can be increased approximately 15 times by replacing the mercury with a low-vapour-pressure oil (because of the reduced density), such as butyl phthalate or Apiezon diffusion-pump oil. This allows pressure differences of the order of 10^{-2} mbar to be detected by eye. There are many descriptions of the practical details of this type of gauge, those by Biondi[1] and Maslach[2] being good examples. Damage to the vacuum system can occur all too easily if fluid is forced out of the manometer by a sudden and large pressure difference set up accidentally across it, for example due to the pressure in one limb rising to atmospheric. The damage can be prevented by simple splash traps put in each of the two arms of the U-tube some distance above the liquid surfaces[3].

The principal sources of error in this manometer are: (i) liquid sticking to the glass, giving a variable capillary depression or elevation; (ii) irregular light refraction in the glass; (iii) a difference of oil composition, and hence density, in the two limbs; (iv) gases or vapours may become dissolved in the oil; and (v) a difference in temperature, and hence density, of the liquid in the two columns.

The surface tension (i) and the uneven refraction effects (ii) can be reduced to a negligible value compared with 0.1 mm by using large-bore thin-walled tubes, that is, not less than 10 mm diameter; (iii) and (iv) are closely related and difficult to estimate accurately. The amount of dissolved gas obviously depends upon the gas composition, and is very considerable for some vapours. Thus wherever the precise control of the purity and the quantity of a small sample of any gas is required, an oil manometer must be excluded from the system. Hickman and his co-workers[4, 5] have described a number of elaborate oil manometers in which it is possible to redistil the oil at frequent intervals inside the gauge itself, thus ensuring a high uniformity and purity of the oil.

The zero error due to a temperature difference is easy to calculate, being

proportional to the temperature difference between the two limbs, the temperature coefficient of expansion and the height of the liquid columns. Taking a realistic value for the temperature coefficient of volume expansion of oil as $10^{-3}\,°C^{-1}$, then for oil columns 300 mm long a temperature difference of 1 °C gives a zero error of 0.3 mm of oil (i.e. 0.02 mbar).

A number of workers have extended the working range of the mercury manometer down to 10^{-3} mbar by using elaborate techniques to detect small changes in liquid levels. The first to do so was Lord Rayleigh[6] in 1901. His design involved a rather elaborate tilting procedure, and used fine glass needles as reference levels. The idea of using a small float, usually a hollow glass bulb resting on the mercury surface, to follow the changes in level was developed later, notably by Newbury and Utterback[7], Shrader and Ryder[8] and Carver[9]. These workers used optical techniques to measure the displacement of the float. The gauge described in some considerable detail by Johnson and Harrison[10] is probably the most sensitive of its type. This gauge was based on the experience of the previous workers and obviously constructed with great care. With it pressure changes as small as 2×10^{-4} mbar could be detected. More recent publications have described a number of well-engineered optical and photo-optical techniques to determine the liquid level differences down to the order of 10^{-3} mm[11–16]. It is interesting to note that various other designs of liquid manometer are used in fluid mechanics work, for example the Chattock gauge[17], the inclined plane manometer[18] and the two-fluid manometer[19].

1.2 The McLeod gauge

The liquid manometer can be used to measure the pressures of permanent gases to below 10^{-2} mbar by compressing the gas by a known amount, thus introducing a kind of pressure amplifier before making the measurement. If, for example, the volume of 100 cm^3 of a permanent gas is reduced by some form of piston to 0.1 cm^3, then the pressure and therefore the manometer reading is increased 1000 times. In 1874, McLeod[20] designed a gauge employing this principle which is used, with only detailed modifications, in present-day applications.

The detailed construction of a modern version of this gauge constructed entirely of glass is shown in Figure 1.1. Normally the pressure in the auxiliary vacuum chamber C is held at a low pressure (a few mbar) so that the mercury levels are as indicated in Figure 1.1. To measure the pressure in the vacuum system, or, to be exact, the pressure in the bulb V, air is slowly admitted to C to force the mercury up the vertical tube A and the capillary B until it reaches the level YY' in B. (The difference in levels between YY' and the base of C must not be greater than the barometric height.) The mercury surface, when it passes the level WW', traps off a definite quantity of gas in the bulb V. This gas is then forced into the closed capillary D (which is of identical cross-section to B), the

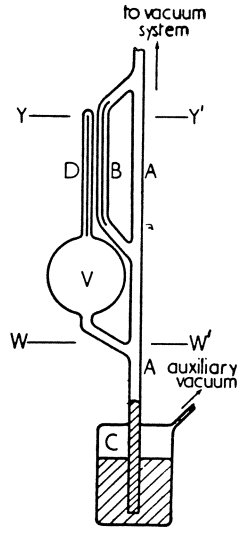


Figure 1.1 The McLeod gauge.

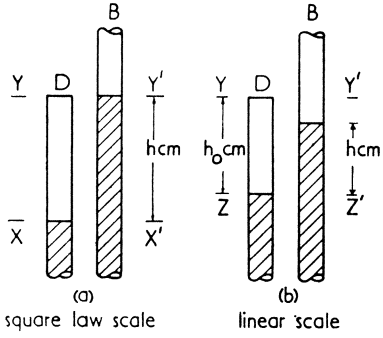


Figure 1.2 Detailed views of the capillary tubes of McLeod gauge. In (*a*), the mercury is raised to the level YY in the open tube (coincident with the top of the closed tube). In (*b*), the mercury is raised so that the level in the closed tube reaches ZZ

mercury acting like a piston as it fills up V. The mercury rises to a level XX' in the capillary D (see Figure 1.2*a* which shows on a larger scale the mercury levels in the equilibrium position). Applying Boyle's law to the trapped gas, which is compressed to a pressure $(p + h)$ in a volume V, gives the following equation:

$$pV_0 = (p + h)Ah \tag{1.1}$$

V_0 is the volume of gas trapped at the unknown pressure p (for convenience p is measured in centimetres of mercury). A cm^2 is the capillary cross-sectional area, and h cm the difference in mercury levels.

Thus
$$p = \frac{Ah^2}{V_0 - Ah} \tag{1.2}$$

As hA is in practice quite negligible compared with V_0

$$p = \frac{10Ah^2}{V_0} \text{ mm Hg} \tag{1.3}$$

i.e., p can be determined in terms of h and the physical dimensions of the apparatus.

The following typical values show the order of magnitude of the amplification achieved. If the capillary D has an area of cross-section $A = 10^{-2}$ cm^2, and a total length of 10 cm, and if $V_0 = 200$ cm^3, then

for $p = 5 \times 10^{-6}$ mm Hg, $h = 0.1$ cm, and
for $p = 5 \times 10^{-4}$ mm Hg, then $h = 1.0$ cm

i.e. 6.5×10^{-6} and 6.5×10^{-4} mbar respectively. The maximum pressure that can be measured is about 0.1 mbar (when h is a little less than 200 mm).

A linear relation between p and h can be obtained if, instead of raising the mercury level in B to YY', the level in D is raised to ZZ' (see Figure 1.2*b*), i.e. the compression ratio is made the same for all pressures.

In this case eqn (1.1) becomes

$$pV_0 = h_0 A(p + h) \tag{1.1'}$$

therefore

$$p = \frac{10h_0 A \cdot h}{V_0} \text{ mm Hg} \tag{1.3'}$$

The first technique, because of the following important advantages over the second, is the more commonly used in spite of its non-linearity. In the first, by raising the mercury level to YY' in B, the compression ratio, and therefore the pressure amplification, is made inversely proportional to p^2. In the second, the compression ratio is fixed at some intermediate value by raising it to a fixed level in D. Thus the first has the advantage of a higher sensitivity at low pressure and a larger useful working scale. Also, raising the mercury column by a fixed amount is of practical importance, especially in some of the operating techniques described below.

It is obvious from eqn (1.3) that this gauge is only reliable for gases that obey Boyle's law to a close approximation. A more accurate representation of the relation between pressure p and volume V for a fixed quantity of any gas is given by the van der Waals' equation, which reads

$$\left(p + \frac{a}{V^2}\right)(V - b) = RT \tag{1.4}$$

Table 1.1 Gas properties derived from tables in *Handbook of Chemistry and Physics*, 64th edn., ed. R.C. Weast (1984), CRC Press Inc., Boca Raton

Gas	T_c (K)	p_c (cm Hg)	a (cm Hg cm^6/mole^2)	b (cm^3/mole)	
	$T_c < T_k$; $T_k = 288$ K				
Ar	151	3.65×10^3	1.02×10^8	3.22×10^1	
CO	134	2.66×10^3	1.13×10^8	3.99×10^1	
He	5.1	1.72×10^2	2.58×10^6	2.37×10^1	
H_2	33.1	9.73×10^2	1.85×10^7	2.66×10^1	
Kr	210	4.11×10^3	1.76×10^8	3.98×10^1	
CH_4	190.5	3.48×10^3	1.71×10^8	4.28×10^1	
Ne	44.3	1.97×10^3	1.60×10^7	1.71×10^1	
NO	179	4.94×10^3	1.02×10^8	2.80×10^1	
N_2	125.9	2.54×10^3	1.05×10^8	3.92×10^1	
O_2	154.2	3.78×10^3	1.03×10^8	3.20×10^1	
	$T_c > T_k$; $T_k = 288$ K				$p_d(T_k)$ (cm Hg)
NH_3	405.4	8.47×10^3	3.17×10^8	3.71×10^1	5.3×10^2
C_2H_2	309	4.71×10^3	3.33×10^8	5.13×10^1	2.9×10^3
C_6H_6	561.5	3.63×10^3	1.36×10^9	1.16×10^2	6×10^0
CO_2	304.1	5.55×10^3	2.72×10^8	4.27×10^1	3.8×10^3
Hg	> 1823	$> 1.52 \times 10^4$	6.12×10^8	1.70×10^1	7.76×10^{-5}
NO_2	431	7.52×10^3	4.01×10^8	4.43×10^1	5.65×10^1
SO_2	430.2	5.91×10^3	5.08×10^8	5.65×10^1	2.1×10^2
H_2O	647	1.64×10^4	4.14×10^8	3.06×10^1	1.28×10^0
Xe	289.6	4.42×10^3	3.10×10^8	5.10×10^1	4.4×10^3

T_c = critical temperature
T_k = measuring temperature
p_c = critical pressure
$p_d(T_k)$ = saturated vapour pressure at T_k K

where R is the universal gas constant, T temperature, and a and b are constants which depend upon the nature of the gas. Equation (1.4) only holds for gases in the McLeod gauge provided that either the gas critical temperature is less than that of the gauge or, if this is not true, then the maximum pressure in the closed limb is less than the saturated vapour pressure at the measuring temperature. The relevant properties for a number of important gases are summarized in Table 1.1.

Jansen and Venema[21] point out that the error introduced by using eqn (1.1), rather than the more accurate eqn (1.4), is given by Δp_0 where

$$\frac{\Delta p_0}{p_0} = \frac{p_0}{RT}\left(\frac{a}{RT} - b\right) \tag{1.4'}$$

p_0 being the final pressure in the closed capillary.

Figure 1.3 shows the magnitude of the error for four representative gases;

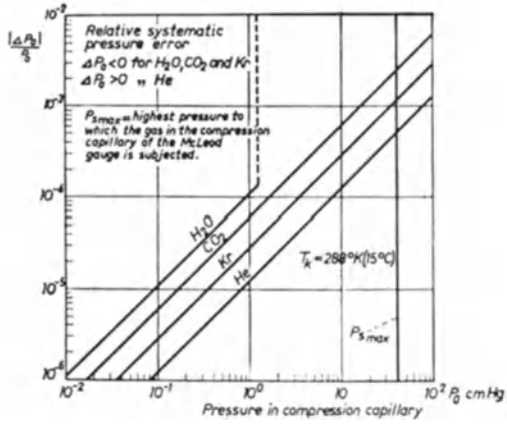


Figure 1.3 Values of the relative systematic error $\Delta p_0/p_0$ for helium, krypton, carbon dioxide and water vapour introduced in the measurement of p_0 when van der Waals' equation is simplified to the ideal gas formula. These calculations were made by Jansen and Venema[21].

helium and krypton (where the critical temperature is above normal ambient) and carbon dioxide and water vapour. The a and b (eqn 1.4) terms are relatively high for krypton and water vapour, but low for helium and carbon dioxide. For all practical purposes, therefore, all the gases listed in Table 1.1 can be assumed to obey Boyle's law with regard to the McLeod gauge. For benzene and water, vapour pressures of 60 and 12 mm Hg respectively must not be exceeded in the closed limb after compression.

Bixler *et al.*[22] have demonstrated experimentally that the McLeod gauge can be used for a wide range of gases and vapours. They found errors of not more than 2% in calibrating a number of gases ranging from CH_3Cl to carbon dioxide with saturated vapour pressures of 4.2×10^3 and 48×10^3 mm Hg respectively. Some of the special problems involved in this type of measurement are described by Armbruster[23], Flosdorf[24], and Hayward[25], who show the practicability of measuring water vapour pressures and gas–vapour mixtures when capillary tubes are heated.

In practice, the following two factors set the limit to the accuracy of the gauge; firstly, the uncertainty in the measurement of h due to capillary depression, and secondly, non-uniformity of the bore of the closed capillary D near the point of seal-off. The first is the more important as, unfortunately, the errors are increased when the bore of the tubes is reduced. The depression of the mercury level in the two tubes may differ even when they have identical cross-sections and the whole of the system is scrupulously clean. Porter[26] has carried out an experimental and theoretical investigation of capillary depression and in particular the importance of tube diameter. He has verified a formula relating depression to the angle of contact between the mercury and glass, showing that it is the variation of this angle which gives rise to the errors.

For the same mercury and the same glass tube (both clean), the contact area can vary from less than 20° to more than 60° depending, for example, on whether the mercury rises or falls to its equilibrium rest position. The curves given in Figure 2 of Porter's publication[26] show the order of magnitude of the errors likely to be encountered in manometer work. For angles of contact of 30° and 60°, the capillary depressions are 2.8 and 1.8 mm respectively for a 1.9 mm bore tube, but only 0.97 and 0.59 mm for a tube of twice the bore. Thus, taking Porter's figures means that in a McLeod gauge with capillary tubes of 1-mm bore the measured values of h are accurate only to ± 0.5 mm. Taking the practical example given on p. 4, this gives a possible error of $\pm 10\%$ at 5×10^{-4} mbar and more than $\pm 100\%$ at 5×10^{-6} mbar. This error cannot be reduced by decreasing A because, while the sensitivity increases in proportion to $1/A$, the error goes up at approximately the same rate.

It is of value to study work of Rosenberg[27] to see the accuracy that can be obtained with a gauge of standard form. He made $V_0 = 1300\,\text{cm}^3$ (the lower portion of his glass bulb had to be sunk into a plaster of Paris bed to support the weight of mercury), and the capillaries 0.63 mm in diameter. The inner surfaces of these tubes were roughened, as it had been shown earlier by both Rosenberg[28] and by Klemperer[29] that this treatment reduced the uncertainty in the capillary depression by a large factor. (Although the tubes became translucent, the menisci could be viewed with the help of strong back-illumination.) Great care was taken to ensure both clean mercury and glass, to the extent of never allowing the pressure inside the gauge up to atmosphere after the initial evacuation. Results accurate and reproducible to ± 0.5, ± 0.6, ± 2 and $\pm 6\%$ at pressures of 10^{-2}, 10^{-3}, 10^{-4} and 10^{-5} mbar were obtained. If we take his results as typical of those obtained 'with good working practice', the performance of other gauges can be estimated. For example, a laboratory instrument with $V_0 = 200\,\text{cm}^3$ cannot be expected to have an accuracy to better than $\pm 6\%$ at $10^{-5} \times 1300/200 = 6.5 \times 10^{-5}$ mbar and $\pm 2\%$ at 6.5×10^{-4} mbar, even with the most careful operation.

More recently, Jansen and Venema[21] have described a carefully designed gauge which is a good representation of the best working practice. Figure 1.4 shows their instrument complete, and Figure 1.5 gives the detailed dimensions of the multiple capillary, showing the close manufacturing tolerances demanded. The use of the multiple capillary allows an extensive range of pressure to be covered in this one instrument. Four scales can be used; three linear with the compression carried to 0_3, 0_2 and 0_1 (Figure 1.4), and a square law scale with the compression carried to 0_0. These zero levels have been carefully chosen so that the scales just overlap, allowing pressures up to 3.5 mm Hg to be measured without the difference in column height exceeding 400 mm. Each section of the closed tube is provided with a parallel comparison tube of the same internal diameter. The top of the closed tube is sealed off perfectly flat with a glass plug (as recommended by Barr and Anhorn[30]), and the bores of all the tubes are roughened as recommended by Rosenberg[28]. The small

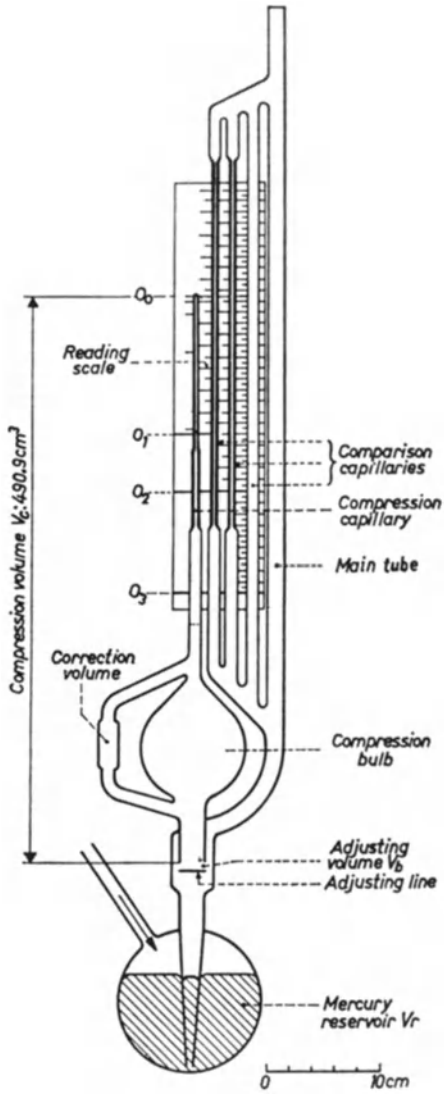


Figure 1.4 McLeod gauge designed by Jansen and Venema[21].

correction volume shown in Figure 1.4 enables the volume of the compression bulb to be adjusted to the required value (490.3 cm^3) in a simple glassblowing operation.

Great care was taken both in the construction and operation of this gauge. Attention was paid to the setting up to ensure that the capillaries were mounted in a vertical plane with the zero lines accurately horizontal. The

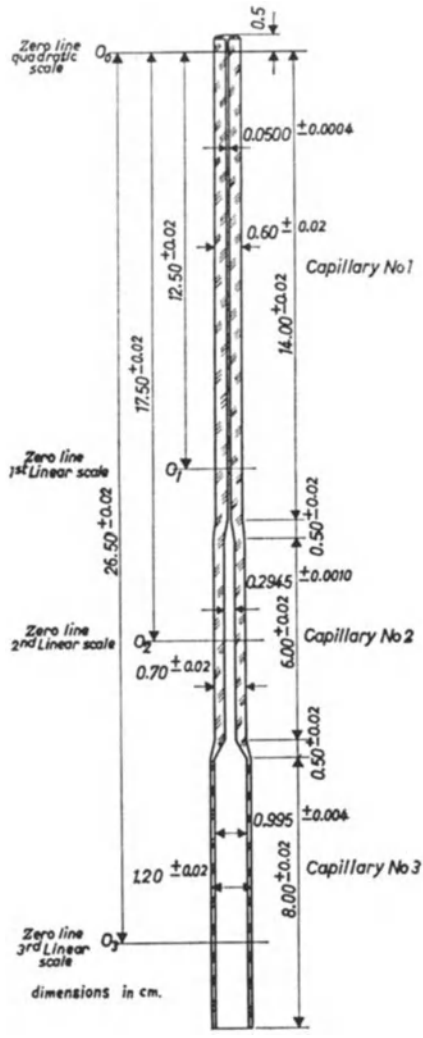


Figure 1.5 Construction of the compression capillary for the gauge designed by Jansen and Venema[21].

adjusting volume (the change in the system volume between starting to raise the mercury piston and sealing off the gas in the compression bulb) was kept to a reasonable minimum and constant at 4.5 cm^3 by always returning the mercury in the rest position to the adjusting line marked, as shown in Figure 1.4. This is obviously important when measuring pressures in a sealed system as the volume, and hence the pressure, changes every time a reading is taken. The systematic error in capillary depression was determined over the

full length of the tubes when the system pressure was very low ($< 10^{-6}$ mbar) by raising the mercury up to the top of the closed capillary in small steps and noting the difference in levels between open and closed tubes at each step. This gave correction values which could be applied in subsequent measurement. Jansen and Venema[21] estimated that the maximum systematic error due to inaccuracies of construction and setting up was below 3% for pressures in excess of 10^{-4} mbar (taking the unlikely worst possible case where all the errors were additive). The error must clearly increase steadily for decreasing pressure, making the gauge unreliable as a standard below 10^{-4} mbar.

In addition to these setting-up errors, there are the so-called 'accidental' errors introduced in measuring the column heights, and also, more importantly, the variation in capillary depression. Jansen and Venema[21] made their measurements both with and without 'tapping the bores', concluding that tapping, with a specially designed mechanical vibrator, reduced the errors by a factor of between two and three. To substantiate this, they quoted figures which show an average error of 0.37 mm (a maximum of 0.6 mm) without, and an average of 0.13 mm (maximum 0.3 mm) with tapping over a series of 14 readings in each case. They estimated that capillary depression was the greatest source of error in their gauge.

The possible overall accuracies of manufacture and operation were checked by comparing four gauges all made to identical specifications. As expected, the maximum deviation between the four was found to increase (from 0.4% to 3%) with decreasing pressure (from 3 to 0.7×10^{-3} mm Hg). It is encouraging to note the very satisfactory agreement of these results with the earlier work of Rosenberg[27].

Podgurski and Davis[31] suggested an alternative construction to reduce the problem caused by variable capillary depression. They followed the design by Keevill *et al.*[32], in which the open-ended comparison capillary of the conventional gauge was replaced by a wide-bore tube sealed off by the mercury piston at a low pressure. The details of their construction are shown in Figure 1.6. In operation the side arm, labelled *A*, is first evacuated by lowering the mercury to level *C*, with the whole system evacuated to a low pressure. Then with the pressure still very low ($< 10^{-6}$ mbar) a correction curve, to take account of the capillary depression tube, is obtained. This correction curve can be checked at intervals during gauge operation to indicate the stability of the gauge. This technique has the advantage of reducing the possible sources of error from two to one, and the further advantage that the correction is measured uniquely along the length of the closed tube. In the conventional gauge (for example of Jansen and Venema[21]) it is measured as the difference in depression at two points at approximately the same level, and subsequently applied to two points at different levels. This is clearly important when the capillary depression varies significantly over the length of the tubes. There is the slight disadvantage that a correction must be made to every reading.

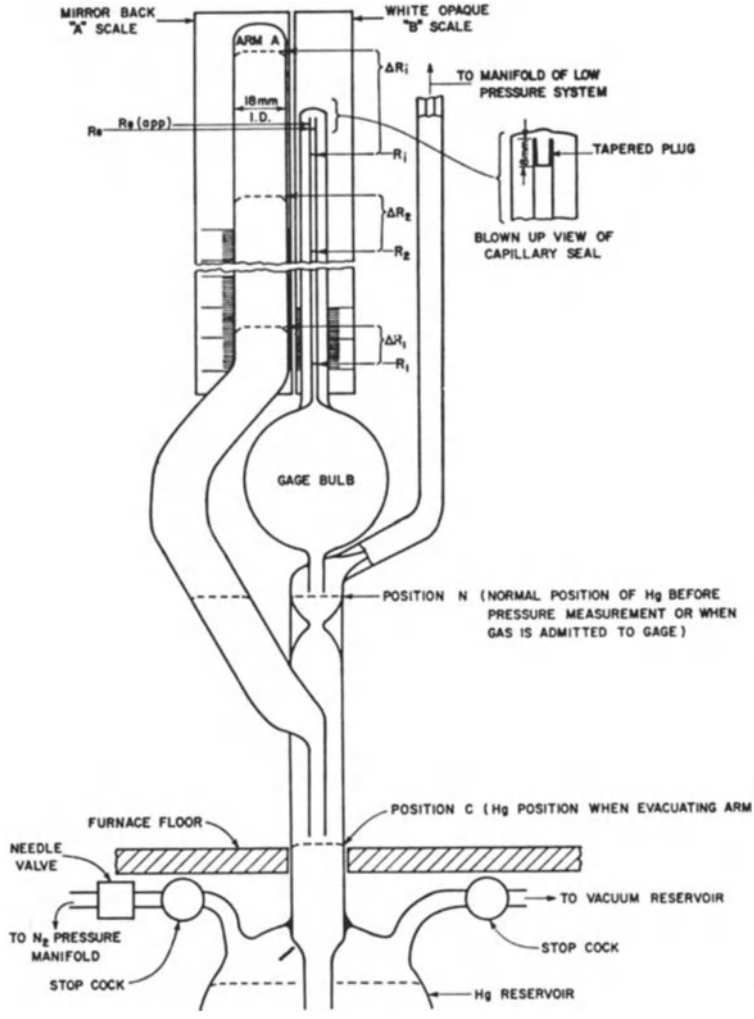


Figure 1.6 McLeod gauge designed by Podgurski and Davis[31].

This technique has been followed and refined in the National Physical Laboratories in London and New Delhi by Elliott *et al.*[33] and by Sharma *et al.*[34] respectively. Both groups demonstrated that it gave a considerable advantage over the standard method of operation by allowing a more accurate correction for capillary depression to be made. They showed that in the limit the reproducibility of measurement was to within $\pm 2\%$. With regard to the care necessary, it is interesting to note the comments of the workers from the National Physical Laboratories in London[33]:

> The McLeod gauge, with careful operation and allowance for the mercury vapour stream effect*, is capable of the calibration of ionization gauges down to pressures approaching 10^{-6} mbar with an accuracy in the region of $\pm 2\%$ or possibly $\pm 1\%$ under favourable conditions. However, for the maintenance of this accuracy, especially at the lower pressures, elaborate calibration procedures are necessary and the actual measurements are lengthy and tedious. The apparatus itself is delicate and very liable to accidental damage. It seems unlikely that much improvement in this situation will be possible, or that the range of the gauge, seen as a reference standard, can, with any confidence, be extended to appreciably lower pressures.

A further point has been made by Podgurski and Davis[31] regarding the design of the compression volume. They noted that, with the conventional method of sealing off the capillary, the heat applied to the plug alters the surface finish at the top of the bore, giving erratic behaviour of the mercury surface in that region. The whole tube had been roughened before mounting, as recommended by Rosenberg[28]. They overcame the difficulty by increasing the length of the tapered plug and sealing at the outer tip as shown in the insert diagram of Figure 1.6. Podgurski and Davis[31] also described a simple technique for determining the effective 'top' of the sealed tube. This followed the method adopted by Clark[35], who had pointed out that the assumption of a perfectly square end cannot be realized.

A further alternative design made specifically to reduce the errors caused by capillary depression has been proposed by Moser and Poltz[36]. In their gauge, compression is constant (as described in Figure 1.2*b*), the gas always being compressed into a small auxiliary chamber at the closed end of the capillary. The lower part of this chamber forms a sharp edge, and when the mercury level is raised it sticks at this point. It is possible to adjust the position of the mercury so that it forms a flat plane at this edge by the following optical method[36]:

> A source of light is placed over the chamber and the mercury level observed through a beam splitter by a horizontally placed microscope. If the level is very curved a little circular light spot may be observed in its centre. The diameter of this spot increases as the curvature of the mercury meniscus decreases. If the meniscus is approximately flat its whole area is bright.

As well as Moser and Poltz[36], Miller[37] and Cespiro[38] have carried out experimental work with this gauge. Miller[37] reported uncertainties of measurement of 4, 8, 25 and 90% at 10^{-6}, 10^{-7}, 10^{-8} and 10^{-9} mbar respectively. Cespiro obtained similar results and in his report highlighted the difficulties in designing the optical system to obtain a true measure of the 'flatness' of the mercury surface.

Akiyima *et al.*[39] developed a gauge with two bulbs (each of volume 1000 cm^3) using the first for the McLeod gauge proper and the second as a fixed compression ratio mercury piston pump (a modified Toepler pump adopting a principle put forward by Groszkowski[40]). It is significant that, even with this modification, they found the accuracy limited to about $\pm 5\%$ at 10^{-5} mbar and $\pm 6\%$ at 10^{-6} mbar.

---

*See p. 14.

An entirely different operating technique in which the whole gauge rotates about a horizontal axis has been successfully used for many years for high-pressure operation (10^{-3} to 10 mbar). The following description, taken from the publication by Flosdorf[41], shows the principles of all these swivel-type gauges.

> [In Figure 1.7] the gauge is shown diagrammatically in position *R* for reading, and (in broken lines) when turned to position *P* for acquiring the pressure of the system to be measured. For pressures above 0.001 mbar in order that the gauge may be turned on swivel *S*, heavy-walled rubber tubing free from sulphur and talc is satisfactory for connection to the vacuum system at point *A*. For lower pressures a ground glass joint is used to permit turning the gauge on the swivel, and glass tubing is brought from point *A* to point *B*, where it is passed out through the back of the metal case. The vacuum connection is attached at *B*, in the centre of the swivel axis, by a glass-to-metal ground joint. In position *P*, which is not exactly horizontal, the mercury will drain completely into bulb *C*. A constriction *D* prevents the mercury from travelling too rapidly into and out of the capillaries so that the gauge may be swung very rapidly to and from the position *R*. The diameter of bulb *C* is such that when a pressure reading is made, the mercury in the right-hand capillary always comes to the top line, irrespective of whether the mercury is high or low in the centre capillary. The safety trap *E* permits the portable gauge to be carried while full of mercury without loss of mercury even when inverted.

This gauge is useful only in the comparatively high-pressure range, because the weight of mercury would be excessive if the bulb volume was made greater than 100 cm^3. Numerous types of swivel gauges are described in the literature. Modifications to the standard type have been made, generally at the expense of

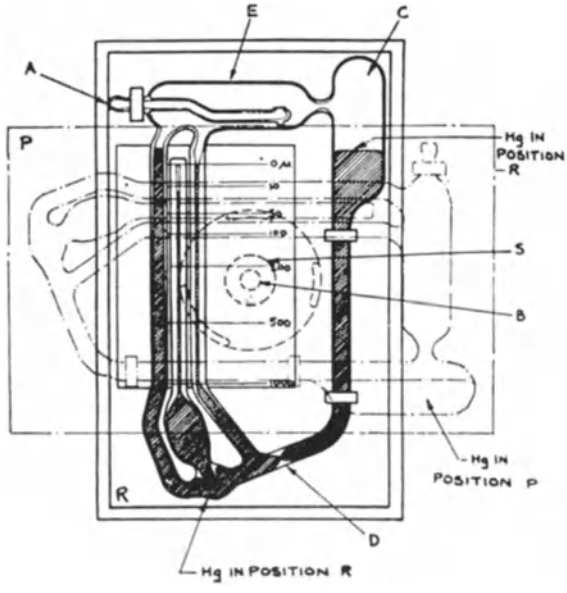


Figure 1.7 Swivel-type McLeod gauge designed by Flosdorf[41].

a more complicated construction, to allow the sensitivity to be increased. The model described by Axelbank[42] is a particularly good example.

For many applications where mercury vapour has to be excluded from the vacuum system, for example in calibrating an ionization gauge, a cold trap must be placed between the system and the gauge. This has a number of disadvantages apart from the obvious increase of cost and complexity. All vapours are condensed to a greater or lesser extent in the trap, and hence the measurements become strictly limited to permanent gases.

A further error introduced by the cold trap, noted by Gaede in 1915[43], received surprisingly little attention from users of McLeod gauges until Ishii and Nakayama[44] published their measurements in 1961. This, despite the fact that Dushman's well-known text *Scientific Foundations of Vacuum Technique*[106], contains a complete description of the process. Neglect of this effect can be serious in certain circumstances. The error is due to a pressure gradient of gas in the pipe connecting the cold trap to the gauge, caused by the continuous stream of mercury vapour moving from gauge to trap. It is interesting to note that Gaede[43] made this point at the time he was measuring the interdiffusion of mercury and nitrogen (work which led to his setting out the principles of the diffusion pump). The error in reading ($\Delta p/p$), due to mercury streaming, can be calculated in terms of the diffusion coefficient for mercury. For the simple case where a gauge and cold trap are connected by a long circular pipe of radius R cm,

$$\Delta p/p = \exp(0.09\, R p_m \sqrt{T}/D) - 1 \tag{1.5}$$

$$\simeq 0.09\, R p_m \sqrt{T}/D \tag{1.6}$$

p_m being the saturated vapour pressure of mercury (mm Hg) at room temperature T K, and D the coefficient of diffusion of the gas against mercury at atmospheric pressure.

Evidence in favour of Gaede's theory came when Ishii and Nakayama[44] observed an apparent change in sensitivity (of about 25%) of a number of ionization gauges over a long period of time. The instabilities could not be due directly to the ionization gauges, as these gave self-consistent readings to within $\pm 1\%$, and so must be attributed to the failure of the measuring technique. The errors were found to be systematic only with the variation in room temperature. After carefully eliminating the various possible links between sensitivity and temperature for the McLeod or ionization gauges (such as shrinkage of the glass, thermal transpiration, etc.), Ishii and Nakayama[44] concluded that only the interdiffusion of nitrogen and mercury could account for the magnitude of the error.

Following the initial report, a number of workers have confirmed the existence of this mercury diffusion effect and have been able to report good agreement between the experimental observations and theoretical predictions. Notable are the papers of Meinke and Reich[45], Elliott *et al.*[33], de

Table 1.2 Experimental and calculated values of $\Delta p/p$ obtained by Elliott and co-workers[33]. The values of diffusion coefficient (D) which they assumed are also given.

Gas	D(cm^2 s^{-1}) (Calculated)	$\Delta p/p$ (tube 10 mm diameter, 23° C) Theoretical	Experimental
Helium	0.6833	0.02	0.02
Neon	0.2399	0.05	0.04
Nitrogen	0.1263	0.10	0.10
Argon	0.1092	0.11	0.09
Carbon dioxide	0.0788	0.16	0.16
Krypton	0.0691	0.19	0.16
Xenon	0.0497	0.27	0.27

Vries and Rol[46], Berman[47] and Sharma *et al.*[34]. Elliott and co-workers[33] made careful measurements for a number of gases ranging from helium to xenon. Their results confirmed the theoretical predictions with respect to (i) the dependence upon the type of gas molecule and (ii) the linear relation between error and the radius of the connecting tube. Table 1.2 gives an indication of the data obtained by these workers for seven gases, and also lists the values of diffusion coefficient chosen for their calculations. These results indicate the strong dependence upon gas composition. The near-perfect agreement between theory and experiment obtained by Berman[47] can be seen from the curves plotted in Figure 1.8, which show results for a 3-mm radius

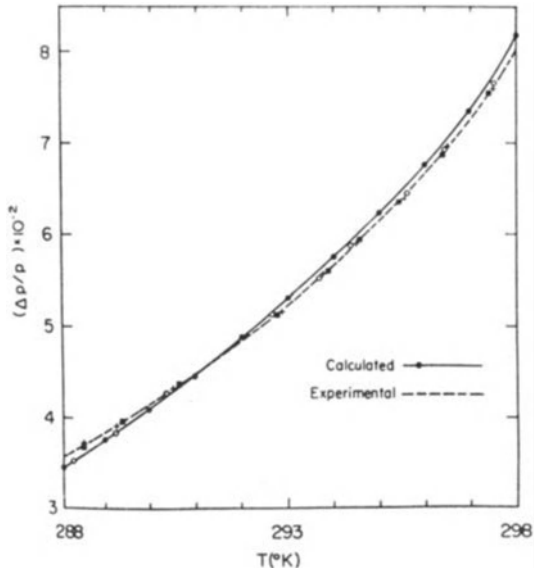


Figure 1.8 Calculated and experimental curves of $\Delta p/p$ as a function of temperature for $r = 3.0$ mm and a measuring pressure of pure nitrogen $< 6 \times 10^{-4}$ mbar (after Berman[47]).

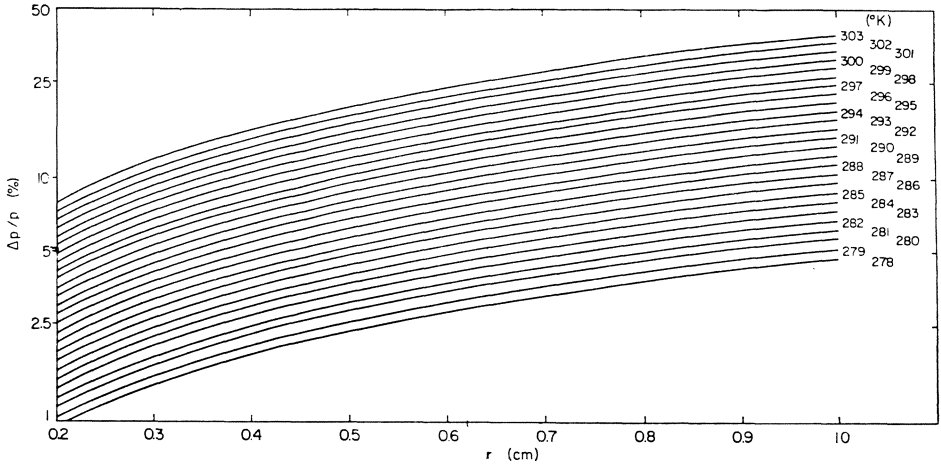


Figure 1.9 $\Delta p/p$ as a function of the tube radius for temperatures ranging from 278 to 303 K at 1 K intervals[47].

connecting tube over the temperature range 288 to 298 K. These curves also indicate the strong dependence of the pressure ratio upon T (i.e. room temperature). Figure 1.9 is the theoretical set of curves published by Berman[47] showing the errors for nitrogen plotted as a function of temperature for a wide range of tube radii.

Ishii and Nakayama[44] noted the mercury streaming effect partly because they worked in an unheated laboratory where room temperature varied from 10 °C in January to 30 °C in July/August. Apart from working in a very cold laboratory, the effect can be reduced to a minimum by making the radius of the McLeod gauge connecting tube as small as practicable. Elliott *et al.*[33] quoted $\sum d_3/\sum d_2$ as the effective diameter of a connecting tube with a non-uniform bore (diameter $d_1, d_2, d_3 \ldots$). Tunnicliffe and Rees[48] noted that care must be taken when applying corrections with the non-uniform bore because of the possibility of mercury evaporation from the wide-bore sections of the tube.

1.3 The diaphragm manometer

Considerable development work has taken place over many years in order to make the diaphragm gauge a useful alternative to the liquid manometer for vacuum work. With the diaphragm replacing the liquid there is no possibility of vapours entering the vacuum system and, again, the sensitivity is the same for all gases and vapours. Reliable and accurate electrical recording techniques have made it comparatively robust and reliable. In principle, the diaphragm gauge is essentially the same as the standard aneroid barometer. In the early models, optical techniques were usually used to measure the diaphragm deflections. With these gauges, pressure differences of less than 1 mbar could

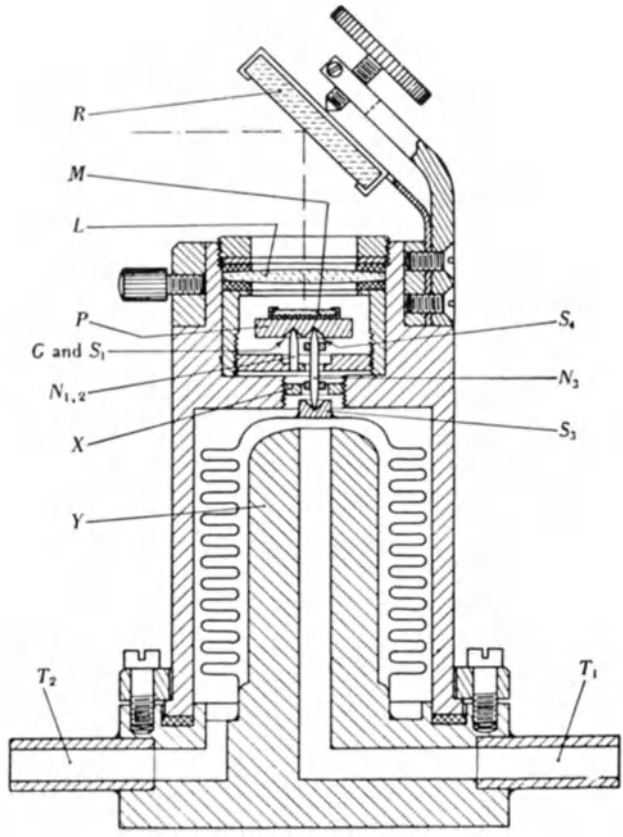


Figure 1.10 Differential bellows manometer designed by East and Kuhn[49].

be detected, but, probably because of their delicate nature, they were not adopted for general vacuum use. The instrument shown in Figure 1.10 is a good example of this type of gauge; in this case a flexible bellows is used instead of a plane diaphragm[49]. East and Kuhn[49] described this gauge as follows:

> The displacement of the bellows causes the small platform P to tilt. A plane mirror M fixed to the top of the platform transmits the tilt to an optical lever similar to that used in mirror galvanometers. The front-silvered mirror R reflects the beam into a horizontal direction. The manometer is mounted on one end of a piece of channel iron, the other end of which carries a lamp and scale. The lens L is fitted into the case with rubber washers and forms an airtight window. The tubes T_1 and T_2 lead to the inside and outside of the bellows respectively. The gauge can thus be used for measurement of pressure differences at any absolute value of pressure, from vacuum up to several atmospheres. $N_{1,2}$ are two needles, one behind the other, on which the platform rests, the position of the latter being fixed by the groove G and a sink S_1. The double-pointed needle N_3 rests in a sink S_3 on the top of the bellows and supports the platform in the sink S_4. This arrangement has the advantage that small lateral displacements, which most bellows show on extension, do not affect the length of the lever which is determined by the distance from S_4 to G. This distance is made small

(3 mm). As a result of this only small displacements of the bellows are used, and any possibility of hysteresis or deviation from Hooke's law is excluded.

With their most sensitive gauge, having a Hydron standard bellows of 50 mm external diameter, East and Kuhn[49] were able to detect pressure changes of 0.5×10^{-3} mbar. Successful Bourdon-type manometers using combincd optical-mechanical levers simpler than that described above, but still satisfactory for work below 1 mbar, have been developed. Crompton and Elford[50] in their instrument, for example, claim an accuracy of 1% at 1 mbar and 0.1% at 20 mbar. More robust gauges with standard mechanical levers are now commercially available which can be used to measure pressures down to 0.1 mbar, with an accuracy of about 10%.

The more recent trends in design have been in the use of electrical, rather than optical, techniques for detecting the small movement of diaphragm or bellows. This has resulted not only in greater sensitivities but, more importantly, vast improvements in the robustness and general ease of operation of the instruments. In most gauges the diaphragm and a fixed electrode form an electrical capacitor. Any movement of the diaphragm, and therefore any pressure change, can be detected as a change in capacitance. (This technique is not new, having been first used successfully by Olsen and Hirst[51] in 1929.)

The lower useful limit to the pressure range of the simple diaphragm gauge is, in general, not set by either changes in elasticity or hysteresis effects in the diaphragm nor by the difficulties of measuring small capacitance changes, but by the distortions caused by temperature variations in the gauge. Pressey[52] considered this problem of temperature instability both theoretically and experimentally, and concluded that a temperature coefficient of capacitance α_c (the fractional rate of change of capacitance with temperature) of the order of 10^{-4} must be expected, even in a well-designed gauge. He showed that two effects contribute to the temperature coefficient; firstly, a change in gap due to the expansion of the body, and secondly a possible diaphragm buckling due to differing expansion rates between different parts of the gauge.

Equations (1.7) and (1.8) below (evaluated by Pressey[52]) give the temperature coefficient α_c for the gauge shown in Figure 1.11, for two extreme cases where the temperature (i) changes slowly, keeping the whole gauge at a uniform temperature; and (ii) changes very quickly so that the diaphragm has not followed the body-temperature change.

For a uniform increase of temperature

$$\alpha_c = \alpha - L(\alpha - \delta)/D \tag{1.7}$$

For increase in body temperature only

$$\alpha_c = \alpha[2 - (L + M + D)/D] \tag{1.8}$$

α and δ are the temperature coefficients of expansion of the metal and insulating material respectively.

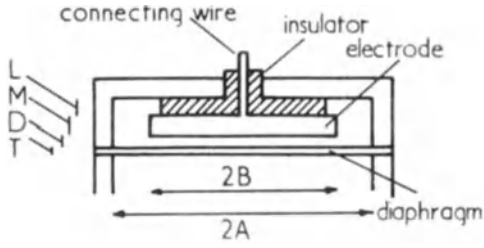


Figure 1.11 Simple diaphragm gauge having a circular plane diaphragm. The dimensions used in the text are shown on this diagram. c_0 is the capacitance between the electrode and the diaphragm when it is in its central position. y_0 and c are the deflections at the centre and the change in capacitance respectively due to a pressure difference p.

Equation (1.9) below gives y_1, the deflection at the centre of the plate, for a uniform temperature rise of $t\,^{\circ}\mathrm{C}$ in terms of γ, the difference between the temperature coefficients of expansion of the body and diaphragm:

$$y_1 = 1.28\,AD\sqrt{\gamma t} \tag{1.9}$$

These three equations show that the depth of insulation L and the difference in thermal expansion γ should be made as small as possible; in practice α_c will be negative. Thus, for the best results, body and diaphragm should be made from the same material.

Pressey[52] constructed the gauge shown in Figure 1.12 in order to check the above formula qualitatively. He chose stainless steel as the material for both body and diaphragm because of its relatively low coefficient of expansion (about 11 ppm $^{\circ}\mathrm{C}^{-1}$); the diaphragm had $T = 0.05$ mm and $A = 23$ mm. With the chosen air gap $D = 0.1$ mm, the total capacitance was 80 pF and the sensitivity 0.3 pF mbar^{-1}. The measured value of temperature coefficient α_c of -1.0×10^{-4} was close to the calculated value of -1.44×10^{-4}. Thus in this gauge a temperature change of 1 °C gave the same output signal as a pressure change of 2.5×10^{-3} mbar.

Since the 1950s very significant advances in this technology have been made, not by any fundamental inventions but rather by steady improvements in engineering design. An excellent account of these developments has been written by Sullivan[53]. He emphasizes the particular attention that has been paid (i) to the problems caused by temperature instabilities in the sensors, and (ii) to the manufacture of the extremely thin diaphragms, and describes the commercially-built gauges now available for general vacuum use which are more reliable, rugged and straightforward in operation than that built by Pressey[52].

In modern gauges, a close match between the coefficients of expansion of metal and insulator parts has been achieved. Usually inconel is used for all the metal parts (including the diaphragm) and ceramic for the insulating sections. An all-welded construction is preferred as it gives close control over

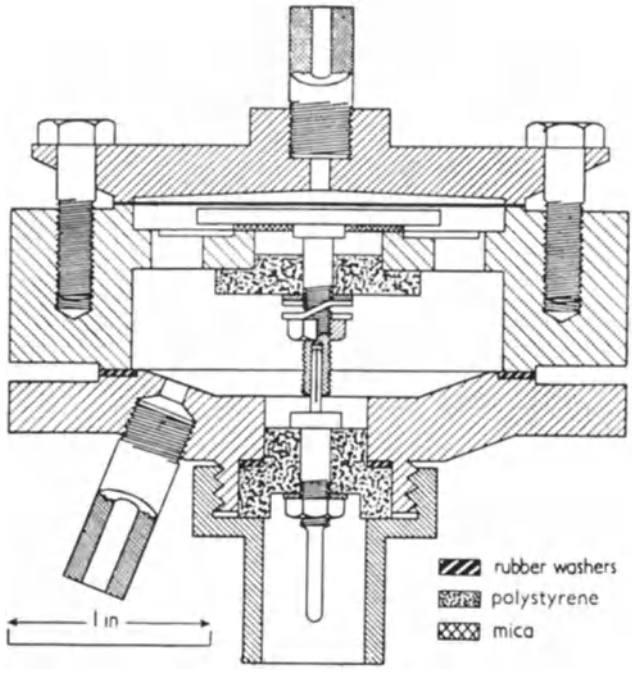


Figure 1.12 Sectional scale drawing of the differential gauge designed by Pressey[52]. The thickness of the diaphragm and the gap are enlarged for clarity. The bolts attaching the lower end-cap to the body do not show in this section.

diaphragm tensioning. This also improves the vacuum seal, so that for many applications continuous pumping of the reference volume is not required, a 'sealed-off volume' with a small getter being adequate to maintain effectively zero pressure. As an alternative to relying upon good thermal lagging to keep temperature-induced stress to a minimum, the whole gauge head is often placed in a temperature-controlled oven operating at about 40 °C. This technique can improve temperature stability by two orders of magnitude. As a further measure to reduce mechanical stress, the design is often modified to include a 'guard volume' which surrounds both vacuum chambers. Various alternative-capacity configurations can be used for measuring diaphragm deflections in addition to that shown in Figure 1.12. A symmetrical system can be used with two identical electrodes (one placed on each side of the diaphragm) so that a differential signal can be detected in a balanced bridge. Placing one electrode in the 'working' vacuum may be a disadvantage, particularly when the gauge operates in a hostile environment. A different double electrode system, the 'bull's eye' configuration, has been introduced to overcome this problem. The electrodes are both on the 'guard' vacuum side of the diaphragm; the first being a central disc, the second a concentric ring. Any

movement causes both capacitances to change in the same sense but at a different rate, hence analysis in a bridge is still realistic.

As an alternative to the bridge, a frequency-change technique can be used which needs only a 'single-sided' electrode system. Here a change in capacity changes the frequency of an oscillator, which can be monitored with a counting method, as shown in Figure 1.13[53]. Because of the sophistication of modern digital and analogue electronic circuits, the ability to measure small movements of the diaphragm is not a serious limit to instrument performance.

These instruments with electronic monitoring have been studied in a number of laboratories. Hyland and Tilford[54], reporting on the performance of a large number of instruments, stressed the importance of zero stability in all measurements where a high precision is needed. Temperature instabilities were found to be the main cause of instability, and in a comprehensive report these authors quoted values of temperature coefficient of zero change from below 4 ppm $°C^{-1}$ to above 100 ppm $°C^{-1}$. It was difficult to correlate the performance of different gauges, even from batches of instruments of identical design. They stressed the need for users to be aware of zero drift problems and to make frequent checks, especially if a gauge is being used in standards work. In this same work at the National Bureau of Standards, measurements were made of stability, and non-linearity, in sensitivity. A report is given of measurements made over a four-year period at the NBS, which Hyland and Tilford[54] claim give results typical of most capacitor diaphragm gauges with respect to the magnitude of the calibration factor shifts and the lack of any apparent linear, or even monotonic, drift with time. Results for one particular gauge are shown in Figure 1.14. No correlation could be found between the shifts in performance occurring from time to time and the handling of the gauges, and in this respect the authors state:

> For example the results [in Figure 1.14] were for a gauge shipped overseas and back in the interval between the August 1980 and March 1981 calibrations, and again between the August 1981 and December 1981 calibrations. For this gauge relatively large shifts were observed after the shipping, while a shift of less than 0.1% was observed between March and August of 1981 when the gauge never left our laboratory. However another gauge shipped and handled at the same time showed small (0.1%) changes after shipment but changed by 0.7% between March and August 1981. As a further example of our inability to correlate calibration shifts with treatment of the gauge we have accidentally over-pressured a gauge in our laboratory but found it to repeat its calibration to within 0.07%, an order of magnitude better than the stability observed for some gauges treated in an exemplary manner.

The average magnitude of the long-term drifts in sensitivity observed by Hyland and Tilford[54] were approximately 0.4%, a figure consistent with the findings of other workers – Poulter[55] and Reich[56] for example. They also found a distinctly superior performance from those gauges which used a temperature-controlled oven operating at 40 °C rather than relying merely on thermal insulation for temperature stability.

When the temperature in a gauge is different from that in the main vacuum

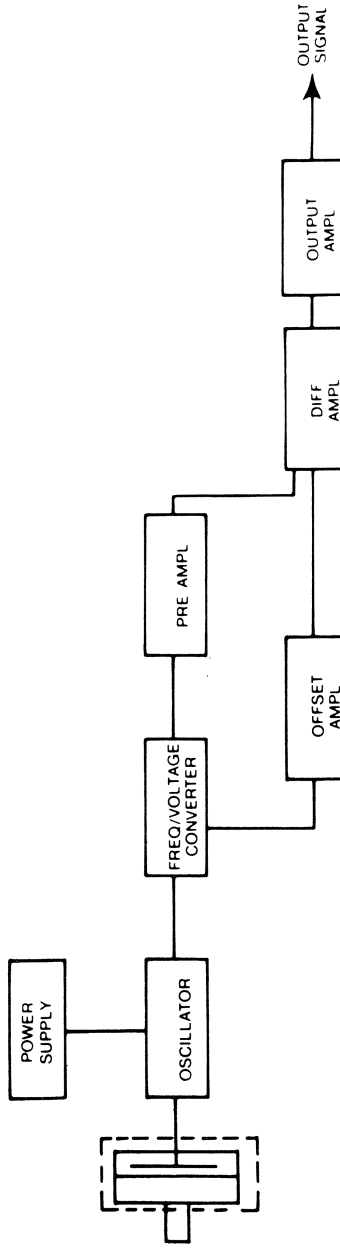


Figure 1.13 Block diagram of the electronic measuring system[53].

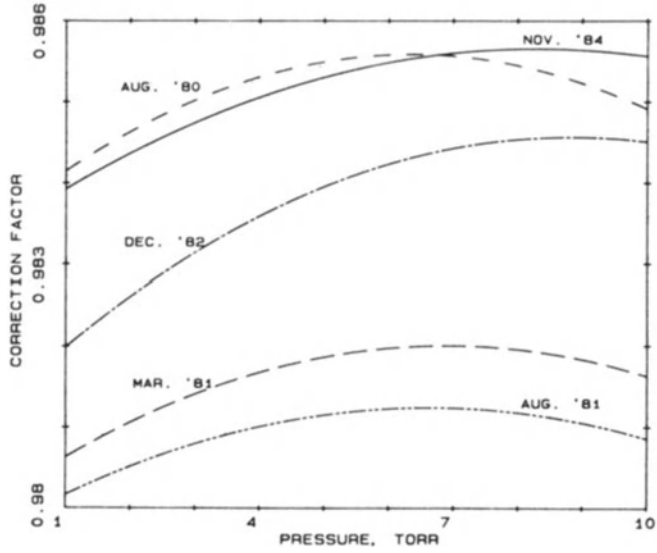


Figure 1.14 Least-squares curves fitted to the ratio of the true to the indicated pressure obtained for a 10-Torr type A gauge calibrated at NBS on the indicated dates. The magnitude of the changes and lack of a linear or monotonic drift with time are typical of results obtained for both type *A* and *B* gauges. The scatter of the data to which these curves were fitted is characterized by a standard deviation of 0.01% or better. This is a measure of the gauge instability over the 4–6 h required for a calibration[54].

system, care must be taken to correct for any pressure difference due to thermal transpiration. It is well known that a pressure difference can exist when any two vessels are connected by a tube or orifice. Simple kinetic theory predicts that if the molecular mean free path is much greater than the diameter of the connecting tube (orifice), then in equilibrium the pressures in the two volumes are related by

$$\frac{p_1}{p_2} = \sqrt{\frac{T_1}{T_2}} \tag{1.10}$$

where T_1 and T_2 are the respective temperatures.

At the other end of the scale, when the mean free path is very small the pressures are equal. At intermediate pressures the ratio p_1/p_2 will lie between unity and the value given in eqn (1.10). Poulter[57] and his group at the National Physical Laboratory, London, measured the magnitude of the pressure difference for the heated capacitor gauge through the intermediate pressure range and made a comparison with theory. The pressure difference increased from zero for all pressures above about 1 mbar, to the steady-state low pressure value at a pressure of about 10^{-2} mbar (4% with the oven at 40 °C). They found good agreement with the theoretical analysis proposed by Takaisi and Sensui[58] and, as expected, some significant differences for a range of gases

(measurements were made for helium, nitrogen, argon and sulphur hexafluoride). This report from the NPL is an invaluable aid to all experimental work where high accuracy is required with heated capacitor gauges in the 'intermediate' pressure range.

1.4 Viscous or friction-type gauges

The basic principles of this type of gauge were recognized as early as the mid-19th century. In 1865, Meyer[59,60], and Maxwell[61] independently in 1866, reported their first experiments aimed at the verification of Maxwell's predictions[62] for the viscosity of gases; subsequently important fundamental work in this field was reported by Sutherland[63] in 1897 and Langmuir in 1913[64]. For reviews of the early work and the considerable developments made subsequently, reference can be made to papers by Drawin[65] and by Fremerey[66].

All gauges of this type operate by virtue of the viscous or 'frictional' forces that gas molecules exert on a moving surface. This force is dependent upon molecular density and temperature over a wide range of pressure. In theory, gauges can be made to cover the whole range of pressures found in vacuum work, from near atmosphere to 10^{-8} or even 10^{-9} mbar. Modern developments have tended towards concentration on gauges useful for the range 10^{-2} to 10^{-7} mbar where there is a linear relation between viscous force and pressure and where the gauge's operating characteristics can be applied most usefully. Like the diaphragm manometer, (i) there is no interaction with the vacuum system, (ii) the gauge does not introduce any foreign gases or vapours to the system, (iii) calibration is absolute for all gases and vapours, and (iv) calibration should be stable, even over very long periods of time. In the main, development work has been directed towards the production of very stable, but of necessity expensive, instruments for use as standards of reference rather than as alternatives to ionization or other gauges for routine measurements.

Two versions of this instrument (oscillatory and rotary) have been developed to measure the damping forces. One is characterized by a swinging vane, pendulum or oscillatory diaphragm; the other by a rotating sphere, cylinder or paddle in the vacuum system. The usual method of operation is to observe the vane swinging freely (or sphere rotating) in the vacuum system and to observe the time taken for the amplitude (or speed of rotation) to fall by a specified amount, for example by 10%. This time interval is dependent upon the viscous damping force and hence pressure. The alternative method of operation is to apply a drive to the vane (sphere) and at the same time continuously monitor the energy required to maintain a steady amplitude (speed of rotation). Clearly in this mode the energy input is a measure of pressure. This second technique has the obvious advantage of giving a 'continuous reading' of pressure.

It is interesting to note that gauges (both oscillatory and rotary) have been

developed from the ideas put forward by Langmuir[64] in 1913. Coolidge[67] made a significant contribution to the vane technology by introducing a bifilar suspension which constrained the pendulum to swing in one plane, a virtual impossibility with a single fibre. Andrews[68] showed it was possible to adapt this gauge to measure vapour pressure (naphthalene) down to 3.4×10^{-5} mbar by using a strip of molybdenum 100 mm long, 2 mm wide and 0.1 mm thick as the pendulum. At about the same time, Bruche[69] also made a notable contribution by using a pendulum consisting of a quartz rectangular paddle 25 mm × 25 mm × 0.04 mm thick, suspended by two flat quartz fibres 75 mm long × 0.8 mm × 0.1 mm. He was able to confirm the $\sqrt{M}$ sensitivity law for hydrogen, air and the inert gases over the range 5×10^{-4} to 2×10^{-2} mbar.

Monitoring techniques for these gauges were in the main optical and, although useful in the specialized laboratories where they were designed and built, they were of no real importance for general vacuum use. Developments since the 1960s have been directed towards improving the methods of control and measurement by introducing modern electronic techniques. Hurd and Corrin[70], for example, made the vibrating fibre form one plate of a capacitor, thus giving a capacity varying in sympathy with the swinging of the pendulum. They showed that this gauge with electrical recording was useful down to 10^{-5} mbar. Anderson[71] determined the amplitude of swing accurately with an ingenious photoelectric technique. He measured the velocity of the tip of the pendulum at the middle of its swing. The bifilar gauge for measurement through the range 10^{-5} to 5×10^{-4} mbar used quartz fibres 70 mm long × 0.08 mm diameter, and it had a small 'paddle' 18 mm × 18 mm × 0.003 mm thick made from quartz sheet attached to its end. A hinge was put in the suspension just below its support by thinning each fibre to about a tenth of its original diameter. This reduced the restoring force and ensured that the suspension vibrated as a true pendulum rather than as a whip; unfortunately it reduced the vibration frequency from 30 to 4 oscillations per second in the typical case. Three gauges were constructed with different sizes of paddle to cover the pressure range 2×10^{-2} to 10^{-5} mbar (non-linearity became evident at a pressure of about 2.5×10^{-2} mbar).

The first practical vibrating diaphragm gauge was introduced by Klumb and Schwarz[72] in 1942. Here a light, taut suspension was mounted vertically and allowed to vibrate at its resonant frequency (between 30 and 300 Hz), the damping forces obviously being a function of gas pressure. By means of an ingenious photoelectric sensing device, an electromagnetic driving signal was created which maintained the oscillation at a constant amplitude. The whole system formed a closed-loop servo-system. The driving signal was equal to the damping losses and hence a function of pressure. Thus, by observing the driving force, a direct reading of pressure was obtained. A commercial version of this gauge (described by Becker[73]) operated over the pressure range 10^{-3} mbar to atmosphere. This gauge was probably most useful below

10^{-3} mbar, where calibration was linear and the $\sqrt{M}$ law was obeyed.

Pacey[74] described a different form of resonant gauge in which a crystal oscillator formed the vibrating system. He found that, when the crystal was kept in resonant oscillation, the driving force varied approximately linearly with pressure from 0.1 to 1000 mbar and that the $\sqrt{M}$ law was followed reasonably well. This idea has been developed by a group of workers in Japan[75–79], who have found the ordinary crystal tuning-fork oscillator (of the type used in a wrist watch) to be a most satisfactory sensor. It has the advantage of being extremely small, typically about 3.00 mm long, 0.33 mm wide and 0.15 mm thick, and to have a 'response' to pressure change measurable from 0.01 to 1000 mbar. The impedance of the crystal increased by about an order of magnitude for this pressure change, the precise relationship between impedance and pressure being a function of the size and shape of the tuning fork. Results published by Hirata *et al.*[76] (reproduced here in Figure 1.15) show the details of this relationship for tuning forks resonating at 26, 33 and 112 kHz. There was a good agreement between theory and experiment, particularly for the $\sqrt{M}$ dependency. Gauges have been built for practical vacuum use, a convenient method of operation being to drive the crystal at a constant voltage and to use the input current as a measure of impedance and hence pressure. Experiments indicate that with the gauge operating at low pressure ($< 10^{-6}$ mbar) transients of the order of 2×10^{-5} mbar can be detected[79]. The designers of the gauge[77,79], while optimistic about its future for pressure monitoring, are cautious in pointing

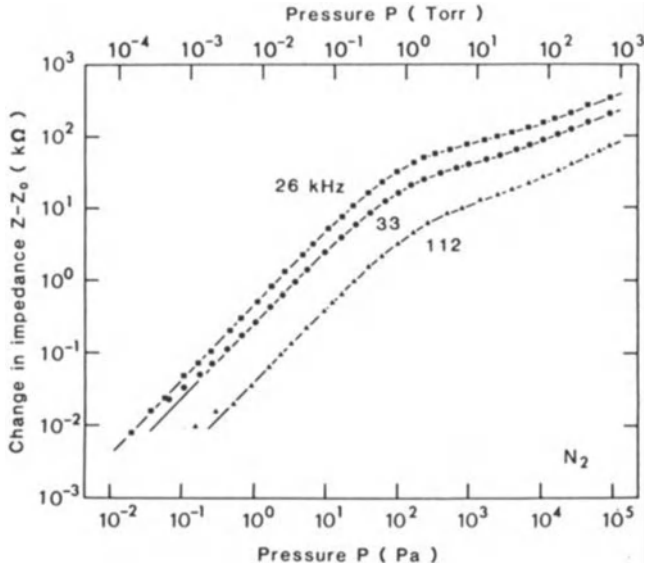


Figure 1.15 Pressure dependence of electric impedances of tuning forks of various lengths and resonant frequencies[77].

out the need for further development in the areas of accuracy and reliability. They note that, whilst the gauge is not affected by mechanical vibrations, it is disturbed by water adsorbed on the crystal after storage and by the presence of oil droplets.

The first practical rotating gauge (constructed by Dushman[80] in 1915) was somewhat different in principle to the modern version. In Dushman's gauge a horizontal disc, of some 30 mm diameter, rotated about its vertical axis at controlled speeds of up to 10 000 rev min^{-1}, with a second disc suspended coaxially a few millimetres above it. The molecular drag, due to the rotation of the lower disc, exerted a torque on the upper, which was balanced by the restoring force in the suspension. At low pressures the torque, and hence the twist in the suspension, was found to be proportional to $up_0\sqrt{MT}$, where u is the velocity of rotation. At higher pressures, where the gaseous mean free path is small compared with the physical dimension, a more complex relation between torque and pressure existed.

A later version of this gauge has been described by Briggs[81], in which the two discs were replaced by concentric cylinders, each with small vanes to improve the high-pressure calibration. The moving disc rotated at a fixed speed of 3600 rev min^{-1}. A pointer was attached directly to the constrained cylinder and indicated the gas pressure on a scale. The scale was linear from zero up to 20 mbar, and over this range the sensitivity followed the $\sqrt{M}$ relation as predicted by theory. Briggs estimated the error in reading to be about $\pm 10^{-3}$ mbar at the low end of the scale and ± 0.75 mbar at the top (20 mbar) end of the scale.

In 1962, Beams and his co-workers[82] used 'magnetic levitation' in their design of a rotary gauge, where the viscous drag forces were measured on a sphere of approximately 1 mm diameter rotating at high speed. The levitating magnetic field, controlled to keep a small soft-iron sphere in a fixed position, was provided by a solenoid. This sphere was driven at operating speeds up to 10^6 rev s^{-1}. By measuring the deceleration rate when the sphere was allowed to spin freely, indications of pressure were obtained throughout the range 10^{-4} to 10^{-9} or possibly 10^{-10} mbar. This spinning rotor gauge (SRG) has been found to have more potential for further development than the swinging vane. This is (as pointed out by Fremerey[66]) 'because the continuous rotation of a body is less affected by ambient shock than is oscillatory movement or static deflection, since surroundings do not supply angular velocity noise amplitudes comparable to the angular velocity of a rotor spinning at several thousand rpm.'

Figure 1.16 is a schematic diagram showing the essential features of the modern spinning rotor gauge developed by Fremerey and his co-workers[83-88]. The rotor (a steel ball-bearing 4.5 mm diameter) is suspended between the pole pieces of a permanent magnet. The two coils shown in Figure 1.16 are connected to a differential sensing circuit for determining the precise position of the rotor on the vertical axis and simultaneously for control

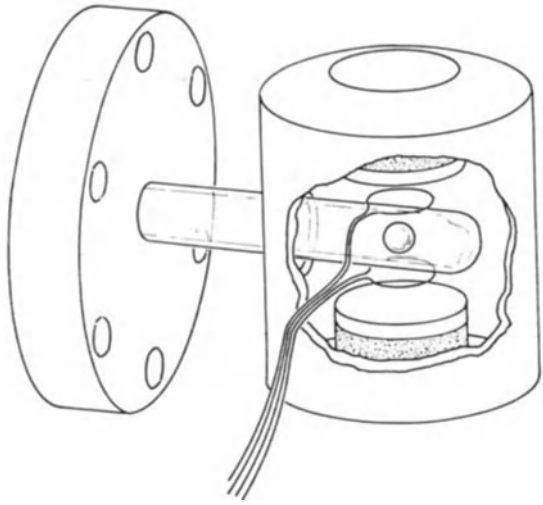


Figure 1.16 Gas friction detector head (symmetric configuration) with 4.5 mm ball-bearing sphere in 8-mm OD glass tube sealed to a 70 mm OD flange[87].

of the vertical restoring force via the magnetic field gradient. Equilibrium provided solely by the permanent magnet must inherently be unstable because any vertical deflection of the ball from its equilibrium position changes the magnetic force in the direction to increase this deflection. The compensating coils control the field so that the magnetic field always acts to restore equilibrium. This changes the suspension from a 'negative' to a 'positive' spring stiffness. The stiffness of the spring can be controlled by varying the gain of the feedback circuit. This is the so-called active permanent magnetic suspension (APMS) which gives stable equilibrium at virtually zero power. The pole pieces of the permanent magnets (shown in Figure 1.16) are made from highly permeable nickel–iron alloy which provides optimum symmetry at the rotor and, as they are good conductors, serve for eddy current damping of the lateral rotor oscillations.

Figure 1.17 is a cutaway diagram of a practical version of this SRG designed for general vacuum use[88]. In addition to the two coils which ensure the stable equilibrium of the ball, Figure 1.17 indicates the position of four coils L, two on each side of the vacuum tube V which are introduced to damp out any lateral oscillations which might occur because of mechanical vibration. When the rotor moves towards (or away from) one of these coils then, due to the strong vertical polarization of the ball, an inductive voltage appears at that coil. This voltage is proportional to the lateral velocity of the rotor. The signal, after amplification, is connected to the control coil of corresponding shape on the opposite side of the rotor. The magnetic movement generated by this coil interacts with the rotor magnet movement in such a way that the lateral

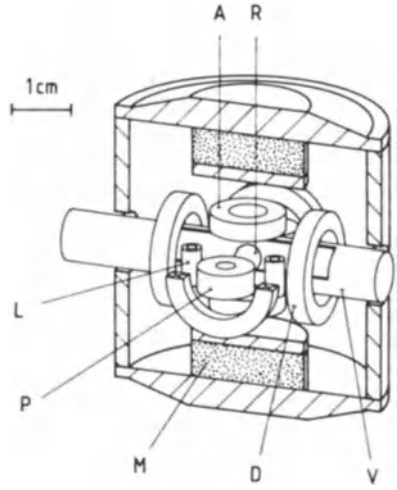


Figure 1.17 SRG head: *R*, rotor; *V*, vacuum enclosure (partially cut away); *M*, one of two permanent magnets; *A*, one of two coils for pickup and control of axial rotor position; *L*, one of four coils of lateral damping system; *D*, one of four drive coils; *P*, one of two pickup coils[88].

movement along the axis of the two coils is damped out. The second pair of coils works similarly to ensure that the rotor is stabilized in all lateral directions. Four coils *D*, connected to the output of a two phase ac generator, provide a rotating field for spinning the rotor about its vertical axis, thus forming a conventional induction motor. From rest, a speed of rotation of 400 revs s^{-1} can be achieved in less than one minute. The rate of deceleration when the rotor power is switched off is determined by observing the zero crossings of an inductive pick-up signal generated by the rotating component of the rotor magnetization. The two pick-up coils *P* are provided for this purpose. Fremerey[66] notes that the electronic lateral damping technique used here, and first described by Nixon and Kenney[89] in 1964, is much better than the techniques used by Beams[82,90] and other workers[91]. In the early work, lateral damping was achieved by the viscous damping of a coaxially mounted piece of iron magnetically coupled to the rotor. Thus the modern gauges, which although having the same fundamental limitations for pressure measurement, are much less sensitive to vibrational instabilities and, therefore, more suited to practical laboratory use.

The calculation of the force exerted by the gas molecules upon the freely rotating ball, and hence the rate of deceleration, is relatively straightforward provided it can be assumed that the tangential momentum exchange at the rotor surface is perfect. This assumption holds if the molecules on arrival at the surface lose all knowledge of their direction of approach and subsequently leave the surface in a completely random manner. When this is the case, the density of the molecular flux re-emitted is proportional to the cosine of the

polar angle, and these molecules have no net transfer of momentum. Thus only the incident molecules need be considered. The tangential momentum transferred by these molecules is, on average, $mr^2\omega$ per molecule, where m is the mass, ω the rate of rotation and r the locus of radial distance of the impingement from the spin axis. By integrating over the whole surface, the total momentum transfer can be calculated. Taking the standard expression for the total angular momentum stored in the sphere as $(8\pi/15)da^5\omega$ allows the deceleration to be expressed as:

$$-\frac{\dot{\omega}}{\omega} = \frac{10}{\pi}\frac{1}{ad}\frac{p}{\bar{c}} \tag{1.11}$$

where a is the radius of the ball, d the material density and $\bar{c}$ the mean molecular velocity, which can be expressed as

$$\bar{c} = 2\left(\frac{2RM}{\pi M}\right)^{1/2} \tag{1.12}$$

with the usual notation.

At any given pressure, the deceleration rate is directly proportional to the square root of the molecular mass M, and inversely proportional to the square root of gas temperature. This relationship between pressure and deceleration rate (expressed by eqn 1.11) is well established and was verified, for example, by Beams and his co-workers in 1946[90].

Although an instantaneous and continuous reading of $\dot{\omega}/\omega$ cannot be obtained for this system, the application of modern electronic techniques for measuring short time intervals and for signal averaging ensures that an acceptably short time delay can be obtained. The standard 'stop-watch' equation is:

$$\frac{\dot{\omega}}{\omega} = \frac{t_n - t_{n+1}}{t_n t_{n+1}} \tag{1.13}$$

where t_n is the time for a specified number of n revolutions, and t_{n+1} is the time for the next n revolutions. (See references 92 and 88, for example, for the details of the measuring techniques.)

The validity of eqn (1.11) has been verified in numerous experimental programmes using this stop-watch technique. As this equation is derived from the classical laws of gas kinetics, the only uncertainty is the assumption of 'complete momentum transfer'. To take account of the possibility of incomplete transfer of momentum, an arbitrary coefficient δ can be introduced, so that eqn (1.11) becomes

$$\frac{\dot{\omega}}{\omega} = \delta\frac{10}{\pi}\frac{1}{ad}\frac{p}{\bar{c}} \tag{1.14}$$

Most experiments made with smooth polished spheres give δ as varying

between the limits of about 0.995 and 1.05 for gases ranging from helium to xenon. The values in Table 1.3 (from the report by Comsa *et al.*[93]) are probably the most extensive for any one series of measurements. Commercial ball-bearings (SKF type 100 Cr class III) without special selection were used for these experiments, giving very stable results. A major accident to the vacuum system, which caused the levitated ball to bounce several times against the walls of the vacuum chamber, fortuitously gave a good indication of the stability of the system. Even though the still-shiny ball showed some rough areas the results changed by only 1.4%. Systematic measurements by Messer[94] at the Physikalisch-Technische Bundesanstalt, Berlin, have confirmed the long-term stability of measurements made with one particular rotor. His results are reproduced in Table 1.4.

McCulloh *et al.*[95] at the National Bureau of Standards have made extensive measurements of the zero stability of this gauge. They proposed a refinement to the deceleration equation by the addition of three terms to give

$$\frac{\dot{\omega}}{\omega} = \frac{p}{\bar{c}} + 2\alpha_0 \dot{T} + RD + RN \tag{1.15}$$

where $p/\bar{c}$ is the pressure-dependent term of eqn (1.15). The second term accounts for the effect of the rate of change of temperature T of the rotor, α_0 being the coefficient of linear expansion of the material. *RD* and *RN* represent respectively the residual drag and the random noise in the digital measuring circuitry. For convenience, each of these terms can be expressed as a 'pressure equivalent'. In normal operation, the second of these three additional terms (*RD*) is the most important. This drag has a number of sources[84]; the dominating ones, however, are the effects of eddy currents induced in the spinning ball by asymmetries in the suspension field and eddy currents induced in surrounding metallic components. The workers at the NBS observed residual drags which varied from 1×10^{-6} to 3×10^{-6} mbar equivalent nitrogen pressure for smooth ball bearings. Noise levels very much depended upon mechanical vibrations, particularly those near to a resonant frequency of the suspension stabilization circuit. Under optimum conditions, stability to within $\pm 1 \times 10^{-8}$ mbar equivalent nitrogen pressure was possible in their 'quiet' basement laboratory. Even the smallest disturbance (such as lightly touching the apparatus) could cause large perturbations, increasing the lower limit of measurement by orders of magnitude. Discontinuities of the order 5×10^{-6} mbar in 'zero' reading were observed on dropping and resuspending the gauge. The effects of temperature instabilities were also measured carefully at the NBS. The effect of stopping the ball and immediately spinning it back to 400 cps was found to be most significant, as this increased rotor temperature by about 4 K. This caused a steady drift in the 'residual' or 'zero' reading, in one typical experiment from 0.3 to 1.1×10^{-6} mbar over a period of 5 h. As the workers point out, it is the rate of change of temperature,

Table 1.3 δ values from different 4.5 mm balls in different gases[93].

Gas	Ball No. 100 (mbar) 6.7×10^{-5}	6.7×10^{-4}	6.7×10^{-3}	Ball No. 101 (mbar) 6.7×10^{-5}	6.7×10^{-4}	6.7×10^{-3}	Ball No. 102 (mbar) 6.7×10^{-5}	6.7×10^{-4}	6.7×10^{-3}
He	1.015	1.018	1.016	1.031	1.029	1.026	1.015	1.012	1.010
Ne	0.992	0.990	0.998	1.001	1.001	0.998	0.983	0.983	0.980
Ar	1.013	1.013	1.008	1.020	1.018	1.011	1.007	1.006	0.996
Kr	1.025	1.021	1.012	1.030	1.027	1.017	1.021	1.019	1.009
Xe	1.025	1.025	1.012	1.035	1.030	1.016	1.027	1.023	1.009
CH_4	1.021	1.018	1.010	1.029	1.026	1.016	1.022	1.015	1.006
N_2	1.018	1.013	1.008	1.023	1.019	1.012	1.010	1.008	1.001
H_2			1.024						1.012
CO			—						1.006
O_2			1.019						1.000
CO_2			1.019						1.007

Table 1.4 Tangential momentum coefficient δ of the spinning rotor gauge at 7×10^{-4} mbar argon[94]. (Messer comments: 'The range of variation of δ is obviously smaller than the uncertainty of calibrations (1%); this is understandable by the fact that the total uncertainty is mainly covered by systematic uncertainties which do not contribute to errors in repeated recalibrations with the same instrument').

Date	Mean value of 10 measurements at constant pressure	Relative standard deviation $\times 10^4$
25. 7.1977	1.003	4
26. 7.1977	1.004	3
17. 7.1978	1.006	5
18. 7.1978	1.005	2
19. 7.1978	1.001	5
12.11.1979	1.006	1
6.12.1979	1.005	1
7. 3.1980	1.006	3
	Mean: 1.005	

rather than temperature, that is significant (1 m K min^{-1} change being equivalent to a nitrogen pressure change of 10^{-8} mbar).

An alternative magnetic levitation technique was developed by Evrard and Boutry[96] (1969). Instead of a ferromagnetic sphere, they used a diamagnetic circular graphite disc suspended with its axis horizontal, and rotating about a diameter; this is a rotating 'vane' or 'paddle' gauge. When enclosed in a carefully designed static vane system, this gauge can operate as either a Knudsen thermomolecular[97] or a viscosity gauge. Although some interesting work has been carried out, the gauge has not yet been developed extensively. A theoretical comparison of the two types of spinning rotor (sphere and vane) is interesting. For the vane, it is again relatively easy to calculate the damping forces from the basic laws of kinetic theory, although now the calculation is slightly more difficult because the mass flow is different at the front and reverse sides of the vane. (See for example the review by Steckelmacher[98]). However, because motion is perpendicular rather than parallel, account must be taken of the efficiency of transfer of energy as well as momentum at the moving surface. In other words, the residual force is a function of both δ and α, the coefficients of momentum and energy transfer respectively. The values quoted in Table 1.5 (showing the limits to the correction factor which must be applied for the extremes of zero and perfect accommodation) suggest a fundamental advantage for the disc, because the maximum variation is only 12%, whereas for the sphere the sensitivity falls to zero as δ approaches zero. Experiments in many laboratories, from those as early as Knudsen's work in 1934[99] (see also references 100 and 101), up to and including the work with the SRG designed by Fremerey (for example the work of Messer[94]), have shown δ to be within 5%

Table 1.5 Limiting values for the correction factor for the spinning sphere and spinning disc as the coefficients of momentum transfer (δ) and energy transfer (α) vary between 1.0 and 0[96].

δ	α	Correction factor Sphere	Disc
1.0	1.0	1.00	$0.50 + \pi/8$
1.0	0	1.00	$0.50 + 9\pi/64$
0	0	0	1.00

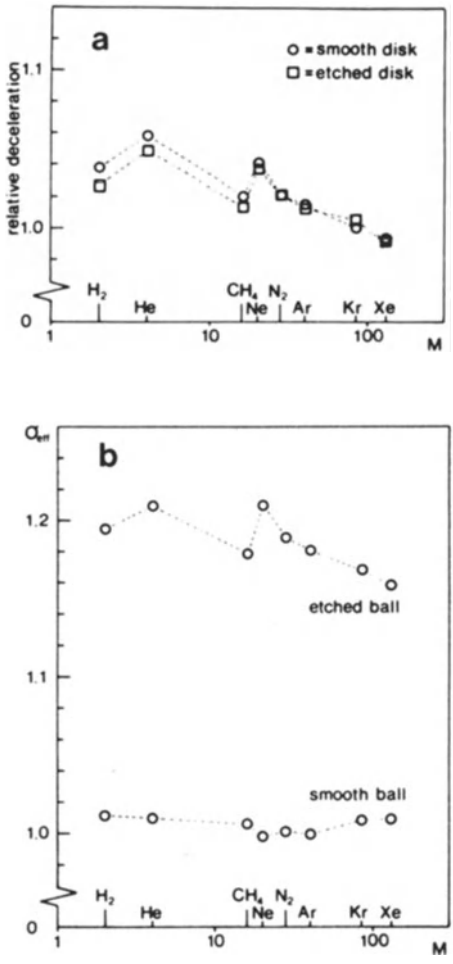


Figure 1.18 Relative deceleration of freely rotating vanes (*a*) and balls (*b*) in different gases[103].

of unity at technological smooth (i.e. polished) metal surfaces. Thus this disadvantage does not appear to be significant.

Careful experimental work by Comsa's group at Julich[102,103] has found that if the steel spheres are roughened (by chemical etch in their case), the frictional drag is increased by as much as 20%, or even 30%. In a specially modified SRG, a comparison was made between spheres and vanes as the rotating elements. The vanes were cut from ball-bearings of the same sort and diameter as were used for the test spheres. (The magnet suspension system was similar, but much larger than the normal, because the vanes rotating about a diameter involve an appreciable disturbance of the magnetic field.) The increase in the apparent or effective value of δ is shown by the results in Figure 1.18. The group at Julich points out the 'striking coincidence of the characteristics obtained for the etched ball and from the vanes – both smooth and etched, whereas the characteristics for the smooth ball are quite different. This suggests that the mechanism of molecular drag on a rough ball is similar to that on the vanes, i.e. most of the molecular surface interactions occur on non-tangential surface elements.' Fremerey[104] analysed the problem of the molecular collisions with a rough surface, calculating the increase in the 'effective' value of δ to be from 1.0 to $4/\pi$ as the surface of the sphere changed from ideally smooth to arbitrarily rough (from all surface elements parallel to perpendicular to the average surface). This is in good agreement with experimental observations reported above.

Details of a typical commercial instrument suitable for use as a reference gauge in vacuum laboratories have been given by Reich[92]. He describes a unit with a linear output characteristic up to pressures of about 10^{-2} mbar. (Sensitivity falls such that the indicated pressure is only 90% of the true pressure at 0.05, 0.02, 0.012 and 0.02 mbar for helium, nitrogen, argon and xenon respectively.) He notes that, to obtain optimum performance, the 'measuring time' t_n in eqn (1.13) should be set up at 40 seconds at 10^{-7} mbar and one second at 10^{-4} mbar. In his paper Reich also points out that, although all viscosity gauges have the disadvantage of the $\bar{C}$ dependency (mean molecular velocity), this means the rate of deceleration is proportional to the mass of gas incident on unit area of surface in unit time independent of the nature of the gas.

McCulloh[105] describes some of the precautions necessary when using the SRG for standards work. In particular, he records the increase in sensitivity of 15% which occurred in his work when the ball-bearing became contaminated with rotary pump oil, a change which was recovered by baking. His overall experience emphasizes the excellent performance that can be obtained with careful work. For example, when he operated one particular ball over a 20-month period he found a change in the 'effective' value of δ of only 4% from the initial value of 0.997 (measured at the PTB in Berlin), to 1.037. The ball suffered some 'adventures' during this time, including 'crashing' and oil contamination (requiring baking). It is interesting to note that McCulloh's initial

calibration[105] at the NBS differed by only 0.7% from the previous calibration at the PTB.

All the operating evidence gathered from the various laboratories indicates the SRG to be an excellent instrument for use as a secondary standard, being particularly suitable for transporting between different laboratories. With a good understanding of its operation and with careful use it is capable of giving a reproducible reading, certainly to better than ± 5%, probably approaching ± 1%. Fremerey[104] goes so far as to suggest that, because the value of δ can be controlled, the SRG can be thought of as an absolute rather than a transfer standard.

References

1. Biondi, M.A. (1953) *Rev. Sci. Instrum.* **24**, 989.
2. Maslach, G.J. (1952) *Rev. Sci. Instrum.* **23**, 367.
3. Maguire, F.S. and Thomas, A.G. (1961) *J. Sci. Instrum.* **38**, 261.
4. Hickman, K.C.D. (1934) *Rev. Sci. Instrum.* **5**, 161.
5. Hickman, K.C.D. and Weyerts, W.J. (1930) *J. Amer. Chem. Soc.* **52**, 4714.
6. Rayleigh, Lord (1901) *Phil. Trans. A* **196**, 205.
7. Newbury, K. and Utterback, C.L. (1932) *Rev. Sci. Instrum.* **3**, 593.
8. Shrader, J.E. and Ryder, H.M. (1919) *Phys. Rev.* **13**, 321.
9. Carver, E.K. (1923) *J. Amer. Chem. Soc.* **45**, 59.
10. Johnson, M.C. and Harrison, G.O. (1929) *J. Sci. Instrum.* **6**, 305.
11. Farquharson, J. and Kermicle, H.A. (1957) *Rev. Sci. Instrum.* **28**, 324.
12. Zigman, P. (1959) *Rev. Sci. Instrum.* **30**, 1060.
13. Elliott, K.W.T., Wilson, D.C., Mason, F.C.P. and Bigg, P.H. (1960) *J. Sci. Instrum.* **37**, 162.
14. Hart, H.R. (1961) *J. Sci. Instrum.* **38**, 300.
15. Hirsch, E.H. (1959) *J. Sci. Instrum.* **36**, 477.
16. Kemp, J.F. (1959) *J. Sci. Instrum.* **36**, 77.
17. Pannell, J.R. (1924) *Fluid Velocity and Pressure*, Edward Arnold, London, 91.
18. Dodge, R.A. and Thompson, M.J. (1937) *Fluid Mechanics*, McGraw-Hill, New York.
19. Sederholm, P. and Benedicks, C. (1940) *Arkiv f. Mat. Astra. och Fys.* **27**, A8.
20. McLeod, H. (1874) *Phil. Mag.* **48**, 110.
21. Jansen, C.G.J. and Venema, A. (1959) *Vacuum* **9**, 219.
22. Bixler, H.J., Michaels, A.S. and Parker, R.B. (1960) *Rev. Sci. Instrum.* **31**, 1155.
23. Armbruster, M.H. (1946) *J. Amer. Chem. Soc.* **68**, 1342.
24. Flosdorf, E.W. (1945) *Industr. engng. Chem. Anal. Ed.* **17**, 198.
25. Hayward, A.T.J. (1962) *J. Sci. Instrum.* **39**, 367.
26. Porter, A.W. (1933) *Trans. Faraday Soc.* **29**, 702.
27. Rosenberg, P. (1939) *Rev. Sci. Instrum.* **10**, 131.
28. Rosenberg, P. (1938) *Rev. Sci. Instrum.* **9**, 258.
29. Klemperer, O. (1944) *J. Sci. Instrum.* **21**, 88.
30. Barr, W.E. and Anhorn, V.J. (1946) *Instruments* **19**, 666.
31. Podgurski, H.H. and Davis, F.N. (1960) *Vacuum* **10**, 377.
32. Keevil, N.B., Errington, R.F. and Newman, L.T. (1941) *Rev. Sci. Instrum.* **12**, 609.
33. Elliott, K.W.T., Woodman, Daphne, M. and Dadson, R.S. (1967) *Vacuum* **17**, 439.
34. Sharma, J.K.N., Dwivedi, H.K. and Sharma, D.R. (1980) *J. Vac. Sci. Technol.* **17**, 820.
35. Clark, R.J. (1929) *J. Sci. Instrum.* **5**, 126.
36. Moser, H. and Poltz, H. (1957) *Z. Instrumentkunde* **65**, 43.
37. Miller, J.R. (1972) *J. Vac. Sci. Technol.* **9**, 201.
38. Cespiro, Z. (1973) *Vacuum* **23**, 277.
39. Akiyama, Y., Hashimoto, H. and Ishii, H. (1967) *Vacuum* **17**, 393.
40. Groszkowski, J. (1949) *Le Vide* **4**, 668.
41. Flosdorf, E.W. (1938) *Industr. engng. Chem. Anal. Ed.* **10**, 534.

42. Axelbank, M. (1950) *Rev. Sci. Instrum.* **21**, 511.
43. Gaede, W. (1915) *Ann. Phys. (Leipzig)* **46**, 357.
44. Ishii, H. and Nakayama, K. (1961) *Vac. Sym. Trans. Amer. Vac. Soc.* **1**, 519 (Pergamon New York).
45. Meinke, C. and Reich, G. (1962) *Vakuumtechnik* **11**, 86.
46. Vries, A.E. de and Rol, P.K. (1965) *Vacuum* **15**, 135.
47. Berman, A. (1974) *Vacuum* **24**, 241.
48. Tunniclife, R.J. and Rees, J.A. (1967) *Vacuum* **17**, 457.
49. East, H.G. and Kuhn, H. (1946) *J. Sci. Instrum.* **23**, 185.
50. Crompton, R.W. and Elford, M.T. (1957) *J. Sci. Instrum.* **34**, 405.
51. Olsen, A.R. and Hirst, L.L. (1929) *J. Amer. Chem. Soc.* **51**, 2378.
52. Pressey, D.C. (1953) *J. Sci. Instrum.* **30**, 20.
53. Sullivan, J.J. (1985) *J. Vac. Sci. Technol.* **A3**, 1721.
54. Hyland, R.W. and Tilford, C.R. (1985) *J. Vac. Sci. Technol.* **A3**, 1731.
55. Poulter, K.F. (1981) *Le Vide* **36**, 521.
56. Reich, G. (1981) *Proc. 9th Int. Vac. Congr., Madrid*, ed. J.L. de Segovia, Madrid.
57. Poulter, K.F., Rodgers, M.J., Nash, P.J., Thompson, T.J. and Perkin, M.P. (1983) *Vacuum* **33**, 311.
58. Takaisi, T. and Sensui, Y. (1963) *Trans. Faraday Soc.* **59**, 2503.
59. Meyer, O.E. (1865) *Pogg. Ann.* **125**, 177.
60. Meyer, O.E. (1873) *Pogg. Ann.* **148**, 203.
61. Maxwell, J.C. (1866) *Phil. Trans. Roy. Soc.* **157**, 249.
62. Maxwell, J.C. (1860) *Phil. Mag.* **19**, 31.
63. Sutherland, W. (1896) *Phil. Mag. Ser. 3.* **42**, 373.
64. Langmuir, I. (1913) *J. Amer. Chem. Soc.* **35**, 105.
65. Drawin, H.W. (1965) *Vacuum* **15**, 99.
66. Fremerey, J.K. (1982) *Vacuum* **32**, 685.
67. Coolidge, A.S. (1923) *J. Amer. Chem. Soc.* **45**, 1937.
68. Andrews, M.R. (1926) *J. Phys. Chem.* **30**, 1947.
69. Bruche, E. (1923) *Phys. Z.* **26**, 717; (1926) *Ann. Phys. (Leipzig)* **79**, 695.
70. Hurd, D.T. and Corrin, M.L. (1954) *Rev. Sci. Instrum.* **25**, 1126; Neher, H.V. (1962) *Rev. Sci. Instrum.* **33**, 808.
71. Anderson, J.R. (1958) *Rev. Sci. Instrum.* **29**, 1073.
72. See, for example, Schwarz, H.J. (1956) *Vac. Sym. Trans. Comm. Vac. Tech.* (Pergamon, New York) **37**; *Vac. Symp. Trans. Amer. Vac. Soc.* (1961) **1**, 467 (Pergamon, New York).
73. Becker, W. (1961) *Vacuum* **11**, 195; (1960) *Vakuumtechnik* **9**, 48.
74. Pacey, D.J. (1959) *Vacuum* **9**, 262.
75. Kokubun, K., Hirata, M., Murakami, H., Toda, Y. and Ono, M. (1984) *Vacuum* **34**, 731.
76. Kirata, K., Kokubun, M., Ono, M. and Nakayama, K. (1985) *J. Vac. Sci. Technol.* **A3**, 1742.
77. Ono, M., Hirata, M., Kokubun, K., Murakami, H., Tamura, F., Hojo, H., Kawashima, H. and Kyogoku, H. (1985) *J. Vac. Sci. Technol.* **A3**, 1746.
78. Ono, M., Hirata, M., Kokubun, K., Mukakami, H., Hojo, H., Kawashima, H. and Kyogoku, H. (1986) *J. Vac. Sci. Technol.* **A4**, 1728.
79. Hirata, M., Ono, M., Kokubun, K., Abe, M., Maruno, N., Shimizu, K. and Ogawa, T. (1987) *J. Vac. Sci. Technol.* **A5**, 2393.
80. Dushman, S. (1915) *Phys. Rev.* **5**, 212.
81. Briggs, W.E. (1954) *Vac. Symp. Trans. Comm. Vac. Tech.* **3**, Pergamon Press, New York.
82. Beams, J.W., Spitzer, D.M.Jr. and Wade, J.P.Jr. (1962) *Rev. Sci. Instrum.* **33**, 151.
83. Fremerey, J.K. (1971) *Rev. Sci. Instrum.* **42**, 753.
84. Fremerey, J.K. (1972) *Rev. Sci. Instrum.* **43**, 1413.
85. Fremerey, J.K. (1972) *J. Vac. Sci. Technol* **9**, 108.
86. Fremerey, J.K. (1973) *Rev. Sci. Instrum.* **44**, 1396.
87. Fremerey, J.K. and Boden, K. (1978) *J. Phys. E. Sci. Instrum.* **11**, 106.
88. Fremerey, J.K. (1985) *J. Vac. Sci. Technol.* **A3**, 1715.
89. Nixon, J.D. and Kenney, D.J. (1964) *Rev. Sci. Instrum.* **35**, 1721.
90. Beams, J.W., Young, J.L. and Moore, J.W. (1946) *J. Appl. Phys.* **17**, 886.
91. MacHattie, L.E. (1941) *Rev. Sci. Instrum.* **12**, 429.
92. Reich, G. (1982) *J. Vac. Sci. Technol.* **20**, 1148.

93. Comsa, G., Fremerey, J.K., Lindenau, B., Messer, G. and Rohl, P. (1980) *J. Vac. Sci. Technol.* **17**, 642.
94. Messer, G. (1980) *Proc. 8th Int. Vacuum Congr. Cannes*, **II**, 191.
95. McCulloh, K.E., Wood, S.D. and Tilford, C.R. (1985) *J. Vac. Sci. Technol.* **A3**, 1738.
96. Evrard, R. and Boutry, G.A. (1952) *J. Vac. Sci. Technol.* **6**, 279.
97. Leck, J.H. (1964) *Pressure Measurement in Vacuum Systems*, Chapman & Hall, London, Chapter 4.
98. Steckelmacher, W. (1973) *Vacuum* **23**, 165.
99. Knudsen, M. (1934) *The Kinetic Theory of Gases*, Methuen, New York, 29.
100. Hurlbut, F.C. (1957) *J. Appl. Phys.* **28**, 844.
101. Auerbach, D., Becker, C., Cowin, J. and Wharton, L. (1977) *Proc. 2nd Int. Symp. Molecular Beams, Noordwijkerhout, Netherlands*, 192.
102. Comsa, G., Fremerey, J.K. and Lindenau, B. (1977) *Proc. 7th Int. Vacuum Congr. Vienna*, I, 157.
103. Comsa, G., Fremerey, J.K. and Lindenau, B. (1980) *Proc. 8th Int. Vacuum Congr. Cannes*, II, 218.
104. Fremerey, J.K., *Proc. 4th Int. Conf. on Solid Surfaces and 3rd Eur. Conf. on Surface Science, Cannes*, II, 869.
105. McCulloh, K.E. (1983) *J. Vac. Sci. Technol.* **A1**, 168.
106. Dushman, S. (1949) *Scientific Foundations of Vacuum Technique*, John Wiley, New York, 176.

2 Thermal conductivity gauges

2.1 Basic principles

Because it is a function of the gas pressure, the rate of heat transfer through a gas can, after a suitable calibration, be used to give an indication of the pressure. This principle was first put to use in the field of high vacuum by Pirani in 1906[1], when it became important to measure with some accuracy the pressure inside electric lamp bulbs. Pirani showed that the heat transfer could best be measured by observing the rate of loss of heat from a thin wire suspended freely in the vacuum chamber. The heat loss from the wire is the sum of the following four components: (i) conduction; (ii) convection; (iii) radiation and (iv) conduction along the wire to the end supports. The first and, in some special gauges, the second of these components are useful in pressure measurement, as both are dependent upon gas pressure.

The relationship between pressure and the heat loss from the wire by conduction through the gas can be evaluated by considering a wire of diameter d cm stretched along the axis of a tube of diameter D cm (Figure 2.1). If the gas molecules have, on the average, temperatures of T'_g on arrival and T'_w on departure from the wire, the rate of energy transfer E from wire to gas, and subsequently to the outer cylinder, is given by

$$E = kn(T'_w - T'_g) \text{ W cm}^{-2} \tag{2.1}$$

The number of molecules n striking unit area of the wire each second is directly

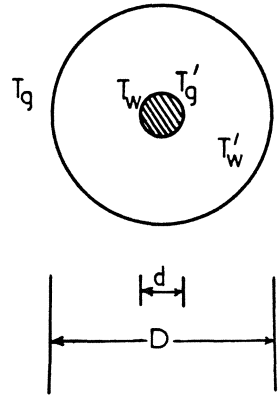


Figure 2.1 Simplest form of heat conductivity gauge element consisting of a thin wire of diameter d mounted along the axis of a cylinder of diameter D.

proportional to the pressure at any given gas temperature; thus eqn (2.1) can be rewritten

$$E = K(T'_w - T'_g)p \tag{2.2}$$

i.e. the rate of heat transfer from wire to wall is proportional to the pressure, provided $T'_w - T'_g$ is constant.

The temperature T'_w depends upon (i) the temperature of the wire surface T_w; (ii) the type of surface; (iii) the temperature of the incident molecules, and (iv) the type of molecule. Knudsen[2] in some early experiments showed the increase in molecular temperatures to be directly proportional to the temperature difference between the surface and the incident molecules for a given gas at a particular surface. He defined a constant α (the accommodation coefficient) as

$$\alpha = \frac{T_r - T_i}{T_s - T_i}$$

where T_s, T_i and T_r are the temperatures of the surface, the incident and the reflected molecules respectively.

Introducing this constant into eqn (2.2) gives

$$E = \alpha K p(T_w - T'_g) \text{ W cm}^{-2} \tag{2.3}$$

K is a basic constant which can be eliminated from eqn (2.3) as follows:

$Kp =$ (heat removed per molecule per unit temperature rise) $\times n = 2k_0 \times \frac{1}{4}N\bar{c}$ where k_0 is Boltzmann's constant, $n = \frac{1}{4}N\bar{c}$ (see Knudsen, p. 3 of ref. 4), $\bar{c}$ is the average molecular speed and N the number of molecules per cm^3.

But $k_0 N = p/T$ (see Roberts, Chapter 3, ref. 3) and $\bar{c} = \sqrt{(8/\pi)}\sqrt{(RT/M)}$ (see Knudsen, p. 4, ref. 4), R, T and M being the universal gas constant, the gas temperature and molecular mass respectively.

Thus

$$Kp = p\sqrt{2R/\pi T_g M} \tag{2.4}$$

This assumes that all the molecular energy is translational. This is true only for the monatomic gases; the others possess rotational and vibrational energy which must be taken into account.[5] Because the ratio of (total energy)/(translational energy) $= (\gamma + 1)/4(\gamma - 1)$ where γ is ratio of the specific heats, eqn (2.4) can be rewritten to cover the general case

$$Kp = p(2R/\pi T_g M)^{1/2}(\gamma + 1)/4(\gamma - 1) \tag{2.5}$$

and the rate of energy transfer given by

$$E = \frac{\gamma + 1}{4(\gamma - 1)} \cdot \left(\frac{2R}{\pi T_g M}\right)^{1/2} \alpha(T_w - T'_g)p \text{ erg. s}^{-1}\text{ cm}^{-2} \tag{2.6}$$

On evaluating constants, eqn (2.6) can be written

$$E = 0.011\frac{\gamma+1}{\gamma-1}\left(\frac{1}{M}\cdot\frac{273}{T_g}\right)^{1/2}\alpha(T_w - T'_g)p \text{ W cm}^{-2} \tag{2.7}$$

(p is expressed in mbar in eqn 2.7, but in dynes cm^{-2} in eqns 2.2–2.6.)

T'_g (the only unknown in eqn 2.7) depends upon the gas mean free path λ and therefore upon the pressure. In estimating T'_g there are three distinct conditions to consider: (i) when $\lambda > D$; (ii) when $D > \lambda \gg d$; and (iii) when λ is of the same order as or less than d. For (i) when $\lambda > D$ each molecule makes many collisions with the outer walls for every one at the wire (provided $D \gg d$) and so is always in temperature equilibrium with the wall on striking the wire, making $T'_g = T_g$. For (ii) when $D > \lambda \gg d$ the molecules, although they may not reach the wall, have many collisions with other molecules between each collision with the wire. T'_g is therefore equal to the mean gas temperature, i.e. $\simeq T_g$. For (iii) when λ is of the same order as or less than d, most molecules return to the wire before reaching equilibrium with the main body of the gas. T'_g is greater than T_g and increases with increasing pressure. Thus the heat conduction is proportional to pressure, and eqn (2.7) can only be applied for cases (i) and (ii), i.e. when $\lambda \gg d$.

To calculate the heat transfer in the high-pressure region where λ is not very much greater than d, Smoluchowski[6] assumed a temperature discontinuity at the surfaces. He proposed the following equation:

$$E = \frac{K_0}{d(\log D/d + g_1/d + g_2/D)} \tag{2.8}$$

where K_0 is the thermal conductivity of the gas at high pressure, and g_1 and g_2 are defined as 'temperature-jump distances' at the wire and wall respectively. The physical meaning of 'temperature-jump distance' can be seen from Figure 2.2, which shows the temperature gradient in the gas near the wall.

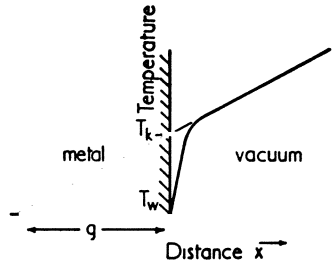


Figure 2.2 Temperature gradient in a gas near to a conducting surface. g is defined as the 'temperature-jump distance'.

g has been calculated to be equal to

$$g = \frac{1}{p}\frac{2-\alpha}{\alpha}(2\pi RT)^{1/2}\frac{\text{constant}}{(\gamma+1)C_v} \tag{2.9}$$

(See Kennard[5] for an evaluation of the constant.)

Since $D \gg d$, g_2/D can be neglected, so that eqn (2.8) may be rewritten

$$g_1 + d\log D/d = K_0/E \tag{2.10}$$

and as

$$g_1 = K_1/p$$

$$K_1/p + d\log D/d = K_0/E \tag{2.11}$$

or

$$K_1/K_0p + 1/E_0 = \frac{1}{E} \tag{2.12}$$

where E_0 is a constant equal to the heat conduction when p approaches infinity. At low pressures where $T'_g = T_g$ and $K_1/K_0p \gg 1/E_0$, eqns (2.7) and (2.12) become identical.

2.2 Measurement of thermal conductivity

The present-day method of measurement is basically the same as that described by Pirani[1] in 1906, the heat loss from the wire being measured electrically with a Wheatstone bridge network which serves both to heat the wire and to measure its resistance. The title 'Pirani gauge' is given to all gauges in which any form of Wheatstone's bridge is used to measure the heat losses.

The electrical circuit in its simplest form is shown in Figure 2.3. The hot wire R_w is included in a bridge with two other arms R_2, R_3, having equal resistance, and a fourth, R_1, variable. The principle of operation is simple. At a very low pressure (say 10^{-6} mbar) the voltage V applied to the bridge is set at some convenient value V_0 and the resistor R_1 adjusted to give balance. An increase of pressure, by increasing the heat loss, lowers the wire temperature and, hence, unbalances the bridge. The wire temperature can be increased, and therefore the bridge rebalanced, if the heat input is increased by increasing the bridge voltage (to some value defined as V_1). When the bridge has been rebalanced,

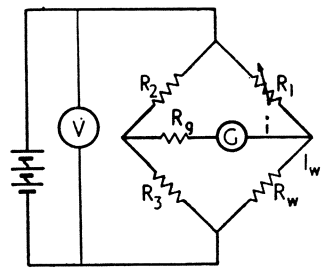


Figure 2.3 Wheatstone bridge control circuit for the Pirani gauge

the resistance of the hot wire, and therefore its temperature, must be at the original value. The thermal equilibrium equations at the two pressures may be written:

$$\text{electrical heat input to the wire}$$

$$= \text{radiation} + \text{conduction} + \text{loss to the end supports} \tag{2.13}$$

At very low pressures where conduction is negligible, this becomes

$$\frac{V_0^2}{R_w}\left(\frac{R_3}{R_2+R_3}\right)^2 = K_R(T_w^4 - T_g^4) + \text{end losses} \tag{2.14}$$

where K_R, the radiation constant, is given by

$$K_R = \sigma\varepsilon(\text{surface area of the wire})\ \text{W}\,^\circ\text{C}^{-4}$$

(see Hunt[7] for comments on the calculation of the constant $\sigma\varepsilon$).

At a pressure p mbar,

$$\frac{V_1^2}{R_w}\left(\frac{R_3}{R_2+R_3}\right)^2 = K_R(T_w^4 - T_g^4) + K_c p(T_w - T_g) + \text{end losses} \tag{2.15}$$

where K_c is the conduction constant for the wire, and

$$K_c = 0.011\,\alpha\frac{(\gamma+1)}{(\gamma-1)}\cdot\left(\frac{1}{M}\right)^{1/2}\left(\frac{273}{T}\right)^{1/2} \times \text{surface area of wire}\ \text{W}\,^\circ\text{C}^{-1}$$

(see eqn 2.7).

Since the wire temperature is constant, we may assume the radiation loss and the conduction of heat through the end supports to be very nearly the same at all pressures.

Thus, subtracting eqn (2.14) from (2.15):

$$\frac{V_1^2 - V_0^2}{R_w}\left(\frac{R_3}{R_2+R_3}\right)^2 = K_c(T_w - T_g)p \tag{2.16}$$

which gives

$$p = \text{constant}\ (V_1^2 - V_0^2) \tag{2.17}$$

i.e. a linear relation between the square of the voltage and pressure.

Some experimental curves of $(V^2 - V_0^2)$ plotted against p for air, hydrogen, argon and helium are drawn in Figure 2.4. In Table 2.1, the extent of the linear range is compared with the mean free path and wire diameter. The results show the linear range to extend approximately to the pressure at which $\lambda = 10d$, and also that the thermal conductivity phenomenon can be usefully employed at pressures up to and above 10 mbar. These curves were obtained with wires 100 mm long, stretched along the axis of a tube 40 mm in diameter ($D = 40$ mm, Figure 2.1).

The above simple theoretical treatment in section 2.1 predicts that the wire-to-wall distance will be unimportant. Von Ubisch[8] has verified this in an

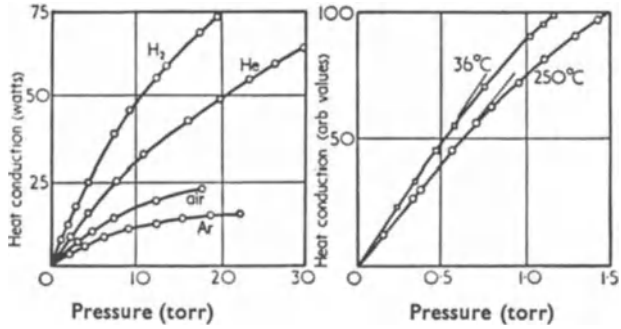


Figure 2.4 Experimental curves of heat loss from a wire 0.01 mm in diameter, 100 mm long, by conduction through the gas as a function of the gas pressure.

Table 2.1 Experimental results showing how the linear operating range p_m of the thermal conductivity gauge depends upon the hot wire diameter (or breadth in the case of a ribbon).

		$d_w = 0.01$		$d_w = 0.025$		$d_w = 0.1$		$w = 0.75$	
Gas	λ_0	p_m	λ_m/d_w	p_m	λ_m/d_w	p_m	λ_m/d_w	p_m	λ_m/w
Air	0.048	650	9.6	200	12.5	50	12.0	25	3.2
H_2	0.090	1950	6.0	455	10.2	110	10.5	60	2.7
He	0.141			715	10.0				

d_w, wire diameter (mm).
w, breadth of a ribbon wire (mm).
λ_0, molecular mean free path (mm) at a pressure of 1 bar and a temperature of 0 °C.
p_m, pressure (mbar × 10^3) at which the heat conductivity falls by 1% below the predicted value (eqn 2.17).
λ_m, molecular mean free path (mm) at the pressure p_m.

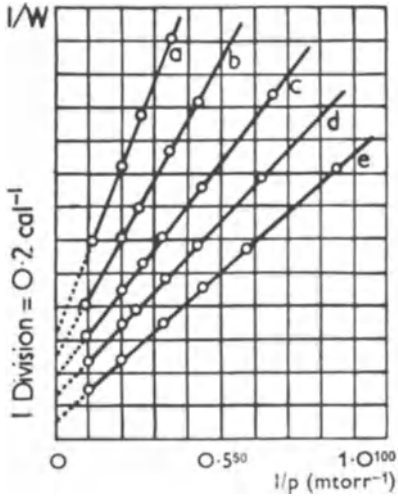


Figure 2.5 Curves of $1/W$ as a function of $1/p$ obtained by Dickins[9] for oxygen with a wire 0.03 765 mm in diameter. (W is the rate of heat conduction through the gas and p the pressure.) Temperature differences between wire and wall (*a*) 7.69 °C; (*b*) 11.71 °C; (*c*) 15.74 °C; (*d*) 19.78 °C; (*e*) 23.82 °C.

experiment where he varied the cylinder diameter D between 2.5 mm and 100 mm and found the characteristics to be unchanged.

Careful measurements in the temperature-jump region were first made by Dickins[9] in 1934. By plotting $1/E$ as a function of $1/p$, he obtained the linear relationship predicted by eqn (2.12). A typical set of Dickins' results is shown in Figure 2.5.

2.3 Sensitivity

The Pirani gauge sensitivity, defined as dT_w/dp, can be calculated as a function of T_w for the linear region by differentiating the basic heat equilibrium, eqn (2.15), with respect to p. This gives

$$0 = 4K_R T_w^3 \, dT_w/dp + K_c p \, dT_w/dp + K_c(T_w - T_g) \tag{2.18}$$

(the end losses are assumed to be independent of pressure.)

Therefore

$$-\frac{dT_w}{dp} = \frac{K_c(T_w - T_g)}{4K_R T_w^3 + K_c p} \tag{2.19}$$

$$\simeq \frac{K_c(T_w - T_g)}{4K_R T_w^3} \quad \text{for } p \simeq 0 \tag{2.20}$$

$$\simeq \frac{3.1 \times 10^8 (T_w - T_g)}{T_w^3} \tag{2.21}$$

(values of K_R and K_c as in the appendix to this chapter).

dT_w/dp has been calculated as a function of p for $T_g = 100$, 200 and 300 K, i.e. for temperatures approximately equal to liquid nitrogen, solid carbon dioxide

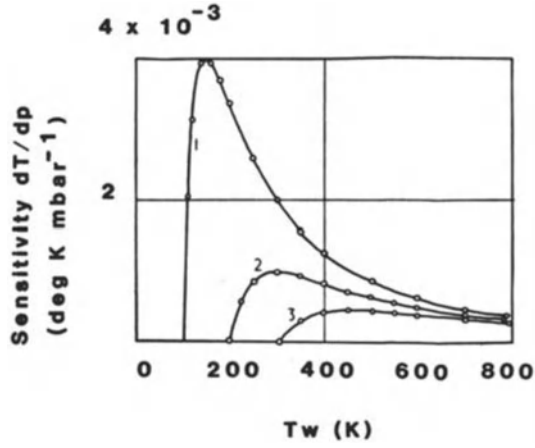


Figure 2.6 The sensitivity of the thermal conductivity gauge expressed as the change in wire temperature per unit pressure change, as a function of the wire temperature: (1) wall temperature $T_g = 100$ K, (2) $T_g = 200$ K, (3) $T_g = 300$ K.

and room, the values being plotted in Figure 2.6. These curves show: (i) sensitivity to be practically independent of T_w when $T_g = 300\,K$ provided $T_w > 400\,K$, and (ii) that an almost tenfold increase in sensitivity can be achieved by reducing T_g to liquid nitrogen temperature.

2.4 End losses

To evaluate eqn (2.16) it has been assumed that both the heat loss to the end supports and the temperature gradient along the wire are independent of pressure. This is exactly true only for a very long wire. The ratio of radial to longitudinal heat loss depends upon the gas pressure, therefore the temperature gradient along the wire and the heat loss to the supports must depend upon the gas pressure. Nevertheless, a straightforward mathematical treatment, as made for example by Weber[10], Von Ubisch[8] and Thomas[11], shows the assumptions to be justifiable for a wire more than a few centimetres long. This can be seen from the curves plotted in Figure 2.7, which show the variation in temperature (actually $T_w - T_g$) along the length of 0.025-mm diameter wires 10 and 100 mm long at pressures of 0.13 and 1.3 mbar. The calculations upon which these curves are based assume[12] a current such that the temperature difference $(T_w - T_g)$ at the centre of an infinitely long wire would be 50 °C.

2.5 Accommodation coefficient and relative sensitivity

Provided the dimensions of the wire are known accurately, the accommodation coefficient α and relative sensitivity for different gases can be calculated

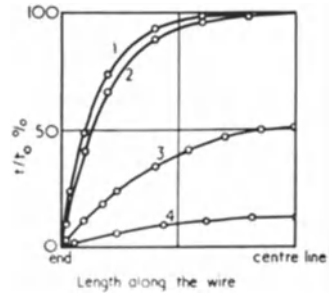


Figure 2.7 Temperature variations along the length of the hot wire 0.025 mm in diameter. All the curves have been drawn for t_0, the temperature at the centre of an infinitely long wire passing the same current, 50 °C.

(1) $l = 100\,mm$, $p = 1.0\,mbar$
(2) $l = 100\,mm$, $p = 0.1\,mbar$
(3) $l = 10\,mm$, $p = 1.0\,mbar$
(4) $l = 10\,mm$, $p = 0.1\,mbar$

from the data presented in Figure 2.4. Measurements have been made by many workers for all of the common gases. There is no exact agreement in the results, probably because they depend so much upon gas adsorption at the surface of the wire. There may be an increase of α by as much as a factor of 10 when a completely 'clean' surface becomes contaminated with an adsorbed layer of gas molecules. This critical dependence on the surface contamination was demonstrated by Roberts[13] in a classic series of experiments performed in the early 1930s. Most of the measurements made by Roberts were for the inert gases and tungsten surfaces. He took care to ensure the best possible vacuum conditions and freedom from foreign gases. The effect of introducing gases such as oxygen or hydrogen, which quickly form an adsorbed surface layer, was also investigated. Figure 2.8 shows the rapid change in α for neon which takes place when an impurity of hydrogen of about 1 part in 1000 is introduced. The slow rise before the deliberate introduction of hydrogen was probably due to a very small residual impurity in the neon. Roberts obtained a 'true' value of α for a clean surface by extrapolating these curves back to zero time. It is interesting to note that changes in accommodation coefficient have been used by Roberts[13], Mann[14] and others as a monitor of surface changes.

The hot wire is almost inevitably grossly contaminated when it is in general use as a pressure gauge. This means that α is high (in many cases approaching unity), but is, fortunately, sufficiently stable for most work. A list of values obtained by various workers is set out in Table 2.2. The measurements under sections *A* and *B* were made in the low-pressure region (0 to 100 mbar); for *A* all possible precautions were taken to ensure clean surfaces, whereas for *B* no special care was taken. The figures obtained for conductivity measurements in the 'temperature-jump region' (1 to 100 mbar, see section 2.2) appear in section *C*. The average values for each gas (ignoring the figures in *A*), together with

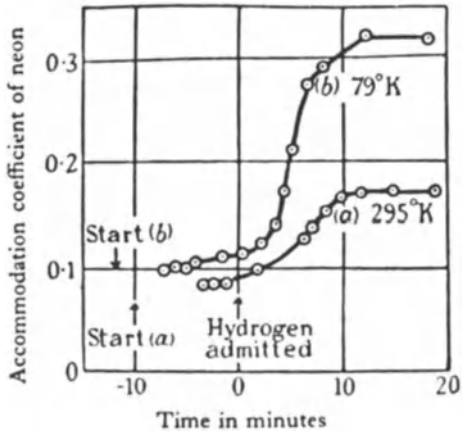


Figure 2.8 Change in accommodation coefficient of neon due to the introduction of an impurity of hydrogen of about 0.1% as measured by Roberts[13]. (Reproduced from Miller, A.R. (1949), *The Adsorption of Glass on Solids*, Cambridge University Press, London.)

Table 2.2 Values of accommodation coefficient obtained by various workers. The measurements under sections *A* and *B* were made in the low-pressure region (0 to 0.1 mbar), but for *A* all possible precautions were taken to ensure clean surfaces, whereas for *B* none of the special techniques was used. The measurements for the 'temperature-jump region' are in section *C*. The average values of the accommodation coefficient (for sections *B* and *C* only), and the corresponding values of the heat conduction in watts per square centimetre per 1 °C per mbar, as calculated from eqn (2.16), are also given.

	A				*B*				*C*				Av. value		Rel.
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	α	Cond. $\times 10^3$	cond.
Helium	0.03	0.057		0.071	0.403	0.38	0.238	0.37	0.50				0.38	11.7	0.97
Neon		0.07	0.056		0.700	0.75	0.57	0.65					0.74	10.5	0.87
Argon	0.55				0.847	0.80	0.89	0.86	0.88				0.86	7.9	0.65
Krypton					0.844								0.84	5.3	0.44
Xenon					0.858								0.86	4.5	0.37
Hydrogen	0.11/0.20				0.312	0.24	0.22	0.28	0.34	0.29	0.28	0.226	0.29	17.9	1.48
Nitrogen					0.769								0.77	12.7	1.05
Oxygen	0.42				0.782	0.62	0.74	0.82	0.82				0.79	12.1	1.00
CO_2						0.52	0.76	0.83	0.78				0.78	13.2	1.09
CO					0.772								0.77	12.7	1.05
Mercury	1.00												1.00	4.1	0.34
Metal surface	Pt	W	Fe	Ni		Pt	Pt		Pt	Pt	Pt	Pt			

(a) Mann, W.B. *Proc. Roy. Soc. A* **146**, (1934) 776.
(b) Roberts, J.K. *Proc. Roy. Soc. A*, **129**, (1930) 146; **135**, (1932) 192; **142**, (1933) 518.
(c) Eggleton, A. E.J. and Tompkins, F.C. *Trans Faraday Soc.* **48**, (1952) 738.
(d) Raines, B. *Phys. Rev.* **56**, (1939) 691.
(e) Amdur, I. and co-workers. *J. Chem. Phys.* **12**, (1944) 159; **14**, (1946) 339.
(f) Soddy, F. and Berry, A.J. *Proc. Roy. Soc. A*, **83**, (1909) 254.
(g) Thomas, L.B. and Olmer, F. *J. Amer. Chem. Soc.* **65**, (1943) 1036.
(h) Knudsen, M. *Ann. Phys. (Leipzig)*, **34**, (1911) 593; **6**, (1930) 129; **46**, (1915) 641.
(i) Dickins, B.G. *Proc. Roy. Soc. A* **143**, (1934) 517.
(j) Archer, C.T. *Proc. Roy. Soc. A* **165**, (1938) 474.
(k) Gregory, H.S. *Proc. Roy. Soc. A* **149**, (1935) 35.
(l) Gregory, H.S. and Dock, E.H. *Phil. Mag.* **25**, (1938) 129.

thermal conductivity in watts per square centimetre per degree Celsius temperature difference, are also given in Table 2.2.

The final column in this table gives values of relative conductivity, which are in effect the values of relative sensitivity in the low-pressure regime. If the accommodation coefficients are known, then the relative sensitivity R_{21} for any two gases can be calculated easily. With reference to eqn (2.7),

$$R_{21} = \frac{\gamma_2 + 1}{\gamma_1 + 1} \cdot \frac{\gamma_1 - 1}{\gamma_2 - 1} \cdot \frac{\alpha_2}{\alpha_1} \left(\frac{M_1}{M_2} \right)^{1/2} \tag{2.22}$$

(see Veis[15] for a discussion of relative sensitivity in the high-pressure regime).

2.6 Alternative methods of bridge control

The above method of operation (the constant-temperature bridge), whilst in many ways satisfactory, suffers two important disadvantages. An adjustment must be made before each reading can be taken, and the relation between pressure and voltage is not linear. There are, however, two alternative methods of operating the Wheatstone bridge, described by Hale[16] in 1911 as (i) the constant-current bridge, and (ii) the constant-voltage bridge. In (i) the hot-wire current, and in (ii) the bridge voltage, are kept constant at all pressures. In both cases pressure is measured as a function of the bridge out-of-balance current i, this current being set to zero at a very low pressure by adjusting R_1 (Figure 2.3). The constant-voltage bridge is used extensively in modern equipment, because no adjustment needs to be made to the bridge after the initial setting. Unfortunately the output current i is directly proportional to pressure only through a limited pressure range.

2.7 Useful range of the constant-voltage bridge

The output current per unit pressure change $\mathrm{d}i/\mathrm{d}p$ for the constant-voltage operation is a function of the temperature difference between wire and gas $T_w - T_g$. Thus, because an increase in pressure reduces the wire temperature (and hence $T_w - T_g$), the sensitivity falls as the pressure increases. In fact the useful pressure range is limited by this fall of temperature rather than by mean free path considerations. The two points to be considered in obtaining the maximum range are (i) to make the sensitivity $\mathrm{d}T_w/\mathrm{d}p$ the minimum practicable, and (ii) to have the operating temperature (the wire temperature at zero pressure) as high as possible.

For a short wire, where the end losses are comparable with the heat conduction, the sensitivity falls as the ratio of end losses to heat conduction increases. (The heat conduction at a given pressure is a smaller fraction of the total heat dissipation for a shortened wire. Thus $\mathrm{d}T_w/\mathrm{d}p$ is reduced and the operating range increased.) The end losses for a 0.025-mm diameter wire are greater than the radiation losses for $l < 10$ mm, thus the useful pressure range

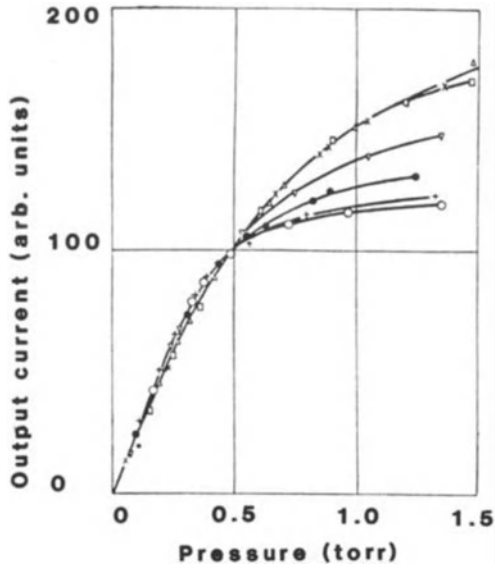


Figure 2.9 Calibration curves for the constant-voltage Wheatstone bridge Pirani gauge. The wire temperature at zero pressure was 200 °C in each case. (All curves have been adjusted to pass through the same point at a pressure of 0.5 mbar.) ○ $R_1/R_w = 10$, △ $R_1/R_w = 0.1$.

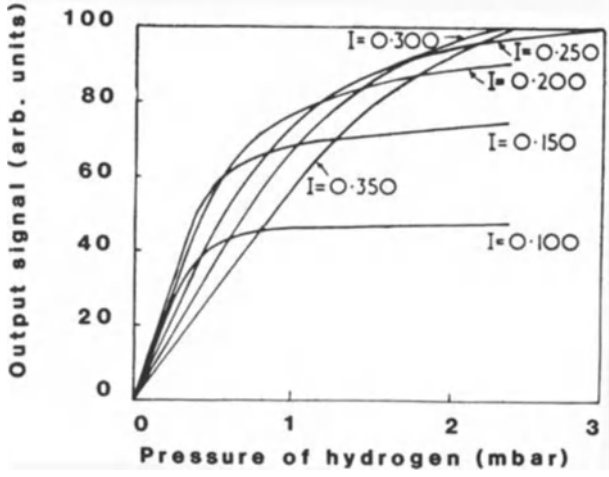


Figure 2.10 Calibration curves of Pirani gauge calculated by Dunoyer[18] for the constant-current Wheatstone bridge. (The accuracy of these curves has been verified by experiment.) The value of the hot wire current is marked on each curve, the operating temperatures being 114, 216, 314, 409, 498 and 581 °C corresponding to wire currents of 0.100, 0.150, 0.200, 0.250, 0.300 and 0.350 A respectively. The calculations were made for a tungsten wire 0.045 mm in diameter and for a gas temperature 0 °C.

of the gauge may be expected to be more or less inversely proportional to length when $l < 10$ mm. This effect has been demonstrated experimentally by Leck and Martin[17]. Their results show very clearly the advantage to be gained by decreasing l.

For any given bridge voltage the power supplied to the wire changes as the pressure changes (the wire can be considered as a variable load fed from a voltage source of internal impedance R_1). If $R_w \gg R_1$, then the power supplied to the wire increases as R_w falls, but if $R_1 \gg R_w$, the power decreases as R_w falls. This means that as R_1 is reduced, the sensitivity decreases, and therefore the workable range increases. Figure 2.9 confirms this by showing the change in the calibration curve of a given wire as R_1 falls from $10R_w$ to $0.1R_w$. It may be noted that the constant-voltage bridge with $R_1 \gg R_w$ is essentially the same as the constant-current bridge mentioned in section 2.6.

The alternative method of improving the working range, by increasing the operating temperature, has been carefully investigated (for example by Dunoyer[18]). The most important results of his work are summed up in Figure 2.10. This figure shows the improvement in the calibration shape of a long tungsten wire when it is operated at high temperature. Similar results have been obtained by Von Ubisch[19] and the present author[20]. It is interesting to note that the dependence on temperature applies only for a long wire. For a short wire, where the end conduction cannot be neglected, the shape of the curves has been found to be independent of temperature.

2.8 The lower limit to the useful pressure range

The smallest pressure change that can be detected is determined by either (i) the smallest signal that can be recorded in the bridge output circuit, or (ii) by the background fluctuations in the signal due to instabilities of the system. For most applications, the limit to the performance at very low pressure is set by the fluctuations in wire temperature due to instabilities in either the supply voltage or ambient temperature (T_g). This is increasingly the case, because with improvements in electronic amplifier design the measurement of the small output signals from the Wheatstone bridge presents little problem. Although considerations of sensitivity need not be important in the design of a gauge and its control circuits, the two main factors are worth considering. These are: (i) the dimensions and the operating temperature of the hot wire, and (ii) the Wheatstone bridge circuit. The design of the hot wire element will be determined by the practical considerations of ease of manufacture and use (sufficiently rugged for all applications) and the need for stable operation. (The choice of metal has little influence on sensitivity, since the temperature coefficients of resistance are similar for the commonly used materials; 0.005 for tungsten and molybdenum and 0.0035 for platinum.) When using the constant-voltage Wheatstone bridge, the resistor values would be chosen to balance the limitations on the operating pressure range with those of

sensitivity (see Figure 2.9). Obviously, if the highest sensitivity is required over a limited pressure range, the ratio R_1/R_w would be chosen between 5.0 and 10.0, approaching the conditions for constant current through the wire.

As stated above, background fluctuations have been found to be the important factors which set the limit to the measurement of low pressure. A change in either supply voltage or gas temperature produces a change of wire temperature, and therefore a signal in the output meter indistinguishable from that due to a pressure change. (An increase in either bridge voltage or gas temperature increases the wire temperature and therefore registers as a fall of pressure in the output meter.) The effect of these voltage and temperature instabilities depends very largely upon the actual wire and gas temperatures, and can be calculated from the basic heat-balance equation. The curves in Figure 2.11 (calculated from eqn 2.15) show the apparent change in pressure reading due to a gas temperature change of 1 °C and also to a 0.1% change of bridge voltage. (The derivation of these results is given in detail in the appendix to this chapter.)

The advantage of cooling the gauge from room to liquid-nitrogen temperature is very marked, being about two orders of magnitude for all wire temperatures. An increase of operating temperature also reduces the importance of temperature fluctuations, but this is only achieved at the expense of increasing the error due to voltage fluctuations, as can clearly be seen in Figure 2.11. The choice of operating temperature must therefore be a compromise.

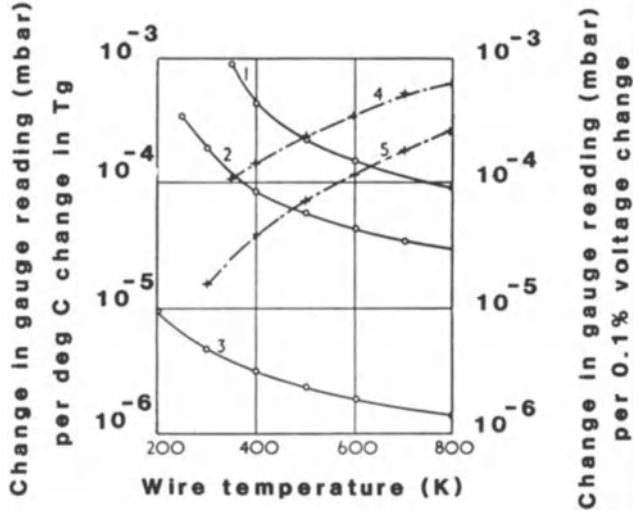


Figure 2.11 The change in gauge reading caused by (*a*) a change in gauge-wall temperature of 1 °C (full curves), and (*b*) a change in supply voltage of 0.1% (chain dotted curves) for various values of T_g: (1) $T_g = 300$ K, (2) $T_g = 200$ K, (3) $T_g = 100$ K, (4) $T_g = 300$ K, (5) $T_g = 100$ K. See appendix to this chapter for the method of calculation.

The smallest detectable pressure change is determined by these fluctuations, rather than by the sensitivity of the measuring circuit. For example, an increase in room temperature of 0.001 °C at 20 °C would register the same output as a pressure decrease from 10^{-6} to 10^{-7} mbar. Ellett and Zabel[21] in their molecular-beam work were able to detect pressure changes of less than 10^{-8} mbar. They used a nickel ribbon 25 mm long (nickel, as well as having a slightly higher temperature coefficient of resistance than tungsten, could more easily be flattened from wire into a thin ribbon by rolling) and operated with the gauge in a liquid-air bath, i.e. $T_g \simeq 100$ K. Even at this low working temperature, Ellett and Zabel[21] found that the background fluctuations, rather than the lack of sensitivity, set the lower limit to their pressure measurements. In most work where $T_g \simeq 300$ K, the background fluctuations are much higher. There is a general agreement that when a compensated gauge is used (see following sections) and the bridge power supply is well stabilized, step changes in pressure of 10^{-6} mbar can be detected.

2.9 The importance of bridge-voltage and temperature fluctuations at high pressure

The errors due to these fluctuations become increasingly important as the pressure rises. Curves showing the error in the gauge reading, due to a 1 °C

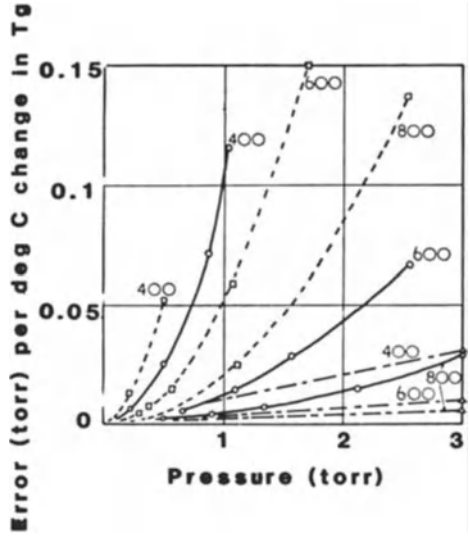


Figure 2.12 The change in gauge reading caused by a change in gauge-wall temperature of 1 °C as a function of the gas pressure (air). The curves are all calculated for a wall temperature of 300 K. ○, constant-voltage Wheatstone bridge; □, constant-current Wheatstone bridge; △, constant temperature. Values of the wire temperature at zero pressure are marked on the curves (K).

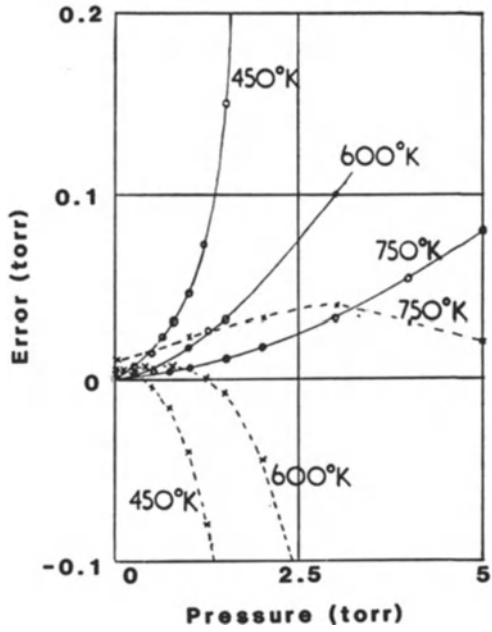


Figure 2.13 The error in gauge reading caused by fluctuations in gauge-wall temperature and Wheatstone bridge supply voltage. For these measurements a 0.025-mm diameter tungsten wire gauge element was used with a constant-voltage bridge ($R_1/R_w = 1/7.5$): —, error due to 1 °C change in temperature; – – –, error due to 1% change in bridge voltage. Values of hot wire temperature are marked on the curves; $T_g = 300$ K.

change in wall temperature, are plotted in Figure 2.12 (see appendix to this chapter for the evaluation). As at low pressure, it is the fluctuations of temperature and voltage which limit the useful range, rather than the sensitivity and stability of the output measuring circuit.

Experiments have been carried out by the author to check these calculations. At various pressures in the range zero to 5.0 mbar, the change in reading was measured as the gauge envelope temperature varied between 15 and 40 °C. The gauge output varied linearly with temperature throughout this temperature range. The change expressed as an error in reading has been plotted as a function of pressure in Figure 2.13. The effect of supply voltage fluctuations has also been measured. The error in gauge reading due to a 1% variation of supply voltage is also plotted as a function of pressure in Figure 2.13.

2.10 Compensation for temperature and voltage fluctuations

It is possible to reduce the effect of the voltage and temperature instabilities by using two identical gauge elements, placed in adjacent arms of the Wheatstone

bridge. The extra element, replacing R_3 in the bridge (Figure 2.3), is pumped out and sealed off at a very low pressure. Provided the pressure is the same in each gauge, a change of either ambient temperature or bridge voltage will have the same effect on each and therefore will not unbalance the bridge. Obviously the two elements must be placed close together so as to be subject to the same fluctuations of temperature.

Compensation is fully effective only when the pressure is approximately the same in each element, because the change in wire temperature (due to gauge envelope temperature fluctuations) depends upon the gas pressure in the element. For example, an envelope temperature change of 1 °C changes the wire temperature from 400 K ($T_g = 300$ K) by 0.4, 0.7, 0.8, 0.9 and 0.95 °C at pressures of 0, 0.1, 0.2, 0.5 and 1.0 mbar respectively (see appendix to this chapter for the calculations). Thus, with the second element sealed off at zero pressure, compensation will be only 50% effective at 0.2 mbar. By sealing off at 1 mbar, compensation can be made effective at this pressure, but only at the expense of magnifying the drift at zero pressure by 50%. In practice, therefore, the compensated bulb technique can only be complete at pressures less than 0.1 mbar. (The limitation of compensation to a narrow pressure range was demonstrated by Hale[16].)

An alternative to using an extra gauge element is to make R_3 a temperature-sensitive resistor mounted in good thermal contact with the original gauge element. This resistor is best made in two sections with differing temperature coefficients of resistance, for example one of copper and the other of nichrome wire. By trial and error the correct ratio between the two sections can be determined, so that the overall temperature coefficient just matches that of the Pirani element itself. Again, just as with the compensating bulb, the matching is correct over a limited pressure range only. English *et al.*[22] described in detail the application of this technique to a gauge operating in the constant-temperature mode. The circuit they used is shown in Figure 2.14. Their hot filament (a 75-mm length of 0.012-mm diameter wire) was incorporated in a Wheatstone bridge, with the adjacent arm either a conventional low-temperature coefficient resistor or a temperature-compensating resistor in direct, and very good thermal, contact with the gauge body. This compensating resistor was designed to have the same resistance and dR/dT value as the gauge filament. The constant-temperature bridge was under automatic control (any out of balance was coupled directly to the series stage of the bridge voltage stabilizer), the bridge voltage being adjusted continuously so that the energy dissipated in the wire maintained the temperature at 400 K. The bridge voltage varied from 0.6 V at the lowest pressure to 8.0 V at atmospheric pressure. The calibrations for three pressure ranges, and at three temperatures of the gauge wall, are shown in detail in Figure 2.15. A comparison of the two sets of results indicates very clearly the value of this compensation technique. The improvement in performance is by factors of about 10 and 2 at the lowest and mid-pressure parts of the scale respectively.

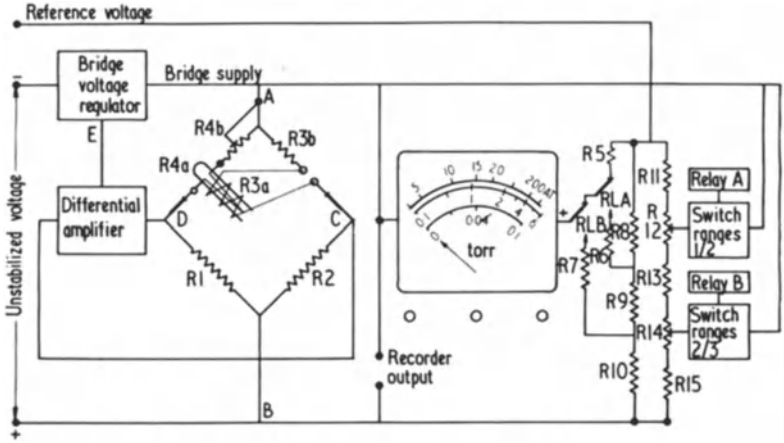


Figure 2.14 Schematic circuit diagram of gauge designed for ambient temperature compensation. *A*, *B*, bridge supply voltage, controlled by the regulator; *C*, *D*, output of the bridge to the differential amplifier, *E*, output of differential amplifier to control the regulator[22].

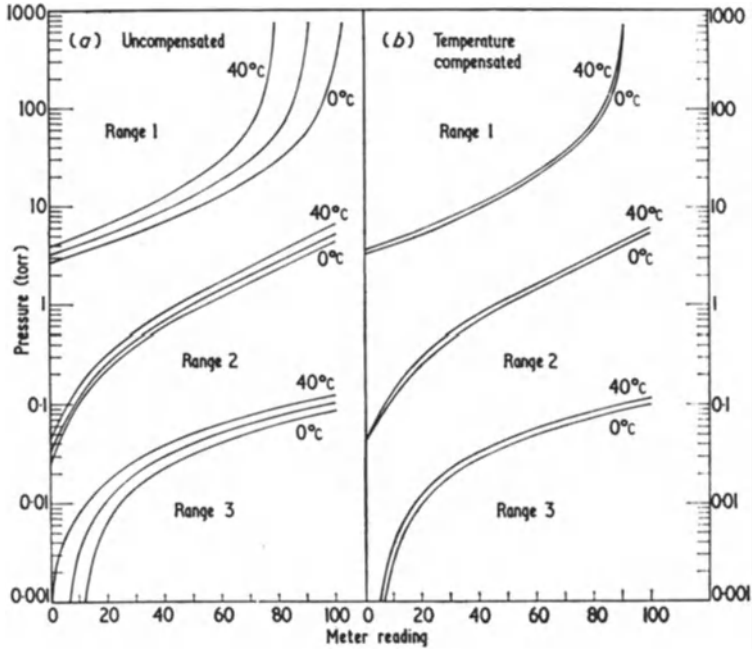


Figure 2.15 Comparison of meter readings on the three ranges, uncompensated and with temperature compensation. In (*a*) the curve between 0 and 40 °C was obtained at 20 °C. In (*b*), the curve for 20 °C lies midway between those for 0 and 40 °C but has been omitted from the diagram to avoid confusion[22].

Above 100 mbar, compensation is obviously vital if the gauge output is to have any real meaning. The authors of this work also point out that compensation reduces very considerably the 'switch-on' transient caused by the small increase in envelope temperature due to the heat dissipation from the wire.

2.11 Physical changes in the gauge wire (ageing effects)

There is unfortunately a lack of detailed experimental evidence from which to estimate the slow 'ageing' of the hot wire. Two different reactions are liable to upset its characteristics: (i) a reduction in diameter of the wire due to oxidation which must increase the resistance of the wire and eventually cause an open circuit, and (ii) a change in surface structure (due for example to an adsorbed gas layer) which may alter the accommodation coefficient. (This latter effect may be expected to be either reversible – dependent upon the operating conditions – or be a drift in a fixed direction which must eventually slow down and reach an equilibrium.) Both oxidation and surface changes can be identified for tungsten and even for platinum.

Experiments[20] carried out for tungsten and platinum in the constant-voltage bridge showed the pressure range 10^{-3} to 10^{-1} mbar to be the most effective in oxidizing the wire. Below 10^{-3} mbar it was the low gas density, and above 10^{-1} mbar the reduction of wire temperature, which slowed down the reaction. The most significant results were obtained by observing the change in zero reading after the gauge had been operated for a long period in air at a pressure of 10^{-2} mbar. The change in zero reading expressed as an apparent change of pressure is plotted in Figure 2.16 as a function of the operating

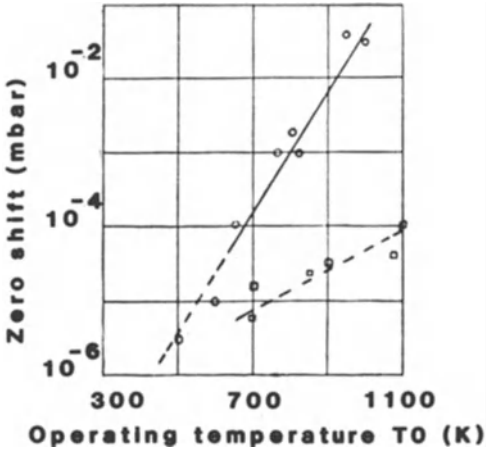


Figure 2.16 Pirani gauge zero drift. The change in zero reading expressed as an apparent change of pressure, at zero pressure caused by operating for 1 h at 0.01 mbar (air), is plotted as a function of the wire operating temperature (temperature at zero pressure). ○, tungsten; □, platinum; $T_g = 300$ K; $R_1/R_w = 1/7.5$.

temperature (the temperature at zero pressure) for both platinum and tungsten wire. The changes were always non-reversible in the direction of increasing resistance, and could have been caused by the wire becoming either thinner or hotter. Later experiments[23] established the drift rate for a quartz-coated wire to be less than 10^{-5} mbar h^{-1} at 10^{-2} mbar pressure, even with an operating temperature of 1200 K.

For platinum operating at a temperature in excess of 450 K, drifts in gauge indication were found (typically from 10^{-2} to 1.0 mbar) after a sudden change of pressure from a low to a high value[20]. These drifts can introduce errors into the gauge reading of the order of $\pm$ 10% at a pressure of 1 mbar. This effect was not observed for tungsten or quartz-coated platinum.

2.12 Extension of working range to atmospheric pressure

It has been known for many years that the range of the Pirani gauge can be extended to higher pressures by making use of the heat transferred in convection currents. This transfer of energy becomes significant above about 50 mbar, and increases as a function of pressure up to 1000 mbar. As the mechanism depends upon gravity, the magnitude of the transfer depends upon the orientation of the wire. (McMillan and Buch[24], Johnson[25] and Flanick and Ainsworth[26] all point out that the calibration changes when the gauge moves from the horizontal to the vertical position. McMillan and Buch[24] go so far as to state that a tilt of 5° from the vertical is significant.)

Heijne and Vink[27] and later Fletcher and Steckelmacher[28] described gauges specially designed to take full advantage of convection cooling. Both groups used thin coiled filaments mounted horizontally to present a large

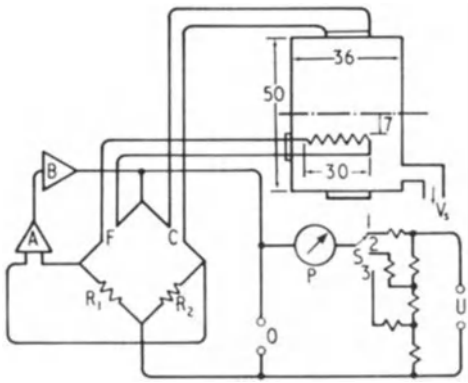


Figure 2.17 Schematic diagram of cylindrical convection gauge head provided with wall temperature sensing, associated control circuit and pressure readout: *A*, differential amplifier; *B*, bridge voltage regulator; R_1, R_2, bridge resistors; *F*, hot wire element; *C*, resistance thermometer (envelope temperature); *O*, recorder output (bridge voltage); *P*, meter with pressure scales; *S*, range switch with zero offset; *U*, reference voltage; V_S, gauge connection to vacuum system[28].

surface area for cooling. In their experimental gauge, Fletcher and Steckelmacher[28] used a 10 mm length of standard 15 W tungsten lamp filament stretched between two supports 30 mm apart. This was mounted horizontally 7 mm below the axis of a metal tube of 50 mm inside diameter, 36 mm long. This arrangement, together with the Wheatstone bridge, is shown in the schematic diagram in Figure 2.17. Clearly the wire operates at constant temperature (i.e. constant resistance) with temperature compensation by the resistance 'thermometer' *C*. Figure 2.18 shows the calibration curve over the pressure range 10–700 mbar for the gauge operating with the filament 200 °C above ambient temperature. For comparison, the comparable characteristic of the normal gauge operating without convection is shown. Steckelmacher, in a comprehensive review[29], discussed other interesting techniques for extending the pressure range of the hot wire gauge to atmospheric pressure. These included, firstly, the use of some form of 'blower' to give relative motion between the hot

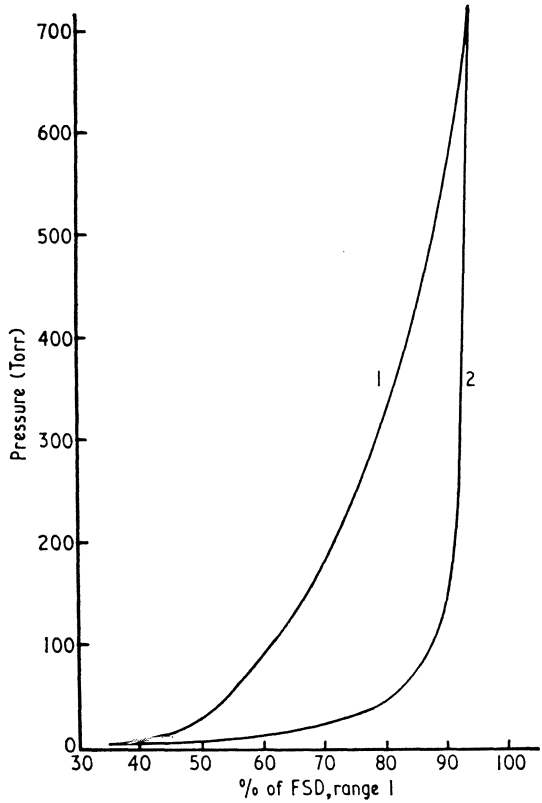


Figure 2.18 Calibration curve for convection gauge and normal Pirani gauge in the range 10–700 mbar; curve 1, convection gauge head; curve 2, Pirani gauge head (Edward High Vacuum Model[28]).

wire and the gas to make the gauge behave as a hot-wire anemometer and, secondly, pulsing the gas with a movable bellows.

2.13 Commercial gauges for laboratory and industrial use

The gauges developed at Edwards High Vacuum International following the investigations of English, Fletcher and Steckelmacher[22,28,29] are representative of the instruments available to vacuum engineers. Many gauges in the Edwards range operate in the constant temperature mode and use a nickel wire resistor for temperature compensation. Most gauges are fitted with a sintered phosphor bronze filter at the inlet to minimize the effect of pressure surges which may damage the filaments. For the range 10^{-3} to 200 mbar, a platinum/10% rhodium hairpin filament, held under tension by a spring attached to its apex, is used as the gauge element. Normally this filament is mounted vertically in an aluminium tube, except in a 'corrosion-resistant' version where stainless steel tubing is used. For applications requiring measurement at the low end of the pressure scale, two modifications have been introduced; firstly, in order to increase the heat loss from the Pirani element, the hairpin is replaced by a helical filament, and secondly, to reduce the drifts

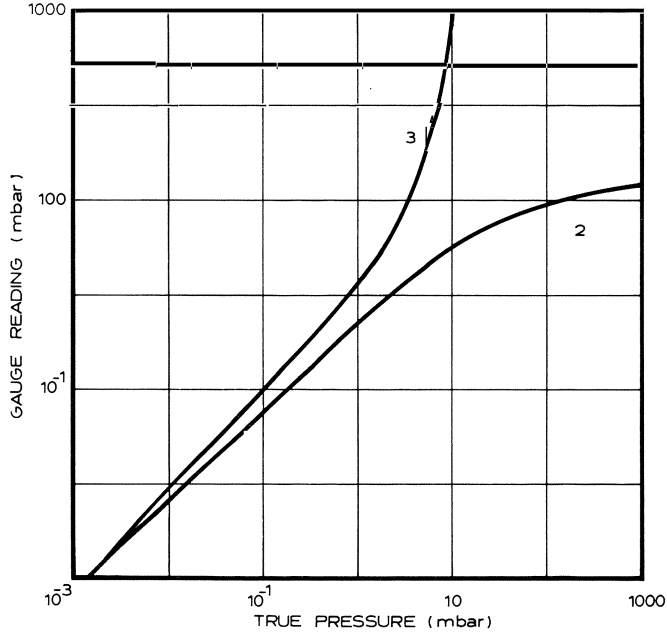


Figure 2.19 Corrections to be applied to the Edwards 'constant-temperature' Pirani gauges when operating in helium (curve 1) or argon (curve 2). The calibration has been carried out in nitrogen.

caused by surface oxidation, a gold-coated tungsten wire is used. Edwards rely upon convection cooling to extend the useful range to atmospheric pressure. They use a long straight gold-coated tungsten filament mounted horizontally in a tube of increased body diameter. This gives characteristics very similar to those described in Figure 2.18.

It is normal practice for commercial gauges to have 'calibration scales' which refer to operation in nitrogen or air. Data published by Edwards emphasize the point that separate calibration scale correction must be applied if Pirani gauges are to be used in a quantitative manner for gases other than nitrogen and air in the non-linear parts of their scales. This is illustrated by the curves given in Figure 2.19, which show the large corrections which must be applied in any gas with a molecular mass and mean free path differing significantly from that of nitrogen.

2.14 The thermocouple gauge

A thermocouple can be used as an alternative to the Wheatstone bridge for measuring the temperature of the hot wire. This has been recognized for many years, having first been suggested by Voege in 1906[30]. In general, the sensitivity of the thermocouple, measured in terms of watts output per unit pressure change, is less than that of the Pirani gauge; obviously the manometer head itself is complicated by the introduction of the extra circuit. The thermocouple may be either spot-welded or silver-soldered to the hot wire. The technique of welding, although probably requiring a little care and practice, is essentially straightforward. Hart and Elkin[31] have evolved a simple procedure for wires down to 0.025 mm in diameter. As an alternative, the joint can be made with a silicate cement, which forms an adequate thermal contact and provides an electrical insulation between the thermocouple and the hot wire.

Numerous practical manometer designs have been described in the literature[32–34]; the following types illustrate the important features. Webber and Lane[35] claim ease and economy of construction for their design. The principles of their design are shown in Figure 2.20. The wires are chromel CC and alumel AA hard-soldered to nickel leads. (The temperature-sensitive element is half chromel and half alumel.) The nickel leads are made rigid by the pyrex spacer *P*. Constant-current operation (150 mA) is used to give a useful working range up to 1 mbar.

Although operating with a short wire reduces the sensitivity in the low-pressure region, the total output swing ΔI is unaffected (ΔI is the change in output current for a pressure change from zero to atmospheric). This means, as well as a reduction of the input power, an increase of the useful range at the expense of a loss of sensitivity in the low-pressure region. The calibration curves of three gauges (those made by Webber and Lane[35], Dunlap and Trump[36] and that by the AEI Scientific Division) are reproduced in

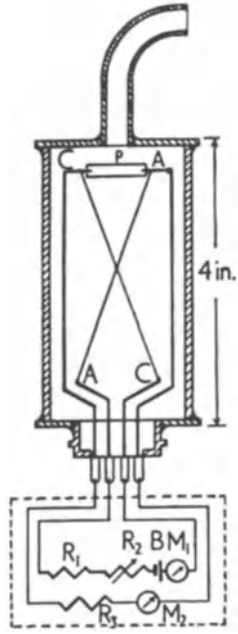


Figure 2.20 Thermocouple gauge and control circuit designed by Webber and Lane[35].

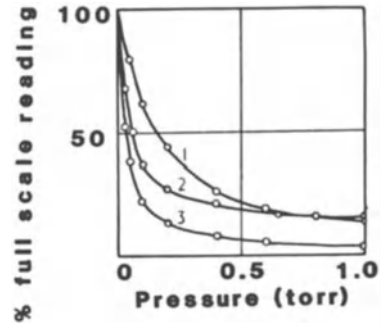


Figure 2.21 Comparison of calibration curves of various thermocouple gauges: (1) commercial unit by AEI; (2) Webber and Lane[35]; (3) Dunlap and Trump[36].

Figure 2.21 to illustrate this point. The hot wire for the AEI gauge was 10 mm long, compared with about 100 mm for the other two.

Benson[37] has described a range of thermocouple elements each designed for a particular operating range. The different scale shapes of these three gauges are shown clearly in Figure 2.22. The gauge DV-4 uses short butt-welded wires that are about as small as can be fabricated economically; this extends the pressure range to the highest limits. The other two gauges have progressively

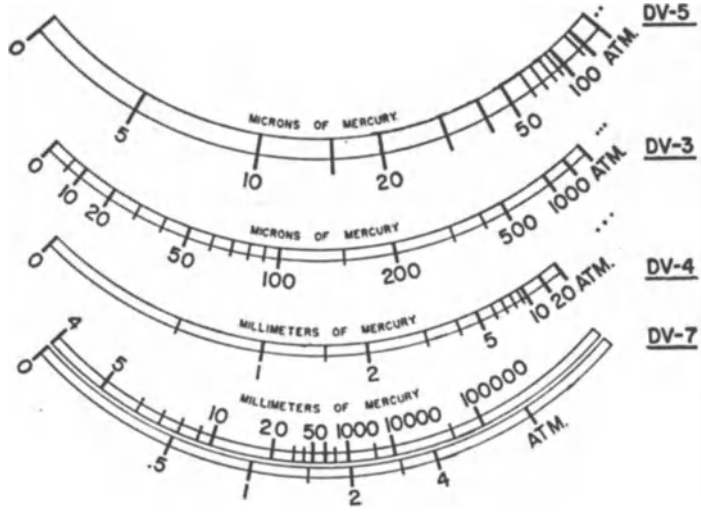


Figure 2.22 Calibration scale shapes for the four gauges designed by Benson[37].

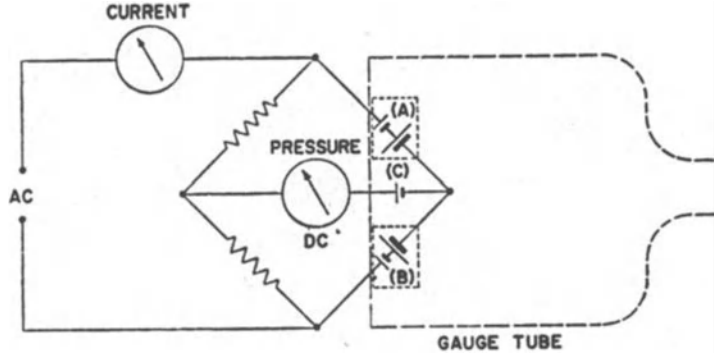


Figure 2.23 Equivalent bridge circuit for the thermopile gauge designed by Benson[37].

thicker wires; the diameters of the wires in DV-3 are eight times those in DV-4. All these gauges have been made with three thermocouple junctions *A*, *B* and *C*, which are arranged electrically as shown in Figure 2.23. In order for this element to operate successfully, the heating current should be ac, with the input transformer centre-tapped so that little or no ac flows through the dc output meter. The output meter thus indicates only the thermal voltages; that is, the mean temperature of the hot wires. The junction *C*, in series opposition to the other two, is unheated so that it responds only to ambient temperature fluctuations and not to pressure. Thus this junction provides temperature compensation. Benson[37] has also given careful consideration to the speed of response of these instruments, and has shown that the time constant is very

much reduced for the thinner wires because of their low thermal inertia. He quotes values of 0.08 and 10 s for the gauges with the thinnest and the thickest wires respectively.

Work carried out in the decade starting in 1945 established the thermocouple gauge as a simple, inexpensive and reliable monitor for pressure measurement in the range 10^{-3} to 1 mbar. It has found application both in the research laboratory and in industry, frequently as a 'pressure sensor' where the vacuum system is under automatic control.

Appendix: Effect of ambient temperature on the gauge reading

1. *Effect of ambient temperature on the gauge reading in the low-pressure region* ($p < 10^{-2}$ mbar). Consider a small change ΔT_g in the gas temperature and at the same time a change of pressure Δp from zero such that the wire temperature is unchanged.

Before the change

$$V^2/R_w = K_R(T_w^4 - T_g^4).$$

After the change in temperature and pressure

$$V^2/R_w = K_R[T_w^4 - (T_g + \Delta T_g)^4] + K_c[T_w - (T_g + \Delta T_g)]\Delta p.$$

K_R and K_c are the radiation and conduction constants respectively for the wire. From these equations

$$\frac{\Delta p}{\Delta T_g} \simeq \frac{4K_R T_g^3}{K_c[T_w - (T_g + \Delta T_g)]} \tag{2.23}$$

$\Delta p/\Delta T_g$ represents the change in pressure which is equivalent to a 1 °C change in ambient temperature, this expression being true for constant-voltage, constant-current and constant-temperature operation.

The curves plotted in Figure 2.11 have been calculated from eqn (2.23) for $K_R = 5.18 \times 10^{-12}$ W cm^{-2} T^{-4} and $K_c = 10^{-3}$ W cm^{-2} °C^{-1} mbar^{-1}. (The ratio K_R/K_c is independent of the length and cross-sectional area of the wire.)

2. *Effect of a change of supply voltage in the low-pressure region* ($p < 10^{-2}$ mbar). Assuming a small change in voltage ΔV, the pressure change required to keep the wire temperature constant can be calculated as follows.

As in the previous case

$$V^2/R_w = K_R(T_w^4 - T_g^4).$$

If V changes by ΔV and p by Δp for T_w and T_g to be unchanged

$$(V + \Delta V)^2/R_w = K_R(T_w^4 - T_g^4) + K_c(T_w - T_g)\Delta p.$$

From these equations

$$\frac{\Delta p}{\Delta V/V} \simeq \frac{2K_R(T_w^4 - T_g^4)}{K_c(T_w - T_g)}. \tag{2.24}$$

Again, this is an expression dependent only on T_w and T_g for a given surface condition.

Values of Δp for $\Delta V/V = 0.001$ have also been plotted in Figure 2.11 for $T_g = 100\,\text{K}$ and $300\,\text{K}$.

3. *Change of ambient temperature at high pressures.* The size of the error depends upon the method of bridge operation so that three separate calculations are necessary to obtain the full picture.

Consider first the constant-current bridge.

In general

$$I^2R_w = K_R(T_w^4 - T_g^4) + K_c p(T_w - T_g).$$

When $T_g \to T_g + \Delta T_g$ and $p \to p + \Delta p$ to keep the bridge balance $I^2R_w = K_R[T_w^4 - (T_g + \Delta T_g)^4] + K_c(p + \Delta p)[T_w - (T_g + \Delta T_g)]$.

Solving these equations gives

$$\frac{\Delta p}{\Delta T_g} = \frac{4K_R T_g^3 + pK_c}{(T_w - T_g)K_c}$$

$$= \frac{4T_g^3 K_R/K_c + p}{T_w - T_g} \tag{2.25}$$

and

$$p = \frac{K_R\left[(T_0^4 - T_g^4)\dfrac{R_w}{R_0} - (T_w^4 - T_g^4)\right]}{K_c(T_w - T_g)}. \tag{2.26}$$

(At zero pressure $I^2R_0 = K_R(T_0^4 - T_g^4)$ where T_0 and R_0 are the wire temperature and resistance values at $p = 0$.)

Equations (2.25) and (2.26) can be used to obtain the relation between p and $\Delta p/\Delta T_g$ for any given value of T_0. This has been worked out in Table 2.3 for $T_0 = 400\,\text{K}$ in order to illustrate the technique.

For the constant-voltage bridge, following the above procedure,

$$V^2/R_w = K_R(T_w^4 - T_g^4) + K_c p(T_w - T_g)$$

so that

$$\frac{\Delta p}{\Delta T} = \frac{4T_g^3 K_R/K_c + p}{T_w - T_g}. \tag{2.27}$$

$$p = \frac{K_R[(T_0^4 - T_g^4)R_0/R_w - (T_w^4 - T_g^4)]}{K_c(T_w - T_g)}. \tag{2.28}$$

Thus a relation between p and $\Delta p/\Delta T_g$ can be obtained as in Table 2.3.

Table 2.3 Apparent change in pressure due to 1 °C change in gauge-wall temperature ($\Delta p/\Delta T_g$) for the constant-current Wheatstone bridge operation.

T_w	R_0/R_w	$(T_0^4 - T_g^4)R_w/R_0 - (T_w^4 - T_g^4)$	p (mbar)	$4T_g^3 K_R/K_c$ 10^{-3}	$\Delta p/\Delta T_g$ (mbar × 10^3) °C^{-1}
400	1	0	0	44.4	0.57
380	1.07	35×10^8	23	44.4	1.01
360	1.15	67×10^8	60	44.4	1.95
340	1.25	89×10^8	118	44.4	4.38
320	1.36	105×10^8	282	44.4	17
310	1.43	111×10^8	594	44.4	65
305	1.47	114×10^8	1217	44.4	255

Wire temperature at zero pressure = 400 K.
Gauge-wall temperature $T_g = 300$ K.
Temperature coefficient of resistance $\alpha_0 = 0.005\,°C^{-1}$.

These expressions are only exactly true when the voltage across the wire is constant. They can be applied in practice when $R_1/R_w < 5$.

For the constant-temperature bridge, the above procedure gives

$$\frac{\Delta p}{\Delta T_g} = \frac{4T_g^3 K_R/K_c + p}{T_0 - T_g} \tag{2.29}$$

which can by itself give $\Delta p/\Delta T_g$ as a function of p since p is now the only variable.

The above treatment has been used to obtain the curves drawn in Figure 2.12.

References

1. Pirani, M. (1906) *Deutsche Phys. Ges. Verk.* **8**, 686.
2. Knudsen, M. (1911) *Ann. Phys. (Leipzig)* **34**, 593.
3. Roberts, J.K. (1943) *Heat and Thermodynamics*, 3rd edn., (Blackie), London.
4. Knudsen, M. (1950) *Kinetic Theory of Gases*, 3rd edn., Methuen, London.
5. Kennard, E.H. (1938) *Kinetic Theory of Gases*, McGraw-Hill, New York.
6. Von Smoluchowski, M. (1911) *Ann. Phys. (Leipzig)* **35**, 983.
7. Hunt, A.L. (1968) *J. Vac. Sci. Technol.* **5**, 61.
8. Von Ubisch, H. (1947) *Arkiv f. Mat. Astra. och Fysik* **34A**, 14.
9. Dickins, B.G. (1934) *Proc. Roy. Soc. A.* **143**, 517.
10. Weber, S. (1917) *Ann. Phys. (Leipzig)* **54**, 165, 325.
11. Thomas, L.B. and Olmer, F. (1943) *J. Amer. Chem. Soc.* **65**, 1036.
12. Leck, J.H. (1964) *Pressure Measurement in Vacuum Systems*, 2nd edn., Chapman & Hall, London, Appendix 2.1.
13. Roberts, J.K. (1930) *Proc. Roy. Soc. A.* **129**, 146; (1932) **135**, 192; (1933) **142**, 518.
14. Mann, W.B. (1934) *Proc. Roy. Soc. A.* **146**, 776; (1937) **158**, 397.
15. Veis, S. (1959) *Vacuum* **9**, 186.
16. Hale, C.F. (1911) *Trans. Amer. Elect. Chem. Soc.* **20**, 243.
17. Leck, J.H. and Martin, C.S. (1956) *J. Sci. Instrum.* **33**, 181.
18. Dunoyer, L. (1949) *Vide* **4** (20) 571; (21) 603; (22) 643.
19. Von Ubisch, H. (1948) *Arkiv f. Mat. Astro. och Fysik* **36A**, 4; (1948) *Nature (London)* **161**, 927.
20. Leck, J.H. (1952) *J. Sci. Instrum.* **29**, 258.

21. Ellett, A. and Zabel, R.M. (1931) *Phys. Rev.* **37**, 1024, 1112.
22. English, J., Fletcher, B. and Steckelmacher, W. (1965) *J. Sci. Instrum.* **42**, 77.
23. Leck, J.H. (1954) *J. Sci. Instrum.* **31**, 226.
24. McMillan, J.A. and Buch, T. (1957) *Rev. Sci. Instrum.* **28**, 881.
25. Johnson, J.B. (1956) *Rev. Sci. Instrum.* **27**, 303.
26. Flanick, A.P. and Ainsworth, J.E. (1961) *Rev. Sci. Instrum.* **32**, 356.
27. Heinje, L. and Vink, A.T. (1969) *Philips Tech. Rev.* **30**, 166.
28. Steckelmacher, W. and Fletcher, B. (1972) *J. Phys. E: Sci. Instr.* **5**, 405.
29. Steckelmacher, W. (1973) *Vacuum* **23**, 307.
30. Voege, W. (1906) *Phys. Z.* **7**, 498.
31. Hart, E.D. and Elkin, W.H. (1946) *J. Sci. Instrum.* **23**, 17.
32. Picard, R.G. and co-workers (1946) *Rev. Sci. Instrum.* **17**, 125.
33. Garrod, R.I. and Gross, K.A. (1948) *J. Sci. Instrum.* **25**, 378.
34. Kenty, C. and Reuter, F.W. (1947) *Rev. Sci. Instrum.* **18**, 918.
35. Webber, R.J. and Lane, C.T. (1946) *Rev. Sci. Instrum.* **17**, 308.
36. Dunlap, G.C. and Trump, J.G. (1937) *Rev. Sci. Instrum.* **8**, 37.
37. Benson, J.M. (1956) *Vac. Symp. Trans. Comm. Vac. Tech.* Pergamon, New York, 87.

3 Thermionic cathode ionization gauges

3.1 Positive ion production in a gas

In a gas at low pressure the number of positive ions produced by the passage of a stream of electrons is directly proportional to the molecular concentration. The linear relation between ionization and density holds from zero gas pressure up to the point at which ion formation is sufficient to alter effectively the current and energy of the electron stream. Since the gas pressure (at a constant temperature) is directly proportional to its density, the positive ion current produced by a steady electron current may be used as an indicator of pressure.

The number of ionizing collisions made by an electron in its passage through a gas has been shown by many workers to depend upon both the type of gas and the kinetic energy of the electron. Systematic measurements of ε (the number of positive ions produced per electron per unit distance travelled at unit pressure in a given gas) were made in the decade commencing in 1930. The work of Tate and Smith[1,2] is noteworthy. Using a carefully designed apparatus, they were able to measure ε for a number of gases at a pressure of about 1 mbar over the range of electron energy from 0 to 600 eV. Their results[2] are reproduced in Figure 3.1, showing ε plotted as a function of electron energy. In all measurements the temperature of the ionization chamber was

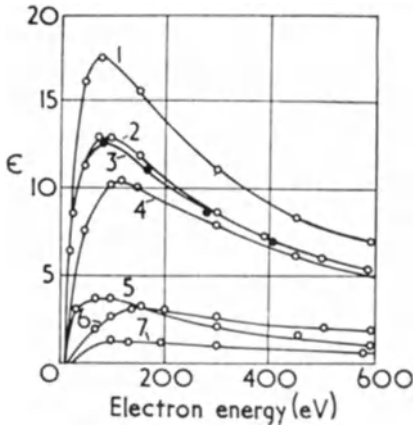


Figure 3.1 The value of ε, the number of positive ions produced per electron per centimetre path at a gas pressure of 1 mbar and a temperature of 0 °C as a function of the electron energy. (1) acetylene, (2) oxygen, (3) nitrogen, (4) argon, (5) hydrogen, (6) neon, and (7) helium.

measured carefully, so that all measurements of ε could be referenced to 0 °C. The results indicate two important characteristics of this electron impact phenomenon: (i) although absolute values are different, the efficiency of ionization passes through a maximum at about 100 eV electron energy for each gas (neon being the one exception), and (ii) the 'shape' of ε versus electron energy characteristics is to a first approximation the same for each gas (neon again being the notable exception).

3.2 The principle of the thermionic cathode ionization gauge

The three-electrode triode valve structure provides a simple and convenient means of producing an electron stream and separating it from the resulting ion current. In its simplest and most convenient form, a cylindrical system is used with a central 'straight wire' thermionic electron emitter. This hot wire is surrounded by two coaxial cylinders of about 15 and 30 mm diameter; the inner cylinder is an open wire mesh grid structure, the outer solid. With the grid held positive and the anode negative to the filament all electrons must travel to the grid. Many, however, first oscillate about the grid, forming an electron cloud distributed throughout the area shown shaded in the diagram (Figure 3.2). All the positive ions produced between grid and anode by this electron cloud travel down the potential gradient to the anode. For a constant electron current and fixed electrode potentials, the number of ions produced between the grid and anode – and therefore the anode current – is directly proportional to the gas density (i.e. pressure, at constant temperature).

The complex relation between the efficiency of ionization and the electron energy means that it is not practicable to calculate the sensitivity for the simple triode gauge, even after making simplifying assumptions about the electron

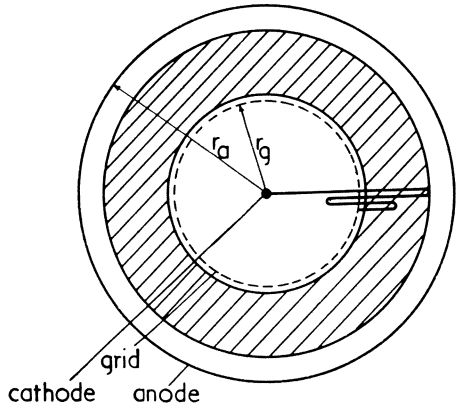


Figure 3.2 Thermionic-cathode triode ionization gauge showing a typical electron trajectory. The useful ionizing region is shown shaded. Grid–cathode voltage $\simeq 200$ V; anode–cathode voltage $\simeq -20$ V.

trajectories. It is more convenient to calibrate the gauge against some standard instrument, for example a spinning rotor gauge. Early descriptions of the ionization gauge, together with calibration characteristics, were published by Buckley[3] in 1916, by Misamichi So[4] in 1919 and by Dushman and Found[5] in 1921.

A typical set of characteristics for a modern gauge is reproduced in Figure 3.3. The anode (positive ion) current is shown as a function of (*a*) electron current, (*b*) grid voltage, (*c*) anode voltage and (*d*) pressure. As may be expected, the anode current is a complex function of the electrode potentials because both the electron trajectories and the ionizing efficiency depend upon these potentials. A minimum negative anode bias of about 5 V is required to prevent any electrons reaching the anode. Beyond this point the current falls steadily with increasing bias because the increased negative field reduces the electron penetration into the grid–anode space. With this gauge, satisfactory operation is obtained for $V_g = 200$ V and $V_a = -10$ V, giving a sensitivity of 12 $mbar^{-1}$ for nitrogen. The method used here to define sensitivity (S) follows

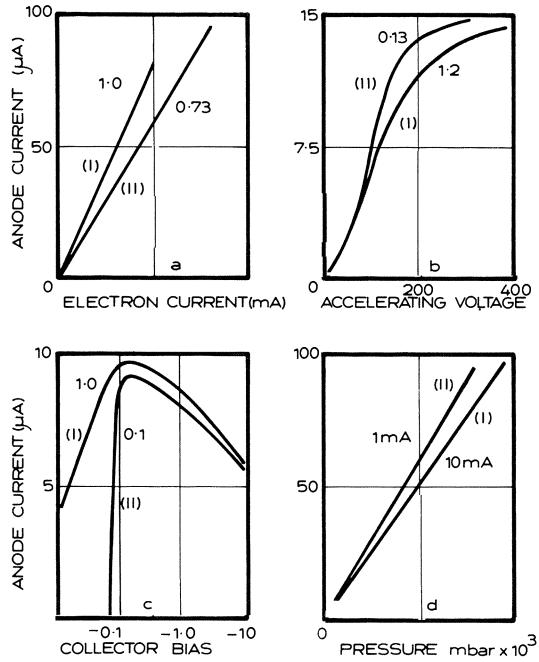


Figure 3.3 Characteristics of a typical cylindrical hot-cathode ionization gauge for nitrogen. The positive ion current is shown as a function of (*a*) electron current, (*b*) electron accelerating voltage, (*c*) positive ion collector voltage and (*d*) pressure. In (*a*), (*c*) and (*d*) the electron accelerating voltage was 170 V. In (*a*), (*b*) and (*d*) the collector negative bias was 10 V. In (*b*) and (*c*) the electron current was 1.0 mA. The values of pressure (mbar $\times 10^{+3}$) are marked on the curves for (*a*), (*b*) and (*c*). In curves (*d*), the electron current is shown on the curves. (For curves marked (I) the scale is given directly in microamperes; for curves marked (II) the ordinates must be multiplied by a factor 0.1.)

conventional practice. Thus

$$S = \frac{I^+}{I^-}\frac{1}{p} \quad \text{mbar}^{-1} \tag{3.1}$$

where I^+ is the positive ion current to the anode and I^- is the electron current to the grid. Alternatively, this can be written

$$p = \frac{I^+}{I^-}\cdot\frac{1}{S} \quad \text{mbar} \tag{3.2}$$

The triode valve can be used as an ionization gauge under different operating conditions from those shown in Figure 3.3. For example, if the anode is made the electron accelerator (approximately + 200 V), then the ion current to the grid (− 10 V approximately) is directly proportional to the pressure. This is not as attractive as the standard connection; the sensitivity is reduced by at least a factor of two and the operation tends to be unstable.

3.3 The relative sensitivity for different gases

The measurements of the probability of ionization (effectively ε as defined in the previous section) show that gauge sensitivity depends upon the gas composition. Furthermore, the relative sensitivity R – the ratio of the sensitivity for a given gas to the sensitivity for a standard gas (usually nitrogen) – is a function of the electron energy. (Argon is sometimes chosen as the reference gas instead of nitrogen.) For example, to take the simplified case of an electron beam of uniform energy, when the electron energy is increased from 100 to 200 eV, the relative ionization efficiency of helium/nitrogen goes up from 0.09 to 0.11 (Figure 3.1). However, as in practice gauges operate with similar potentials (the grid normally between the extremes of 120 and 200 volts), the differences are not expected to be too significant.

In early work, Dushman and Young[6] and also Reynolds[7] measured R for a number of gauges of the same design, operating under identical conditions. This work indicated differences which they could not attribute to experimental errors. With two gauges, Dushman and Young[6] found differences of the order of 10% and Reynolds[7] a maximum of 20% between seven almost identical gauges. These differences were of the same order of magnitude as the differences in absolute sensitivity. Reynolds compared the sensitivities of nitrogen and neon as the grid voltage was changed over a wide range. Over the range 140–180 V, R changed by only about 10%. This is likely to be a 'worst case', since the ionization characteristics for neon stand out as being quite different from those of the other common gases. Thus it seems likely that much of the scatter observed in the early work was due to experimental difficulties in measurement and gauge operation, rather than to any fundamental physical effect. Significant errors could arise in a number of ways from (i) uncertainty in

the pressure standard used for the comparison, (ii) modifications of the gas composition in the test cell due to reactions with the hot filament and electron beam, and (iii) reactions between gas molecules and the electrode surfaces of the gauge actually changing sensitivity in a manner difficult to reverse. (These physical and chemical phenomena are discussed in section 3.9.)

In 1973, Holanda[8] analysed the published data from many laboratories[9–25] in his search for a particular molecular property that could be related closely to relative sensitivity. For his analysis he referred back to the investigation made by Summers[26] in 1969, who had concluded that the best correlation was with ionization cross-section (δ) and, in particular, with the maximum value (δ_{max}) for each individual gas. Holanda confirmed cross-section to be a good guide, and in a comprehensive analysis concluded that for the normal triode and Bayard Alpert (BA) gauges (see section 3.4 for a description of the BA gauge), there is no significant difference in any of the three following techniques when used as a basis of comparison: (i) cross-section value for 100 eV electrons; (ii) δ_{max} for each particular gas; or (iii) δ calculated at $\frac{2}{3}$ of the potential of the electron collecting grid of each gauge. However, for the high-pressure gauges δ calculated at $\frac{2}{3}$ of the grid potential was found to be the best choice (see section 3.8).

In his analysis of published data in 1975, Nakao[27] concentrated upon a comparison of R with δ values measured for 100 eV electrons. Taking data for R from some twenty authors and for δ from eight authors (refs. 9, 10, 12–15, 20, 22, 28–38), he again found good correlation between R and δ within the limit set by the 'scatter' in the published data. As he pointed out, there is considerable practical interest in the heavy hydrocarbons, for which, when he carried out his analysis, there were very few recorded values of R, but considerable information available on cross-sections. Data for 32 of the more common of the 44 gases investigated by Nakao[27] are set out in Table 3.1. R_δ, the relative value of the molecular cross-section for 100-eV electrons, and the corresponding values of relative sensitivity R, are shown in columns 1 and 2 respectively. All values are referred to nitrogen.

In an attempt to find an alternative to ionization cross-section for correlation purposes, Bartmess and Georgiadis[39] measured and carefully analysed R for 74 different gases, mostly heavy hydrocarbons. Their motive was to overcome the problem of the comparative lack of cross-section data for those heavy hydrocarbons of interest to the chemical industry. Their results (where appropriate) are plotted in column 3 of Table 3.1, and present a useful comparison with the earlier data. As expected, Bartmess and Georgiadis[39] found that, where they were able to make a direct comparison, the best correlations were given first by cross-section and second by polarizability. Figure 3.4 (reproduced from their paper) shows the correlation between R and polarizability (values for polarizability were obtained following the technique of Savchik and Miller[40]).

Like a number of other workers, Bartmess and Georgiadis[39] returned to the

Table 3.1 Relative ionization cross-section R_δ (referred to 100 eV electrons) and ionization gauge relative sensitivity R measured in a number of laboratories.

	Nakao	Nakao	Bartmess	Young	Holanda	Summers
Gas	R_δ			Relative sensitivity R		
Helium	0.13	0.19	0.20	0.17	0.18	0.18
Neon	0.25	0.33	0.33		0.32	0.30
Argon	1.23	1.37	1.32	1.39	1.42	1.30
Krypton	1.84	1.91	1.92		1.94	1.90
Xenon	2.62	2.79	2.78		2.81	2.90
Hydrogen	0.38	0.44	0.44	0.38	0.41	0.46
Deuterium	0.41	0.40				
Ammonia	1.23		1.12			1.23
Water	1.03	1.25	0.97			1.10
Carbon monoxide	1.06	1.02		0.99	1.01	
Nitrogen	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
Oxygen	0.96	0.87	0.87	0.86	0.78	0.87
Hydrogen sulphide	2.03		1.82			
Hydrochloric acid	1.61					
Carbon dioxide	1.39	1.36	1.30	1.46	1.39	1.4
Nitrous oxide	1.30		1.20			
Sulphur hexafluoride	2.41	2.5			2.20	
Mercury	2.07	3.30				
Methane	1.63	1.49	1.62	1.60		1.4
Ethane	2.74	2.53	2.84	2.64		2.6
Propane	3.64	3.80	2.92			4.2
Butane	4.57	4.37	4.46			
Pentane	5.60		5.21			4.9
Hexane	6.77		5.90			
Heptane	7.72		6.94			
Iso-octane	8.18		7.21			
Acetylene	2.06	1.66				
Ethylene	2.27	2.14		2.08		
Propene	3.25	3.16	2.92			
Butene	3.82	3.60				
Pentene	4.81					
Hexene	6.49		5.81			
Benzene	5.19	5.18	4.29			

suggestion made in 1945 by Dushman and Young[6] to relate R to the total number of electrons in the molecule. They confirmed Nakao's report[27] of a good correlation for each of a number of groups of molecules, but no simple single relationship for all gases. Curves labelled (1) in Figure 3.5 show the magnitude of the differences between the various groups, the lower curve being for the noble gases, the upper for the alkanes. To illustrate the broad agreement between the different workers, corresponding curves by Nakao[27] (2), and the curve obtained by Young[37] for the heavy hydrocarbons (3) are also drawn.

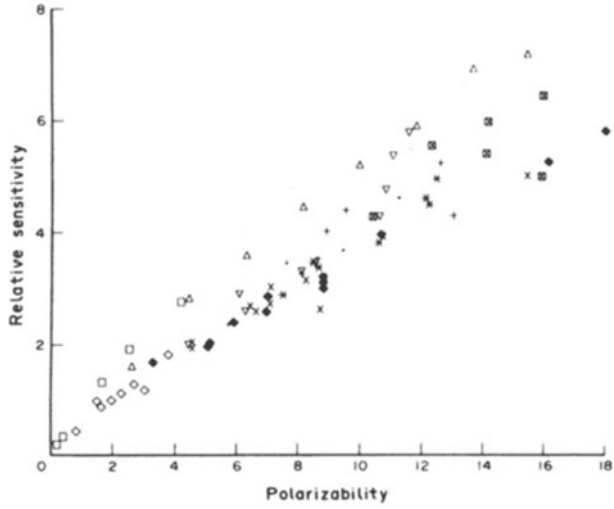


Figure 3.4 Relative sensitivity for ionization gauge shown as a function of polarizability. ◇, small inorganics; □, noble gases; △, alkanes; ▽, alkenes; ⊠, aromatics; ◆, alcohols; +, ethers; ×, carbonyl compounds, *, chlorocarbons; ●, thiols (Bartmess and Georgiodis[39]).

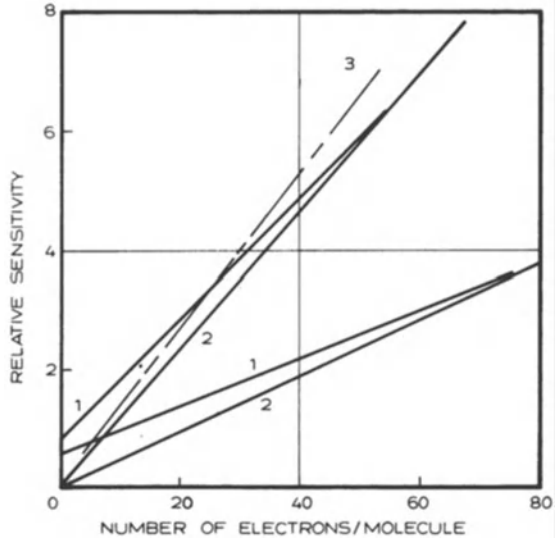


Figure 3.5 Sensitivity for the ionization gauge relative to nitrogen shown as a function of the numbers of electrons in the molecule. The upper curves (1) and (2) for the alkanes, the lower curves (1) and (2) for the noble gases and curve (3) for the heavy hydrocarbons.

For general vacuum use, the data presented in Table 3.1 can be applied with some confidence. The discrepancy in reported measurements is not greater than about $\pm$ 10% for the common gases, rising to a little above $\pm$ 20% for the less common gases, where measurement is more difficult and there are fewer results to allow comparison. Clearly where greater precision is required, gauges must be calibrated individually and under as near as possible the actual operating conditions of the system. For the noble gases, the spread of 10% for the more recent results probably represents the true differences caused by the different gauge geometries and operating potentials.

For gases where little or no data are available the recent work by Holanda[8] Nakao[27], and Bartmess and Georgiadis[39] has shown that reasonable approximations can be obtained by correlating with data on cross-section, polarizability and, provided proper caution is used, with the number of electrons in the molecule.

3.4 The measurement of low pressures

At first sight, there appears to be no lower limit to the pressure range of any of the gauges described in section 3.2, provided a sufficiently sensitive instrument is available to measure the ionization current. In practice, however, a definite lower limit is set by spurious anode currents due, in the main, to photoelectron emission from the anode. Electrons ejected from the anode travel to the positive grid and are indistinguishable from the positive ions arriving at the anode. All the experimental evidence obtained, for example by Apker[41], Anderson[42], and Nottingham[43], shows that even at a very low pressure ($< 10^{-10}$ mbar) the gauge never indicates a pressure below 10^{-8} mbar. It can be assumed, therefore, that for normal operating conditions the photocurrent introduces a zero error of about 10^{-8} mbar into the gauge reading. The exact magnitude of the error cannot be predicted accurately, so that measurement of absolute pressure is not possible below about 10^{-8} mbar with the conventional gauge.

The undesirable photocurrent must be caused by radiation from the following sources: (i) direct illumination from outside the apparatus, (ii) light from the cathode and (iii) X-rays emitted from the grid by the electron bombardment. Although (i) can be suppressed easily, it is more difficult to reduce the effect of (ii) and (iii).

Lander[44] was able to achieve partial success by using a small plane anode and by increasing the grid–anode spacing to almost 75 mm. In this gauge, although the anode collects virtually all the positive ions, the photocurrent is reduced by at least one order of magnitude because of the reduced angle subtended at grid and filament by the anode. Bayard and Alpert[45] made a much greater improvement by abandoning the conventional electrode assembly and adopting that shown in Figure 3.6. They made the grid *B* in the form of a helical coil, the hairpin filament *A* being placed outside the grid, and

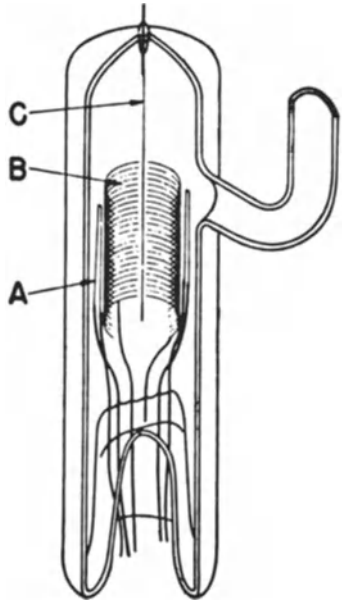


Figure 3.6 A Bayard–Alpert-type ionization gauge. *A*, cathode; *B*, electron accelerating grid; *C*, positive ion collector[45].

the anode – a fine wire *C* approximately 0.125 mm in diameter – placed along the axis of the cylinder. With the normal electrode potentials (approximately + 150 and − 20 V grid-filament and anode-filament potential respectively), an electron cloud forms between the grid and anode. All positive ions formed in this region must travel down the potential hill to the anode. The sensitivity is approximately the same as that of the standard gauge but, because of its small surface area, the photocurrent from the anode must be reduced by at least two orders of magnitude. A fairly high sensitivity can be expected, because the potential gradient is high in the space close to the anode. This ensures that the electrons travel almost to the anode with an approximate constant energy of about 100 eV, i.e. the energy for most efficient positive ion production. This is a distinct advantage over the conventional gauge where the potential, and therefore the electron energy, falls monotonically between grid and anode, so that efficient ion production cannot take place throughout the whole volume.

Bayard and Alpert[45] have made an interesting comparison of the conventional and the 'inverted' triode gauges by comparing the grid voltage–anode current characteristics. These are shown at different pressures in Figure 3.7. At a high pressure, where the current is predominantly due to ionization, the anode current increases only slowly with voltage above 200 V. (This is the characteristic shown in Figure 3.3.) The relationship is, however, quite different at low pressure, where the current continues to increase with grid

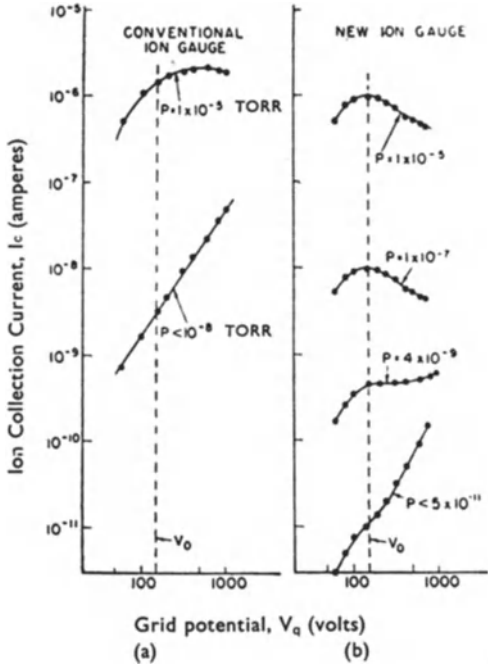


Figure 3.7 Characteristics of the Bayard–Alpert (BA) gauge (*b*) compared with those of the 'normal' gauge (*a*) at low pressure[45].

potential, the slope of the log curve being between 1.5 and 2.0. This characteristic can be explained by assuming the current to be due entirely to photo-electrons from the anode released by the soft X-rays from the grid. In the intermediate range of pressures, the characteristics correspond to a superposition of the 'residual' and 'ionization' curves. These results show a reduction by a factor of 100 in the photocurrent for the inverted gauge. The residual current does not predominate until a pressure of the order of 10^{-10} mbar is reached. Thus the gauge can be used to measure pressures down to 10^{-9} mbar, and even lower. For many applications it is permissible to simplify the construction shown in Figure 3.6 by taking the anode connection through the main 'pinch'. This, of course, reduces the electrical leakage path, but experience has shown that this is not important even for measurement of the lowest currents. Because of the practical advantages of this gauge, the simplicity of the construction, the ease of outgassing, and the practicability of fitting one or more 'spare' filaments, one can agree with Alpert when he stated '... its overall simplicity indicates that it may well have advantages not only in the special field of ultra-high vacuum, but also in the more general field of high vacuum'.

Steady progress in the design of Bayard–Alpert gauges has been made since 1950. Nottingham[46] pointed out that, in the basic design (Figure 3.6), ions could be lost through the open ends of the collector cylinder. He showed that end plates at grid potential (effectively closing off the cylinder) prevent this loss and consequently improve sensitivity. The effect of the electrical potential of the glass envelope upon the characteristics can be serious because in this gauge, as opposed to the earlier type with solid cylindrical anode, the discharge 'sees' the glass and will be influenced by its potential. To prevent this interference Nottingham[46] suggested an additional cylindrical grid to surround the electrode structure. Alternatively, in a number of designs a metal conducting layer, which can subsequently be 'earthed', is evaporated or sputtered on to the inner surface of the glass to ensure stable operation. Carter and Leck[47] pointed out the possible instabilities of the gauge when such precautions are not taken. They showed that the surface of the glass will stabilize in one or other of two equilibrium states, at approximately either cathode or anode potential. These represent equilibrium between electron and positive ion collection, and between the primary electrons to and the secondary electron emission from the surface, respectively (at the higher potential the electrons strike the glass with considerable energy). Fortunately the changes from one stable state to the other are relatively infrequent, but when they do occur they are quite sudden and give rise to considerable changes in gauge characteristics.

In order to take the limit of operation to the lowest practical pressure, Van Oostrom[48] has reduced the diameter of the ion collector to 4 μm, which must be near the limit of practical construction. Operation with a thin collector wire is difficult because (as Alpert[49] pointed out earlier) the sensitivity is greatly reduced with collector wire diameters below 100 μm. Van Oostrom[48] verified that the loss of sensitivity is caused by an excessive number of ions drifting axially out through the ends of the cylindrical structure. For very fine collector wires, the drift is exaggerated due to the 'orbiting effect' of the positive ions as they are drawn towards the collector. Because of the small gain in kinetic energy in the formation process and the effect of distortions in the radial field, the ions will have a small angular momentum with respect to the central axis. This gives them a rotary motion, and means that they will tend to move around rather than travel directly to the collector. Since angular momentum must be conserved, there is a minimum radius of curvature to this motion. With very fine wires, this may well be of the same order of magnitude as the collector radius. Thus ions will 'corkscrew' around the central axis, causing them to drift out axially from the electrode structure.

Van Oostrom[48] showed that this effect could be reduced if 'end caps' in the form of grids are added to the structure. He presented a comprehensive set of characteristics showing that by choosing carefully the electrode potentials, and by using end shields, the thin wire gauge can give a sensitivity of the order of $10\,\text{mbar}^{-1}$. He observed significant improvements in sensitivity by

increasing the collector negative bias considerably beyond the conventional values of 20–50 V. The increase in collection efficiency with increasing negative bias is obviously due to the increased radial field, and hence acceleration of positive ions to the central collector. For practical operation of this gauge, Van Oostrom[48] recommends operating potentials of +60 V on the grid and end shields, and a collector negative bias of 210 V (both with respect to the filament), to give a sensitivity of 10 $mbar^{-1}$ for nitrogen. Commercial gauges are now available from a number of companies which have X-ray limits of the order of 10^{-11} mbar, about an order of magnitude lower than the original BA design.

With the X-ray induced current largely suppressed, a further component to the background or 'leakage' current becomes evident. This arises from the ionization and subsequent release of gas molecules from the surface of the grid, and is the so-called 'electron stimulated desorption' effect. The important point is that these surface ions cannot easily be distinguished from those produced in the gas phase. Surface ion desorption is a very selective mechanism occurring with the desorption of oxygen (and oxygen-containing molecules such as water vapour) and, to a lesser extent, hydrogen and carbon monoxide from the grid surface.

Because it depends upon the cleanliness of the grid surface, the intensity of the surface ion current is a function of the past history of gauge operation. If, for example, a burst of oxygen gas is introduced to a 'clean' vacuum system, increasing the pressure from (say) 10^{-10} mbar to 10^{-6} for only one minute, then the reading of a Bayard–Alpert gauge will be spurious for many hours or even days. Typically the pressure indication will slowly fall from the low 10^{-9} mbar range to the true 10^{-10} mbar, with a time constant between one hour and one week depending upon the operation of the gauge. The explanation of this effect is that a significant fraction of the oxygen becomes tightly bound to the grid surface and is subsequently desorbed as positive ions by the electron bombardment.

In many vacuum systems, a slow drift in gauge reading is noted when the electron current is changed during gauge operation in a high vacuum. Ackley *et al.*[50], for example, observed an 'apparent' slow increase in pressure by as much as a factor of 10 when they reduced the electron emission in their gauge from 4 mA to 20 μA. This, as the authors pointed out, was due not to a change in pressure but simply a shift in the equilibrium oxygen contamination of the grid to a higher level on the reduction of the electron current. However, if the grid in their gauge was cleaned either by ohmic heating or by electron bombardment, the effect was eliminated completely. This indicates the advantage of operating with a relatively high electron current (i.e. in the range 2–10 mA) in order to ensure a constant and high electron bombardment cleaning action. It is interesting to note that the orbitron effect (a disadvantage from the point of view of the collection of gaseous ions) helps in suppressing 'surface' ion currents. The ions produced by the electron bombardment of the

grid are ejected with a kinetic energy in the range 5–10 eV. This results in a large angular momentum about the central collector and hence a low collection efficiency, even in a gauge where the grid is closed at the ends.

Figure 3.8 illustrates three versions of the hot cathode gauge. The first two, a conventional triode and a BA gauge with twin filaments, have glass envelopes; the third is a flange-mounted BA gauge for direct insertion into a metal vacuum system. These are typical of the many commercial gauges available for general use.

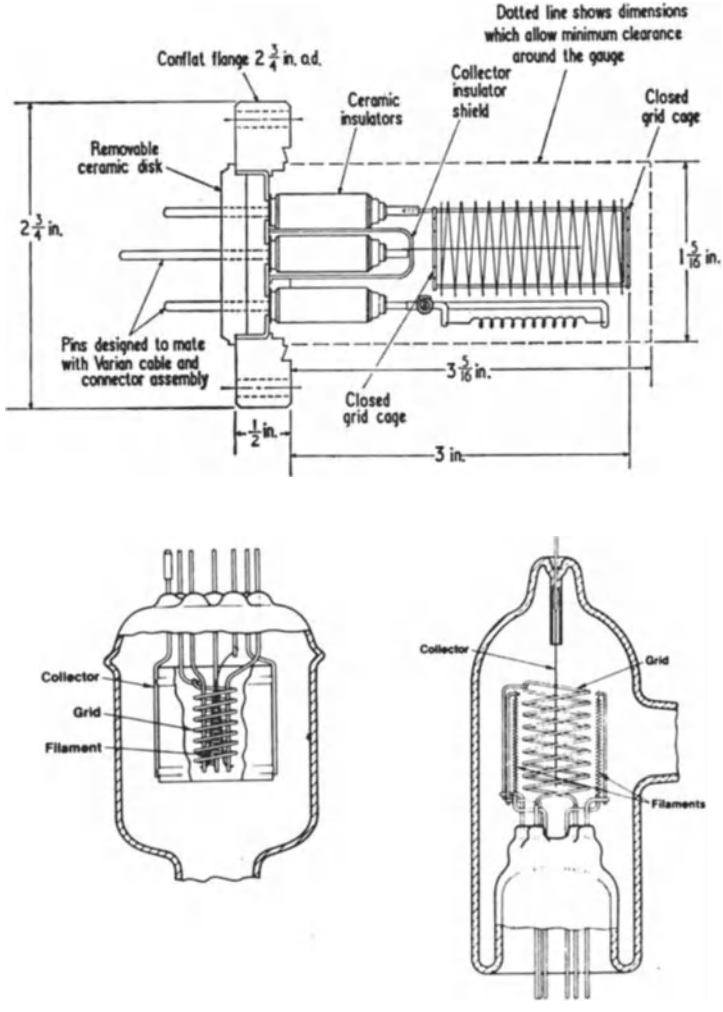


Figure 3.8 Outline drawings to illustrate the design of commercial thermionic cathode ionization gauges.

3.5 Extension of the range of the BA gauge to very low pressures

The discussion in the previous section indicated that the simple BA gauge could not be used for pressure measurement much below 10^{-10} mbar, this lower limit being set by the background current at the ion collector. By using the smallest practical diameter for the collector wire and end caps to the grid cage to improve collector efficiency, this background can be reduced under ideal operating conditions to give a zero offset of the order of 10^{-11} mbar. Thus, if quantitative measurements are required in the 10^{-11} mbar region and below, a more refined technique must be introduced to overcome the problems set by the X-ray generated current. To date, three alternative approaches have been made:

(i) To measure accurately the X-ray generated background current so that it can be taken into account in subsequent gauge operation
(ii) To suppress further the background current by a radical change in gauge geometry
(iii) To increase significantly the gauge sensitivity (by lengthening the electron paths) without changing the level of the background current.

The first two of these techniques have been developed and found reliable and relatively easy to use in laboratory applications. The greatest advance has been made by designing systems in which the background current is both reduced significantly and made relatively easy to measure. In this way, the lower limit of the standard BA gauge can be extended by at least one, and perhaps two, orders of magnitude. Unfortunately, developments of the third technique have been disappointing because (although it has been shown to have potential) reliable operation has not yet been demonstrated. Probably the fundamental disadvantage of the ideas so far developed is that, in order to gain a significant increase in sensitivity, an instability has been introduced. This makes operation inaccurate and difficult to predict. Thus, although some elegant designs have been proposed, there has been no widespread commercial development.

The different principles involved can easily be seen from the following descriptions of the three 'suppression' techniques.

The first technique evolved in 1960 when Redhead[51] suggested ion current modulation as a simple means for measuring the residual current, and hence extending the gauge operation to lower pressures. He pointed out that, with an extra electrode in the grid space, it is possible to change the gauge sensitivity without altering the X-ray induced current. This allows the two components of collector current to be separated. Adequate modulation can be achieved by fitting an extra wire inside and close to the grid and parallel to the ion collector. With this wire at grid potential, there is little or no effect on the gauge operation, but when its potential is lowered by about 100 V it seriously distorts the ion trajectories, with an inevitable loss of sensitivity. The change in

potential does not, however, have any significant effect on the residual current. At a given pressure, the total ion collector current (I_1 or I_2) can be expressed in terms of the components due to 'X-ray' and the 'positive ion' generations by the simple equations (3.3) and (3.4).

With the modulator at grid potential:

$$I_1 = I_R + I_P \tag{3.3}$$

I_R and I_P are the components due to 'X-rays' and 'positive ions' respectively.

With the modulator at low potential, the true ion current falls from I_P to $m I_P$ thus:

$$I_2 = I_R + mI_P \tag{3.4}$$

Therefore

$$I_P = \frac{I_1 - I_2}{1 - m} \tag{3.5}$$

and

$$I_R = \frac{I_2 - mI_1}{1 - m} \tag{3.6}$$

The constant m can be determined by observing the change in ion current which occurs when the modulator potential is switched at a pressure sufficiently high for I_R to be neglected. Values of m are usually in the range 0.4–0.6. To make a measurement at low pressure, both I_1 and I_2 must be determined and eqn (3.5) applied. In practical operation, modulation is usually effected by manually switching the grid potentials; however, ac techniques have been developed by a number of workers.

Appelt[52], and later Hobson[53], have indicated the limited value of this modulation technique by pointing out that the residual current I_R is affected to a certain extent by the modulation. Equation (3.4) should for greater accuracy be written as

$$I_2 = mI_P + \gamma I_R \tag{3.7}$$

giving

$$I_P = \frac{\gamma I_1 - I_2}{\gamma - m} \tag{3.8}$$

Making the simplifying assumption of γ equal to unity limits the correction that can be made for the residual current to about an order of magnitude; this takes the lower limit for ionization gauge operation into the 10^{-11} to 10^{-12} mbar range. Hobson[53] made estimates of the factor γ when working at very low pressure in order to extend the measuring range to below 10^{-12} mbar. He expressed the view that the factor γ could only be obtained by calibration at

extremely low pressure and that, for his gauge, assuming $\gamma = 1$ would lead to errors of the order of 3×10^{-12} mbar when applying the modulation technique.

Lange and Singleton[54] found that the change in electron current distribution between the grid and modulator when switching the modulator potential caused bursts of gas to be emitted from the gauge. They also found that an acceptable 'depth' of modulation could still be achieved when the voltage swing from grid potential was limited to 20 or 30 volts. This reduced the electron current redistribution and, hence, the pressure bursts. Poulter[55] observed similar pressure transients due to gas bursts, and suggested, as an alternative to the Lange/Singleton technique, a swing in modulator potential from ground to 20 or 30 volts below grid potential. For his gauges this gave the best compromise between maximum 'depth' of modulation and minimum disturbance to the electron distribution. Poulter[55] also noted that the constant m was, to a small extent, dependent upon both electron current and gas composition. It changed by 15% when the electron current increased from 1 mA to 10 mA, and by 12% when the gas composition changed from argon to neon. These changes of m are almost certainly due to the differing pattern of ion formation in the grid collector space.

Edwards and Lanni[56] have described a method of modulating the ion current which does not require an additional electrode. By cutting short the central collector of a commercial BA gauge they made ion current dependent upon collector potential. Modulation could now be effected by switching the filament and grid potentials with respect to ground (i.e. collector potential) while keeping the potential between grid and filament constant. Subsequently, Chen *et al.*[57] showed that this technique could be applied to the axial emission gauge they had described earlier. They claimed that with this technique the residual current is constant during the modulation process (i.e. $\gamma = 1$).

The second technique involves using opaque screens to shield the collector from the grid. To this end, Watanabe[58] has carried out experiments with a geometry different from the standard BA. The details of his design are clear from the drawings in Figures 3.9 and 3.10. He used a collector in the form of a needle point, extremely well shielded by a cylindrical sleeve held at grid potential. The design was based around a spherical construction in order to improve ion collection efficiency at the very fine tip. Watanabe[58] also introduced a conventional modulator electrode which proved very effective, having a factor $m = 0.05$ (see eqn 3.4). He constructed a number of gauges to the new design in order to find the optimum dimensions, which are those shown in Figure 3.10. With this gauge he measured a sensitivity of 40 mbar^{-1} and an X-ray background current equivalent to a pressure of 2.5×10^{-13} mbar (nitrogen equivalent).

A more radical approach for reducing the background X-ray effect is to remove the ion collector from direct line of sight of the filament. A simple lens must now be introduced between the grid and collector to draw out the ions to

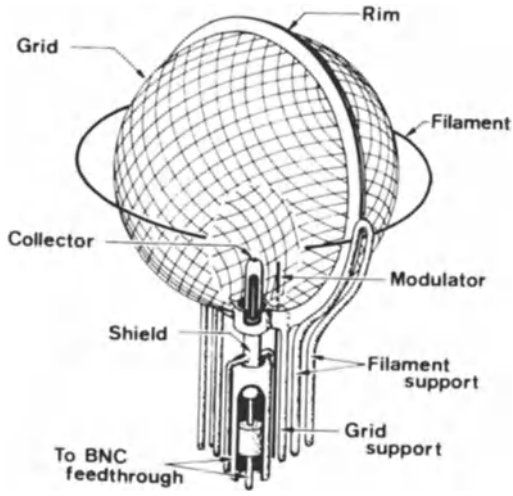


Figure 3.9 Cutaway drawing of the small point collector gauge designed and constructed by Watanabe[58].

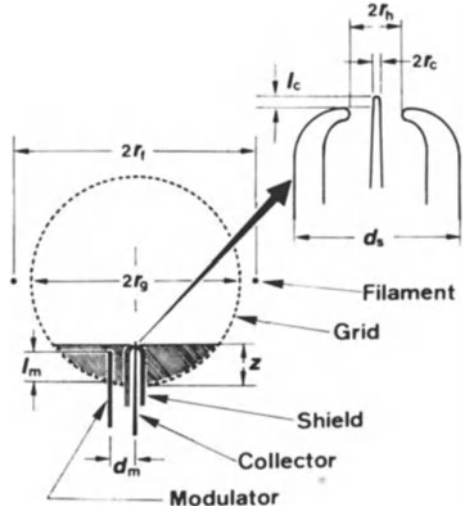


Figure 3.10 Schematic diagram of the Watanabe[58] gauge with details of the collector assembly. $r_g = 13$ mm; $r_c = 0.015$ mm; $l_c = 0.05$ mm; $d_s = 2.0$ mm; $l_m = 2.0$ mm;

maintain a reasonable sensitivity. The principle of this so-called 'buried collector' or 'extractor' gauge is shown in Figure 3.11[59]. The ions formed inside the closed grid space are drawn out from one end of the cylindrical structure to the very small ion collector, the potentials of the principal electrodes being similar to those in the conventional gauge. With this technique it is possible to reduce the X-ray background to the order of 10^{-12} mbar. An extractor gauge

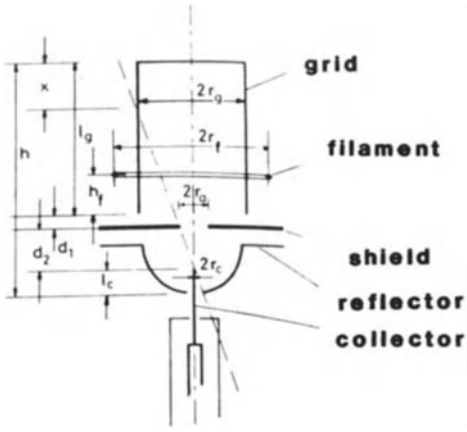


Figure 3.11 Schematic drawing of the miniature extractor gauge designed by Blechschmidt[59] with an X-ray limit $\sim 10^{-11}$ mbar. $r_g = 4.5$ mm; $h = 18.9$ mm; $r_c = 0.025$ mm; $r_a = 1.125$ mm; $r_f = 13$ mm.

of this type, manufactured and marketed by Leybold Ltd, is shown in Figure 3.12. This gauge is of a simple and rugged construction with a linear measuring range from 10^{-12} to 3×10^{-4} mbar. It is mounted on a standard stainless steel flange of 70 mm (2.75 in) overall diameter, and is typical of a number of gauges of this type which are suitable for general UHV operation.

Probably the most sophisticated gauge of the extractor type is that designed by Pittaway[60,61] in 1974, who combined the principles of extraction and modulation in one instrument. His work was based upon an extensive computer study of the electron and ion paths in the extractor gauge. The principal features of his design are shown in Figure 3.13. The grid cylinder is formed from a fine wire mesh with a high optical transmission. This allows very little field penetration into the central ionization region. The cloud of oscillating electrons forms, by space charge, a shallow potential well from which the positive ions cannot escape. They must, however, in their random motion travel to the base of the cylinder, where they are drawn out by the weak penetration field of the extractor. (There is a simple analogy with ball-bearings moving gently around in a shallow, frictionless saucer with a lip at one point. The balls move in stable orbits, until their random motion takes them to the lip, when they fall gently from the trap.) By careful design of the extractor system, Pittaway ensured that virtually all the ions reached the collector, thus giving a sensitivity similar to that of the conventional BA gauge. With the very small ion collector the X-ray induced current is extremely low, giving a background current below 10^{-12} mbar equivalent pressure. (Note the glass bead in Figure 3.13, this being important to shield the collector lead from reflected X-rays.)

In Pittaway's design[61] the extractor and modulator techniques can be combined to exceedingly good effect. A small wire tip, penetrating only a few

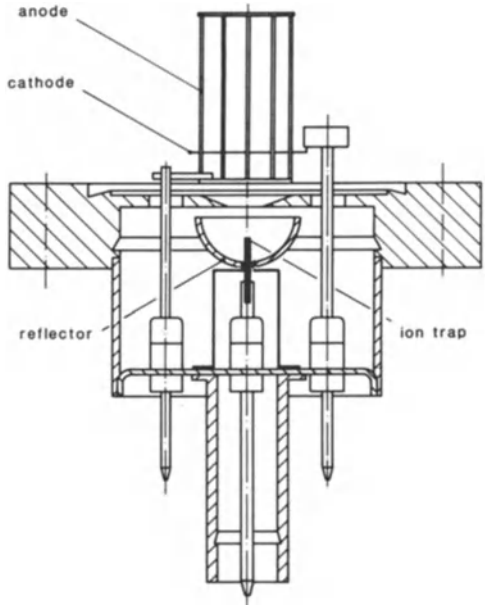


Figure 3.12 Extractor gauge manufactured by Leybold Ltd for pressure measurement down to 10^{-12} mbar.

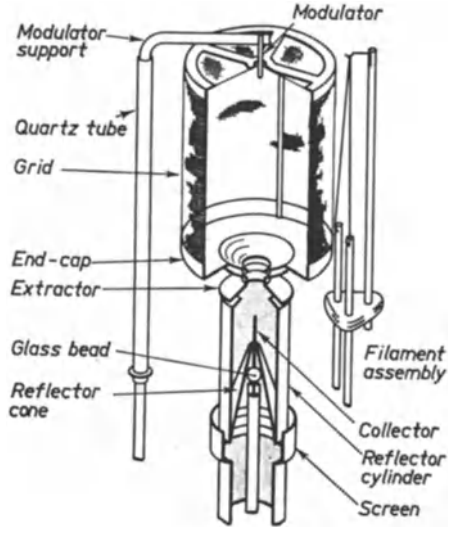


Figure 3.13 A cutaway drawing of the extractor gauge designed by Pittaway[60], showing the positioning of the modulator, external extractor, modified reflector and collector electrodes.

millimetres into the grid cylinder, is adequate to modulate the gauge. Changes in the wire potential by only a few volts are sufficient to reduce sensitivity by more than 95%, i.e. $m < 0.05$ (eqn 3.4). Pittaway[61], and later Watanabe[58], combined the Bayard–Alpert technique (described in section 3.4 above) with the modulator operation to differentiate the X-ray background from the true ion current signals. They achieved this by comparing the collector current as a function of grid to filament potential with and without modulation. The power of this approach can be seen from the curves in Figure 3.14, which have been taken from the publications of Bayard and Alpert[45], Pittaway[61] and Watanabe[58]. In Figure 3.14*a*, curve 1 follows the typical ion current characteristic and is similar to those shown in Figure 3.3*c*, while for curve 3, taken at the lowest pressure, the current is governed almost entirely by the soft X-rays. At the intermediate pressure, curve 2, the collector current is the sum of the 'ionization' and 'residual X-ray' components. Curves 1 in Figures 3.14*b*, *c* are the corresponding 'intermediate' characteristics for the Pittaway and Watanabe gauges. In these gauges the residual currents have been measured directly using the modulator technique (expressed in mathematical terms by eqn 3.6). The data are shown in curves 2 of Figures 3.14*b*, *c*. When operating with optimum grid to filament potentials, the residual currents are approximately 10^{-15} A and 10^{-14} A for the Pittaway and Watanabe gauges respectively. These currents correspond to pressures of 10^{-13} and 2.5×10^{-13} mbar. Thus under ideal operating conditions it is possible with these gauges to make measurements down to, and possibly below, 10^{-13} mbar.

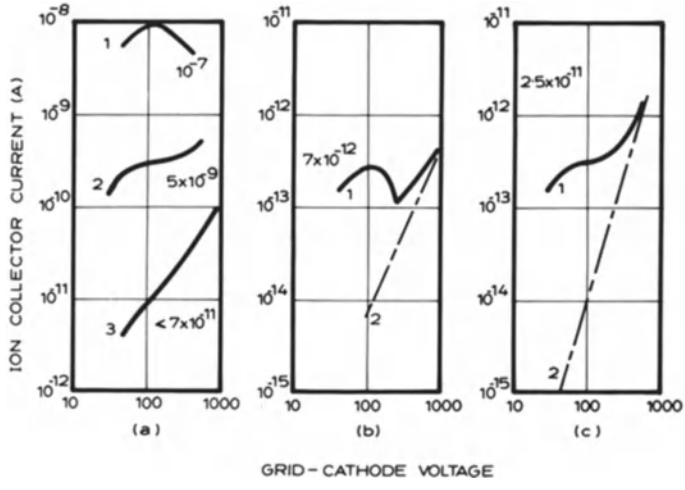


Figure 3.14 Ion collector current shown as a function of grid-filament potential difference over the range 10–1000 V for three ionization gauges: (*a*) Bayard–Alpert[45] (reproducing Figure 3.7*b*); (*b*) Watanabe[58] and (*c*) Pittaway[60]. The pressure (mbar) at which the measurements were made is marked on each curve. The dotted lines on (*b*) and (*c*) are the values of the residual current measured using the modulator technique.

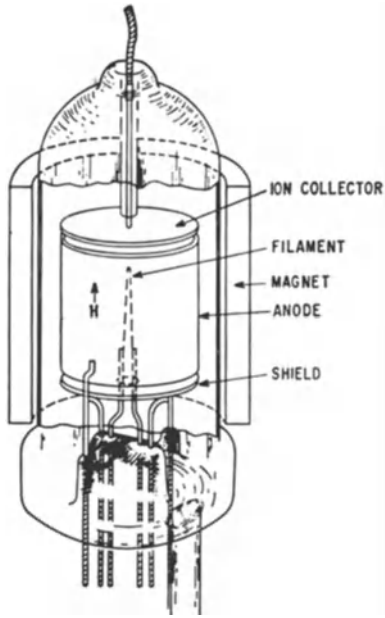


Figure 3.15 Thermionic cathode magnetron ionization gauge designed by Lafferty[62].

For the third technique, the simplest and most obvious means of increasing electron path length is to operate a triode structure inside a magnetic field, arranging the geometry so that the electrons follow circular paths and are, therefore, prevented from travelling directly to the electron collector. The gauge described by Lafferty[62] is one of the best examples of the application of this principle. Lafferty[62] adopted the diode magnetron principle by placing the electron-emitting filament along the axis of the cylindrical electron-collecting anode. The geometry of his gauge is illustrated in Figure 3.15. In these gauges an axial magnetic field prevents the electrons from travelling directly between filament and anode, forcing them instead to follow essentially circular paths travelling initially towards the cylinder but quickly returning to the region of the filament. Electrons make about one thousand excursions before finally reaching the anode. Ions produced from the gas phase are attracted axially outwards to the ion collector, which is held at a few volts negative to the cathode. The sensitivity is found to be two or three orders of magnitude higher than that of the conventional ionization gauge.

The inherent efficiency, expressed in terms of the number of ions produced for each electron emitted from the cathode, is two or three orders of magnitude greater than in the conventional ionization gauges. Hence a corresponding improvement in the ratio of the true to the background signal can be expected. A practical operating disadvantage is the need to keep the electron current

very low, certainly less than 1 μA, in order to ensure that the electron path lengths are not reduced by space charge distortions. Sensitivity increases slowly with increasing electron current but, unfortunately, the background current increases at a faster rate. This limits the actual sensitivity of the gauge to the order of 10^{-2} A.mbar^{-1}. Although a number of workers have (like Lafferty) considered the possibility of replacing the simple plate collector with an electron multiplier, because of the resulting complication, together with the uncertainty of multiplier gain, this idea has not been developed. Davis[63], with a gauge similar to that shown in Figure 3.15, has demonstrated linear operation down to about 10^{-13} mbar and estimated the background signal to be equivalent to a pressure of 10^{-14} mbar. In his work he noted some instabilities in gauge operation; in particular, some dependence upon small changes in magnet position and small negative currents to the collector unless a large negative voltage was applied to the electrode. Developments have been made in a number of laboratories with the aim of better understanding and improving the performance of this magnetron gauge. Visser[64], for example, introduced a ring suppressor at a large negative voltage, placed immediately in front of the ion collector, to return to the collector all electrons emitted by X-ray bombardment. By removing the magnet and measuring the background current with and without the suppressor, with a large electron current and at a very low pressure, Visser estimated that this reduces the X-ray generated current effectively to zero. This allows a higher electron current to be used and hence a higher sensitivity of the order of 0.1 A.mbar^{-1} with an electron current of 1 μA. The modifications made by Chen *et al.*[65] can be seen in Figure 3.16.

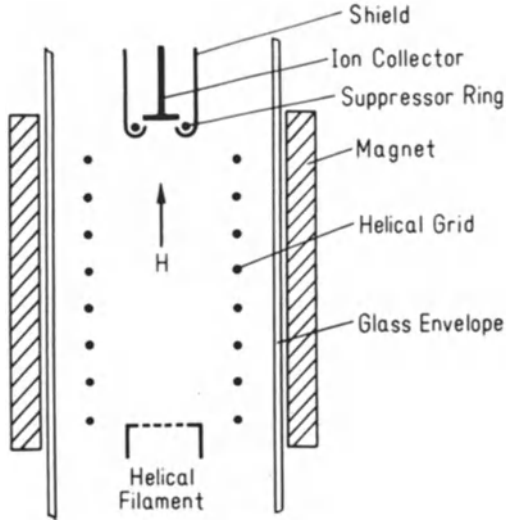


Figure 3.16 Schematic diagram of the axial emission magnetron suppressor gauge designed by Chen *et al.*[65].

They have adopted the suppressor ring, moved the filament out of the ionization chamber and made the anode (electron collector) into an easily outgassable grid. They also find the suppressor to be completely effective in eliminating background X-ray-induced current. With potentials of 200 V at the cathode, 500 V at the anode (electron collector) and − 500 V at the screen, all with respect to the ion collector, and a magnetic field of 0.03 tesla, they measured a sensitivity 0.065 A.mbar^{-1}. Thus if the smallest ion current that can be measured is 10^{-15} A, this sets the low pressure limit to the gauge to about 2×10^{-14} mbar.

Gabor[66], in proposing the orbitron principle, demonstrated that a very long path length can be achieved without a magnetic field. He, and later Herb and co-workers[67–69], showed that electrons injected into a cylindrical system (in which a fine wire down the central axis is held positive with respect to the outer cylinder) oscillate indefinitely, provided that their angular momentum at entry is greater than a certain critical value. The principle of operation is shown in Figure 3.17. Because the electric field is directed radially inwards, the initial angular momentum of the electrons must be conserved, hence they are forced

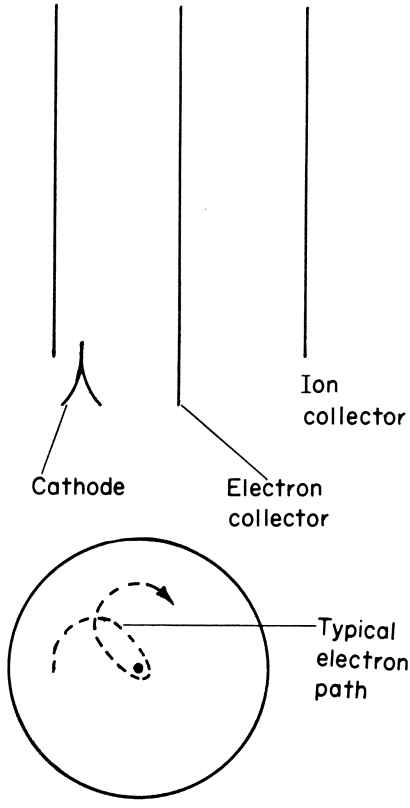


Figure 3.17 Schematic diagram of the orbitron gauge showing a typical electron path.

to circulate around the wire anode. Usually the electrons are injected either from a small filament mounted inside the cylindrical space, or from an electron gun external to the cylinder. By varying the potentials of the electron source, it is relatively easy to ensure that the electrons have the required angular momentum. As with the Lafferty gauge[62], sensitivities three orders of magnitude higher than those in conventional gauges can be obtained. Thus, there is a potential available for an improvement in the signal-to-noise ratio as compared with the BA gauge. It is difficult to estimate the actual improvement that has been achieved in practice. To date, no clear limit to the orbitron has been measured, although Gosselin *et al.*[70] have reported a rather disappointing measurement, suggesting a background current equivalent to a pressure of about 5×10^{-11} mbar. They could not improve this limit by modifying the geometry to suppress secondary electron emission from the ion collector. Other workers, notably Fitch and associates[71,72], have demonstrated that gauges of this type can be constructed with quite different gauge geometries (in at least one investigation a field emission cold cathode was used to provide the electron current). Unfortunately, the commercial viability of these gauges has not yet been demonstrated.

In comparing all the various types of hot cathode gauge from the point of view of sensitivity to background current, a number of fairly complex factors must be taken into account. For example, in comparing the X-ray limits of BA and orbitron gauges, it is not sufficient to equate their relative sensitivity with the area of the ion collector surfaces. Equating in this way suggests comparable X-ray limits, as the orbitron gauge has the greater sensitivity by a factor of 1000, but an ion collector surface area greater by about the same factor. It must be remembered, however, that in the orbitron some photoelectrons will have sufficient angular momentum to permit them to travel round the central anode wire and return to the ion collector surface. Thus, the orbitron may gain an order of magnitude advantage over the BA gauge on this count. On the other hand the BA gauge is (fortunately) a very inefficient device from the point of view of trapping the 'surface ions' at the central collector. These ions, formed by the electron bombardment of the surface adsorbed layer of oxygen and hydrogen on the grid, are generated with up to 8 eV energy, and many have sufficient angular momentum to orbit the central wire and return through the grid wires to the outer shield or glass envelope. In fact, in this gauge only about 1% of the surface ions reach the central collector. It is interesting to note that, without this orbitron effect, the BA gauge would be less effective in its application to low-pressure measurements.

3.6 The precision to which measurements can be made with the hot cathode gauge

This section will concentrate on the two factors which are of paramount importance in any consideration of the accuracy and reproducibility of

measurement. The first of these is the 'spread' of characteristics, in particular, sensitivity for a group of commercial gauges which are, so far as manufacturing technology will allow, of identical construction. The second is the changes that occur in the characteristics of a particular gauge during its operating life. This section will not be concerned with the errors in measurement which are due to instabilities in the background signal (for example X-ray generated residual currents), nor will account be taken of the chemical and physical reactions in the gauge which can materially alter the vacuum environment.

Because of the nature of the problem, it is not possible to give precise information on either spread in characteristics in a group or the drift in a particular gauge. Nevertheless, a definite pattern has emerged from the experimental evidence, in particular from work at various National Standards laboratories, for example from the NBS in Washington and the NPL in London. Results from these laboratories confirm the general qualitative pattern that has been built up from observations by many workers.

Tilford and his colleagues[73–75] have reported on a thorough investigation carried out at the NBS on various types of BA and conventional triode gauges. The results shown in Figure 3.18[75] and the following notes provide a broad summary of their work. The spread in sensitivity for each of a number of groups is given, in all cases measurements having been made for relatively new gauges typically with only 200 hours' operating life. The conventional triode gauges were of a relatively old design with operating parameters not compatible with modern commercial controllers, therefore no direct com-

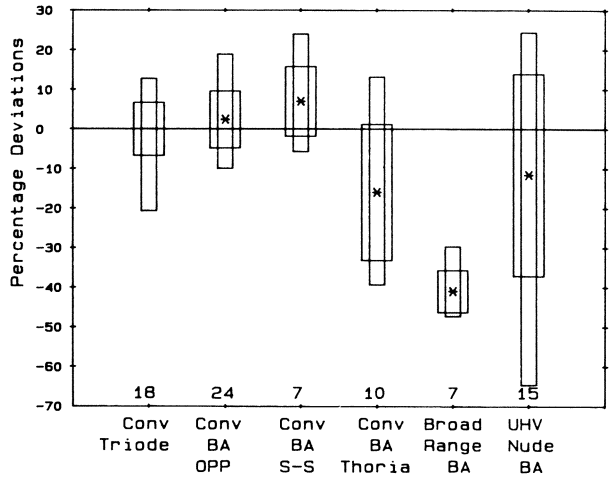


Figure 3.18 Average offset from specified nitrogen sensitivity, standard deviation of sensitivities about the mean, and range of sensitivities for six different gauge types. The gauge types are designated at the bottom of the figure and explained in the text. The numbers at the bottom indicate the number of filaments tested. The mean sensitivity offset is indicated by (*), ± one standard deviation by the wider box, and the range of sensitivities measured by the narrower box[75].

parison could be made with manufacturers' published characteristics. The range of sensitivities was dominated by three gauges that had sensitivities significantly outside the range of the other fifteen. Day-to-day instabilities with these gauges (which had tungsten filaments) were of the order of 2%. It is interesting to note that six other gauges with thoria-coated filaments (not quoted in Figure 3.18) had comparable sensitivities, but day-to-day instabilities a factor of 2 or 3 larger. The 'CONV BA OPP' gauges were of conventional design (the grids were open at the ends), having two filaments 180° apart; the 'CONV BA SS' gauges were of similar design with the filaments side by side. The 'BROAD RANGE' gauges were specially designed to extend the upper limit of the range to beyond 10^{-3} mbar. Tilford[75] noted that sensitivity values for the 24 filaments of the 12 'opposed filament gauges' showed a degree of uniformity better than any other type tested. A systematic difference in sensitivity between the two filaments was found in the side-by-side gauges, which was believed to be due to the field differences caused by an asymmetric filament support structure. The thoria-coated filament gauges had a significantly greater variation in sensitivity. (Some gauges showed significant high pressure non-linearities extending down to 10^{-4} mbar.) The instabilities were believed to be due to a changing thermal contact between the filament coating and the substrate, with consequent changes in emission characteristics; a problem which may not be present in other gauges.

Poulter and Sutton[76] at the NPL found a standard deviation of 5% for fourteen conventional triode gauges and 14% for five BA gauges (all gauges had tungsten filaments). An investigation at CERN in 1972[77] gave a similar figure of 11% as the standard deviation for BA gauges, although later in 1977 a much smaller spread in sensitivity (standard deviation 3.3%) was found for a batch of 300 specially designed gauges[78]. These latter gauges all had grids closed at both ends. Redhead[79] observed a spread of about $\pm 15\%$ in the sensitivities about the mean for a batch of 'nominally identical' gauges constructed at the NRC in Ottawa.

Evidence from the various laboratories that the spread in sensitivity from a group of nominally identical gauges is smaller for the conventional triode than the BA (or at worst equal to that of the BA) is not unreasonable, in view of the difference in construction of the two types of gauge. The obvious important difference between the two is the positioning of the electron-emitting filament, which must give rise to quite different electron trajectories between filament and grid. In the conventional triode the filament is symmetrically placed at the centre of the cylindrical electrode system, in effect at the bottom of a potential well. Electrons emitted from this filament are drawn outwards to the grid along radial paths (as indicated in Figure 3.2). Small differences in either the position of the filament or in the field near its surface will not change these paths to any significant extent. The situation is quite different for the 'inverted' or BA structure, where the filament lies outside the grid and, therefore not in a potential well or at the centre of a symmetrical system.

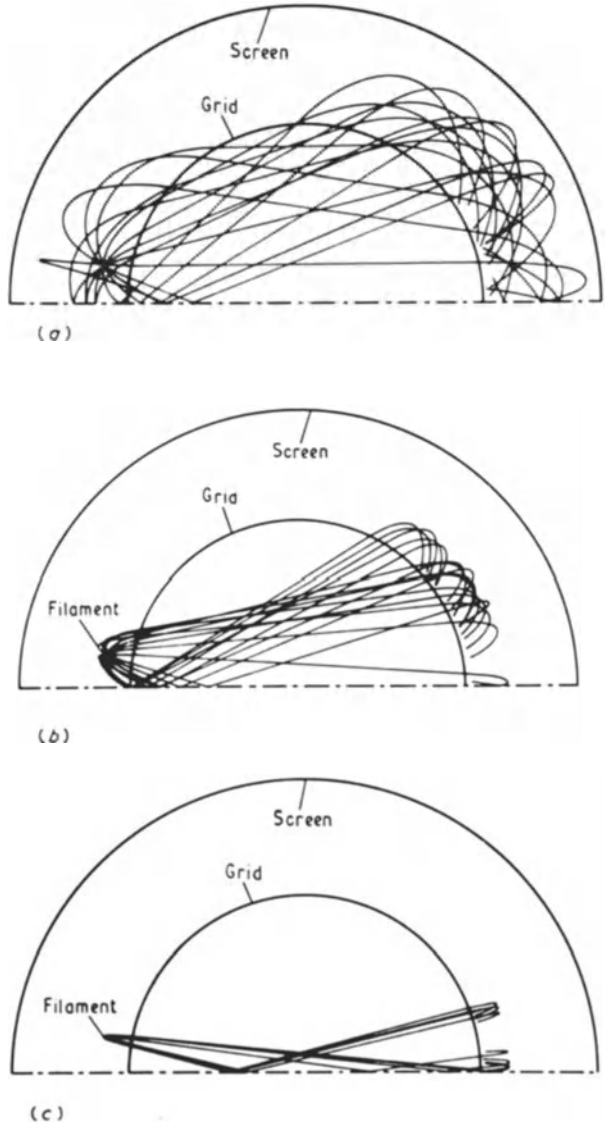


Figure 3.19 Electron trajectories across the grid region of the BA gauge and their return paths in the grid-to-screen field for low, mean and high sensitivity. (*a*) $V_{gs} = 120$ V; $V_{fs} = 20$ V; (*b*) $V_{gs} = 250$ V; $V_{fs} = 150$ V; (*c*) $V_{gs} = 370$ V, $V_{fs} = 270$ V (after Pittaway[80]).

Redhead[79] used a simple membrane model to illustrate graphically the significance of the precise position and potential of the filament. His model showed clearly how the electron paths inside the grid depended upon the 'launch angle' of the electrons from the filament surface. Redhead[79] calculated that the electron path lengths inside the grid decreased by a factor of two when the launch angle changed from 45° to 90°. Later, Pittaway,[80] using a sophisticated computer analysis, was able to provide a more detailed picture, which again emphasized the importance of the filament. He showed the electron path length through the grid to be a maximum (and hence sensitivity a maximum) when the filament is placed 'in such a position that it does not distort the radial field between grid and screen cylinders'. The electron trajectories calculated by Pittaway[80] for three different operating conditions are reproduced in Figure 3.19. For (*c*), optimum conditions have been chosen (i.e. to give maximum sensitivity), while for (*b*) and (*a*) respectively, an increasing field distortion in the region of the filament has been built in. (For convenience, trajectories which would pass into the lower half of the cross-section are shown as though reflected from the axis of symmetry.) Arnold and Bills[81] also carried out simple experiments to show that changes in the position of the filament could have significant effects upon the electron trajectories, and hence sensitivity. Their experiments highlighted the importance of relative movement between the filament and any 'field-distorting' grid supports. They pointed out that their results showed that any changes in the electron emission density pattern along the length of the filament could cause the electron paths, and hence sensitivity, to vary widely.

The emphasis placed upon the field in the region of the filament, by both the membrane and computer analysis, should be borne in mind when designing and operating these BA gauges. For example, it probably provides the explanation for the observation by Tilford[75] that the spread in characteristics is greater for 'nude' than for 'tubulated' gauges. This indicates that it is good practice in nude gauges to ensure 'positive screening' of the electrode assembly to eliminate, as far as possible, field variation in the filament region.

In view of the rather ill-defined electron trajectories in the BA gauge, a number of design studies have been carried out, and prototype gauges constructed in attempts to obtain more precisely defined performance. Redhead[79], for example, observed the beneficial effects of a small reflector placed immediately behind the filament to stabilize (and optimize) the field in this region. He reported a gauge with a high sensitivity (40 $mbar^{-1}$ for nitrogen), relatively independent of both the precise position of the shield and the potentials applied to the various electrodes.

Work has been carried out in a number of standards laboratories with the aim of constructing a gauge sufficiently stable for use in general calibration work. Particular attention has been paid to the range 10^{-2} to 10^{-6} mbar. Work has concentrated on the precise control of the electron trajectories in the ionization chamber. Of particular note is the work by Gentsch *et al.*[82] and

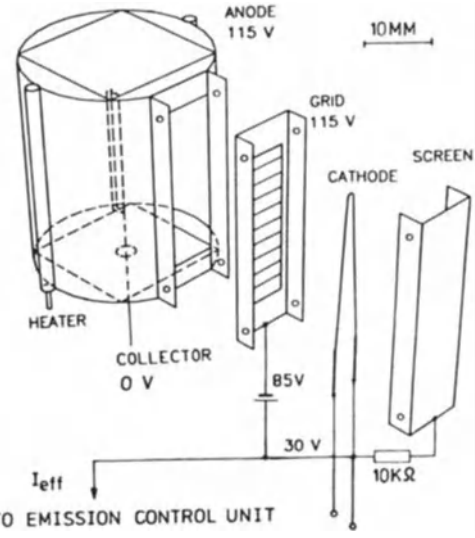


Figure 3.20 Exploded drawing of the screened ion gauge. The grid is insulated from the anode cylinder in order to separate the grid current from the ionizing electron current. The connection of the grid and the cathode by means of a potential source permits the stabilization of the effective ionization electron current with a suitable emission control unit. Dimensions: anode cylinder length 35 mm, diameter 24 mm; the collector, a gold foil of width 1.0 mm and length 16 mm, is V-shaped and supported on a gold-plated tungsten wire (diameter 0.08 mm[82]).

Choumoff and Lapteff[83]. The study by Gentsch *et al.*[82] illustrates a number of important principles. Their gauge, shown schematically in Figure 3.20, has a double electron collector, a grid directly in front of the hot filament and a cylindrical anode at the grid potential. Only those electrons which pass through the grid and enter the cylinder are effective in the ionization process. These electrons make a single traverse across the ionization chamber before being collected at the inner wall. Thus the path of each electron is precisely equal the cylinder diameter. In order to obtain stable control, this electron current is monitored by the emission regulator, as can be seen from Figure 3.20. To ensure the greatest possible stability of operation, all the electrode surfaces were gold-plated. There was a spread of only 5% in the sensitivity of a batch of 12 gauges made to this design.

Choumoff and Lafteff[83,84] had earlier investigated gauges designed with the similar objective of ensuring a precise path for all the ionizing electrons. The principles of these gauges are illustrated in Figure 3.21. All electrons effective in the ionization process pass through the ionization chamber and are collected on the high potential plate. None of the positive ions produced in this region impinges on the walls of the chamber. This gauge has been used in standards work. An idea of its suitability when used as a reference can be gained from a report issued by workers from the French, German, Italian and British Standard Laboratories[85]. This referred to a calibration exercise in

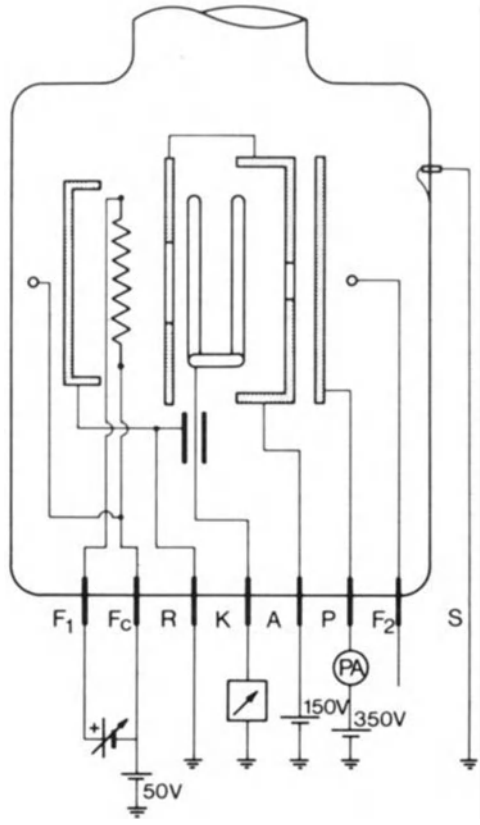


Figure 3.21 Diagram of the electrode configuration of the JHP gauge[85]: *A*, anode box; *K*, ion collector; *P*, plate 'electron collector'; *R*, reflector; F_c, common filament terminal; F_1, thoria-coated iridium filament used when measuring the pressure; F_2, hooped tungsten filament for outgassing; *S*, metallized screen on the gauge envelope. The potentials shown on the diagram are those applied when the gauge is used to measure pressure.

which these gauges were circulated around the various laboratories, in order that a comparison of the absolute standards of pressure measurement (10^{-6} to 10^{-3} mbar) could be made. Some 12 gauges were involved in the exchange; all operated under well-defined and rigorous out gassing cycles. The fact that at the end of the exercise the workers were able to report discrepancies of about 1–2% between the various standards indicates the stability and general reliability of this type of gauge.

It is known that all hot cathode ionization gauges can exhibit general drifts in sensitivity, usually downward when operated for long periods. The supporting evidence for this statement is not well documented, for, as might be expected, there are wide variations in performance, depending upon both the type of gauge and its treatment in the vacuum system. Systematic measurements on commercial conventional triode and BA gauges have been made

over long operating periods by Poulter and Sutton[76] and Wood and Tilford[86]. Considering the type of experiment, the agreement in the results from the two laboratories was satisfactory for the conventional triodes, but for the BA gauges, Wood and Tilford reported a significantly more stable operation. Poulter and Sutton's main findings[76] (summarized by the curves in Figures 3.22 and 3.23) show the drift in sensitivity of two conventional and one BA gauge. The results in Figure 3.22 are typical of the five conventional triodes tested. The data show no sudden changes in the sensitivity, but rather a scatter of results around a value which gradually decreases with the number of operating hours. During the periods *A* to *B* and *C* to *D*, the gauges were removed from the calibration system and run on the ion-pumped system and then returned to the calibration system. The mean sensitivity for the five gauges was 4.17 $mbar^{-1}$ (standard deviation 5%) for these gauges. For four of the gauges, the sensitivity drift was downward at a rate between 0.5% and 0.1% per 100 h. For the fifth gauge sensitivity increased at a rate of 0.3% per 100 h. For the six BA gauges (mean sensitivity 8.25 $mbar^{-1}$, standard deviation 14%) the performance was noticeably less stable, as can be seen from the data for one particular gauge which are shown in Figure 3.23. Poulter and Sutton[76] make the following comments:

> During the periods *A* to *B* and *C* to *D* the gauge was operated on the ion-pumped system and at time *F* it was removed from the calibration system, stored in a laboratory cupboard for fifteen months, before being returned to the system and recalibrated at time *G*. It can be clearly seen that distinct changes in the sensitivity of the gauge have taken place during some of the intervals when the gauge was removed from the system. With some gauges the change could be as large as 25%. At point *E* the gauge was operated for about 2 h in nitrogen at a pressure of 0.1 Pa, and this caused a 15% increase in the sensitivity. During the next few hours the sensitivity fell by about 5%. When this treatment was repeated on this gauge or on the other BA gauges an increase in sensitivity of about 5% was typically found and this change was always reversible by continued operation. The sensitivity had always returned to its previous value after 50 h of operation and usually much sooner. This reversible effect was not found when any of the triode type gauges were given the same treatment.

Poulter and Sutton[76] have no ready explanation for their observations. They do, however, refer to the work of Messer[87] who described an experimental procedure for stabilizing gauge sensitivity. This consisted of annealing the collector to a high temperature (1500–2000 °C for tungsten) which removed surface layers rich in oxygen, exposing a surface more typical of the bulk material. This is a permanent change which is not affected by the subsequent history of the gauge. The metal surface will have an increased work function compared with the original oxygen-rich surface, and this results in a reduced number of secondary electrons being released during the neutralization of an incident positive ion. Thus a reduced ion current and sensitivity will be measured. Becker and Messer[88] showed that operating the gauge for a long period has the same effect on the surface as the stabilization treatment. However, the change in surface, and hence sensitivity, takes place at a much lower rate.

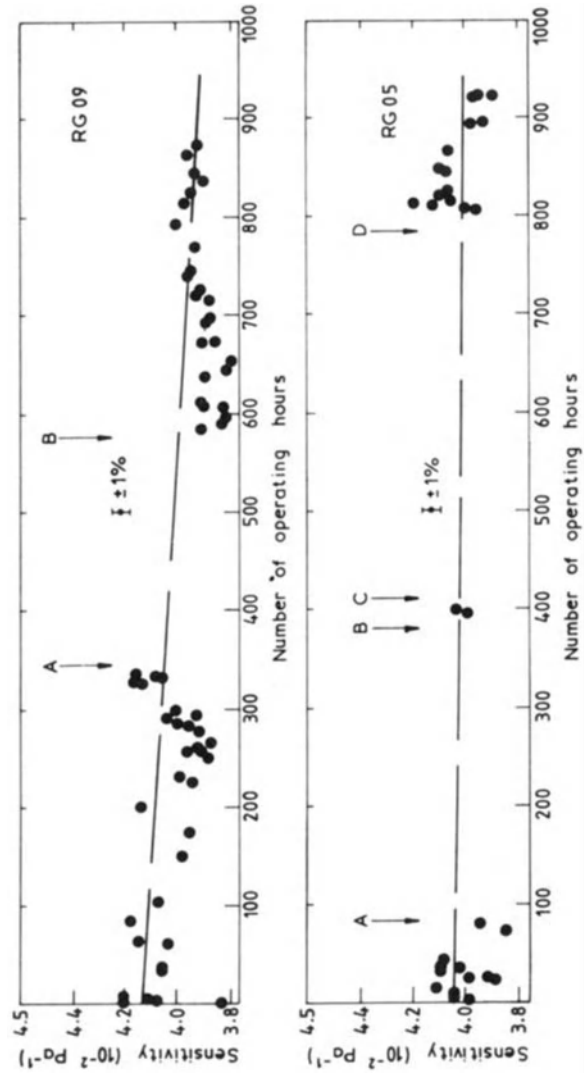


Figure 3.22 Variation of the nitrogen sensitivity at 5×10^{-5} mbar with the number of operating hours for two of the conventional triode type ionization gauges[76].

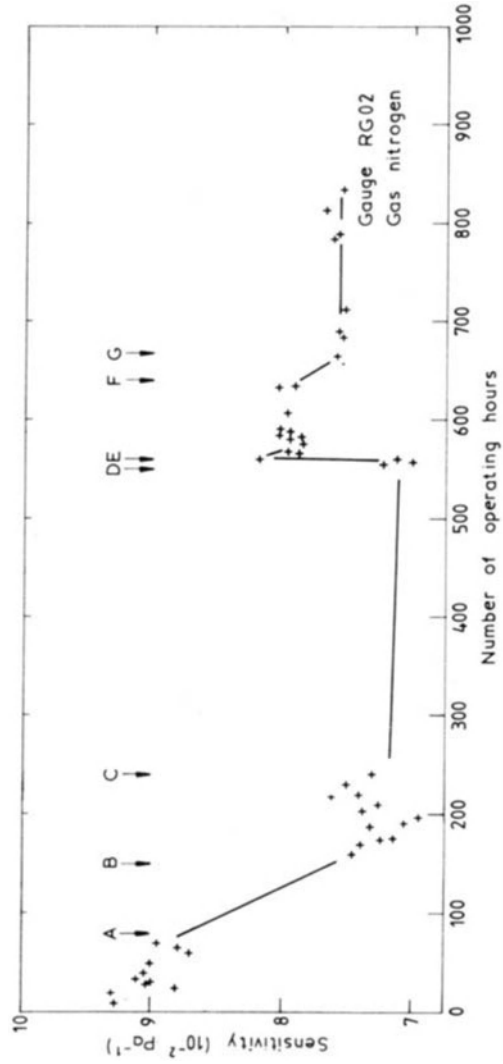


Figure 3.23 Variation of the nitrogen sensitivity at 5×10^{-5} mbar with the number of operating hours for one of the Bayard–Alpert type ionization gauges[76].

Although the materials for the collectors are different (nickel and molybdenum for the gauges in Figures 3.22 and 3.23 respectively), the mechanism proposed by Messer[87] can still be applicable to the observations of Poulter and Sutton[76]. It would be reasonable to assume that the BA gauges would show the greatest rate of change, because of the relatively high ion bombardment density of the collectors in these gauges. No explanation could be advanced to explain the sudden changes that occurred in the BA gauges after exposure to atmosphere. The point about the critical importance of the geometry in the filament region made by Redhead[79] and by Pittaway[80] may be significant, because of the possibility of changes in the effective emission areas of the filament.

The above observations were made for gauges operating in relatively clean vacuum systems. The performance may be much worse when these gauges are exposed to a hostile environment. In carefully conducted experiments, Young[37] confirmed the earlier observations of Utterback and Griffith[89] that operation in methane is extremely detrimental to the BA gauge. For example, in one experiment he found a dramatic drop in sensitivity, by nearly an order of magnitude, when the gauge operated at a partial pressure of methane of 3×10^{-4} mbar for 30 h. He was able to show that the deterioration was due to the formation of high-resistance hydrocarbon layers on the collector surface. These layers not only reduced sensitivity, but made performance unreliable when operating in other gases and, in particular, made sensitivity dependent upon the electron current. Young[37] found that the changes could be reversed and the original performance regained by outgassing the collector at a very high temperature. It is interesting to note that Poulter and Sutton[76] found changes in sensitivity for their BA gauges after operation in carbon monoxide and methane (the effects were particularly large for methane), but no significant changes in sensitivity were obtained when operating the conventional triodes under similar conditions.

Gentsch *et al.*[82] focused attention upon another important surface in the ionization gauge, the electron collector, when they noted the measured sensitivity of their gauge to be about 50% greater than expected from simple calculations. They suggested that the enhanced sensitivity was due to the elastic reflection of a considerable fraction of the electrons when they first strike the inner surface of the electron-collecting cylinder. The reflected electrons re-cross the ionization chamber and, therefore, given an added component to the ionization current. The magnitude of this effect is likely to depend upon the cleanliness of the electron-collecting surfaces. Werner and Leck[90] had also noted that the sensitivity of the standard BA gauge was increased by the order of 10% when the electron-collecting grid became contaminated with a surface layer of adsorbed hydrogen. Again the most likely explanation was the increased electron reflection due to the contaminating layer.

3.7 Gauges specially designed to operate at high pressure

The upper limit to the pressure range for all gauges, including the BA construction, must be reached when the secondary currents caused by the ion bombardments approach the same order of magnitude as the primary electron current. Most workers agree that non-linearity starts at some pressure between 10^{-4} and 10^{-2} mbar. Dushman and Found[5], in an early paper, reported the shape of the calibration curve in this region to be dependent on electron current. They estimated the upper limit of linearity to be approximately 5×10^{-2} mbar for an electron current of 0.5 mA and 10^{-3} mbar for an electron current of 15 mA. This dependence on electron current has been confirmed by later workers[91]. It is almost certainly due to the increased ion space charge effects associated with the increased electron emission. It does, therefore, seem prudent to keep the electron emission to the lowest practical value, certainly not more than 0.1 mA for all work at about and above 10^{-3} mbar (as recommended by Nottingham and Torney[92] and by Schulz[93]).

Schulz and Phelps[94], who investigated thoroughly the behaviour of ionization gauges at the high-pressure end of the scale, have listed three important limiting parameters, and from a study of these designed a number of gauges particularly suitable for high-pressure operation. The limiting factors they defined are as follows.

(i) The loss in sensitivity when the secondary electrons and ions produced by ionization become a significant fraction of the primary electron stream. These secondary electrons and ions contribute little to the ionization, but count as far as the emission stabilizer is concerned.
(ii) The sensitivity may increase at high pressure due to an increase in the efficiency of ion trapping by the collector. (At low gas pressure, many ions may be lost by drifting axially out from the discharge space. At high pressure there is less chance of loss in this way, because elastic collisions between ions and gas molecules cut down axial movement.)
(iii) Non-linearities may occur due to changes in the path and mean energy of the electron stream caused by non-elastic collisions between electrons and molecules at high pressure.

Schulz and Phelps[94] designed gauges with the special intention of keeping these three factors to a minimum and so giving linear operation to the highest possible pressure. The simplest of their designs is shown in Figure 3.24. This consists of a straight wire filament between two parallel rectangular plates, one plate acting as electron collector and the other as ion collector. Factor (i) is reduced to a minimum because of the low sensitivity (the pressure at which the effect of secondary electrons and ions becomes significant is inversely proportional to sensitivity); (ii) is satisfied by making the surface area of the ion collector large compared with the filament diameter, and (iii) by making the electrons from the filament travel in straight lines to the electron collector.

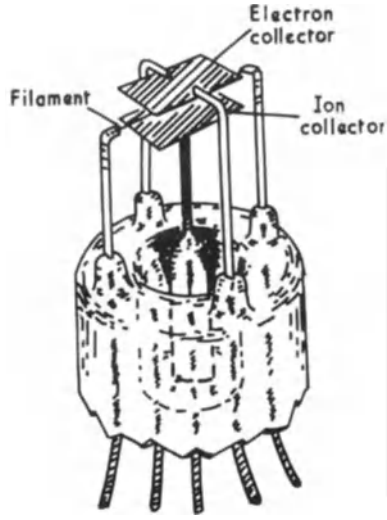


Figure 3.24 The ionization gauge designed by Schulz and Phelps[94] for high-pressure operation.

This gauge has a sensitivity of the order of 0.45 $mbar^{-1}$ for gases such as nitrogen and carbon dioxide. For these gases, a 10% deviation from the linear calibration does not occur until a pressure of 0.8 mbar is reached. The performance of this gauge is poor at low pressure because the X-ray induced background sets the lower limit of operation to 10^{-6} mbar.

Workers in a number of laboratories have taken the essential point made by Schulz and Phelps[94] that the sensitivity of a gauge must be made low if operation is to be extended to high pressure. Weinmann[95] in 1966, Cleaver[96] in 1967, and Beeck and Reich[97] in 1973 have described gauges with quite different geometries, but all giving sensitivities of approximately 0.5 $mbar^{-1}$. These gauges operated with near-linear sensitivity up to a pressure of 1 mbar. Commercial gauges are now available designed specifically with low sensitivity in order to extend the operating pressure range up to 1 mbar. Leybold, for example, offer a conventional triode which, because of its low sensitivity (0.5 $mbar^{-1}$ for nitrogen) is able to give useful readings up to a pressure of 1 mbar.

In 1980, Kudzia and Słówko[98,99] described a carefully conducted experimental project, backed by a sound theoretical analysis, to show that by using a small sphere as the anode (the electron collector) the pressure range could be extended a further order of magnitude to 10 mbar. In their analysis these workers presented the basic equation which sets out the relation between sensitivity and pressure. In this equation, I_c/I_f (where I_c is the ion current to the collector and I_f the electron current emission from the cathode) is shown to be

given by the product of four terms:

$$\frac{I_c}{I_f} = \frac{I+\gamma}{1-\gamma\phi_c(p)} \cdot \frac{1}{1+\delta\phi_f(p)} \cdot \frac{1-\delta}{1} \cdot \phi_f(p) \tag{3.9}$$

where $\phi_f(p)$ is the fundamental 'sensitivity term', being the number of positive ions produced by an electron in travelling from cathode to anode; $\phi_c(p)$ is the corresponding term for the positive ion production produced by an electron travelling from the ion collector to the electron collector (the 'grid' in Figure 3.2); δ is the fraction of the total ion production which travels to the cathode; and γ is the coefficient for secondary electron production by ion bombardment of the collector.

The first term in eqn (3.9) represents the amplification of the ion current by the ionization from the secondary electrons emitted from the ion collector. Clearly there is the onset of a self-sustaining glow discharge when the factor $\gamma\phi_c(p)$ approaches unity. The second term represents the decrease in the ion current caused by the decrease in electron emission from the cathode due to the arrival of positive ions. (This is the first factor in the Schulz and Phelps analysis discussed above.) The third term represents the loss in sensitivity due to the diversion of a fraction of the ion current from the collector to cathode.

At low pressure, when the term $\phi_c(p)$ is small and represented by $\phi_c(p) = Kp$, eqn (3.9) reduces to the conventional

$$\frac{I_c}{I_f} = (1-\delta)Kp \tag{3.10}$$

At high pressures, the function $\phi_c(p)$ ceases to be linear and, in fact, eventually must reach and pass through a maximum due to the loss of energy by inelastic electron collisions (the third factor in the Schulz and Phelps analysis). It is clear that, to obtain optimum performance, that is as near as possible to eqn (3.10), all four factors δ, γ, $\phi_c(p)$ and $\phi_f(p)$ should be as small as possible.

Kudzia and Słówko pointed out that the pressure at which the function $\phi_c(p)$ reaches saturation depends upon the electric field in the region where ionization is taking place. They made the important point that, by increasing the electric field, the linear range can be extended and the saturation point pushed to higher pressures. They rejected the two obvious ways of increasing the field, by increasing the anode (electron collector) potential, and by reducing the interelectrode spacing, the former because of the strong probability of also increasing γ and therefore the onset of the glow discharge, and the latter because it would decrease the efficiency of ion collection and therefore increase δ. Instead, they proposed the novel idea of making the anode small and spherical. In fact, for both analysis and experimental gauge, they chose a sphere diameter of only 0.64 mm. Because of field enhancement in the region of a sharp point – in this case a sphere of small radius – almost all the

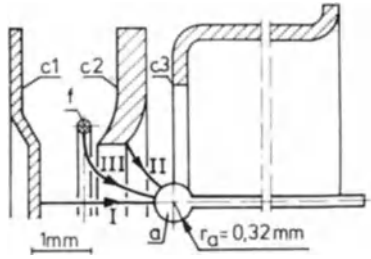


Figure 3.25 Diagram of the 'spherical' ionization gauge described by Kudzia and Słówko[98]. *f*, filament; *a*, electron collector; *c*1, *c*2, *c*3, the three separate ion collectors.

potential drop between anode and cathode is in the thin spherical shell immediately in front of the anode. The thickness of this shell is virtually independent of the geometry of the rest of the electrode system. For the system chosen, the ionizing region extended only 0.5 mm from the surface of the sphere. Kudzia and Słówko, using standard data, calculated the ionization function $\phi_c(p)$ for various anode to cathode potential differences. (The method was verified experimentally for argon in a planar electrode geometry.) Elastic collisions, ionizing collisions and two types of excitation collision (3^4S_0–4^3P_1 and $2p_8$) were taken into account. They found the factor $\phi_f(p)$ to increase approximately linearly with pressure to about 1 mbar, and to reach a maximum at just below 20 or 40 mbar (for potential differences of 150 and 300 V respectively). As predicted, they found the pressure for maximum ionization to be proportional to the potential difference. The gauge constructed to verify the advantages of the spherical anode is shown in half-section in Figure 3.25. A circular filamentary cathode was placed symmetrically between two screening collectors *c*2 and *c*3 and the actual ion collector *c*1. The anode was a small sphere formed on the end of a molybdenum rod. This geometry gives a low electric field in the cathode (*f*) region and at the surface of the ion collector *c*1. The factor δ was calculated to be 0.08. Because of the concentration of the electric field near to the surface of the anode, ionization is almost identical for electrons travelling on paths I, II or III, i.e. from either the filament or from collectors *c*1 or *c*2. The measured characteristics are in good agreement with those calculated from basic theory. In particular, the experiment confirms that the ion current generated is dependent only upon the term $\phi_f(p)$ when the pressure is less than 1 mbar. Kudzia and Słówko show that when the anode–cathode potential difference does not exceed 250 V, the only additional term to take into account at the higher pressures is $(1 + \delta\phi_f(p))$. However, at high voltages, certainly above about 400 V, the glow discharge term $(1 - \gamma\phi_c(p))$ dominates at the higher pressures. These points can be seen from the set of characteristics (for air) shown in Figure 3.26. The linear relation between the collector current *c*1 and pressure up to 1000 Pa (10 mbar) is particularly significant. The theoretical maximum in this gauge of the function

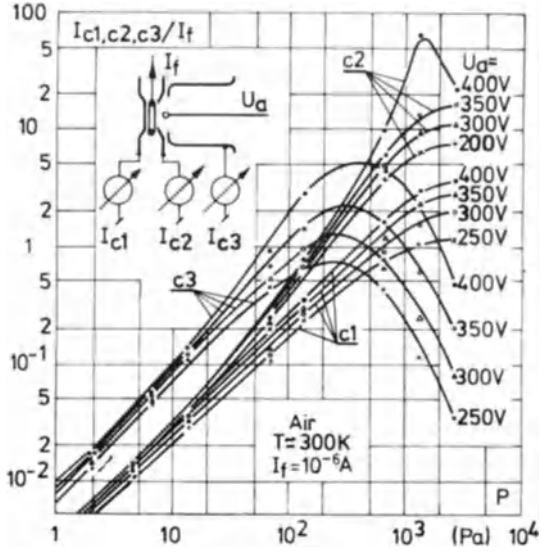


Figure 3.26 Calibration curves for the 'spherical' ionization gauge designed by Kudzia and Słówko[98]. *c*1, current to ion collector; *c*2, current to screening ion collector; *c*3, current to auxiliary ion collector.

I_c/I_f given by $(1-\delta)/\delta$ is 12, $\delta \simeq 0.08$ when the glow discharge term is neglected. Figure 3.26 shows this value is exceeded at the higher voltages, indicating the onset of a glow discharge.

It is interesting to note that Kuo[100], at approximately the same time as Kudzia and Słówko[98], proposed an entirely different gauge geometry to extend the range to 10 mbar. With the simple electrode structure shown in Figure 3.27, he covered the two important design principles. In this gauge, all the ionization occurs in the small cylindrical sheath surrounding the anode, because of its small radius of curvature. As Kuo pointed out, this gives both

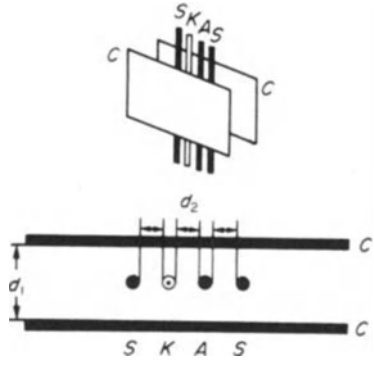


Figure 3.27 Configuration of the gauge electrodes in the gauge designed by Kuo[100].

ionization in a high field and a low overall sensitivity, the primary requirements for high-pressure operation. This is confirmed by the experimental results, which show gauge characteristics very similar to those in Figure 3.26.

It is interesting to note that both Kudzia and Słówko[98] and Kuo[100] changed potentials for low-pressure operation (defined as $< 10^{-3}$ mbar), so as to extend the range down to 10^{-5} or 10^{-6} mbar.

Thoria-coated cathodes are usually used for high-pressure gauges, as they are especially resistant to oxidation, allowing operation for long periods at any point in the operating range of the gauge.

3.8 Chemical and physical reactions in the hot cathode ionization gauge

There are two important and distinctly separate mechanisms which have to be considered in any investigation of the reactions in the gauge electrode system. These are, firstly, those essentially 'thermally induced' at the hot cathode surface, and secondly the effect of the bombardment of the gas molecules by the relatively high-energy electrons. The phenomena are many and varied, some simple, but some extremely complex and not completely understood. The net effect is that any gauge, depending upon its past history of operation and the precise atmosphere in the vacuum system, can act as either a source or sink of gas. Further, its operation can cause significant changes to the gas composition in the system. Clearly the magnitude and practical importance of the reactions depend upon the overall vacuum characteristics as much as, or more than, upon the action of the gauge. For example, the changes in pressure and gas composition are more significant in a small clean UHV system with a low pumping speed and very little outgassing, than in a large industrial plant with large turbomolecular or diffusion pumps. Similarly, any pressure difference between the gauge and the main chamber will depend upon the conductance of the pipe connecting the two, and will be zero only when the gauge is inserted directly into the chamber (i.e. the nude gauge).

Consider first the reactions at the hot cathode. It is important to note that there is a great dependence upon both the material chosen for the cathode and the type of gas in the gauge environment. The different behaviour is due primarily to the wide range of cathode operating temperatures, from about 1200 K for some oxides, to above 2200 K for tungsten, and, of course, the quite different chemical activity of the gases in the systems (such as oxygen and argon). The very thorough investigation by Langmuir[101–104] of the reaction between oxygen and a hot tungsten surface, reported in a comprehensive series of papers published before 1920, set the standard for investigation in this field. There are two essentially different reactions: (i) chemical combination with the tungsten (oxidation being the most important example), and (ii) the dissociation of diatomic molecules into single atoms at the hot surfaces and the subsequent adsorption of these atoms on to the walls. The 'clean-up' of atomic hydrogen is a good example of this process.

In 1913, Langmuir[101] carried out a series of experiments which showed that oxygen would combine with tungsten to form tungsten oxide. In this work, a glass bulb containing oxygen at a pressure of a few mbar was sealed off from the vacuum pumps and, on heating the filament, the fall of pressure was observed by means of a McLeod gauge. At moderately low temperatures (in the range 900–1250 K) a straw-coloured oxide film formed on the tungsten filament. This film could be driven off to the glass walls by increasing the temperature to 1250 K. Above this temperature, the wire always remained bright. By comparing the loss of weight of the filament with the amount of gas taken up, the reaction was shown to be consistent with the formation of WO_3. At a given filament temperature, the oxygen pressure always fell exponentially, the rate of removal of gas being directly proportional to the rate at which molecules struck the filament, i.e. the ratio β (number of molecules removed from the gas phase)/(number striking the filament) is constant. β was found to increase with increasing tungsten temperatures. The values obtained by Langmuir at various wire temperatures are given in Table 3.2. It can be seen from these figures that β tends to a limiting value of approximately 0.15 at very high temperatures.

In practical vacuum technology terms, the effect of this chemical activity can be considered as a small gettering action, removing molecules from the gas phase at a rate proportional to the pressure inside the gauge electrode system. For the purpose of calculation, this gettering can be represented as a vacuum pump with a constant speed S normally expressed in units of litre second^{-1}. It can easily be shown, from the fundamentals of the kinetic theory of gases, that the pumping speed is given by

$$S = 0.11\, A\beta \quad \mathrm{l\,s^{-1}} \tag{3.11}$$

where A is the surface area of the filament (mm^2).

Usually A ranges between 1 and 10 mm^2, from the smallest to the largest gauge. Thus the practical values of S, the 'pumping speed', will lie between 10^{-3} and 10^{-2} l s^{-1} at 1600 K, and 10^{-2} and 10^{-1} l s^{-1} at 2300 K.

The validity of the data given in Table 3.2 has been confirmed in many laboratories. Riddiford[105], for example, found the experimentally measured 'pumping speeds' of the gauge to agree closely with the values calculated from the above parameters. The principles of his experiment are shown schematically in Figure 3.28. The gauges used to measure the pressures in the vacuum chamber are connected in the two conventional ways: directly in the system

Table 3.2 The ratio of the number of oxygen molecules reacting with a hot tungsten filament to form WO_3 to the total number striking the tungsten surface, as a function of filament temperature.

Temp. (K)	1070	1270	1470	1570	1820	2020	2290	2520	2770
β	0.0003	0.0011	0.0053	0.0094	0.0255	0.049	0.095	0.12	0.15

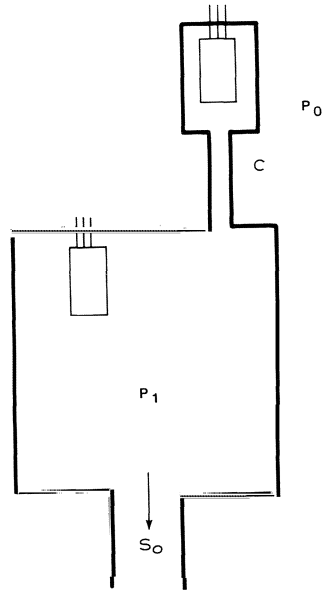


Figure 3.28 Vacuum chamber with tubulated and 'nude' gauges. C, conductance of the gauge tubulation; S_0, effective 'pumping speed' from the chamber (litre s^{-1}).

(nude), and in a glass envelope connected to the system via a short tube. Taking the conductance of this tube as $C\,1\,s^{-1}$ and the effective pumping speed of the gauges as $S\,1\,s^{-1}$, it is clear that the fractional error in the reading of the tubulated gauge is given by

$$\frac{P_0 - P_1}{P_0} = \frac{S}{C} \tag{3.12}$$

Riddiford found that the difference between the measured error in reading and that calculated by use of eqn (3.12) differed by no more than 10%.

The reduction of the oxygen pressure in the vacuum manifold by the filament gettering depends directly upon the effective conductance from the manifold ($S_0\,1\,s^{-1}$ in Figure 3.28). Heating the filament will reduce the system to a new, and lower, equilibrium; the fractional fall in pressure is given by $S/(S + S_0)$. Thus, depending upon the temperature of the filament, the reduction in equilibrium must seriously be considered.

The reactions between oxygen and the carbon impurities in the tungsten filament were first observed with certainty when studies of the residual gases were made by mass spectrometry (for example by Blears[106]). It has now been established that oxygen reacts with the carbon impurities of any cathode material. Young[107] showed that carbon monoxide and, to a lesser extent, carbon dioxide are formed in oxygen reactions with cathode materials of

molybdenum, tungsten, rhenium and tantalum. He showed that in each case the intensity of the reactions is directly proportional to the carbon impurity. He also confirmed that the reaction took place at the hot filament itself, the walls or other surfaces playing no active part. Schlier[108] observed the same reaction at hot surfaces, and also showed (a note added in proof to his paper) that carbon monoxide was formed only when a carbon impurity was present in a filament. Schissel[109] noticed the production of carbon monoxide and carbon dioxide in considerable quantities (20% and 10% respectively of the oxygen intensity) when a thoria-coated iridium filament was used in an ionization gauge. Grayson[110], and later Podor[111], have observed complex reactions at the surface of a rhenium filament, initiated by oxygen molecules in the vacuum atmosphere. This has, like the reaction between oxygen and carbon impurities in tungsten filaments, special implications for gas analysis applications using mass spectrometers with thermionic cathodes.

The take-up of hydrogen has been investigated in a similar manner to that of oxygen, either by observing the fall of pressure in a sealed system or by observing the steady flow along a tubulation, as indicated in Figure 3.28 and eqn (3.12). There is a most important practical difference between the two; whereas the oxidation continues indefinitely until all the tungsten has been converted to oxide, the adsorption of hydrogen ceases when a surface layer has been formed on the glass envelope. A further build-up cannot take place because the atoms recombine and are desorbed as molecular hydrogen. In a bulb of 1-litre volume, Langmuir found virtually complete saturation (the reaction rate had fallen by a factor of 20) after the take-up of approximately 20×10^{-3} l.mbar of gas. This take-up took place with the glass envelope at liquid nitrogen temperature. A much smaller quantity of gas was held on the glass at room temperature. The initial equivalent 'pumping speed' in the Langmuir experiments is of the order of 0.001 to 0.01 l s^{-1}. This adsorption, which can be detected at a temperature as low as 1300 K, increases only slowly with filament temperature.

Schwarz[112] in a similar, but more recent, experiment with modern measuring techniques observed hydrogen clean-up at an initial rate of 0.02 l.s^{-1} due to the action of the tungsten filament. Carter *et al.*[113] also observed a similar initial pumping speed in their experiments. Werner and Leck[90] noted that the formation of atomic hydrogen became significant, with consequent sorption at glass and metal surfaces, as the temperature of a tungsten filament reached 1400 K. The rate of gas removal increased steadily with increase in filament temperature through the range 1400–2100 K. Hickmott,[114] working with a sealed-off vacuum system, showed how reaction of the atomic hydrogen at the glass wall of the ionization gauge can produce carbon monoxide, water vapour, and also methane. In certain experiments the residual pressure of carbon monoxide actually exceeded the hydrogen pressure.

Reaction with the filament is not restricted to hydrogen and oxygen, but

also takes place with gases containing either of these elements. The reaction with water vapour, for example, is particularly interesting. This was shown by Langmuir[115], in a particularly elegant series of experiments, to consist of two distinct steps: Firstly, the breaking of the H_2O molecule at the hot tungsten surface to form WO_3 and atomic hydrogen, and secondly the reaction between the atomic hydrogen and the WO_3 on the glass walls to form tungsten and water vapour. Thus the water vapour acts as a vehicle for transporting tungsten from cathode to wall without itself suffering any loss.

Materials with a lower work function, and hence a lower operating temperature, have replaced tungsten for many applications[116]. This reduces not only the chemical activity but also the evaporation of metal from the hot surface. Thoriated tungsten, rhenium, lanthanum hexaboride, thoria-coated iridium or tungsten have all been used to reduce the operating temperature into the range 1300–1500 K. Cathodes of ThO_2 on tungsten or iridium have reasonable emission efficiencies and are relatively unaffected by chemically active gases. ThO_2 on iridium has the advantage that it can be exposed to air when hot without damage. For this reason, many commercial ionization gauges are available with thoria-coated iridium cathodes.

For gases other than those containing oxygen or hydrogen, it is the collision between the relatively high-energy electrons (~ 100 eV) and the gas molecules which is responsible for the sorption of gas. Extensive reports[117–120] have been made on this subject for the pressure range from 10 to below 10^{-6} mbar. It is now well established that the positive ions are chiefly responsible for the sorption and consequent pumping action. Ions are driven with comparatively high energies (of the order of 100 eV) into the walls of the gauge and are trapped either at, or just under, the surface. The binding is sufficiently strong to require heating to over 300 °C to release those most tightly bound. However, something more than this simple ion bombardment mechanism must be introduced to explain the pumping action in certain gases. Helium, for example, is only sorbed strongly when the walls have a film of evaporated or sputtered metal[121,122]. In the case of nitrogen, the pumping rate is too large to be explained by the pumping of ions. Metastable molecules or atomic species must play some part in the action.

The effective initial rate of sorption is directly proportional to the rate of ion production, that is proportional to the electron emission current. For an electron current of 1 mA, the effective pumping speed varies from about 10^{-3} $l.s^{-1}$ for the inert gases to 10^{-1} $l.s^{-1}$ for nitrogen. Table 3.3 summarizes the results reported from various laboratories for the noble gases and for nitrogen. This gives a good indication of the magnitude of the initial pumping speed in relatively clean systems. All the results are for tubulated gauges. Since the phenomenon depends upon the history of gauge operation (cleanliness of the walls and electrode surfaces for example), some considerable variation in sorption rate must be anticipated in practical operation. It is interesting to note the relatively large values of sorption rate for nitrogen, which approach

Table 3.3 Effective 'pumping speed' of ionization gauge for noble gases and for nitrogen. All measurements have been made with tubulated gauges. Data presented in review papers by Berman[121].

Reference	Gas	Electric current (mA)	Effective pumping speed (litre $s^{-1} \times 10^3$)
122	Helium	10	0.1
119	Helium	10	6
124	Helium	8	12
125	Helium	4	6
119	Neon	10	7
119	Argon	10	17
126	Argon	8	7
126	Argon	1	10
119	Krypton	10	50
119	Xenon	10	60
128	Nitrogen	1	20
129	Nitrogen	10	500
113	Nitrogen	1	125
131	Nitrogen	8	250

those for oxygen and hydrogen. In all cases the walls become saturated, usually after about 10^{-3} to 10^{-2} l.mbar of gas has been sorbed onto the walls of the system, at which time the pumping speed falls to zero. However, there is an exchange mechanism between the ions bombarding the walls and those already sorbed, so that the zero pumping speed represents dynamic rather than static equilibrium.

The sorption mechanism for the residual heavy hydrocarbon vapours in vacuum systems is significantly different from that of gases and other low-molecular-mass vapours. Blears[132] first investigated the phenomenon using two ionization gauges connected to a large bell-jar evacuated with an oil diffusion pump. One gauge was mounted directly inside the bell-jar (nude), the other in a glass envelope with a relatively small-diameter tube connection. For partial pressures of the normal gases and vapours (such as air, hydrogen, water vapour and carbon dioxide), the two gauges gave approximately the same reading. But for the residual vapours of the oil diffusion pump (for example Apiezon B), the nude gauge gave the higher reading by a factor of about 10. The pressure differences are almost certainly caused by the sorption of gas on the walls of the glass envelope and connecting pipe of the tubulated gauge. Blears[132] showed that saturation did not occur even after four or five weeks' operation and, as might be expected in this direct sorption phenomenon, increasing the surface area of the glass envelope (by inserting glass wool) increased the pressure difference. The effect was unchanged when the tubulated gauge had a metal instead of a glass envelope. Observations in many laboratories have confirmed the above measurements. For example, Haeffer and Hengevoss[133] observed that a nude and a tubulated gauge differed by a

factor of 10 when working at a much lower pressure under ultra-high vacuum conditions.

This effect is particularly significant in measurements of the 'background' or 'ultimate' pressure in any vacuum system where there is a component of heavy hydrocarbon vapour. The measurements serve to emphasize the care that must be taken to ensure that the ionization gauge is correctly positioned if errors in measurement are to be avoided. Ideally, where hydrocarbon or other condensable vapours are present in the residual atmosphere, the gauge should be placed inside the working chamber or via a very high conductance tubulation. Even if hydrocarbons are not present, the gauge tubulation should be at least $20 l s^{-1}$, in order to negate the pumping effects of gases such as oxygen, nitrogen and water vapour.

References

1. Smith, P.T. (1930) *Phys. Rev.* **36**, 1293; (1931) **37**, 808.
2. Tate, J.T. and Smith, P.T. (1932) *Phys. Rev.* **39**, 270.
3. Buckley, O.E. (1916) *Proc. Nat. Acad. Sci.* USA **2**, 683.
4. Misamichi, So. (1919) *Proc. Phys. Math. Soc. (Japan)* **1**, 76.
5. Dushman, S. and Found, C.G. (1921) *Phys. Rev.* **17**, 7.
6. Dushman, S. and Young, A.H. (1945) *Phys. Rev.* **68**, 278.
7. Reynolds, N.B. (1931) *Physics* **1**, 182.
8. Holanda, R. (1973) *J. Vac. Sci. Technol.* **10**, 1133.
9. Found, C.G. and Dushman, S. (1924) *Phys. Rev.* **23**, 734.
10. Dushman, S. and Young, A.H. (1945) *Phys. Rev.* **68**, 278.
11. Downing, J.R. and Mellen, G. (1946) *Rev. Sci. Instrum.* **17**, 218.
12. Riddiford, I., (1951) *J. Sci. Instrum.* **28**, 375.
13. Wagener, S. and Johnson, C.B. (1951) *J. Sci. Instrum.* **28**, 278.
14. Moesta, H. and Renn, R. (1957) *Vakuumteknik.* **6**, 35.
15. Schulz, G.J. (1957) *J. Appl. Phys.* **28**, 1149.
16. Schulz, G.J. and Phelps, A.V. (1957) *Rev. Sci. Instrum.* **28**, 1051.
17. McGowan, W. and Kerwin, L. (1960) *Can. J. Phys.* **38**, 567.
18. Cobic, B., Carter, G. and Leck, J.H. (1961) *Vacuum* **11**, 247.
19. Ehrlich, G. (1961) *J. Appl. Phys.* **32**, 4.
20. Rothe, E.W. (1964) *J. Vac. Sci. Technol.* **1**, 66.
21. Shaw, M.L., (1966) *Rev. Sci. Instrum.* **37**, 113.
22. Utterbach, N.G. and Griffith, T. Jr. (1966) *Rev. Sci. Instrum.* **37**, 866.
23. Cleaver, J.S. (1967) *J. Sci. Instrum.* **44**, 969.
24. Walters, W.L. and Craig, J.H. Jr. (1968) *J. Vac. Sci. Technol.* **5**, 152.
25. Holanda, R. (1972) *NASA Rept* No. TH D-6815.
26. Summers, R.L. (1969) *NASA Rept* No. TN D-5285.
27. Nakao, F. (1975) *Vacuum* **25**, 431.
28. Langmuir, I. and Jones, H.A. (1928) *Phys. Rev.* **31**, 357.
29. Morgulis, N. (1934) *Z. Physik* **5**, 407.
30. Schwarz, H. (1944) *Z. Physik* **122**, 437.
31. Tominaga, G. (1950) *J. Vac. Soc. (Japan)* **1**, 27.
32. Metson, G.H. (1951) *Br. J. Appl. Phys.* **2**, 46.
33. Tominaga, G. (1955) *Br. J. Appl. Phys.* **2**, 19.
34. Oda, Z. and Arata, Y. (1957) *J. Vac. Soc. (Japan)* **7**, 197.
35. Ishii, H. and Nakayama, K. (1960) *J. Vac. Soc. (Japan)* **3**, 77.
36. Anderson, H.V. (1963) *Rev. Sci. Instrum.* **34**, 703.
37. Young, J.R. (1973) *J. Vac. Sci. Technol.* **10**, 212.
38. Nakayama, K. and Hojo, H. (1974) *6th Int. Vacuum Congr., Kyoto.*
39. Bartmess, J.E. and Georgiadis, R.M. (1983) *Vacuum* **33**, 149.

40. Savchik, K.J. and Miller, J.A. (1979) *J. Am. Chem. Soc.* **101**, 7206.
41. Apker, L. (1948) *Industr. Engng. Chem.* **40**, 846.
42. Anderson, P.A. (1935) *Phys. Rev.* **47**, 958.
43. Nottingham, W.B. (1937) *J. Appl. Phys.* **8**, 762.
44. Lander, J.J. (1950) *Rev. Sci. Instrum.* **21**, 672.
45. Bayard, R.T. and Alpert, D. (1950) *Rev. Sci. Instrum.* **21**, 571.
46. Nottingham, W.B. (1954) *Vac. Symp. Trans. Comm. Vac. Tech.*, Pergamon, New York, 76; (1961) *Vac. Symp. Trans. Amer. Vac. Soc.* **1**, Pergamon, New York, 494.
47. Carter, G. and Leck, J.H. (1959) *Br. J. Appl. Phys.* **10**, 364.
48. Van Oostrom, A. (1961) *Vac. Symp. Trans. Amer. Vac. Soc.* **1**, Pergamon, New York, 443.
49. Alpert, D. (1958) *Handbuch der Physik* **12**, Springer, Berlin, 609.
50. Ackley, J.W., Lothrop, C.F. and Wheeler, W.R.W. (1962) *Vac. Symp. Trans. Amer. Vac. Soc.* **9**, Pergamon, New York, 452.
51. Redhead, P.A. (1960) *Rev. Sci. Instrum.* **31**, 343.
52. Appelt, G. (1962) *Vakuumtechnik* **11**, 174.
53. Hobson, J.P. (1964) *J. Vac. Sci. Technol.* **1**, 1.
54. Lange, W.J. and Singleton, J.H. (1966) *J. Vac. Sci. Technol.* **3**, 319.
55. Poulter, K.F. (1970) *Vacuum* **20**, 385.
56. Edwards, D. and Lanni, C. (1980) *J. Vac. Sci. Technol.* **17**, 355.
57. Chen, J.Z., Suen, C.D. and Kuo, Y.H. (1984) *Vacuum* **34**, 641.
58. Watanabe, F. (1987) *J. Vac. Sci. Technol.* **A5**, 242.
59. Blechschmidt, D. (1974) *J. Vac. Sci. Technol.* **11**, 1160.
60. Pittaway, L.G. (1974) *Philips Res. Repts.* **29**, 261.
61. Pittaway, L.G. (1974) *Philips Res. Repts.* **29**, 283.
62. Lafferty, J.M. (1961) *J. Appl. Phys.* **32**, 424.
63. Davis, W.D. (1968) *J. Vac. Sci. Technol.* **5**, 23.
64. Visser, J. (1967) *Vacuum* **17**, 73.
65. Chen, J.Z. Suen, C.D. and Kuo, Y.H. (1978) *J. Vac. Sci. Technol.* **A5**, 2373.
66. Gabor, D. (1962) British Patent No. 887251.
67. Herb, R.G. Pauly, T. and Fischer, H.J. (1963) *Bull. Amer. Phys. Soc.* **8**, 336.
68. Mourad, W.G., Pauly, T. and Herb, R.G. (1964) *Rev. Sci. Instrum.* **35**, 661.
69. Meyer, E.A. and Herb, R.G. (1967) *J. Vac. Sci. Technol.* **4**, 63.
70. Gosselin, C.M., Beitel, G.A. and Smith, A. (1970) *J. Vac. Sci. Technol.* **7**, 233.
71. Fitch, K.R. and Rushton, G.J. (1970) *Vacuum* **20**, 535.
72. Fitch, K.R., Mulvey, T., Thatcher, W.J. and McIlwraith, A.H. (1971) *J. Phys. E.* **4**, 533.
73. McCulloh, K.E. and Tilford, C.R. (1981) *J. Vac. Sci. Technol.* **18**, 994.
74. Tilford, C.R. (1983) *J. Vac. Sci. Technol.* **A1**, 152.
75. Tilford, C.R. (1985) *J. Vac. Sci. Technol.* **A3**, 546.
76. Poulter, K.F. and Sutton, C.M. (1981) *Vacuum* **31**, 147.
77. Angerth, B. (1972) *Vacuum* **22**, 7.
78. Laurent, J.M., Benvenuti, C. and Scalambrin, F. (1977) *Proc. 7th Int. Vacuum Congr., Vienna*, **1**, 113.
79. Redhead, P.A. (1969) *J. Vac. Sci. Technol.*, **6**, 848.
80. Pittaway, L.G. (1970) *J. Appl. Phys.* **3**, 1113.
81. Arnold, P.C. and Bills, D.G. (1984) *J. Vac. Sci. Technol.* **A2**, 159.
82. Gentsch, H., Tewes, J. and Messer, G. (1985) *Vacuum* **35**, 137.
83. Choumoff, P. and Iapteff, B. (1974) *Electron Fisc. Apli.* **17**, 71.
84. Choumoff, P. and Iapteff, B. (1974) *Proc. 6th Int. Vac. Congr., Kyoto.*
85. Poulter, K.F. Calcatelli, A., Choumoff, P.S., Iapteff, B., Messer, G. and Grosse, G. (1980). *J. Vac. Sci. Technol.* **17**, 679.
86. Wood, S.D. and Tilford, C.R. (1985) *J. Vac. Sci. Technol.* **A3**, 542.
87. Messer, G. (1977) *Phys. Bull. (Germany)* **33**, 343.
88. Becker, H.-U. and Messer, G. (1980) *Vide, Couches Minces*, **2**, Suppl. 201, 234.
89. Utterback, N.G. and Griffith, T. Jr. (1966) *Rev. Sci. Instrum.* **37**.
90. Werner, J.G. and Leck, J.H. (1969) *J. Sci. Instrum.* **2**, 861.
91. Redhead, P.A. and Hobson, J.P. (1965) *Br. J. Appl. Phys.* **16**, 1555.
92. Nottingham, W.B. and Torney, F.L. (1960) *Vac. Symp. Trans. Amer. Vac. Soc.*, Pergamon, New York, 117.

93. Schulz, G.J. (1957) *J. Appl. Phys.* **28**, 1149.
94. Schulz, G.J. and Phelps, A.V. (1957) *Rev. Sci. Instrum.* **28**, 1051.
95. Weinman, J.A. (1966) *Rev. Sci. Instrum.* **37**, 636.
96. Cleaver, J.S. (1967) *J. Sci. Instrum.* **44**, 969.
97. Beeck, U. and Reich, G. (1974) *Vacuum* **24**, 27.
98. Kudzia, J. and Słówko, W. (1981) *Vacuum* **31**, 9.
99. Kudzia, J. and Słówko, W. (1981) *Vacuum* **31**, 359.
100. Kuo, Y.H. (1981) *Vacuum* **31**, 303.
101. Langmuir, I. (1913) *J. Amer. Chem. Soc.* **35**, 105.
102. Langmuir, I. (1915) *J. Amer. Chem. Soc.* **37**, 1139.
103. Langmuir, I. (1915) *J. Amer. Chem. Soc.* **37**, 417; (1912) **34**, 860; (1912) **34**, 1310; (1914) **36**, 1708.
104. Langmuir, I. (1913) *J. Amer. Chem. Soc.* **35**, 931.
105 Riddiford, L. (1951) *J. Sci. Instrum.* **28**, 375.
106. Blears, J. (1950) *Rev. Sci. Instrum. Suppl.* No. 1, 36.
107. Young, J.R. (1959) *J. Appl. Phys.* **30**, 1671.
108. Schlier, R.E. (1958) *J. Appl. Phys.* **29**, 1162.
109. Schissel, P.O. (1962) *J. Appl. Phys.* **33**, 2659.
110. Grayson, M.A. (1979) *Int. J. Mass. Spectr. Ion Phys.* **30**, 383.
111. Podor, B. (1983) *Vacuum* **33**, 67.
112. Schwarz, H. (1944) *Z. Phys.* **122**, 437.
113. Carter, G., Cobic, B. and Leck, J.H. (1961) *Br. J. Appl. Phys.* **12**, 384.
114. Hickmott, T.W. (1960) *J. Chem. Phys.* **32**, 810.
115. Langmuir, I. (1913) *Trans. Amer. Inst. Elect. Engrs.* **32**, 1893.
116. Gear, P.E. (1975) *Vacuum* **26**, 3.
117. Alpert, D. (1958) *Handbuch der Physik* **12**, Springer, Berlin, 609.
118. Carter, G. (1959) *Vacuum* **9**, 190.
119. Cobic, B., Carter, G. and Leck, J.H. (1961) *Br. J. Appl. Phys.* **12**, 282.
120. James, L.H. and Carter, G. (1962) *Br. J. Appl. Phys.* **13**, 2.
121. Berman, A. (1982) *Vacuum* **32**, 497.
122. Varnerin, L.J. and Carmichael, J.H. (1955) *J. Appl. Phys.* **26**, 782.
123. Young, J.R. (1955) *J. Appl. Phys.* **26**, 1302.
124. Hobson, J.P. and Edmonds, T. (1963) *Can. J. Phys.* **41**, 827.
125. Byvik, C.E. and Bradford, J.M. (1966) *Space Sim. Conf., NASA Langley Research Center, Houston*, 209.
126. Ishikawa, K. (1965) *Jap. J. Appl. Phys.* **4**, 461.
127. Leck, J.H. and Carter, G. (1960) *Trans. 1st Int. Congr. Vac. Technol.*, 463.
128. Alpert, D. (1953) *J. Appl. Phys.* **24**, 860.
129. Bills, D.G. and Carleton, N.P. (1958) *J. Appl. Phys.* **29**, 692.
130. Cobic, B., Carter, G. and Leck, J.H. (1961) *Br. J. Appl. Phys.* **12**, 288.
131. Hobson, J.P. (1961) *Vacuum* **11**, 16.
132. Blears, J. (1947) *Proc. Roy. Soc. A.* **188**, 62.
133. Haefer, R.A. and Hengevoss, J. (1960) *Vac. Symp. Trans. Amer. Vac. Soc.*, Pergamon, New York, 67.

4 Cold-cathode ionization gauges

4.1 The development of cold-cathode (crossed-field) gauges

The practicability of using the cold-cathode glow discharge to measure gas pressure in the region below 10^{-3} mbar was first realized by Penning[1], who used a strong magnetic field to increase the electron path length from cathode to anode, so raising the ionization to a measurable level. In 1937 he described a practical gauge in which a discharge took place between an anode (made in the form of a loop), and two zirconium discs connected together electrically (one on each side of the ring) acting as cathode. The geometrical arrangement is shown in Figure 4.1. A permanent magnet was placed with its field parallel to the axis of symmetry. With a field strength of about 0.04 tesla and a constant anode–cathode voltage, the discharge current was found to be proportional to pressure in the range 10^{-5} to 10^{-3} mbar. At constant pressure, the current increased with increasing voltage; at 2000 V (the recommended operating value) the current was approximately 1 mA at a nitrogen pressure of 10^{-3} mbar.

Although the exact mechanism of this discharge is not completely understood, the general principles are well defined. An electron originating from either section of the cathode is prevented from going directly to the anode by the magnetic field. It travels instead in a helical path back and forth in the potential trough between the cathodes, eventually drifting out to the anode. (An electron of energy 2000 eV travelling at right angles to a magnetic field of 0.05 tesla moves in a circle of diameter 4 mm). Due to the long path length, there is a high probability of an ionizing collision with a gas molecule, even at

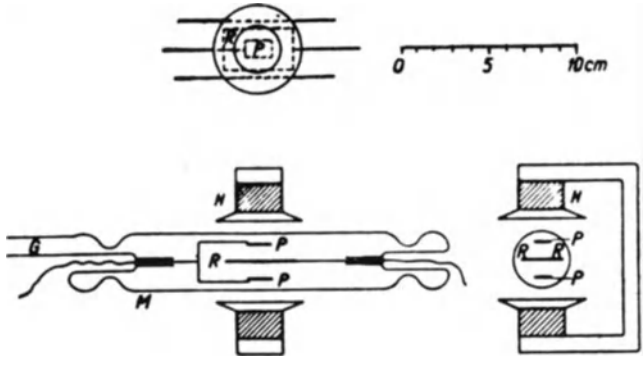


Figure 4.1 Cold cathode ionization gauge with ring electrodes designed by Penning in 1937[1].

very low pressure. Because of their large mass, the positive ions are virtually unaffected by the magnetic field and therefore travel directly to the cathode. The electrons, however, move in helical paths with the original electron and eventually, possibly after producing further ionization, reach the anode. Above a threshold (or 'striking') pressure, secondary electrons released from the cathode by the positive ion bombardment serve to build up and maintain the discharge. Thus, with the magnetic field a glow-type discharge is maintained, even when the electron mean free path in the gas is many times larger than the distance between anode and cathode. Investigating the mechanism of this discharge, Dumas[2] and Redhead[3] detected electromagnetic radiation which occurred at a frequency dependent only upon the magnetic field strength. This radiation arises largely from the rotation of the electrons in circular paths perpendicular to the magnetic field. Since the time of revolution T depends only upon the magnetic field B ($T = 2\pi m/Be$ for a particle with a mass-to-charge ratio m/e), the radiation must be at a discrete frequency. In the experiments made by Dumas,[2] the measured values of frequency corresponded with those predicted by this simple theory.

The discharge is unfortunately not perfectly stable, there being occasional unpredictable and sudden changes of between 2 and 5% in the current intensity. This instability appears to be present to a greater or lesser extent in all gauges, and is a serious disadvantage, as it sets a limit to the accuracy of measurement. Furthermore, with this early design, the discharge becomes erratic below 10^{-5} mbar, and is often extinguished completely at 10^{-6} mbar. In 1949, Penning and Nienhuis[4] modified the electrode structure with the aim of increasing the active volume of the discharge, and consequently the sensitivity. They replaced the ring anode with a cylinder extending almost to the cathodes; the electrodes thus formed a closed box. In the new gauge, the sensitivity is increased by a factor of 10 (approximately 1.0 mA at 10^{-4} mbar for 2000 V), the useful linear range extended to below 10^{-6} mbar, and the erratic behaviour reduced. Serious discontinuities of the type common in the ring anode gauge were only observed at pressures above 10^{-4} mbar.

Workers in other laboratories quickly followed the lead given by Penning in developing gauges of the same basic type, but with quite different electrode geometries. The investigations of Beck and Brisbane[5], Conn and Daglish[6] and Reynolds and Lipson[7] give a good indication of this wide range. Beck and Brisbane[5], for example, showed that the stable discharge could be set up with a long straight fine wire anode surrounded by a complicated cylindrical cathode. In this gauge, the magnetic field (again of about 0.05 tesla) was parallel to the wire anode. It was reported to give a truly linear relation between output current and pressure, and to have a much better 'starting characteristic' than gauges with the Penning-type geometries.

Hobson and Redhead[8–10] in their thorough investigation of cold cathode gauges have paid particular attention to behaviour at very low pressures. They have modified Penning and Beck and Brisbane gauges. The two gauges

constructed by Hobson and Redhead are shown in Figures 4.2 and 4.3. These have, for obvious reasons, been designated the inverted magnetron gauge (i.m.g.) and the magnetron gauge (m.g.). One of the modifications in design (important for very low-pressure measurement) is the introduction of cathode guard rings (auxiliary cathodes). These guards (held at cathode potential) prevent field emission currents from the cathode reaching the anode. Without this modification, the field emission would be comparable with the main discharge current at very low pressure, and thus set a lower limit to the

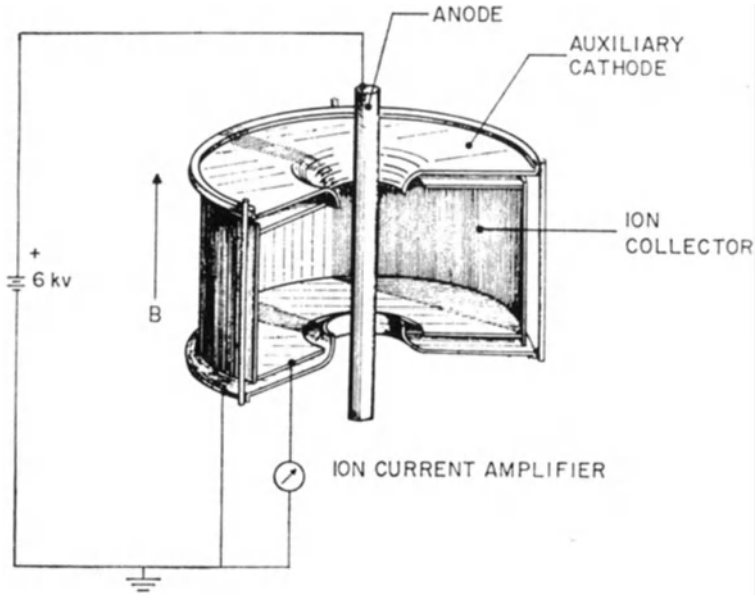


Figure 4.2 Cutaway diagram of the inverted magnetron gauge designed by Hobson and Redhead[9].

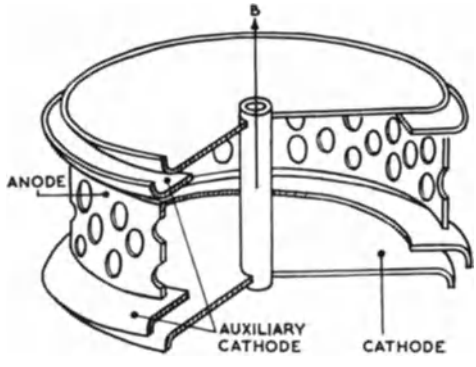


Figure 4.3 Cutaway diagram of the magnetron gauge designed by Redhead[10].

operating range. This would be analogous to the 'X-ray' background in the thermionic cathode gauges. The cathode in the inverted magnetron gauge is a short cylinder 30 mm in diameter with its axis parallel to the magnetic field, and the anode is a rod 1 mm in diameter passing axially through the holes in the cathode cylinder. As can clearly be seen from Figure 4.2, the auxiliary cathode is in the form of a box, to which have been welded short tubes to extend inside the end plates of the cathode. This gauge operates with a magnetic field of 0.2 tesla and an anode–cathode potential difference of 6000 V. In the alternative (magnetron gauge) design, the anode consists of a cylinder 20 mm long by 30 mm in diameter perforated to improve gas flow. The spool-like cathode, consisting of an axial cylinder 20 mm long and 3 mm diameter, is welded on to circular end discs. This gauge is normally operated with a magnetic field of 0.1 telsa and an anode–cathode potential difference of 6000 V.

These gauges operate on the basic Penning principle: electrons, trapped by the crossed electric and magnetic fields, thus travelling large distances between cathode and anode and setting up an electron plasma in the ionization chamber. The sensitivities are all of the same order of magnitude (approximately 10 μA cathode current at 10^{-6} mbar) with the general behaviour following the Penning principle. Because the electrons are more efficiently trapped than in the early Penning gauges, two important advantages follow. Firstly, the starting characteristics are improved, as was pointed out by Beck and Brisbane[5]. In fact, it is reported by Redhead that at no time have these gauges failed to 'strike', even at pressures below 10^{-8} mbar. Furthermore, the relations between pressure, magnetic field and striking voltage follow theoretical predictions. Secondly, the discharge is stable to much lower pressures; experiments have shown stability to extend to far below 10^{-10} mbar. Hobson and Redhead[8] carried out indirect calibrations on an ultra-high vacuum plant with liquid helium refrigerant traps, and are confident that the calibration of the magnetron gauge can be extrapolated to 10^{-13} mbar. In their opinion, this gauge can be used as a measuring instrument over the whole pressure range of 10^{-4} to 10^{-13} mbar.

A number of commercial instruments have been designed for use both in industry and the research laboratory, where cold cathode gauges are used in large quantities to become one of the most popular of all ionization gauges. Some have a particularly rugged construction especially suitable for industrial use, and others can be used in ultra-high vacuum applications. For the former application, attention has been paid to ensure that the gauge electrodes can be cleaned relatively easily should they become contaminated. Metal and carbon deposits always constitute a problem if these gauges are operated for long periods at high pressure (say above 10^{-5} mbar). For ultra-high vacuum, in addition to being bakeable the gauge must be designed in such a way that stray field emission currents do not interfere with the discharge current measurement. Also, the discharge must be maintained down to the lowest pressures.

4.2 Commercial gauges for high- and ultra-high-vacuum applications

Good examples of commercial gauges are the Edwards High Vacuum International range CP25K (for industrial medium- and high-vacuum use) and the Balzers type IKR020 (for high- and ultra-high-vacuum use). One of the gauges from the Edwards range is illustrated in Figure 4.4. All gauges in this range are 'inverted magnetrons' designed to operate over the range 10^{-7} to 10^{-2} mbar. They have a central rod anode held at about 2.5 kV positive to the earthed disc cathode. The stainless steel spring to hold the cathode in place, the disc at the end of the anode to shield the glass-insulated feed-through from metal spluttering and the permanent magnet providing an axial magnetic field of 0.1 tesla can be seen clearly in Figure 4.4. A special feature of the design illustrated is the open construction of the electrodes, giving a high conductance from the discharge chamber to limit the pressure transients caused by

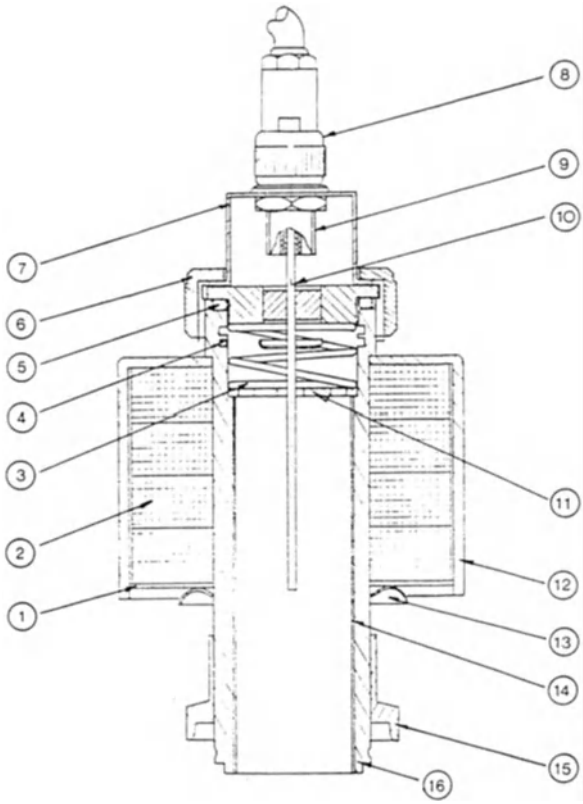


Figure 4.4 Cold cathode gauge type CP25.3 manufactured by Edwards High Vacuum International. (1) Magnet plate (thin); (2) magnets 4; (3) spring; (4) radioactive insert; (5) 'O' ring; (6) plastic nut; (7) cathode plate; (8) lead assembly; (9) connector type; (10) anode assembly; (11) cathode plate; (12) moulded cover; (13) starlock washer; (14) insert S/S sleeve; (15) muff coupling; (16) body tube.

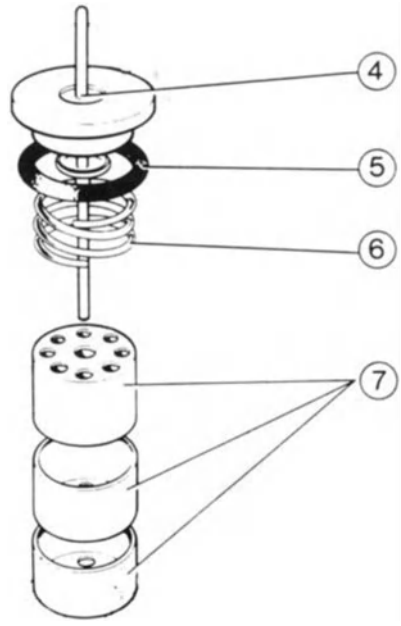


Figure 4.5 Electrode assembly of the Edwards High Vacuum International cold cathode gauge type CP25K. (4) Anode assembly; (5) 'O' ring; (6) spring; (7) cathode cups.

gas bursts. Other, slightly more complex, versions of this gauge have three cathode discs with the configuration shown by the expanded diagram (Figure 4.5). Because of its better stability, this gauge is recommended for general vacuum use, although, in spite of the rings of holes in the cathode discs, gas conductance from the ionization chamber is relatively low. If the gauges are switched on at low pressure (below about 10^{-6} mbar) there may be a delay of many minutes before the gauge 'strikes' to establish the discharge. To reduce this delay, and therefore the inconvenience and the possibility of serious errors being made (before striking, the gauge indicates a pressure of less than 10^{-7} mbar), Edwards offer the option of a small radioactive source of about 1 μCi in the discharge region. This takes the form of a nickel ring containing a nickel-63 radioactive insert mounted into the gauge body, as shown in Figure 4.4. A special feature of the rugged design of the CP25K series is the ease with which the electrode assembly can be dismantled, cleaned and re-assembled after it has become contaminated by prolonged use.

The cross-section drawing in Figure 4.6 shows the precise concentric construction of the Balzers IKR020 gauge, with the magnetic field, derived from the cylindrical permanent magnet, well aligned with the geometrical axis. This inverted magnetron gauge has a molybdenum anode mounted along the central axis of the stainless steel vacuum chamber. When copper gaskets are

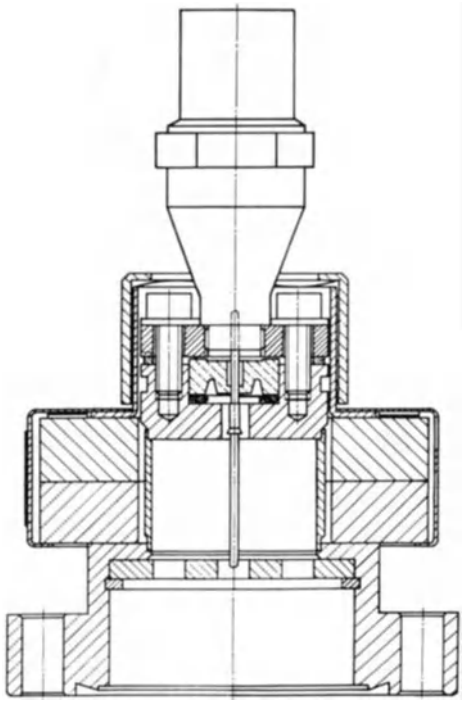


Figure 4.6 Drawing, in cross-section, of the Balzers IKR020 inverted magnetron gauge.

used both for the ceramic insulator and for connection to the vacuum system (via a standard 70 mm o.d. UHV flange) the baking can be to 250 °C. The measuring range is 5×10^{-3} mbar to 1.0×10^{-11} mbar (nitrogen equivalent). For pressures below 2.5×10^{-5} mbar, the current is measured at a constant voltage of 3.3 kV. Above 2.5×10^{-5} mbar, the voltage is measured at constant current (100 μA) in order to reduce contamination. The curves in Figure 4.7 show the corrections that must be applied to give a true pressure reading when this, or other, cold cathode gauges are operated in gases other than air, nitrogen, oxygen or carbon monoxide. The operation of the IKR020 in a practical UHV system has been described by Angerth *et al.*[11] who operated in the range 10^{-5} to about 10^{-10} mbar.

The absence of thermionic filaments in these gauges is an obvious advantage, making them particularly attractive for many applications. It leads to a simplification of the protective circuits and to the general operation and control. Modern permanent magnets have helped to reduce considerably the overall size of the gauge head. Set against these advantages, it must be recognized that some discontinuities in output signal must be expected. Consequently, the cold-cathode are less precise than the hot-cathode gauges.

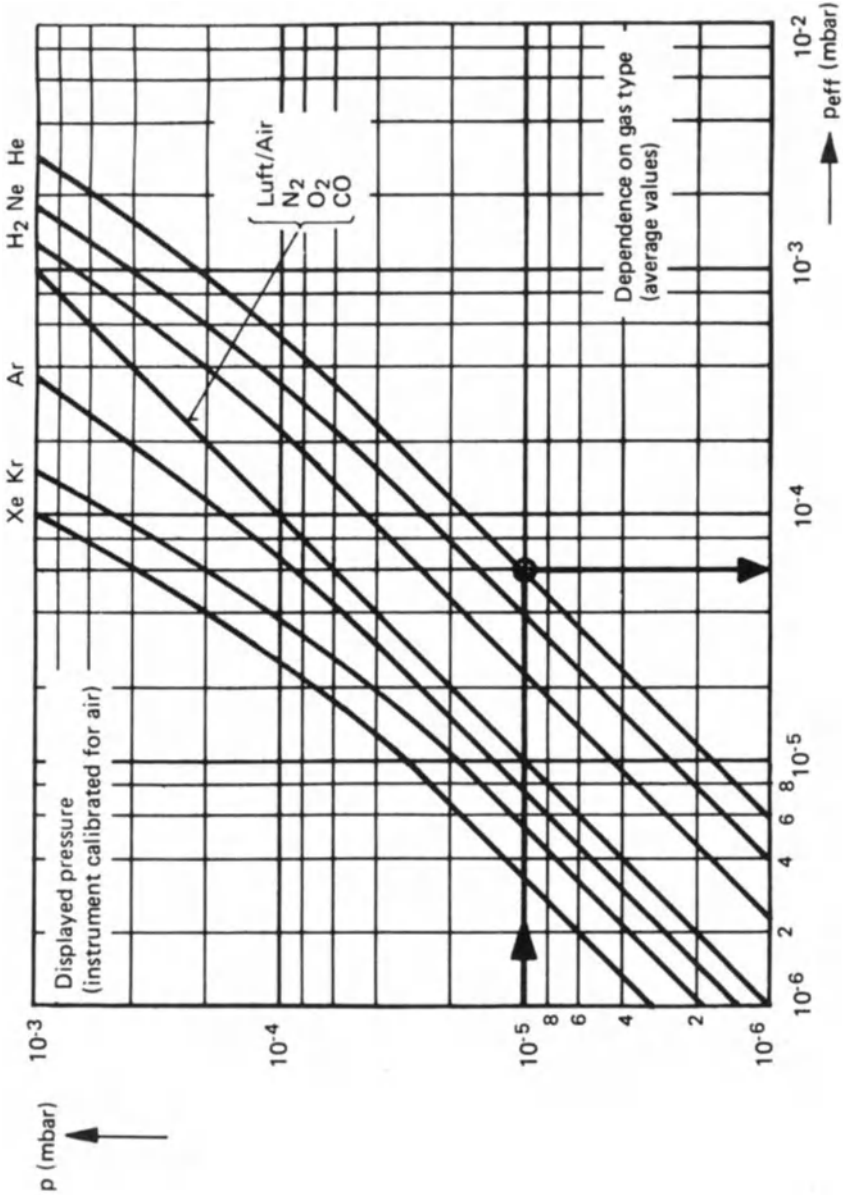


Figure 4.7 Correction factors to be applied to the Balzers IKR020 gauge when it is operated in xenon, krypton, argon, hydrogen, neon or helium (calibration in air). For the example illustrated on the graph, the gauge is operating in helium with a 'displayed pressure' of 10^{-5} mbar. The true pressure is obviously 6×10^{-5} mbar.

Table 4.1 Effective 'pumping speed' of various types of cold cathode ionization gauge.

Reference	Gas	Type of gauge	Effective pumping speed (litre $s^{-1} \times 10^3$)
12	Helium	Inverted magnetron	30
	Helium	Modified magnetron	200
13	Nitrogen	Modified magnetron	140
	Oxygen	Modified magnetron	150
14	Helium	Magnetron	170
	Nitrogen	Magnetron	250
15	Air	Penning	250

It is not possible to quote figures for long-term stability with any authority. Also, Angerth *et al.*[11] warned that the gauge should not be operated above 10^{-5} mbar for long periods if a reliable performance is required at very low pressure. It is also interesting to compare the reported measurements of 'effective pumping speed' for the cold cathode gauges, set out here in Table 4.1, with the corresponding values for the hot cathode gauges as presented in Table 3.3 (Reference 121 of Chapter 3).

References

1. Penning, F.M. (1937) *Physica* **4**, 71; (1937) *Philips Tech. Rev.* **2**, 201.
2. Dumas, G. (1955) *Rev. Gen. Elect.* **64**, 331.
3. Redhead, P.A. (1958) *Can. J. Phys.* **36**, 255.
4. Penning, F.M. and Nienhuis, K. (1949) *Philips Tech. Rev.* **11**, 116.
5. Beck, A.H. and Brisbane, A.D. (1952) *Vacuum* **2**, 137.
6. Conn, G.K.T. and Daglish, H.N. (1954) *J. Sci. Instrum.* **31**, 433.
7. Reynolds, J.H. and Lipson, J. (1954) *Rev. Sci. Instrum.* **25**, 1029.
8. Hobson, J.P. and Redhead, P.A. (1960) *Adv. Vac. Sci. Technol.* **1**, 384.
9. Hobson, J.P. and Redhead, P.A. (1958) *Can. J. Phys.* **36**, 271.
10. Redhead, P.A. (1960) *Adv. Vac. Sci Technol.* **1**, 410.
11. Angerth, B., Gontero, G. and Wahl, H. (1984) *Vacuum* **34**, 749.
12. Kornelsen, E.V. (1960) *Trans. 7th Nat. Symp. Vac. Technol.* 29.
13. Rhodin, T.N. and Rovner, L.H. (1960) *Trans. 7th Nat. Symp. Vac. Technol.*, 228.
14. Barnes, G., Gaines, J. and Kees, J. (1962) *Vacuum* **12**, 141.
15. Leck, J.H. (1953) *J. Sci. Instrum.* **30**, 271.

5 Gauge calibration

5.1 Basic considerations

There are three alternative techniques available for use in industrial and research laboratories where the calibrations can be carried out (i) by comparison with a 'transferable' or 'reference' gauge which is subject to periodic calibration in a National Standard or equivalent laboratory; (ii) against a fundamental standard which may be a true pressure (for example a McLeod) gauge; (iii) on a plant especially designed to produce precise pressures. Due to the extensive developments in the design and construction of the reference gauges (the capacitance manometer, section 1.3, and the spinning rotor gauge, section 1.4) it is the former technique which has the greatest attraction for most vacuum workers. With the comparison techniques, calibration of practical working gauges, such as the ionization and thermal conductivity, can be made down to pressures of 10^{-4} mbar, with uncertainties of the order of $\pm 5\%$. It requires a considerable investment in both skill and finance to obtain a significant improvement in this performance.

5.2 Calibration against the transfer gauge

This is in principle a straightforward technique, the only basic requirement being for a 'test chamber' in which the pressure can be varied throughout the working range for the gas or gases of interest. The obvious requirements for an accurate calibration are:

(i) The ultimate pressure in the test chamber should be low, significantly less than the lowest point in the calibration scale
(ii) The pressure inside the test and reference gauges should be the same throughout the calibration
(iii) No gauges should introduce impurities to the system to any significant level
(iv) The temperature of all parts of the system should be known accurately
(v) The flow of test gas should be controlled, so that the pressure in the vacuum chamber can be varied or held constant as required over the whole pressure range of calibration.

Nash and Thompson[1] indicated the important design features in their plant intended to cover the operating range 10^{-6} to 1.0 mbar. They followed the constant-flow principle, and designed equipment similar to that used in other laboratories[2,3] to provide calibrations traceable to a standard laboratory.

They used a 100-litre chamber of conventional stainless steel construction evacuated with a liquid-nitrogen-trapped oil-diffusion pump. A variable conductance valve between the chamber and the cold trap allows the diffusion pump to be throttled at the higher calibration pressures. There is an odd number (nine) of gauge ports around the circumference of the chamber so that gauges do not face one another; this limits gauge interaction. These ports are on elbows equally spaced in a ring around the top of the vacuum chamber, so that the gas in the vicinity of the gauges is at a uniform pressure and in thermal equilibrium with the walls of the chamber. The chamber is pumped by a side port so that there is no 'line of sight' between surfaces at liquid nitrogen temperature and the main parts of the vacuum vessel. The gas input is on the centre line of the chamber, and arranged so that molecules will have to undergo at least three collisions with the walls before they can enter any gauge tubulation. The test gas flows from an auxiliary chamber at a pressure of about 500 mbar via a servo-controlled piezoelectric leak, to maintain a constant predetermined pressure of test gas superimposed upon the base pressure of about 3×10^{-8} mbar. Commercial conventional triode ionization gauges are used as the reference at low pressures. For routine work, Nash and Thompson[1] recommended that these be calibrated against 'the standard' every six months. To cover the high-pressure end of the scale, a capacitance manometer type 220 Baratron (MKS Instruments, Burlington, Mass.) is used as reference.

Nash and Thompson[1] were able to check the performance of their plant using the NPL series expansion system (see section 5.4) over a period of two years, and as a result have been able to make a good estimate of the uncertainties of their plant. The figures are reproduced in Table 5.1. It is interesting to note the authors' caution in estimating the errors in the ionization gauge in view of the experience in their own, and other, Standards Laboratories. As they point out, 'by using a spinning rotor gauge as reference the "reference gauge instability term" can probably be reduced for the range 10^{-3} to 10^{-4} mbar.' (Reference to Chapter 1, section 1.4, indicates a figure of about $\pm 1\%$ for this term in the range 10^{-3} to 10^{-4} mbar). Naturally, to maintain the values of uncertainty quoted here, great care must be taken in the operation of the reference gauges; for example, with ionization gauges precision instrumentation must be used to measure electron and ion currents,

Table 5.1 Estimate of uncertainties by Nash and Thompson[1].

Source	Pressure (mbar) 10^{-6}	10^{-3}	1.0
Non-uniformity of pressure distribution	$\pm 1\%$	$\pm 1\%$	$\pm 1\%$
Base pressure and gas purity	$\pm 1\%$	–	–
Reference gauge instability	$\pm 4\%$	$\pm 4\%$	$\pm 2\%$
Pressure fluctuations	$\pm 2\%$	$\pm 1\%$	$\pm 1\%$
Total (linear sum)	$\pm 8\%$	$\pm 6\%$	$\pm 4\%$

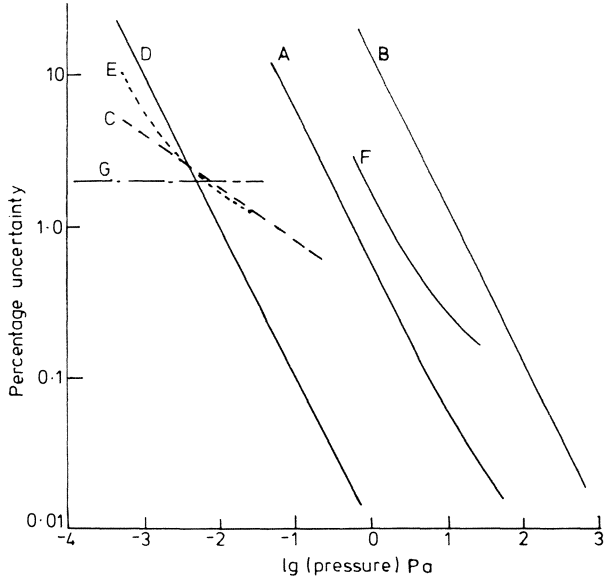


Figure 5.1 Uncertainty as a function of pressure for various absolute gauges[4]. *A*, oil U-tube manometer; *B*, mercury U-tube manometer; *C*, McLeod gauge; *D*, interferometric oil manometer; *E*, piston manometer (low range); *F*, piston manometer (high range); *G*, levitation gauge (after Poulter).

and with the spinning rotor gauges there must be a full understanding of 'zero offset errors'.

5.3 Comparison with absolute gauges

A number of types of gauge can be regarded as absolute when they are carefully constructed and proper precautions are taken in their use. Poulter[4] lists the most important and indicates the errors in operation for each gauge. These are summarized in Figure 5.1. He makes special reference to the McLeod gauge, indicating the precautions that must always be taken. He notes that performance to the level quoted in Figure 5.1 will be obtained only with gauges specially designed for calibration work.

5.4 Series expansion techniques

The principle of operation is simple, because a precise reduction in pressure can be obtained by expanding a small volume of gas at a known relatively high pressure into a larger volume. The pressure after expansion can be calculated, provided the dimensions of the two volumes are known and the gas can be assumed to obey Boyle's law. Either single or multiple expansions can be used, the technology being essentially the same in other cases. Elliott *et al.*[5] have

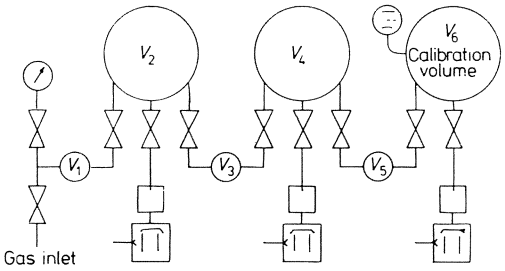


Figure 5.2 Schematic diagram of the multiple expansion plant[4,5].

described a triple expansion calibration plant built at the NPL in London, following experience gained by Barton and Chubb[6] at the UKAEA. A total expansion of about 10^{+7} was obtained over three stages, using large and small volumes of 6000 cm^3 and 25 cm^3 respectively. Thus, by starting with an initial pressure of 50 mbar, they could carry out calibration measurements down to the 10^{-6} mbar range. Figure 5.2 is a schematic diagram of this plant, showing the relationship of the various volumes and control valves. Each large volume is attached to a vacuum pump and separated from it by an isolation valve. An operating sequence is started by evacuating the volumes until a specified base pressure is reached and the rate of rise of pressure is acceptably low. The volumes are then isolated from the pumps. The first small volume V_1 is then filled with gas to the pressure p_1, which is measured with high precision on an absolute or precision calibrated gauge. The gas is then allowed to expand into the first large and second small volume. The second small volume V_3 is then isolated from the other volumes and expanded into the next large and small volumes. After the process has been repeated for the third time, the pressure increment p that has been produced in the final calibration volume V_6 is given by

$$p = p_1 \cdot \frac{V_1}{V_1 + V_2 + V_3} \cdot \frac{V_3}{V_3 + V_4 + V_5} \cdot \frac{V_6}{V_5 + V_6} \tag{5.1}$$

Poulter in his review[4] describes the procedures that must be carried out to obtain the basic parameters, in particular the precise values of the volumes V_1 to V_6. He estimates that, with meticulous attention to detail, pressures down to 10^{-8} mbar can be generated, with an uncertainty not exceeding 1%. Because of the departure from the ideal gas laws for the majority of gases, and problems with sorption and desorption, the list of preferred gases for this technique is limited to helium, neon, argon, krypton, xenon, nitrogen and methane.

5.5 Dynamic flow techniques

Dynamic expanders or orifice flow devices pass a gas at a known rate through a known conductance (usually an orifice), to generate a known pressure

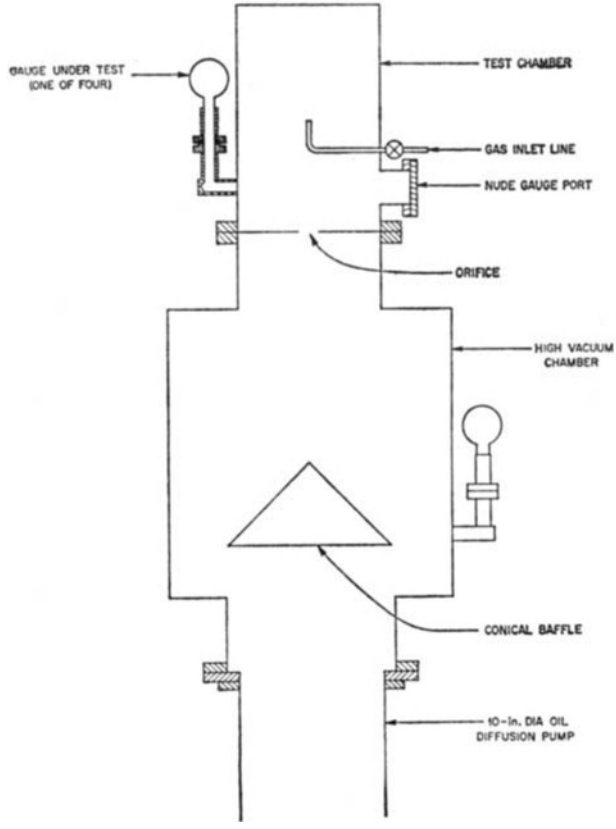


Figure 5.3 Dynamic calibration plant designed by Normand[7].

difference quite analogous to generating a voltage difference by passing a current through a resistor. Figure 5.3 shows the experimental system described by Normand[7], which is built around an orifice of accurately known dimensions. It illustrates the technique used successfully by many other workers, for example Florescue[8], Bannenberg and Tip[9], Hayward and Jepson[10], Bennewitz and Dohmann[11], Owens[12], Roehrig and Simons[13], Hultzman and Krause[14], Hojo *et al.*[15], Poulter[16], and McCulloh *et al.*[17]. The system has the virtue of being relatively easy to set up; also, it can be operated to give background base pressures of 10^{-10} mbar or lower, using standard ultra-high vacuum techniques. When test gas flows at a rate Q litre mbar s^{-1} through the system, the equilibrium pressures above and below the orifice p_1 and p_2 respectively are related by the fundamental equation

$$Q = c(p_1 - p_2) \tag{5.2}$$

giving

$$p_1 = \frac{Q}{C} \cdot \frac{1}{1 - p_2/p_1} \tag{5.3}$$

neglecting, of course, the background pressure. C represents the conductance of the orifice. The ratio p_2/p_1 can be determined directly with two identical gauges, one mounted above and the other below the orifice, as shown in Figure 5.3. Alternatively, p_2/p_1 can be measured with a single gauge switched between the two sections of the chamber. In normal operation it is practicable to use a diffusion or turbomolecular pump, with a speed large compared with the conductance of the orifice in order to make $p_1 \gg p_2$, so that small errors in determining p_2/p_1 are unimportant.

The conductance C can be calculated from the fundamental laws of gas kinetics (see for example Knudsen, introductory chapters of ref. 18, or Dushman, Chapter 2, section 2 of ref. 19) and is given by

$$C = \left(\frac{RT}{2\pi M}\right)^{1/2} A \quad \text{l.s}^{-1} \tag{5.4a}$$

$$= 0.03638(T/M)^{1/2} A \quad \text{l.s}^{-1} \tag{5.4b}$$

Where T is the gas temperature (K), M the molecular mass of the gas (reference O = 16), and A mm^2 the cross-sectional area of the orifice. Equation (5.4) is only exactly true when the diaphragm containing the orifice is both infinitely large and infinitely thin. In practice, the greatest error arises because the diaphragm is not very large compared with the diameter of the orifice, but is, as in the apparatus shown in Figure 5.3, contained in a tube of finite length and diameter. Bureau *et al.*[20] have discussed the necessary modification to eqn (5.4) for the particular, but common, case where an orifice of diameter d mm is at the centre of a circular diaphragm, stretched across the centre of a tube of diameter D mm and length L mm. They introduced a correction factor K such that

$$C = K\left(\frac{RT}{2\pi M}\right)^{1/2} A \quad \text{l.s}^{-1} \tag{5.5}$$

where

$$K = \left\{\frac{1 - d/D}{K_0} + \frac{3}{4}\frac{Ld}{D^2}\right\} \tag{5.6}$$

K_0 being a dimensionless constant increasing in value from 1.000 for $d/D = 0$ to 1.017 for $d/D = 0.3$. (The values for K_0 for a complete range of d/D are given in tabular form by Bureau *et al.*[20].) This correction obviously holds for the limiting case; for when $d/D \to 0$, $K \to 1.0$ and when $d/D \to 1.0$, eqn (5.6) becomes simply the conductance of a tube of length L and diameter D.

If, as can easily be arranged in practice, the ratio d/D is kept small, the

correction factor K is also small. For example for $d/D = 0.2$ and $L = 2D$, $K = 1.01$. Thus even if eqn (5.5) is not exactly correct, there will be a negligible error introduced into eqn (5.4). Any error due to the finite thickness of the diaphragm may be neglected, as it is in practice easy to ensure that the thickness is never more than 1% of the diameter of the orifice. By careful design it is possible to make a system where conductance $C = 10$ litre s^{-1} for nitrogen can be calculated with an uncertainty of much less than $\pm 1.0\%$.

The system installed at the NBS[17] in Washington is a good example of a sophisticated orifice gas flow primary standard used in the high and ultra-high vacuum ranges. Multiple injection points are arranged for the test gas inlet so that the range can be extended to the lowest pressures. The vacuum chamber has two cylindrical halves 270 mm in diameter and 340 mm long, separated in the middle by a wall containing the orifice. The upper half includes eight ports for the attachment of the gauges to be calibrated, and the gas inlet with baffle. The lower half includes a second gas inlet and baffle, and is pumped by a nominal 0.5 m^3 s^{-1} turbomolecular pump. The chamber and all attached gauges are routinely baked between 200 and 250 °C at the beginning of each calibration series. The orifice is fabricated in a separate orifice plate, which is sealed in the wall between the chamber halves by a gallium seal. During pump-down and bake-out the orifice plate can be lifted away from the seal, opening a 125-mm diameter hole between the chamber halves. Base pressures for the system are typically 10^{-10} mbar, almost exclusively hydrogen. The 11.25-mm diameter orifice was constructed by drilling a 11.2-mm diameter hole through a 0.74-mm thick stainless steel plate and then lapping the edges of the hole with a 0.625-inch diameter ball bearing from both sides until the concave spherical surfaces met at the centre of the plate. The two lapped surfaces form a sharp burr-free edge. The existence of spherical surfaces simplifies the calculation of the transmission probability above and below the orifice. The total uncertainty in the calculation of the conductance is estimated to be 0.8%, most of which is due to uncertainties in the measured dimensions of the orifice and duct. A spinning rotor gauge is connected through bakeable valves to the upper and lower chamber halves to measure alternately the pressures above and below the orifice, to give an accurate value of the ratio p_2/p_1 for substituting in eqn (5.3). The accuracy of the ratio p_2/p_1 is limited by the pressure gradients in the lower chamber; however, as the value is approximately 0.04 the error introduced is negligible.

The lower useful limit of this standard is determined by (i) the hydrogen base pressure, and (ii) the difficulties of measuring very small gas flow rates. This second limitation can partly be circumvented by introducing the gas flow downstream of the orifice, thus extending the range by a factor of 26 (i.e. the ratio of p_1/p_2). However, because of the uncertainties of pressure gradients in the lower chamber, the overall accuracy must be reduced considerably. The workers at the NBS estimate the total uncertainties of measurement to be 2.5%, 1.5% and 3.4% at pressures of 10^{-3}, 10^{-6} and 10^{-8} mbar respectively.

They make the point that the problems of measuring the gas flow rate are most important in setting the above limits. A brief description of the flow meters used at the NBS, together with the other various techniques, are set out in the following section.

Interesting variations of this technique have been reported from a number of laboratories. Choumoff *et al.*[21] developed the technique of using a variable and accurately calibrated conductance in the gas flow to eliminate the need for a precise measurement of gas flow. Grosse and Messer[22] and Poulter[23] placed a cryopump immediately behind the orifice to ensure that any molecule passing through was immediately trapped at the cold surface (4.2 K). Fowler and Brock[24] and Grosse and Messer[25] extended operation into the ultra-high vacuum range by using a molecular beam and cryopump technique to inject very small quantities of gas into the calibration chamber. Grosse and Messer[25] estimated an accuracy of measurement with their system of $\pm 10\%$ at 10^{-10} mbar and $\pm 30\%$ at 10^{-12} mbar.

5.6 The measurement of gas throughput

If the dynamic flow calibration technique is to be used successfully, equipment must be available to measure gas flow rates, ideally from 10^{-2} to 10^{-8} litre mbar s^{-1}. Many systems have been developed from the basic 'variable volume–constant pressure' flow meter reported by Kaye[26] in 1927. The operating principles of his equipment are obvious from the diagram in Figure 5.4. Gas flow into the vacuum system is controlled by the variable leak valve, and the rate $PA\,ds/dt$ determined by measuring the velocity (ds/dt) of the mercury pellet as it is drawn along the capillary tube of cross-section A. The pressure P in the closed volume is normally atmospheric. The practical disadvantage of this apparatus is the erratic movement of the mercury, making small gas throughputs difficult to measure. An alternative technique, using oil instead of mercury as the piston, is indicated in Figure 5.5. Here a correction must be made to account for the change in pressure in the closed volume as the oil rises up the small-bore glass tube. (The type of meter shown in Figure 5.5 is not suitable for the measurement of the lowest flow rates, but because of its

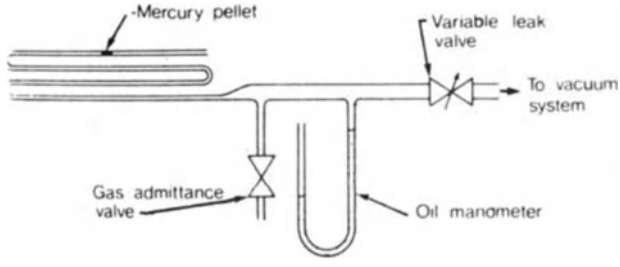


Figure 5.4 Simple 'mercury piston' flow meter designed by Kaye[26].

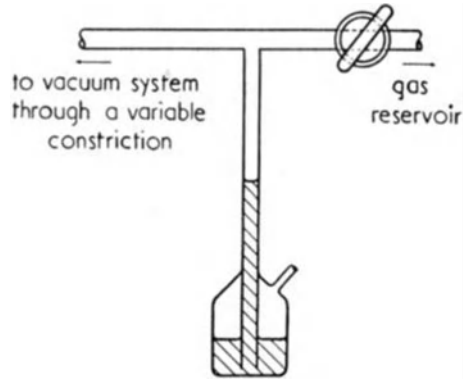


Figure 5.5 Simple flow meter in which the 'piston' is a low-vapour-pressure oil.

simplicity, it is extremely useful in other applications, in particular when measuring the characteristics of vacuum pumps.)

The range of flow meters is described in a comprehensive review by Peggs[27], who highlighted the work of Hayward and Jepson[10] in developing a reliable variable-volume meter to measure very low flow rates. The general arrangement of their flow meter is shown in Figure 5.6. The reservoir, consisting of two sections separated by the diaphragm of a sensitive differential pressure manometer, contains the test gas at a pressure p_0. Before a gas throughput measurement, the bypass valve is opened to equalize the pressure in the two sections. A measuring sequence is initiated by opening the variable leak valve to establish a suitable gas throughput into the vacuum system. The bypass valve is then closed, isolating the reference section of the flow meter which contains the pressure gauge, from the working section which contains the volume displacers. The ensuing loss of gas from the working section is compensated by varying the volume of this section to maintain a null pressure differential across the diaphragm of the manometer. Hayward and Jepson[10] estimate that, with this equipment, flow rates down to 10^{-5} litre mbar s^{-1} can

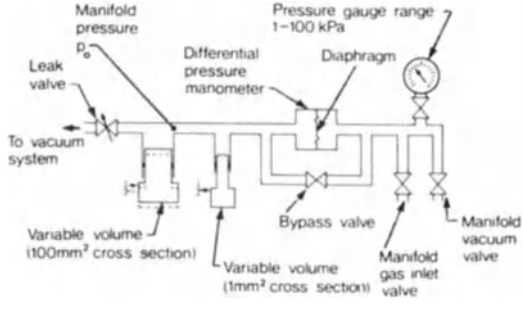


Figure 5.6 Schematic diagram of the flow meter designed by Hayward and Jepson[10,27].

be measured with an uncertainly of not more than about 1% using the 1 mm^2 plunger displaced 25 mm over a 20-minute period.

Poulter[16] reported a similar precision with equipment of the same basic design, over an uncertainty of $\pm 1.0\%$ at 10^{-5} litre mbar s^{-1} and $\pm 0.4\%$ for flow rates in excess of 10^{-4} litre mbar s^{-1}. He noted that, even with the greatest care taken to achieve as near perfect temperature stability as possible, it is still the fluctuations in temperature which set the limit to the accuracy of measurement. Flow meters of the same type developed at the NBS to cover the range 10^{-2} to 10^{-6} litre mbar s^{-1} operate to an estimated uncertainty of 0.8% at the high and 2.0% at the low end of the scale. They are described in detail by McCulloh *et al.*[28]. An indication of the complexity of this equipment can be seen from Figure 5.7, which shows a cross-section of the 'variable volume', including the piston and its drive mechanism. The micrometer for the piston is advanced by a stepping motor to keep the pressure in the closed volume

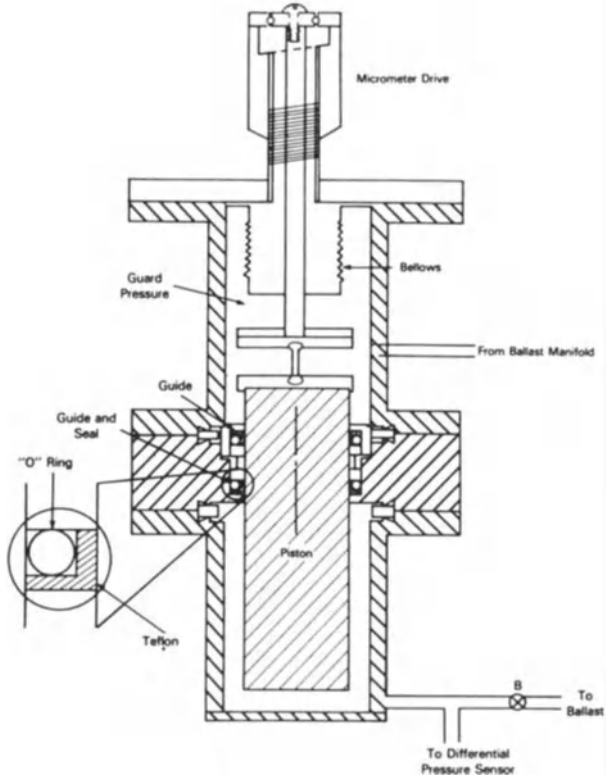


Figure 5.7 Detail of the variable volume, including piston and piston drive mechanism, used in the NBS flow meters. To minimize leakage past the seal, a guard pressure close to the pressure in the variable volume is maintained on the back side of the seal by the line connected to the ballast manifold[28].

constant. A feedback circuit, operating on the output of a 10-mbar differential capacity diaphragm gauge connected between the reference volume and the variable valve, controls this motor.

An obvious practical disadvantage of this family of flow meters is the fact that gas flow must be held constant for a considerable time to enable accurate measurements to be made. The uncertainties introduced by the need to measure accurately the pressure in the closed volume are almost certainly considerably less than those caused by small temperature drifts in the system. Peggs[27] stressed the need to take the utmost care to maintain a constant temperature during the whole of an experiment. He pointed out that, in a typical system, a temperature drift of 0.01 K min^{-1} will indicate an apparent gas throughput of about 5×10^{-6} litre mbar s^{-1}.

The flow meters described in Figure 5.6 can be simplified greatly if the variable (uncalibrated) leak valve between the manifold and the vacuum chamber is replaced by a fixed 'leak' of known conductance. Flow rates can then be calculated directly in terms of the pressure p_0. Practical instruments using this technique have been reported from a number of laboratories. As early as 1944, Blears used a small cylindrical porous plug as a fixed leak of very low conductance, both for gauge calibration and as a precise element in mass spectrometer handling plants. In experiments described in 1966, Christian and Leck[29] (following the example of Blears), used a cylindrical bar of silicon carbide 3 mm long and 3 mm in diameter which was found to have a conductance of 1.19×10^{-6} litre s^{-1} for argon. (For cylindrical porous plugs of this type, conductance is approximately proportional to the square of the diameter divided by the length.) They measured the conductance U by sealing a sample of gas at a pressure of about 200 mbar in a closed manifold, and then observing the exponential fall of pressure as gas was evacuated to high vacuum through the porous plug. As the pressure in the higher vacuum was completely negligible compared with that in the manifold p_0, the rate of flow was always given by $U.p_0$ (this is analogous to measuring high resistance by observing the discharge of a capacitor). A true exponential decay was noted for all the noble gases for manifold pressures p_0 not greater than 100 mbar, and further, conductance was found to be of the form $U = U_0(M)^{-1/2}$, giving a strong indication of true molecular flow. No significant changes in conductance were observed over a period of many months. Lawson[30] measured the temperature coefficient of U_0 as approximately $-0.001\,K^{-1}$ at room temperature.

Close *et al.*[31] designed and operated a dynamic flow calibration plant operating down to 10^{-8} mbar, also using a silicon carbide porous plug as the basic element in their flow meter. They indicate that, when measuring the manifold pressure p_0 with a precision instrument, errors in calibration should be no more than $\pm 2\%$.

The limitations of this technique are set by the stability of the porous plug and the precision to which its conductance can be measured. A judgement of these factors can be made from the values of U_0 quoted in Table 5.2, which

Table 5.2 Conductance of porous plug U and U_0 measured for the noble gases and nitrogen by Close *et al.*[31] and Christian and Leck[29]. U_0 defined by the relationship $U = U_0(M)^{-\frac{1}{2}}$ with the usual notation.

	Close *et al.*[31]			Christian and Leck[29]		
Gas	U $\times 10^6$	U_0 $\times 10^6$	% deviation from mean	U $\times 10^6$	U_0 $\times 10^6$	% deviation from mean
Helium	6.86	9.65	−1.4	3.77	7.54	−0.7
Neon	4.91	9.76	−0.3	1.70	7.65	+0.8
Argon	1.58	9.97	+1.8	1.19	7.51	−1.0
Krypton	1.07	9.78	−0.3	0.815	7.46	−1.7
Xenon				0.674	7.70	+1.4
Nitrogen	1.85	9.78	−0.1			

indicate not only the operation of the 'square root M law', but also the accuracy to which measurements can be made. Close[31] and his co-workers make the point that their measurements for the various gases continued over a period of approximately one year, giving a good indication of the long-term stability of the porous plug. Extremely good long-term stability has also been observed at Liverpool with silicon carbide plugs; for example, during 25 years' operation, the conductance of the 'reference plug' has not differed by more than $\pm 5\%$ from the initial value of 4.9×10^{-5} litre s^{-1} (Austin, unpublished). Fowler and Brock[24] found a plug of porous vycor glass both linear to 5000 mbar and stable to within 0.5% during two years' operation. This is inconsistent with the observations of Hultzman and Krause[14], who observed a steady drop in the conductance of their porous platinum plugs of 0.15% per day. They were, however, using very fine material giving low conductances (of the order of 10^{-8} litre s^{-1}).

The convenience of the simple and relatively inexpensive continuous flow calibration plant with gas input through a porous plug has been recognized in a number of laboratories. Owens[12] wrote in 1965: 'a major convenience of this method is the relatively short time required for gauge calibration; for example up to six gauges may be placed on the system simultaneously for calibration. After pumpdown and overnight bakeout the calibration of six gauges can be completed in one day. The fact that the system can be used for different gases is worthy of some merit'. Experience at Liverpool over many years has shown the technique to be reliable, simple to operate and to require only the minimum precautions in order to get reliable results. The long-term reproducibility of pressure over the range 10^{-4} to 10^{-6} mbar is estimated to be $\pm 5\%$.

References

1. Nash, P.J. and Thompson, T.J. (1983) *J. Vac. Sci. Technol.* **A1**, 172.
2. Buckingham, J.D. (1976) *Vacuum* **26**, 143.

3. Reich, G. (1980) *Proc. 8th Int. Vacuum Congr., Cannes*, **11**, 222.
4. Poulter, K.F. (1977) *J. Phys. E. Sci. Instrum.*, **10**, 112.
5. Elliott, K.W.T., Woodman, D.M. and Dadson, R.S. (1967) *Vacuum* **17**, 439.
6. Barton, R.S. and Chubb, J.N. (1965) *Vacuum* **15**, 113.
7. Normand, C. (1962) *Trans. 8th AVS Vacuum Symp. 1961*, **1**, Pergamon, Oxford, 534.
8. Florescue, N. (1962) *Trans. 8th AVS Vacuum Symp. 1961* **1**, Pergamon, Oxford, 504.
9. Bannenberg, J.G. and Tip, A. (1968) *Proc. 4th Int. Vac. Congr., Manchester 1968, Inst. of Phys. and Phys. Soc. Conf. Ser.* **6**, Addlard & Son, Surrey, 609.
10. Hayward, W. and Jepson, R. (1962) *Trans. 9th AVS Vacuum Symp. 1962*, Macmillan, New York, 459.
11. Bennewitz, H. and Dohmann, H. (1964) *Vakuumtechnik* **14**, 8.
12. Owens, C. (1965) *J. Vac. Sci. Technol.* **2**, 104.
13. Roehrig, J. and Simons, J. (1962) *Trans. 8th AVS Vacuum Symp. 1961*, **1**, Pergamon, Oxford, 511.
14. Hultzman, W., and Krause, L. (1974) *J. Vac. Sci. Technol.* **11**, 889.
15. Hojo, H., Ono, M. and Nakayama, K. (1977) *Proc. 7th Int. Vac. Congr., Vienna 1977*, Berger & Sohne, Vienna, 117.
16. Poulter, K.F. (1978) *Vacuum* **28**, 135.
17. McCulloh, K.E., Tilford, C.R., Wood, S.D. and Martin, D.F. (1986) *J. Vac. Sci. Technol.* **A4**, 362.
18. Knudsen, M. (1950) *Kinetic Theory of Gases*, 3rd edn., Methuen, London.
19. Dushman, S. (1949) *Scientific Foundations of Vacuum Technique*, John Wiley, New York.
20. Bureau, A.J., Laslett, L.J. and Keller, J.M. (1952) *Rev. Sci. Instrum.* **23**, 683.
21. Choumoff, P., Bernardet, H., Mativet, J. and Sauneuf, R. (1970) *J. Vac. Sci. Technol.* **7**, 270.
22. Grosse, G. and Messer, G. (1970) *Vacuum* **20**, 373.
23. Poulter, K. (1974) *J. Phys. E., Sci. Instrum.* **7**, 39.
24. Fowler, P. and Brock, F.J. (1970) *J. Vac. Sci. Technol.* **7**, 507.
25. Grosse, G. and Messer, G. (1981) *Vakuumtechnik* **30**, 226.
26. Kaye, G.W.C. (1927) *High Vacua*, Longman & Green, New York.
27. Peggs, G.N. (1976) *Vacuum* **26**, 321.
28. McCulloh, K.E., Tilford, C.R., Ehrlich, C.D. and Long, F.G. (1987) *J. Vac. Sci. Technol.* **A5**, 376.
29. Christian, R.G. and Leck, J.H. (1966) *J. Sci. Instrum.* **43**, 229.
30. Lawson, P.R.W. (1975) *Vacuum* **25**, 377.
31. Close, K.J., Vaugham-Watkins, R.S. and Yarwood, J. (1977) *Vacuum* **27**, 511.

6 Gas analysis in vacuum systems: magnetic, crossed-field and time-of-flight analysers

6.1 Introduction

The significance of gas analysis in vacuum systems is becoming more important as increasing demands are made upon the system designer. Some form of mass spectrometer capable of measuring the partial pressure of all the constituent gases and vapours from 10^{-4} mbar down to below 10^{-10} mbar should be attached to the system. Care must be taken to ensure that this instrument does not significantly modify the pressure equilibria. Mass spectrometers suitable for this work are ionization gauges in which positive ions formed by electron bombardment are resolved into a mass spectrum, the intensity of each component being measured separately. None of the instruments so far developed gives a true gas analysis, as all suffer from three fundamental disadvantages:

(i) They resolve into an 'ion mass' spectrum, not a 'gas composition' spectrum. For example, the gases nitrogen, carbon monoxide and ethylene all have their most intense signal at mass 28 amu (on the O = 16 atomic scale). Also, many gases, for example large organic molecules, give complex ion patterns which are not readily distinguishable. (See Figure 6.1[1], which presents the ion spectrum of butane, a relatively simple hydrocarbon.)

(ii) The sensitivity, as in all ionization gauges, is different for each gas, and unfortunately the relative sensitivity between gases is a function of the individual instrument design and operation.

(iii) The resolution of the mass spectrometer is not perfect; it usually decreases at the high end of the mass scale. In general, for a particular instrument an improvement in resolution is obtained only at the expense of a reduction in sensitivity.

As there are a number of published designs for various types of spectrometer suitable for general vacuum use, and numerous instruments are available commercially, it is worthwhile to analyse the advantages and limitations of each. The instruments developed for this work can be classified under one of the three following headings.

Magnetic deflection: the path followed by a beam of positive ions in a magnetic field is a function of the mass of the individual ions.

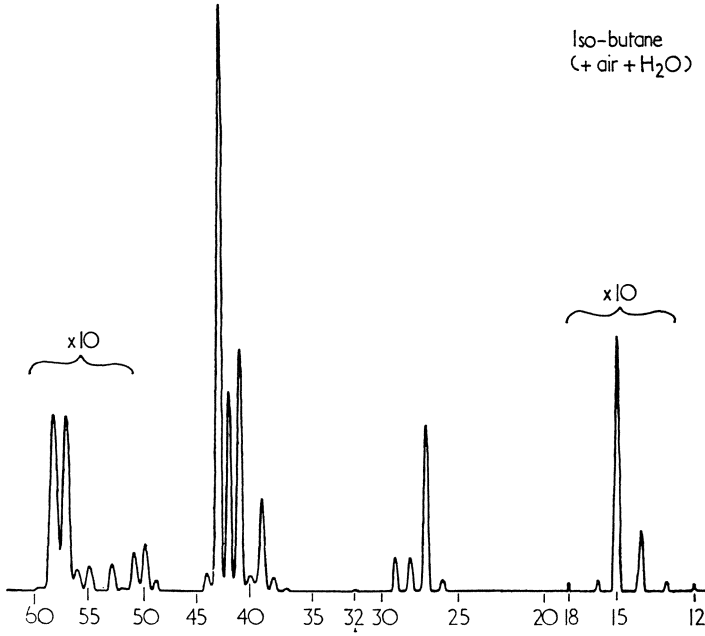


Figure 6.1 Mass spectrum of iso-butane with small impurities of air and water vapour[1].

Resonance: the resonant frequency of ions in an alternating electric field is dependent on the ion mass.

Time-of-flight: the time taken for positive ions of fixed energy to travel between two points is a function of their mass.

A brief description of the principles of each type, together with some practical examples of their operation, is given below. The quadrupole spectrometer, because of its special importance and widespread general use, is described separately in Chapter 7.

6.2 The magnetic deflection mass spectrometer

This instrument has undergone almost continuous development since the pioneering work of Aston and Dempster in the 1920s. Today it is used extensively in many branches of chemistry, physics and engineering as a gas analyser. A brief description of the principles will, however, be given, in order that the operating characteristics can be presented for comparison with the other techniques. Figure 6.2*a*, *b* shows the basic electron and ion optics for the two instruments which are used for low-resolution applications. There can be variations in detail, but these are relatively unimportant and do not affect the characteristics in any fundamental manner. The process starting point is the

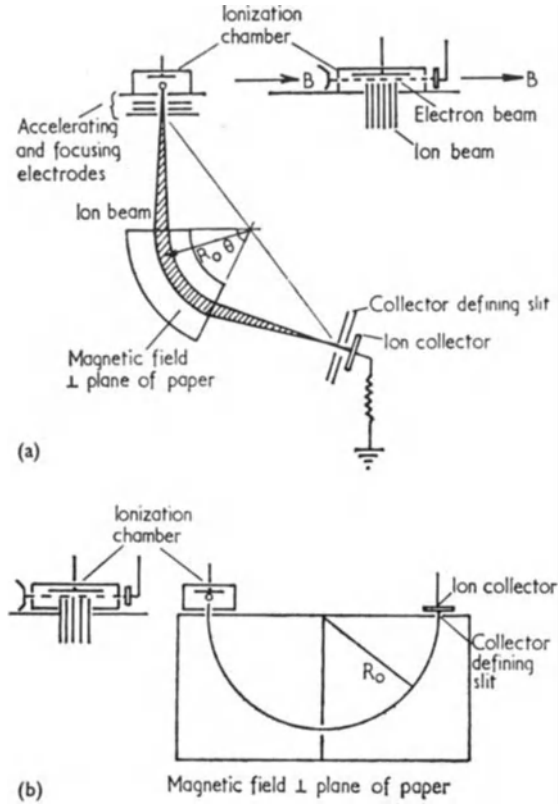


Figure 6.2 Electron and ion optics for the magnetic deflection mass spectrometer: (*a*) sector instrument; (*b*) 180° deflection instrument.

ionization chamber, which is a small box where positive ions are formed by electron impact as in the conventional ionization gauge. Provided the gas conductance of the 'box' to the rest of the vacuum system is adequately high, these ions will be typical of the residual gas in the system. Unlike the ionization gauge, the ionizing electrons are mono-energetic and confined to a narrow beam which passes through the box without any collisions with the walls. (If the ionization chamber lies outside the main magnetic field, focusing is usually assisted by a small permanent magnet, as shown in Figure 6.2*a*.)

The positive ions formed by the electron beam drift out of the base of the ionization chamber through a slit parallel to, and immediately below, the centre line of the electron beam. This drift is assisted both by field penetration into the box from the accelerating electrodes, and by a small positive potential applied to the ion repeller electrode. A mono-energetic parallel beam of rectangular cross-section is formed by the accelerating and focusing electrodes and directed downwards away from the ionization chamber. The design of this

focusing system can vary considerably. In the 180° deflection instrument, the single electrode illustrated in Figure 6.2*b* gives less precise control than the complex lens shown in Figure 6.2*a*.

The ion beam is deflected into an arc of a circle by means of the homogeneous magnetic field, the radius of curvature R of the path being

$$R = \frac{1.47 \times 10^{-4} \sqrt{V/M}}{B} \quad \text{metre} \tag{6.1}$$

where V is the potential difference in volts through which the ions have been accelerated, B is the magnetic field in tesla, and M is the atomic mass number (atomic mass numbers are referred to O = 16 throughout this work unless otherwise stated). Therefore there is mass discrimination for these mono-energetic beams. Those beams which satisfy the condition that $R = R_0$ (Figure 6.2) pass through the collector defining slit and are registered at the ion collector. (Additional electrodes are often introduced in the collector assembly to suppress secondary emission from the collector surface.) The geometrical relations between the two defining slits and the centre of curvature of the path are shown for the ideal case in Figure 6.2. In the sector machine, the angle θ is usually set at either 60° or 90°. In operation, the whole mass spectrum is swept across the collector slit by varying the intensity of either the magnetic field B or the accelerating voltage V. Ideally, it is best to make B the variable in order to keep the electrical potential gradients constant when the spectrum is scanned, hence keeping the extraction efficiency from the ionization chamber constant. For small gas analysers, V is usually made the variable, so that a small inexpensive permanent magnet can be used to provide the field B. This has the disadvantage that the voltage V is very small when heavy ions are extracted from the source, giving a much reduced sensitivity for these ions.

A portion of a typical spectrum (in this instance the ion fragments of butane) is shown in Figure 6.1. Because the defining slits have finite width, the peaks cannot be perfectly sharp. The broadening at mass M is defined by the resolution which is measured by the ratio $M/\Delta M$ where ΔM is the width of the base of the peak. In the ideal case, where there is perfect alignment of the slits with respect to the magnetic field, resolution is given by

$$\frac{M}{\Delta M} = \frac{0.5 R_0}{S_1 + S_2} \tag{6.2}$$

where S_1 and S_2 are the widths of the source and collector slits respectively.

This is only true for ions which enter the system normal to the plane of the first slit. In practice, the ions form a divergent beam so that some can follow paths with larger radii and still pass through the slits, thus giving a larger ΔM than predicted by eqn (6.2).

In operation, where ideal conditions cannot exist, there must be additional broadening and distortion of the peaks. Inevitably the sides of the peaks meet

the baseline asymptotically, which means that the above simple definition of ΔM cannot be applied. A more precise method of measurement is used, in practice, defining ΔM as the width of the peak at either 1.0%, 10% or 50% above the baseline. For RGA work it is normal to take either the 10% or the 50% definition (i.e. for the latter the peak is measured at 'half height' – the so-called 'half height peak width' definition).

For the magnetic deflection instruments, the three most important factors which prevent the theoretical limit being achieved are:

(i) Alignment errors in the analyser tube
(ii) Energy spread in the beam entering the analyser
(iii) Instabilities in the control voltages and in the magnetic field
(iv) Angular divergence of the beam.

The results shown in Figure 6.1 are typical of those obtained with a small mass spectrometer; in this instance obtained from a 180° deflection instrument (Figure 6.2*b*) with $R_0 = 50$ mm, S_1 and S_2 1.0 and 0.5 mm respectively, and $B = 0.2$ tesla. The mass scan from 100 to 2 amu was achieved by increasing the accelerating voltage from 40 to 2000 V.

6.3 The trochoidal (or cycloidal) mass spectrometer

This instrument is closely allied to the magnetic deflection spectrometers described above. The significant difference is that separation takes place in crossed electric and magnetic fields, these fields being of uniform intensity and mutually perpendicular. Ions moving in these fields follow trochoidal paths, the individual loci depending upon the ratio M/e, thus giving mass separation. The basic principles have been outlined in a number of publications, notably those of Bleakney, Hipple and Mariner[2,3] who confirmed the focusing properties to depend only upon the ratio M/e of the charged particles and not upon their velocity or direction of entry into the analyser.

The projection of the ion path in a plane perpendicular to the magnetic field is the prolate cycloid shown in Figure 6.3. The distance between the entry and

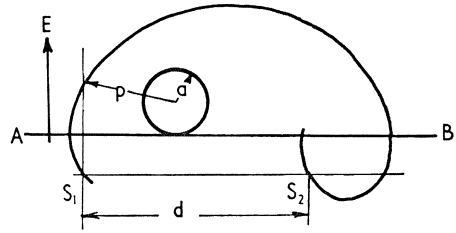


Figure 6.3 Path of the prolate cycloid which is the trace of the end point of the radius arm p. This radius arm is rigidly attached to the cylinder of radius a which rolls along the plane AB. The source and collector defining slits are at S_1 and S_2 respectively. Note: the plane containing S_1 and S_2 need not be AB.

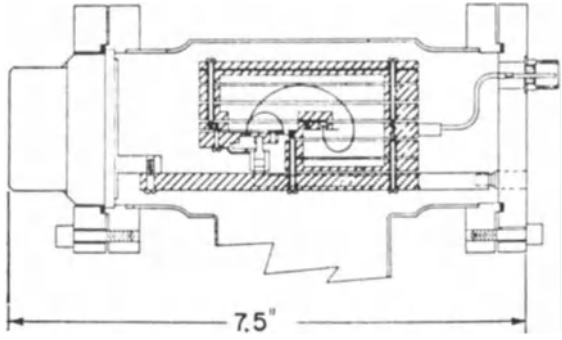


Figure 6.4 Trochoidal mass spectrometer designed by Robinson and Hall[4].

collector slits d is given by

$$d = 2\pi a = \frac{2\pi EM}{eB^2} \quad \text{metre} \tag{6.3}$$

where E is the electric field (V m^{-1}), B the magnetic field (tesla), e the ionic charge (C) and M the ion mass (kg).

Although the length of the radius arm p in Figure 6.3 depends upon the energy of the ion at injection, this is unimportant in determining the distance d travelled in one revolution.

Practical instruments have been described by Robinson and Hall[4] and by Huber and Trendelenburg[5]. The instrument constructed by Parkins and Charpentier[6] and Kornelson[7] and described by Robinson and Hall[4] is shown in Figure 6.4. The ions move inside a circle of 42 mm diameter to give a resolution of approximately 100 (1% valley definition). The relevant constructional data and operating characteristics are as follows:

Slit widths 0.075 mm (entrance), 0.2 mm (collection)
Pitch of cycloid 27.5 mm
Depth of instrument 25 mm
Diameter of magnetic pole pieces 70 mm
Magnetic field strength 0.33 tesla
Overall ion collection efficiency 2–4%
Sensitivity 7.5×10^{-3} mbar (55 peak of butane)
Linear in the pressure range $0–10^{-4}$ mbar
Linear in electron emission current range 0–100 μA.

6.4 The omegatron

The first suggestion for the construction of a mass spectrometer using the principle of cyclotron resonance came from Hipple *et al.*[8] in 1949. In this instrument, positive ions move perpendicular to a magnetic field, and are

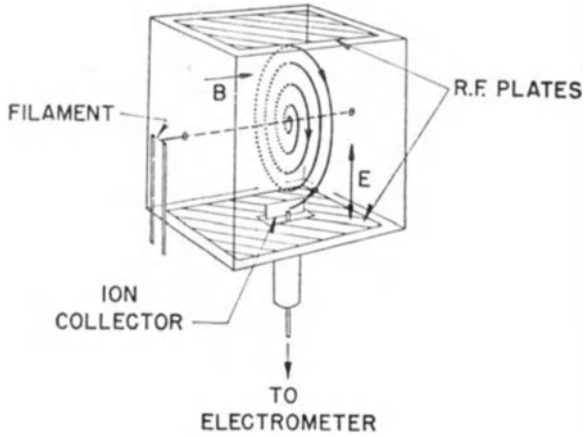


Figure 6.5 Omegatron designed by Alpert and Buritz[9].

accelerated along spiral paths of ever-increasing radii (Archimdes spiral) by a sinusoidally alternating electric field. This is similar to the cyclotron, where ions move in circular paths, being accelerated with a sudden increase of radius twice per revolution at the edges of the 'dees'. A simplified diagram of the omegatron is shown in Figure 6.5[9]. A narrow beam of electrons passes from the filament to the electron collector parallel to a magnetic field B. Above and below the beam are the two plates which provide the ac field. Ions formed along the central axis by electron impact are accelerated in this field. If the resonant frequency of the ions in the magnetic field is the same as the field alternating frequency, they will gain energy continuously and therefore move with ever-increasing radius until they strike the collector. Ions not in resonance have no continuous build-up of energy, and hence remain in the vicinity of the central axis.

The principal characteristics of this spectrometer can be evaluated by following the analysis put forward by Sommer *et al.*[10].

Singly charged positive ions, moving in a plane perpendicular to a magnetic field, follow circular paths with an angular frequency ω_c independent of their energy given by:

$$\omega_c = \frac{eB}{M} \quad \text{rad.s}^{-1} \tag{6.4}$$

where e is the electron charge (C), M is the ion mass (kg) and B is the magnetic field strength (tesla).

If in the plane of motion there is superimposed an alternating electric field of strength $E = E_0 \sin \omega t$, then provided $|\omega - \omega_c| \ll \omega_c$, the particles follow approximately spiral paths with an angular frequency $(\omega + \omega_c)/2$ and with a

radius given by:

$$r = \frac{E_0}{B\varepsilon} \sin \frac{\varepsilon t}{2} \tag{6.5}$$

where $\varepsilon = |\omega - \omega_c|$.

The radius of the path thus passes through successive maxima and minima, except in the special case of 'resonance', where $\omega = \omega_c$, when the radius increases indefinitely. Ions will reach the collector (i.e. have a maximum value of r greater than or equal to R_0, the radial distance from the central axis to the tip of the collector) only when

$$\varepsilon = |\omega - \omega_c| \leqslant \frac{E_0}{R_0 B} \tag{6.6}$$

Thus for ω, fixed collection occurs throughout a range $\Delta\omega_c$ given by

$$\Delta\omega_c = \frac{2E_0}{R_0 B} \tag{6.7}$$

It can be shown from eqn (6.4) that numerically the resolution is given by

$$\frac{M}{\Delta M} = \frac{\omega_c}{\Delta\omega_c} = \frac{eR_0B^2}{2E_0M} \tag{6.8}$$

The total path length of the resonant ions L before collection is given by[10]

$$L = \frac{eR_0^2B^2}{E_0M} \tag{6.9}$$

$$= \frac{2M}{\Delta M} R_0 \tag{6.10}$$

Using a cubic structure of approximately 25 mm side for the electrode system, Klopfer and Schmidt[11] designed an instrument to cover the range 2–100 amu, with side plates and a number of guard rings to ensure uniform electric fields. This is probably one of the most sophisticated omegatrons that has been described in the scientific literature[12–16]. It used a permanent magnet (0.5 tesla) and applied an AC voltage of 1.0 V rms across the RF plates. An operating frequency from 4.0 to 0.1 MHz was needed to scan from 2 to 100 amu. The measurement $\Delta M = 1$ at 30 amu was in satisfactory agreement with that calculated from eqn (6.8). A careful study of the ion motion in this instrument has been made by Schluchhardt[16].

It appears from a detailed study of the literature that the characteristics of the above instrument are in line with those reported by other workers[17–21]. Its performance can therefore be taken as typical of the best that can be obtained when great care is taken in design, construction and operation. Many instruments have been simplified by eliminating the plates and/or the guard

rings. This invariably results in some loss of performance. Zdanuk[20] and his co-workers, for example, report that with the simplified system there is a background ion current to the collector of 10^{-14} A, which limits the minimum detectable signal. The same workers found that, although stability was relatively easy to obtain, there were large differences between individual instruments. Stark[21] also found that with the simplified construction there may be some loss of accuracy and reliability after baking or opening to atmosphere. He observed background currents of the order of 10^{-14} A, and reported that the dc potentials had to be reset from day to day to achieve optimum conditions, and that therefore neither the mass scale nor the sensitivity calibration could be relied upon for long periods.

A serious disadvantage of all omegatrons is the very rapid decrease in resolution at high mass (eqn 6.8 shows $\Delta M \propto M^2$). Whilst improvements can be made either by increasing the magnetic field strength or by decreasing the ac field, unfortunately neither is practicable. Any increase in magnetic field must be prohibitively expensive, and any reduction in ac field must increase the difficulties due to contact potential variations. This limitation on resolution, plus the inconvenience of the large magnet and the instabilities due to contamination on the electrode surfaces, probably explains why the omegatron has not been adopted for routine vacuum analysis.

An interesting application which takes advantage of the very high resolution at low mass has emerged in the field of hydrogen plasma engineering. Here leak detection presents a special problem, because the presence of deuterium means that the conventional helium leak detector with a small magnetic sector mass spectrometer cannot be used effectively. The omegatron can, however, be substituted for the magnetic sector instrument to take advantage of the very high resolution at low mass to distinguish between deuterium and helium with masses of 4.03 and 4.004 amu respectively. Winkel and Hemmerich[22] have found in a prototype instrument that it is possible to detect a partial pressure of helium of 10^{-10} mbar in a total pressure of 3×10^{-6} mbar deuterium.

6.5 Time-of-flight (TOF) mass spectrometer

The early development of the time-of-flight technology was based upon the pioneering work of Bennett described in 1950[23]. Instruments of this type were particularly attractive for space research, because at the time they were the only practical mass spectrometers which did not require strong magnetic fields. (Recently a small reflection instrument, designed for the analysis of dust particles, made a trip to Halley's comet.)[24] A particular feature of all TOF mass spectrometers is their fast response, allowing complete spectra to be recorded at rates of up to 10 000 per second. For this reason, their main application for many years was for the study of fast reactions such as explosions. Their other attractive features, which may have some value for

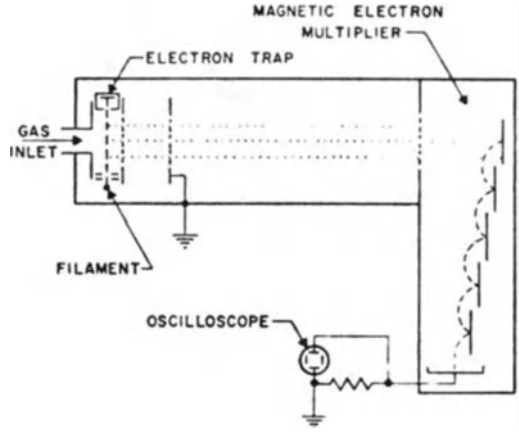


Figure 6.6 Time-of-flight mass spectrometer designed by Wiley and McLaren[25].

residual gas analysis, are (i) the inherently high transmission efficiency (50% in some instruments), (ii) the fact that performance is not critically dependent upon the mechanical precision of the assembly, and (iii) when repetitively scanning, the sensitivity is relatively high, because all the ions are detected, there being no loss due to scanning from mass to mass.

In the basic TOF spectrometer shown schematically in Figure 6.6, pulses of ions, formed by electron bombardment, are accelerated out of the source region by a series of electric fields, individual ions entering the field-free region with velocities dependent upon their charge-to-mass ratio. The transit time down the 'drift tube' is therefore dependent upon this charge-to-mass ratio[25,26]. If only singly charged ions are present, the lightest reach the collector first, followed by groups of successively higher mass. Thus each pulse

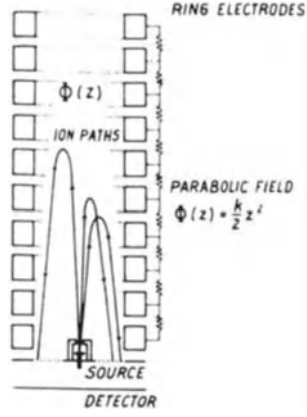


Figure 6.7 Parabolic reflection field used to obtain perfect time focusing.

results in a spectrum which can be amplified, stored and displayed by standard electronic techniques. The instrument with a drift tube 400 mm long, described by Wiley and McLaren[25], shown schematically in Figure 6.6, illustrates the basic technology. Here the electron beam pulse was controlled by a 100 V 'draw out' potential applied to the electrode nearest to the filament to give a pulse width variable between 0.1 and 1.0 s. In operation, this instrument had a resolution of about 150 (1% peak height definition).

Many advances have been made in the design of these instruments, in particular by the introduction of alternative techniques for ion creation in analytical applications[27]. More sophisticated drift tubes have also been designed; for example, it has been demonstrated that the loss of sensitivity due to the spread of energy in the ion pulse can be reduced by introducing electric fields into this tube[28,29]. The schematic diagram (Figure 6.7) shows a flight tube designed as a mirror where, in theory, there is perfect focusing independent of the initial energy of the ions if the field is parabolic.

6.6 Interpretation of mass spectra

Because of the relatively high kinetic energy involved in the ionizing collisions in the source of a mass spectrometer, molecular dissociation often occurs, which means that a group, rather than a single ion species, is formed by the passage of the electron beam. The pattern produced is, to a first approximation, a unique 'fingerprint' for a particular molecule. This can be seen from Figure 6.1, where electrons impacting with butane molecules have produced a group of ions, the most intense being $C_3H_5^+$, $C_3H_6^+$, $C_3H_7^+$ and $C_2H_3^+$. As it is important to measure both the position and relative intensities of the principal peaks in any interpretation of mass spectra, the cracking patterns of the gases commonly found in vacuum systems are listed in Table 6.1. For convenience, and because of the limited mass range of the small RGA, this table is restricted to the mass range 2 to about 100 amu. Abundances are compared by scaling the largest peak in each pattern to 100 units. For a number of gases, in particular for the larger, more complex molecules, the largest is not necessarily the parent peak. For example, in trichloroethylene the parent at mass 134 is relatively small. It must be stressed that all data in Table 6.1 must be treated with caution and used only as approximations, as there are considerable variations from instrument to instrument. The break-up of the molecules under electron bombardment depends to a considerable extent upon the detailed geometry of each individual source assembly. It is a well-known mass spectrometer operating practice to recognize the dependence of the cracking patterns upon such factors as source cleanliness, filament temperatures and gas impurities. Many workers, for example Breth *et al.*[30], have pointed out the complex nature of the interactions between the gas molecules and the vacuum system components (in particular the break-up of large molecules by mass spectrometer, ionization gauge or any other hot filament). Because the source

dictates the details of the cracking patterns, it is not unreasonable to apply the values quoted in Table 6.1 to all types of spectrometer (including the quadrupole), even though most of the data were obtained with magnetic deflection instruments.

For some of the gases listed in Table 6.1 (notably the noble gases), the clusters of ions are formed not by 'cracking', but by either the generation of doubly charged particles or the presence of isotopes. Argon is a good example of the former process, where under the impact of 100 eV electron there is about a 15% chance of the loss of two rather than one electron from the parent atom. This gives a doubly charged argon ion represented as Ar^{++}, which is indistinguishable from a singly charged ion of 20 amu. It is, therefore, often referred to – quite wrongly – as a 'mass 20 ion'. (The ion 'at mass 14' in the spectrum of carbon monoxide is another example of double charge.) Neon is an example of the latter process; the ions at mass 20 and 22 being due to the presence of isotopes. The pattern for krypton arises from a combination of the two effects; the cluster of ions around mass 84 represents the isotope pattern for singly charged ions, while that around 42 stems from the doubly charged ions. It may be noted that doubly charged ions of the mass 81 isotope of krypton cause the anomaly of 'an ion at mass 40.5'.

The fragmentation patterns listed in Table 6.1 can be used in conjunction with the relative sensitivity factors for the ionization gauge (Table 3.1) to calculate the corresponding data for the RGA. The final two columns of Table 6.1 list these relative sensitivity factors. Where reliable data are available for the ionization gauge, the figures are tabulated in the penultimate column. R_i for the ionization gauge is defined in the conventional manner as (sensitivity

Table 6.1 Cracking patterns for some of the gases commonly encountered in vacuum systems.

Gas	amu	Peak height	Sensitivity factors R_i	R_m
Helium	2	0.1	0.15	0.16
	4	100		
Neon	20	100	0.3	0.3
	22	10		
Argon	20	15	1.3	1.25
	40	100		
Krypton	40	1	1.9	0.95
	41	5		
	41.5	5		
	42	20		
	43	6		
	80	4		
	82	25		
	83	25		
	84	100		
	86	30		

(contd.)

Table 6.1 (*contd.*)

Gas	amu	Peak height	Sensitivity factors R_i	R_m
Hydrogen	1	3	0.4	0.43
	2	100		
	3	0.3		
Nitrogen	14	10	1.0	1.0
	28	100		
Oxygen	16	15	1.0	0.95
	32	100		
Water vapour	16	3	1.5	1.25
	17	27		
	18	100		
Carbon monoxide	12	6	1.05	1.05
	14	1		
	16	3		
	28	100		
Carbon dioxide	12	10	1.4	1.15
	16	9		
	22	2		
	28	10		
	44	100		
	45	1		
Ammonia	14	2	1.2	0.7
	15	7		
	16	80		
	17	100		
Nitrous oxide	14	13	1.2	0.8
	16	5		
	28	11		
	30	31		
	44	100		
Silane	28	28		
	29	32		
	30	100		
	31	80		
	32	7		
Methane	12	2	1.4	0.7
	13	7		
	14	15		
	15	83		
	16	100		
Ethane	1	3	2.6	1.3
	13	1		
	14	3		
	15	4		
	25	5		
	26	22		
	27	33		
	28	100		
	29	30		
	30	20		

Table 6.1 (*contd.*)

Gas	amu	Peak height	Sensitivity factors R_i	R_m
Ethylene	1	6	2.1	0.9
	2	1		
	12	2		
	13	4		
	14	8		
	24	3		
	25	12		
	26	61		
	27	59		
	28	100		
	29	3		
Acetylene	1	4	1.7	1.3
	12	4		
	13	8		
	24	7		
	25	23		
	26	100		
Methyl alcohol	28	6		
	29	67		
	31	100		
	32	67		
Ethyl alcohol	18	5		
	26	8		
	27	24		
	28	7		
	29	23		
	30	6		
	31	100		
	43	8		
	45	34		
	46	16		
Acetone	26	6		
	27	8		
	29	4		
	37	2		
	38	2		
	39	4		
	41	2		
	42	7		
	43	100		
Trichloroethylene	35	40		
	37	13		
	47	26		
	60	65		
	62	21		
	95	100		
	97	64		
	130	90		
	132	85		
	134	27		

for test gas)/(sensitivity for nitrogen). For the spectrometer the definition of R_m is similar: (sensitivity for the test gas, determined from the largest peak in the spectrum)/(sensitivity for nitrogen, determined by the mass 28 peak).

The data quoted in Table 6.1 are 'ideal', because all cracking patterns have been measured with high-quality analytical mass spectrometers where sensitivity is independent of the ion mass-to-charge ratio. The small residual gas analysers do not always match the performance of the analytical instruments, as often sensitivity can vary considerably over the mass range. (This can be particularly important in quadrupole filters operating at high resolution; see for example Chapter 7, Figure 7.17.) To take account of this non-linearity, an extra relative efficiency term η_m must be introduced when using the RGA to determine partial pressures. The sensitivity S for a given gas can be defined as

$$S = S_0 R_m \eta_m \tag{6.11}$$

where S_0 is the instrument sensitivity for nitrogen and η_m is defined as (efficiency of ion transmission at the mass for which R_m is measured)/(efficiency of ion transmission at mass 28).

In using eqn. (6.11) to calculate sensitivity, it must be borne in mind that, unless calibrations are carried out under rigorous conditions, the greatest source of error is likely to occur from uncertainties in the instrument constants S_0 and η_m.

In many vacuum systems, it is necessary to identify and determine the concentration of the major components in a gas mixture. If the major peaks

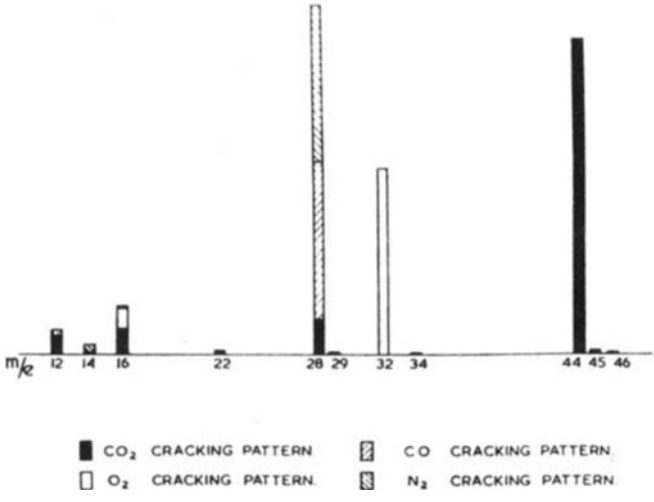


Figure 6.8 Spectrum[31] of mixture of carbon dioxide, oxygen, carbon monoxide and nitrogen (2:2:1:1).

occur at the same mass, the problem can be complex, requiring careful analysis. There are further difficulties if the minor peak of one component occurs at the same mass as the major peak of a small (but important) impurity. In general, the cracking patterns of all the components must be known to a high accuracy, in order that a meaningful set of simultaneous equations can be drawn up. The residual gas analyser is not suitable for any sophisticated analysis of this type unless extreme care is taken in setting up and calibration. The simple instrument can, however, be of real value for many of the routine analyses needed in high vacuum systems. Craig and Harden[31] discussed the problem in detail, and showed the simple technique applied to a mixture of carbon dioxide, oxygen, carbon monoxide and nitrogen. In this example (which is illustrated in Figure 6.8) the contributions of the heavier gases (carbon dioxide and oxygen) are first subtracted from the measured 12, 14, 16 and 28 peaks. This is possible because the 44 peak is uniquely representative of carbon dioxide and the 32 peak of oxygen. Simultaneous equations based on the 12, 14, 16 and 28 peaks can then be solved to determine the relative contribution of carbon monoxide and nitrogen. This is an acceptable technique when the carbon dioxide partial pressure is of the same order as, or smaller than, that of carbon monoxide and nitrogen. Obviously the errors in the analysis may be unacceptable when carbon dioxide is the major constituent. The differentiation between nitrogen and carbon monoxide is always difficult, because any analysis must depend upon the correct interpretation of the 14 (N^+) and 12 (C^+) intensities. Unfortunately these peaks are often corrupted by the general background. Craig and Harden[31] are careful to point out the importance of using accurate cracking pattern data and they discuss the dependence of the data on such factors as the temperature and electric field distribution in the ion source.

References

1. Yarwood, J. (1959) *Br. J. Appl. Phys.* **10**, 383.
2. Bleakney, W. and Hipple, J.A. (1938) *Phys. Rev.* **57**, 521.
3. Mariner, T. and Bleakney, W. (1949) *Rev. Sci. Instrum.* **20**, 297.
4. Robinson, C.F. and Hall, L.G. (1956) *Rev. Sci. Instrum.* **27**, 504.
5. Huber, W.K. and Trendelenberg, E.A. (1961) *Vac. Symp. Trans. Amer. Vac. Soc.* **1**, Pergamon, New York, 592.
6. Perkins, G.D. and Charpentier, D.E. (1957) *Vac. Symp. Trans. Amer. Vac. Soc.*, Pergamon, New York, 125.
7. Kornelsen, E.V. (1959) *Proc. 19th Physical Electronics Conf., M.I.T.*, Cambridge, Mass.
8. Hipple, J.A., Sommer, H. and Thomas, H.A. (1949) *Phys. Rev.* **76**, 1877.
9. Alpert, D. and Buritz, R.S. (1954) *J. Appl. Phys.* **25**, 202.
10. Sommer, H., Thomas, H.A. and Hipple, J.A. (1951) *Phys. Rev.* **82**, 697.
11. Klopfer, A. and Schmidt, W. (1960) *Vacuum* **10**, 363.
12. Charles, D. and Warnecke, R.J. (1959) *Vac. Symp. Trans. Amer. Vac. Soc.*, Pergamon, New York, 34.
13. Peper, J. (1957–58) *Philips Tech. Rev.* **19**, 218.
14. Wagener, J.S. and Marth, P.T. (1957) *J. Appl. Phys.* **28**, 1027.
15. Lawson, R.W. (1962) *J. Sci Instrum.* **39**, 281.

16. Schluchhardt, G. (1960) *Vacuum* **10**, 373.
17. Brubaker, W.M. and Perkins, G.D. (1956) *Rev. Sci. Instrum.* **27**, 720.
18. Berry, C.E. (1954) *J. Appl. Phys.* **25**, 28.
19. McNarry, L.R. (1958) *Can. J. Phys.* **36**, 1710.
20. Zdanuk, E.J., Bierig, R., Rubin, L.G. and Wolsky, S.P. (1960) *Vacuum* **10**, 382.
21. Stark, D.S. (1959) *Vacuum* **9**, 288.
22. Winkel, T. and Hemmerich, J.L. (1987) *J. Vac. Sci. Technol.* **A5**, 2637.
23. Bennett, W.H. (1950) *J. Appl. Phys.* **21**, 723.
24. Kissel, J. (1986) in Todd, J.F.J. ed., *Proc. 10th Int. Mass Spectrometry Conf., Swansea, 1985*, John Wiley, London, 175.
25. Wiley, W.C. and McLaren, I.H. (1955) *Rev. Sci. Instrum.*, **26**, 1150.
26. Walcher, W. (1950) *Rev. Sci. Instrum.* **21**, 578; Varadi, P.F. and Sebestyen, L.G. (1956) *J. Sci. Instrum.* **33**, 392; Diels, K. and Moesta, H. (1958) *Vac. Symp. Trans. Amer. Vac. Soc.*, Pergamon, New York 115; Wherry, T.C. and Karasek, F.W. (1955) *J. Appl. Phys.* **26**, 685; Kendall, B.R.F. (1962) *J. Sci. Instrum.* **39**, 267.
27. Brunnee, C. (1987) *Int. J. Mass Spectrom. & Ion Processes* **76**, 125.
28. Yoshida, Y. (1986) US Patent 4625112.
29. Rockwood, A.L. (1986) *Proc. 34th Ann. Conf. on Mass Spectrometry & Allied Topics*, 173.
30. Breth, A., Dobrozemsky, R. and Kraus, B. (1983) *Vacuum* **23**, 73.
31. Craig, R.D. and Harden, E.H. (1966) *Vacuum* **16**, 67.

7 Gas analysis in vacuum systems: quadrupole mass analysers

7.1 Introduction

This instrument has been subject to continuous development over the past 20 years[1] to the point where it is used successfully in many applications ranging from sophisticated chemical analysis to general vacuum monitoring*. It is, in fact, the most extensively used mass spectrometer in high- and ultra-high vacuum systems for routine gas analysis. In its simple form it has become recognized as a reliable and easy-to-use instrument in the heavy industrial, as well as the research, laboratory. The description of this quadrupole mass spectrometer (or mass filter), often referred to as the residual gas analyser (RGA), and its use in the general vacuum laboratory is emphasized in this chapter. In many applications these RGA mass spectrometers substitute for the ionization gauge, in effect covering the same pressure range but with the advantage of indicating the partial pressure of the main constituents in the vacuum system. The geometry of the quadrupole mass spectrometer is very similar to that of the extractor ionization gauge developed by Pittaway[2] (described in detail in Chapter 3, section 3.5), with a mass filter interposed

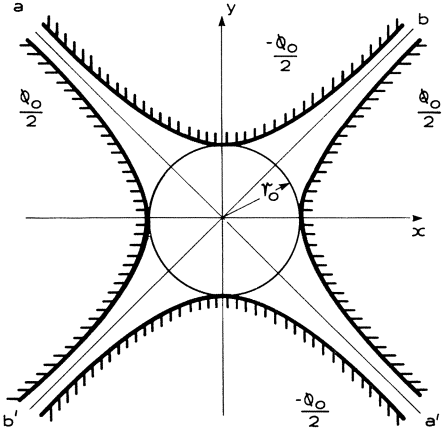


Figure 7.1 The two pairs of electrodes with hyperbolic profiles at potentials of $\phi_0/2$ and $-\phi_0/2$ which gives the electrical potentials described by eqns (7.3) and (7.4).

*For a comprehensive review of the subject of quadrupole mass spectrometry and its applications, the reader is referred to the text edited by P.H. Dawson[1].

between source and collector. The filter is an electrostatic lens consisting of an array of four parallel rods arranged symmetrically around a central axis along which the ions travel. The 'quadrupole' structure is illustrated in Figure 7.1 which shows that the inner surfaces of the electrodes have hyperbolic profiles. The principles of using quadrupole magnetic or electric fields for ion beam control were developed before the application to the mass filter[3-5]. They were first introduced in the early 1950s. However, it was Paul and his colleagues at the University of Bonn who pioneered the work in adapting and applying the general principles to the application of mass analysis[6], their paper published in 1958[7] being of fundamental importance.

7.2 Principles of the quadrupole mass filter

An understanding of the operation of this filter can best be gained by considering the motion of charged particles in a symmetrical two-dimensional electric field E expressed in cartesian co-ordinates by the equation

$$E = E_0(\lambda_1 x + \lambda_2 y) \tag{7.1}$$

where E_0 is a constant with the dimensions of field strength independent of position, but not necessarily of time, and λ_1, λ_2 are simple non-dimensional constants. The field in the z direction is zero.

To satisfy the Laplace equation for this two-dimensional case (i.e. $\text{grad}E = 0$), eqn (7.1) reduces to

$$E = E_0\lambda_1(x - y) \tag{7.2}$$

From this equation the electric potential ϕ over the XY plane, obtained by integration, is given by

$$\phi = -\frac{E_0}{2}\lambda_1(x^2 - y^2) \tag{7.3}$$

Four electrodes, with hyperbolic profiles, can be arranged (as shown in Figure 7.1) to provide this electric field, the opposite electrodes being connected together electrically. With the x and y directions defined (arbitrarily) such that a positive potential $\phi_0/2$ is applied to the 'X electrodes' and a negative potential $-\phi_0/2$ to the 'Y electrodes', potential in the inter-electrode space is given by

$$\phi = \frac{\phi_0}{2}\frac{x^2 - y^2}{r_0^2} \tag{7.4}$$

where r_0 is the radius of the inscribed circle (Figure 7.1).

The electrode profiles in three dimensions required to satisfy the additional condition that the electric field is zero in the z direction are those shown in Figure 7.2.

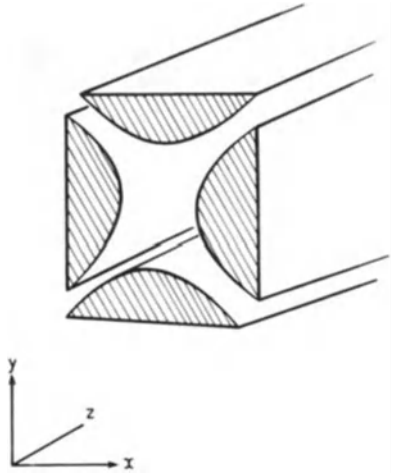


Figure 7.2 Electrode profiles needed to produce hyperbolic field for the quadrupole mass filter.

The equation of motion for a charged particle of mass m kg and e coulomb, injected into the inter-electrode space approximately parallel to the z axis, can be obtained from Newton's fundamental equation of motion force = mass × acceleration

$$m\ddot{x} = eE_x$$

For the x direction

$$\ddot{x} + \frac{e}{mr_0^2} \cdot \phi_0 x = 0 \tag{7.5}$$

For the y direction

$$\ddot{y} - \frac{e}{mr_0^2} \cdot \phi_0 y = 0 \tag{7.6}$$

For the z direction where the field E_z is zero

$$\ddot{Z} = 0 \tag{7.7}$$

If the special, almost trivial case, is considered where ϕ_0 is constant (i.e. not varying with time), the resulting 'saddle' field is easy to visualize. It is clear that the motion of a positively charged particle in the x direction is simple harmonic about the origin, and therefore inherently stable. In the y direction the motion is unstable because the force is always acting away from the origin. As the z field is zero, the velocity is constant in this direction. A simple model can be used to give a visual appreciation of the motion of the ions in the lens. The beams sketched in Figure 7.3 are useful in this respect. Here the movement of the small sphere in the gravitational field is analogous to that of the positive

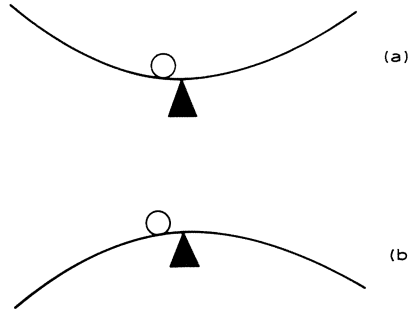


Figure 7.3 Mechanical analogy indicating motion of positive ions in the quadrupole field. The beams are 'clamped' in the horizontal position to represent the application of dc potentials to the rods. (*a*) then represents motion in the XZ plane and (*b*) motion in the YZ plane.

ions in the electric fields. In (*a*) the sphere is shown lying in the central region of a simple trough, whereas in (*b*) it is balanced on the ridge of an inverted trough. Operation of the filter with constant, time-invariant fields corresponds to the troughs being clamped rigidly in the symmetrical position. Clearly in (*a*) the sphere will oscillate about the centre line unless the perturbations are extreme, whereas in (*b*) any small perturbation will cause the sphere to be lost. (*a*) represents the XZ plane and (*b*) the YZ plane in the filter. Thus when only constant potentials are applied between the pairs of electrodes, the positively charged particles are ejected rapidly to the Y electrodes. (Electrons or negative ions will be ejected to the X electrodes.) Particles injected into the filter approximately parallel to the z axis penetrate only very short distances down the rod system.

The behaviour becomes more interesting and also of practical value when the field is made dynamic by adding an oscillatory component so that the potential ϕ_0 has the form

$$\phi_0 = U - V\cos\omega t \tag{7.8}$$

where U and V are constants.

The basic equations of motion (7.5) and (7.6) must now be rewritten:

$$\ddot{x} + \frac{e}{mr_0^2}(U - V\cos\omega t)x = 0 \tag{7.9}$$

$$\ddot{y} - \frac{e}{mr_0^2}(U - V\cos\omega t)y = 0 \tag{7.10}$$

For ease of presentation, it is conventional in theoretical analyses of the quadrupole filter to write these equations in the following form:

$$\frac{\mathrm{d}^2u}{\mathrm{d}\xi^2} + (a_u - 2q_u\cos 2\xi)u = 0 \tag{7.11}$$

defining $$\xi = \omega t/2$$

and

$$a_u = a_x = -a_y = \frac{4eU}{m\omega^2 r_0^2}$$

$$q_u = q_x = -q_y = \frac{2eV}{m\omega^2 r_0^2}$$

where u represents either x or y.

Equation (7.11), the Mathieu equation, has no simple analytical solution, but because of its importance to this filter, it is worthy of detailed study.

After the trivial, time-invariant, operation considered above where $V = 0$, the next condition to consider is the other extreme where U is zero and V is finite. Here only alternating potentials are applied, giving the constant $a = 0$. For this case only one direction need be considered since the solutions for x and y are identical except for a phase shift of $\pi/2$. It is relatively easy to compute the trajectories of ions injected approximately parallel to the z axis with a small displacement from the centre of the three-dimensional lens shown in Figure 7.2.

A selection of such computed trajectories for increasing values of applied alternating voltages (i.e. increasing q) is shown in Figure 7.4*a*. The oscillatory nature of the motion and the dependence of the fundamental frequency upon q is quite clear. The trend towards instability for $q = 0.908$ is clear from Figure 7.4*b* where the increase in the amplitude of oscillation is plotted as a function of q. (The amplification is defined by the ratio X_m/X_0 where X_m is the maximum excursion from the central axis and X_0 is the displacement of the ions on entry to the filter.) Although the paths are changed in detail if the ions are injected with different phase angles, and also if not parallel to the z axis, the change from stability to instability always takes place when $q = 0.908$. Thus the quadrupole lens has a more stable pattern of behaviour when it is driven with an alternating rather than a steady voltage signal. Ion motion is stable in both the XZ and YZ planes for low values of q. Thus ions injected parallel to the Y axis will follow stable paths and can reach the exit provided the mass-to-charge ratio is less than a critical value given by $m/e = 2V/0.908\,\omega^2 r_0^2$. For m/e greater than this critical value, motion will be unstable in both the x and y directions, with the ions being driven radially outwards to either the X or Y electrodes. In this configuration the lens acts as a 'low (mass) pass filter'.

When a combination of the direct and alternating voltages is applied across the pairs of rods (both U and V non-zero), the motion of the charged particles in the hyperbolic field becomes more complex. A convenient way of obtaining an overall view of the characteristics is to consider the XZ and YZ planes when initially a steady potential difference is applied with only a small alternating component superimposed, and then to consider the change that will take place in the ion motion as the sinusoidal voltage is increased in stages

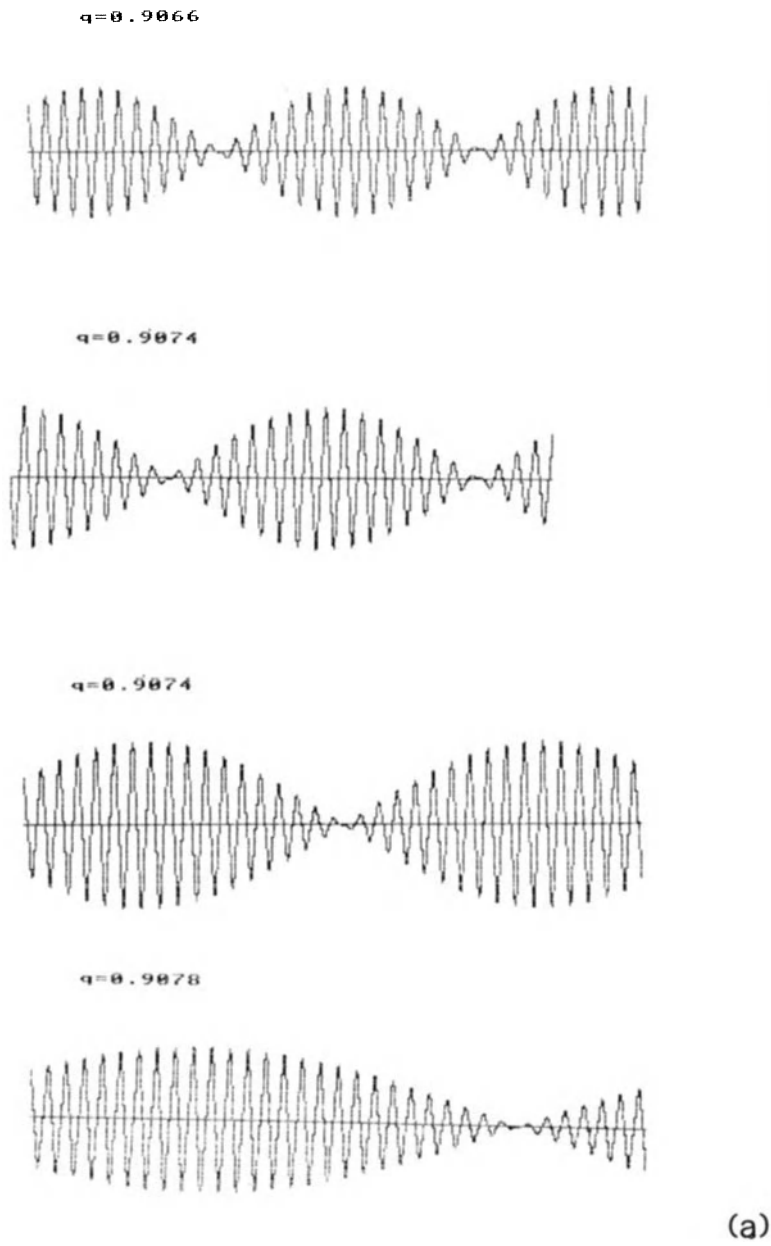
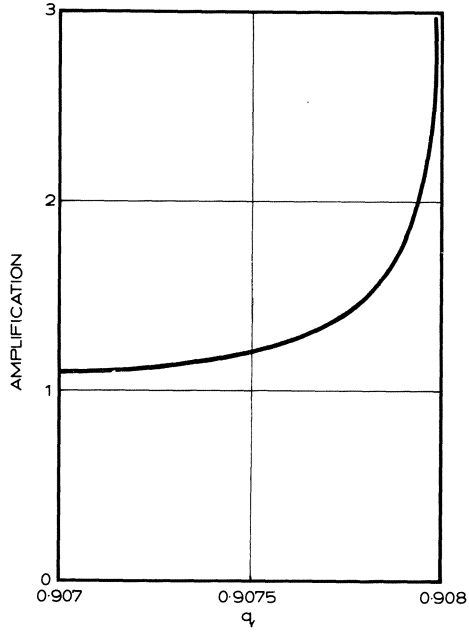


Figure 7.4 Motion of ions in the XZ and YZ planes: (*a*) with pure alternating fields, values of q being indicated on each trace ($a = 0$ for each curve); (*b*) graph showing the increase in the excursion of ions from the central axis with increasing q for oscillations of the type shown in (*a*); (*c*) oscillation of ions in fields where both a and q are non-zero. a is constant in each case, but q increased from (I) to (II) and from (II) to (III).

to the point where it dominates completely. In the XZ plane the motion must be stable when the steady field predominates, but as the alternating component is stepped up, the amplitudes of oscillation increase. The point must be reached where the alternating field predominates; then the motion becomes unstable, as is shown in Figure 7.4*a*. In the YZ plane, the pattern is completely different. Here, for very low values of alternating component (where the steady field predominates), the motion must be unstable. It is, however, possible to imagine a limited regime where the forces due to the alternating component just balance, on average, those due to the steady-state field. This suggests a range of values of alternating voltage where there is stability in the YZ plane. This is illustrated by the trajectories in Figure 7.4*c* which have been computed by Batey[8]. These show (I) stability in the XZ but not in the YZ plane when there is a low value of alternating field, (II) the increase in the alternating field to the point where there is stability in both planes, and (III) instability in both planes after a further increase in the alternating field.* Thus it is possible for there to be a limited range of applied voltages which lead to

*In the analogy presented in Figure 7.3*b*, it is not difficult to visualize the frictionless ball being prevented from falling off the convex surface by oscillating the surface about the centre at 'just the right frequency and amplitude'. The motion is then erratic but symmetrical about the central axis.



(b)

Figure 7.4 (*contd.*)

ION TRAJECTORIES

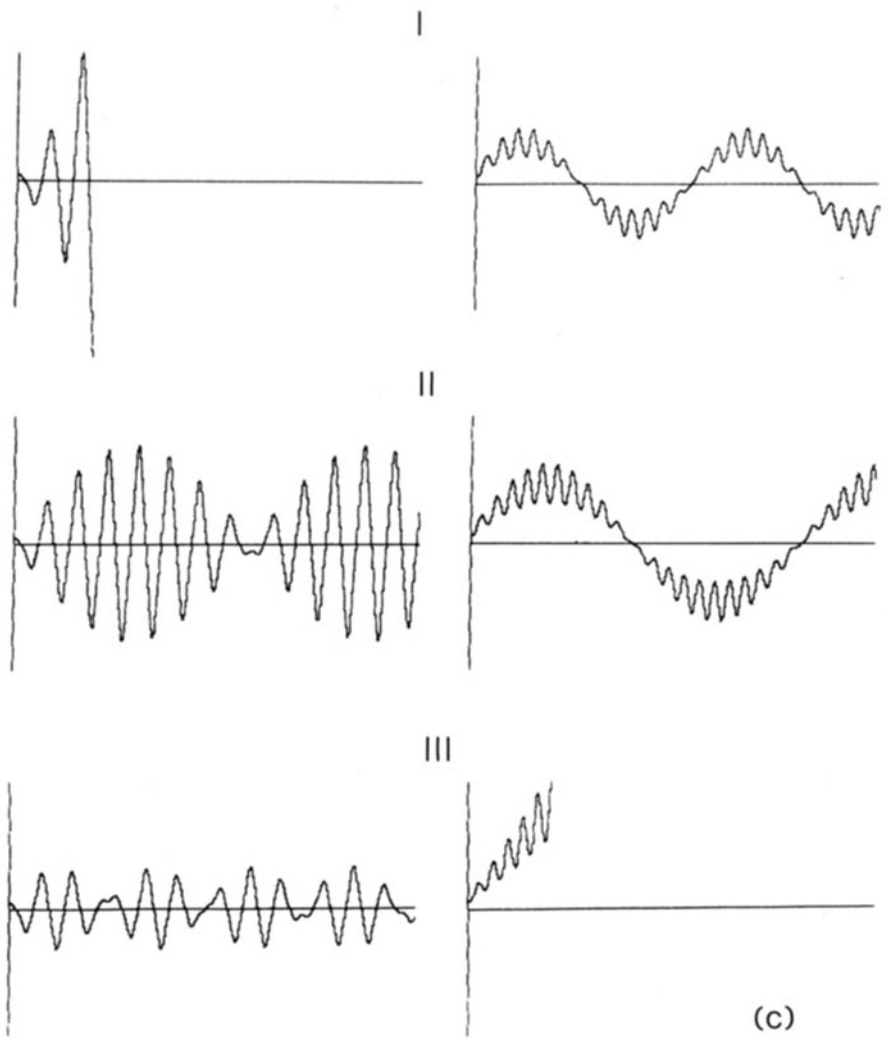


Figure 7.4 (*contd.*)

stable trajectories in both the XZ and YZ planes, which allows ions to pass through the filter.

For a quantitative analysis of the characteristics of motion in the combined alternating and steady hyperbolic fields, a solution must be sought to the Mathieu equation set out in eqn (7.10). Dawson[1] has presented a comprehensive analysis from which the salient points, as they apply to the RGA, can be extracted. He expressed the solution to eqn (7.11) in the form

$$u = \alpha' e^{\mu\xi} \sum_{n=-\infty}^{\infty} C_{2n} e^{2in\xi} + \alpha'' e^{-\mu\xi} \sum_{n=-\infty}^{\infty} C_{2n} e^{-2in\xi} \tag{7.12}$$

where α' and α'' are integration constants depending upon the initial conditions, that is on $u_0, \dot{u}_0$ and ξ_0. The constants C_{2n} and μ depend upon the values of a and q and not on the initial conditions. The first important point to emerge from this solution of the Mathieu equation is that the ion motion depends upon a and q but not upon the initial conditions. This means that all ions with the same a and q value will have the same periodicity of motion and all be in either 'stable' or 'unstable' orbits. The second point is that stability (i.e. u always finite) depends upon the exponential terms $\exp(\mu\xi)$ and $\exp-(\mu\xi)$.

Since stable motion is possible only when both these exponential terms remain finite, as $\xi \to 0$ (i.e. for very long periods) the constant μ must take the form $\mu = i\beta$ where β is real but not a whole number. When β is an integer, the solutions to eqn (7.12) are called the 'Mathieu Functions of Integral Order', and form the boundaries in the aq space between the stable and unstable regimes. Motion on these boundaries is periodic but unstable. The condition for stability can be represented on a two-dimensional a–q or 'stability' diagram, as shown in Figure 7.5 for the 'first' or 'lowest' zone between the limits of $\beta = 0$ and 1.0. Contours for constant β values are drawn on this diagram. This is the only zone which, to the present time, has been found to have practical value for the mass filter.

Ion motion in the quadrupole field is linked directly to the constant β, the fundamental frequency being given by ω_0 where

$$\omega_0 = \beta\omega/2 \tag{7.13}$$

and higher frequencies ω_1, ω_2, etc. where

$$\omega_1 = (1 - \beta/2)\omega \tag{7.14}$$

$$\omega_2 = (1 + \beta/2)\omega \tag{7.15}$$

The first quadrant in Figure 7.5 defines stability in the XZ plane, and the fourth quadrant stability in the YZ plane with the sign convention adopted in Figure 7.1. (Clearly there is no advantage to be gained by considering negative values of q, as the curves are symmetrical about the a axis. A change from $+q$ to $-q$ simply changes the phase of the motion by 90°.) While it is possible to

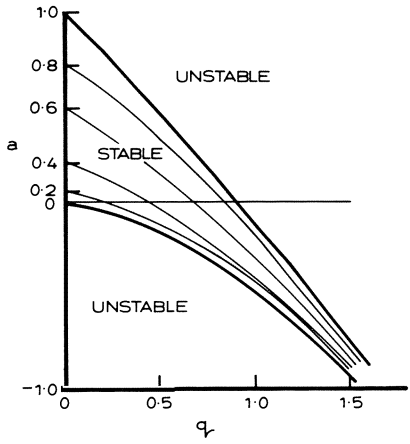


Figure 7.5 The Mathieu functions in a–q space plotted for the 'first stability zone', i.e. between the limits of β 0.0 → 1.0. Values of $\beta = 0$, 0.2, 0.4, 0.6, 0.8 and 1.0 indicated on the curves.

determine pairs of values of a and q in Figure 7.5 where motion is stable in the two planes, stability can best be visualized by folding the diagram about the a axis to superimpose the stability criteria for the two planes. The result of this folding—the familiar triangular stability diagram which forms the base for all quadrupole mass filter designs—is shown in Figure 7.6. Ions are stable in both the XZ and YZ planes when the operating point, described by the particular a and q values, lies inside the triangle. Trajectories become unstable in either the

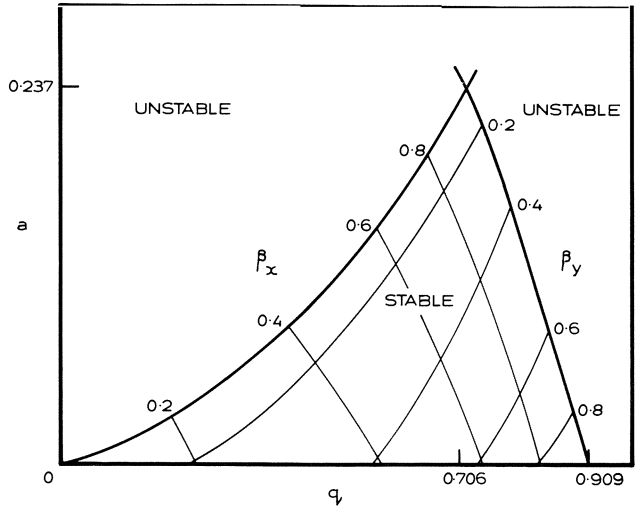


Figure 7.6 The 'first stability zone' which indicates the region in a–q space which gives stable motion of the ions in both XZ and YZ planes.

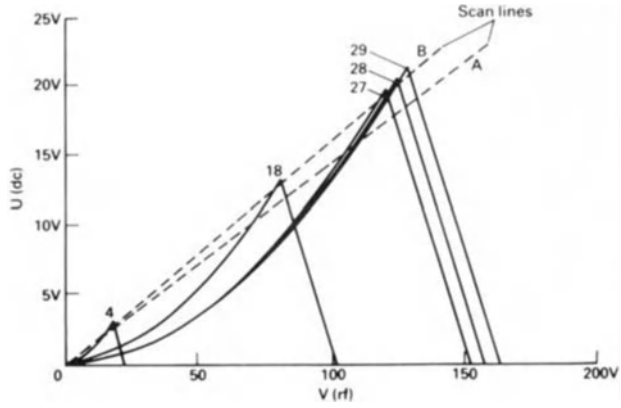


Figure 7.7 Stability diagram plotted in terms of U and V, dc and ac voltages, for ions at masses 4, 18, 27, 28 and 29 amu[8].

XZ or YZ planes when the a and q values change, so that the operating point moves out of the 'stability triangle' as indicated in Figure 7.6.

The basic stability diagram of Figure 7.6 can be redrawn with the axes changed from a and q to U and V, the direct and alternating applied voltages, to indicate the operation of the practical ion filter. The pattern which results is presented in Figure 7.7, which has been drawn for a typical analyser with $r_0 = 2.75$ mm and ω in the RF range at 2 MHz for masses 4, 18, 27, 28 and 29 amu (28 corresponds to 4.68×10^{-26} kg). The charge e has been taken as the electron charge, 1.609×10^{-19} coulomb, on the assumption that positive ions have lost one electron. Because the 'tips' of the triangles have the U and V co-ordinates $a_0 = K_1 M$ and $q_0 = K_2 M$, then, for a given analyser geometry and a fixed frequency, the ratio K_1/K_2 is a constant; thus all the tips must lie on a straight line passing through the origin with a slope $K_1/K_2 = 0.167$.*

Paul[7] in his initial pioneering work, pointed out that a spectrum from low to high mass can be generated by scanning from the origin along a line OB so that the operating point just cuts through the tips of all the triangles. Obviously this means increasing the voltages applied to the pairs of rods, taking care to keep the ratio U/V constant. By varying the slope of this scan line (the ratio U/V), the resolution can be increased or decreased, the limit being set when OA has a slope of 0.167. There are two limitations which mean that the sharp boundaries between stability and instability implied in Figure 7.7 cannot be realized in practice. These are that (i) the ions stay in the hyperbolic field for only a finite time, whereas in the analysis $\xi \to \infty$ is assumed, and (ii) the field is

* Batey[8] computed the trajectories for the masses 29, 28 and 27 in both the XZ and YZ planes with U and V adjusted to the top of the mass 28 stability triangle. The motion is represented by the curves plotted in Figure 7.4*c*. The upper pair of curves are for the mass 29 ions where there is instability in the YZ plane, and the lowest for the mass 27 where instability is in the XZ plane.

bounded by the surfaces of the rod electrodes and therefore not infinite in either the x or y direction. This means a gradual stability/instability transition as the operating point moves across the boundaries $\beta = 0, 1.0$.

Paul[7] and his co-workers have analysed the trajectories of the ions with the operating point near the tip of the stability region, paying particular attention to the two parameters u_0 and ξ. A significant result from this analysis, showing the importance of u_0 in determining ion transmission, is shown in graphical form by the curves in Figure 7.8. Contours are drawn which define the aperture size (i.e. the maximum value of u_0) which allows 100% transmission. The numbers on the contours define the maximum values of u_0/r_0 for which transmission is 100%. These values become progressively more critically dependent upon resolution as the tip of the stability region is reached. These curves for a fixed aperture (u_0 maximum) show how rapidly transmission efficiency is likely to decrease as the stability tip is approached. Analysis[7] indicates that the maximum resolution attainable in practice is proportional to N^2, N being the number of cycles of RF field to which the ions are subject in the filter. This prediction has been confirmed in a number of laboratories[9–11], for example by the results shown in Figure 7.9 which cover a wide range of operating conditions. There must be some uncertainty in the value of the constant h in the relationship $R_m = N^2/h$, because of the dependence on such factors as aperture size, the focusing of the ions at the filter entrance, and field imperfections. Experience indicates that values of h approximately equal to 25 are common for small RGAs, although values of 10 can be obtained for large analytical instruments[9].

Two important features of the mass spectrometer emerge from the above analysis. The first is the simple linear relationship in the scan between applied voltage and ion mass and the second the 'trade-off' between sensitivity and resolution by the simple electrical technique of varying the slope of the scan line. This is in contrast to the small magnetic sector mass spectrometers with

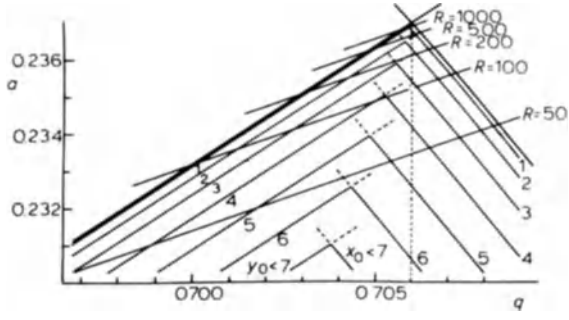


Figure 7.8 The tip of the stability diagram showing operating lines nominally giving resolutions of 50, 100, 200, 500 and 1000[7]. The lines parallel to the stability boundaries represent the maximum allowable initial x or y displacements, expressed as percentages of r_0, if there is to be 100% ion transmission for all initial RF phases. The calculations were for ion entry parallel to the instrument axis and in the absence of fringing fields.

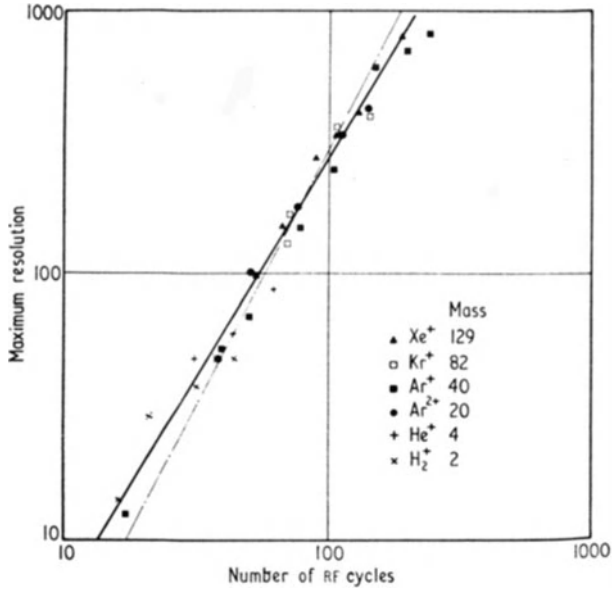


Figure 7.9 The maximum resolution obtained in individual experiments shown as a function of the number of RF cycles which the ion experiences in the quadrupole field[11]. The solid line is the best fit to the experimental results; the broken line the square law approximation.

permanent magnets which have a 'square law relationship with mass' in the scan and where resolution can only be changed by physically altering slit widths (Chapter 6, section 6.2).

7.3 Design of small residual gas analysers (RGAs)

To develop this type of instrument, the aim is not to extend the capabilities of the filter to its limits, but instead to produce a mass spectrometer of limited performance but small, reliable and inexpensive. The design generally adopted is shown in Figure 7.10, which is a schematic diagram of the whole analyser showing the ion source, mass filter and ion collector. For RGA applications, filters usually have rods ranging in length from 50 to 80 mm and with an inscribed radius r_0 about 3 mm. The design of the ion source and its associated electrostatic lens system is vital to the whole operation of the instrument. The construction and operation of the ion source, which is similar to that of the extractor BA gauge[12], can be seen from the photograph in Figure 7.11. The ionization chamber is cylindrical, usually of wire mesh construction mounted co-axially with the filter assembly. There is considerable variation in detail in the design of the many sources in general use. In most instruments the thermionic cathode is a short straight wire parallel to the grid, but in some cases a circular construction is chosen. In practice there seems to be little

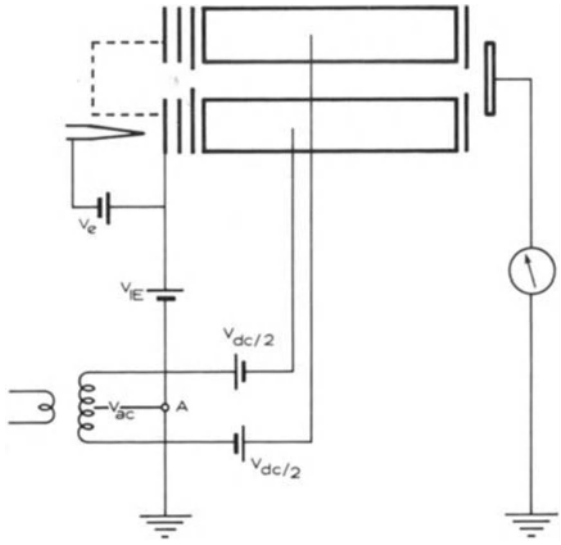


Figure 7.10 Schematic diagram of quadrupole mass filter showing ion source, filter and ion collector. The electron energy in the source and the ion injection energy into the filter are determined by V_e and V_{ie} respectively.

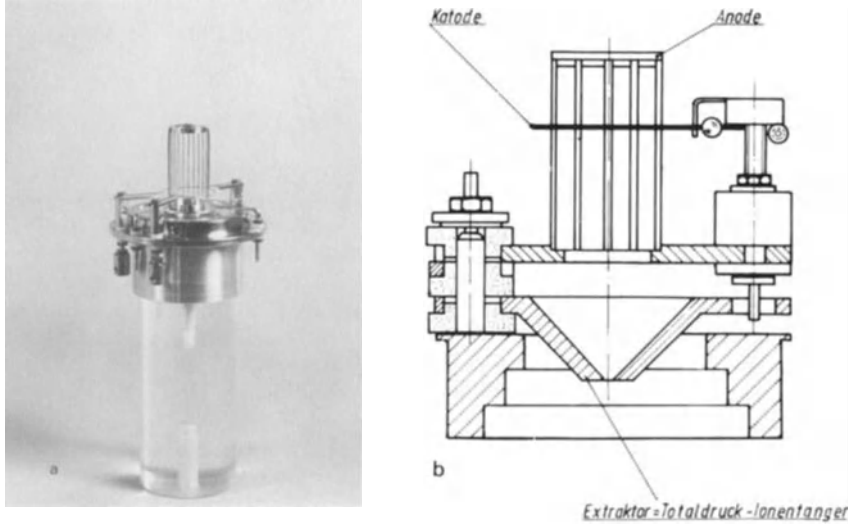


Figure 7.11 Typical electrode arrangements for ion source of commercial quadrupole mass analysers designed by (*a*) Balzers Ltd; (*b*) Leybold Ltd.

difference between the performance of the two designs. The linear system does, however, have the practical operating advantage that two filaments can easily be mounted one on each side of the cylindrical grid, thus allowing quick changeover in the event of a failure. Sources are coupled directly to the rods with a simple lens system which extracts ions from the central ionizing region and forms them into a parallel beam injected into the filter along the central axis. The entrance aperture, and hence the value of u_0 maximum, is defined by the orifice in the final lens plate.

In most instruments the hyperbolic field is approximated by four circular rods rather than by the profiled electrodes shown in Figure 7.1. Dennison[13] showed that the best approximation to a hyperbolic field is obtained when the radius of the rods $r = 1.147\, r_0$. The approximation is extremely good inside the central circle radius r_0. The photograph (Figure 7.12) shows the simple construction, and in particular it illustrates the type of ceramic spacer used to clamp the four rods together. For these small units, the errors in the field due to component tolerances and misalignments in assembly are probably greater than those due to the use of circular rods.

The fringing fields at the entrance to the filter are important because ions encounter strong defocusing forces in travelling the short distance from the exit orifice of the lens to the entrance plane of the filter. As they approach and first come under the influence of the filter, the fringing fields are small, so that the 'effective' values of both U and V are very small. As the ions approach the filter, these parameters increase steadily and attain their true values as the entrance plane is reached. Consider the pattern on the stability diagram, Figure 7.7, for (say) mass '40' ions with the filter potential adjusted to put the operating point near the top of the stability diagram. At the exit orifice, the operating point is near the origin of the stability diagram, but as the ion travels through the fringing field it moves along, or close to, the scan line OB, reaching the 'stability tip' as the ion enters the filter. Thus the ion has to pass through a field which is defocusing in the YZ plane. Although a number of workers, notably Brubaker[14] and later Fite[15], have introduced short pre-filters with a zero, or very low, fixed field component to reduce the defocusing action, this complication has not been found worthwhile for RGA applications.

At the exit of the filter, the simplest structure is usually adopted with a plate collector covering the aperture. Although ions 'spray out' from the filter with considerable radial velocity, experimental evidence suggests that more than 50% of the exit ions are collected. Measurements by Holme (unpublished) showed that the ion current could be increased by about 30% by applying a potential of -10 V to the collector. However, further increase in bias caused no significant additional ion current. This suggests that, with the simple system shown in Figure 7.10, at least two-thirds of the ion beam at exit reaches the collector. In a number of commercial instruments the point A is raised 50 to 100 V positive with respect to ground. This has the same effect as biasing the collector negative, and hence tends to improve collection efficiency.

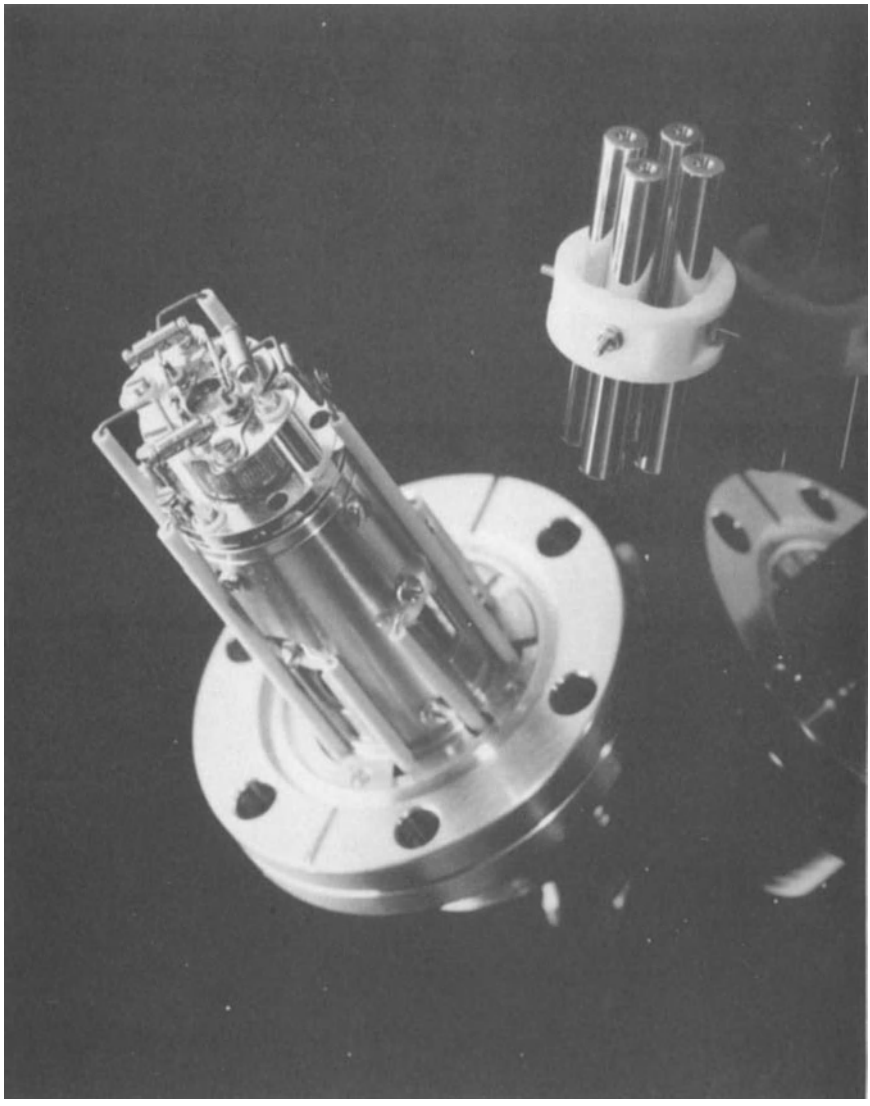


Figure 7.12 Construction of ion source and rod assembly for quadrupole mass analysers designed by VG Quadrupoles.

If the analyser is considered as a 'partial pressure ionization gauge' it can be compared directly with the 'total pressure ionization gauge'. When the source is operated with electron currents in the range 1–5 mA, i.e. in a similar manner to the ionization gauge, the sensitivity is about 10^{-4} A mbar^{-1}, which means that with good conventional electrometer amplifiers partial pressures down to

10^{-10} or 10^{-11} mbar can just be detected. (The sensitivity is about two orders of magnitude lower than in the extractor gauge designed by Pittaway[12]. This is not unexpected, because (i) the ionization chamber volume is smaller by almost a factor of 10, and (ii) the overall efficiency of ion transmission from lens output to collector is probably about 10%.)

Although the basic requirements for the small RGA are easy to formulate, considerable attention must be given to the detailed specification. This is essential because of the interrelation between all the various operating parameters, which means that a change in design to improve one feature almost inevitably causes a deterioration in at least one other. The following is a list of those parameters which are important in a typical application.

(i) Overall cost
(ii) Size
(iii) Mass range
(iv) Sensitivity (or minimum detectable partial pressure)
(v) Resolution
(vi) Maximum operating pressure
(vii) Stability of operation.

The basic relationships to take into account when fixing the operating parameters are:

Resolution $\propto N^2$

Sensitivity $\propto r_0^2 I_e$

$$M_m \propto \frac{V_m}{\omega_0^2 r_0^2}$$

where I_e is the electron current in the ion source, V_m the maximum output voltage from the RF power supply, and M_m the upper limit to the mass range, i.e. the maximum ion mass to which the quadrupole can be tuned.

The link between the various parameters is obvious. For example, if the resolution is to be improved by increasing the length of the filter, there is a consequent increase in cost and complexity; if by stepping up the RF frequency, there is a reduction in the mass range M_m. Sensitivity can be increased by increasing r_0, but again there is a cost penalty as well as a contraction of the mass range. Extending the mass range M_m by increasing V_m again carries a cost penalty because of the larger RF power supply.

Due to space charge effects the electron current plays a complex part in the gauge operation. The electron space charge depresses the potential in the ionization chamber with respect to the cage, and consequently reduces the ion injection energy. In order of magnitude terms, the ion injection energy is depressed by about 1 eV for every 1 mA increase in electron current. Positive ion space charge is also important because it affects the sensitivity at the high-pressure end of the scale. The magnitude of this effect for a commercial

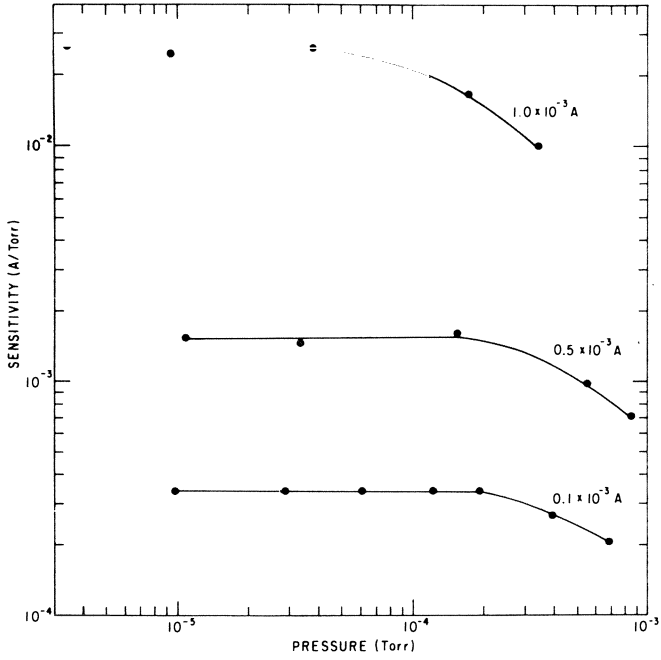


Figure 7.13 Graphs showing sensitivity for hydrogen as a function of total pressure (hydrogen) in the mass filter. Curves are plotted for electron current in the source of 1.0, 0.5 and 0.1 mA[17].

analyser is illustrated by the graphs of sensitivity plotted as a function of pressure for hydrogen in Figure 7.13[16]. These typical curves indicate the need to operate at low electron currents if quantitative measurements are required at pressures in the region of 10^{-5} mbar or above. As might be expected, the effect increases in importance at the high mass end of the scale. This is indicated by Table 7.1, which lists the values of pressure p_0 at which sensitivity falls to zero for the mass range 4–84 amu. Hu and Qiu[17] have demonstrated that, with specially designed ionization chambers, the useful range can be extended up to 10^{-3} mbar. By using short electron paths, and with draw-out fields to overcome space charge problems, their sources gave a linear operation up to 10^{-3} mbar, thus making ion – molecule collisions in the filter the limiting factor for high-pressure operation. In this way, when using a filter 80 mm long, they found a linear operation to 5×10^{-4} mbar and a maximum output signal at a pressure of 10^{-3} mbar (p_0 as defined in Table 7.1) for Ar^+.

The ion injection energy (IE) must always be chosen with care because of its influence upon both sensitivity and resolution – a reduction in IE, and hence the velocity with which the ions traverse the filter, obviously increases N according to the relation $N \propto (\mathrm{IE})^{-1/2}$. Interrelation between resolution, sensitivity and injection energy is shown by the results plotted in

Table 7.1 The pressures, $p_0 \times 10^{+5}$ mbar, at which the quadrupole output signal reaches the saturation level. Values are given for electron currents of 0.2 and 2.0 mA.

	Helium	Argon	Krypton	Xenon
0.2 mA	33.0	6.0	4.5	3.6
2.0 mA	6.5	2.8	1.6	0.9

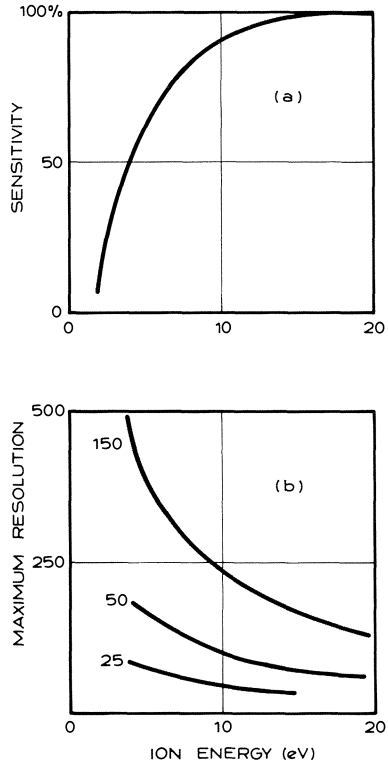


Figure 7.14 Graphs showing (*a*) sensitivity and (*b*) resolution of mass filters 150, 50 and 25 mm long as a function of the ion injection energy.

Figure 7.14[18], which were obtained for small prototype instruments with different filter lengths.

7.4 The operating characteristics of the RGAs designed for general laboratory and industrial use

The measured characteristics of most small instruments exhibit the features predicted in the above theoretical analysis. For example, there is a precise

linearity between 'mass and applied RF voltage' in the spectrum obtained by the normal scan, and the facility to 'trade' sensitivity against resolution by varying the slope of the scan line[19]. It is easy to obtain a precise calibration of the mass scale to within $\pm$ 0.1 amu over the whole range. One important factor which, because of its complexity, cannot easily be taken into account, is the degradation in performance due to any departures of the electric fields from the theoretical ideal. Usually these imperfections affect instrument sensitivity and resolution, but not the calibration of the mass scale. In practice the problems arise mainly from (i) mechanical imperfections in the individual components and misalignments in the assembly process, and (ii) electrical charges on the electrode surfaces caused by the build-up of contaminating layers (usually hydrocarbon)[20].

Paul's contour map[7] of the stability region, Figure 7.8, indicates the precision to which the dc potentials must be controlled when the filter is operated as a quantitative measuring instrument. For example, when the filter is tuned to the centre of the stability zone ($q = 0.706$), and on to the operating line giving $R = 100$, an increase in the variable a by 0.001 (i.e. an increase in V_{dc} by 0.4%) reduces the effective orifice diameter from 0.03 r_0 to 0.02 r_0. This reduces the efficiency of ion transmission and hence sensitivity by a factor of about two. The experimental results shown in Figure 7.15, which are typical of many made at the University of Liverpool on small prototype instruments,

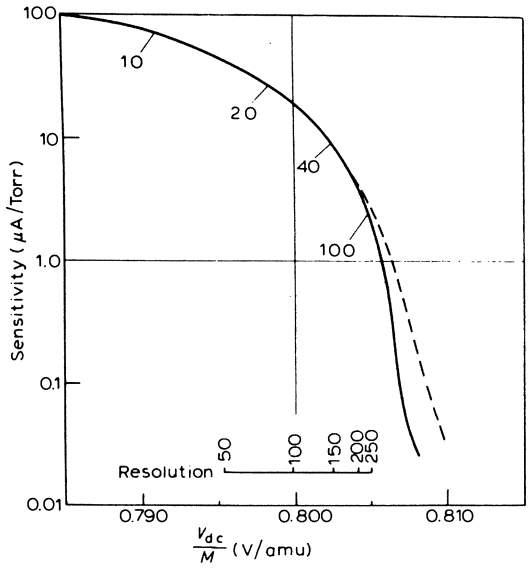


Figure 7.15 The relationship between the normalized sensitivity and V_{dc}/M for Xe^+ (solid line) and Xe^{++} (broken line). Values of resolution for the Xe^+ ions are indicated by the auxiliary scale. Also at particular points on the curve the % change in sensitivity caused by a 0.1% change in dc voltage are marked.

confirm this prediction. In this graph, sensitivity is plotted as a function of V_{dc}/M, a factor directly proportional to a when ω and r_0 are constant (the independent variable moves along the dotted line in Figure 7.8, maintaining $q = 0.706$). Measurements were made for xenon, the two ions $^{129}Xe^+$ and $^{129}Xe^{++}$ being chosen for analysis. The curves show a slow steady fall in sensitivity as V_{dc} is increased up to the point where $R = 200$, beyond which the sensitivity falls extremely rapidly for the singly charged ion. These curves indicate considerable mass discrimination at high resolution, and their steadily increasing gradients at higher values of V_{dc} show the greater dependence of sensitivity upon the applied voltages at the higher resolutions. In general terms, the measured values in Figure 7.15 are in good agreement with values calculated from Figure 7.8*.

Departure from the ideal characteristics in small quadrupole filters designed for residual gas analysis can be seen from the study carried out by Reid and James[21], who made a careful analysis of eight nominally identical RGA heads. In one experiment, calibrations were made for the noble gases, ranging from helium to xenon, over the pressure range 10^{-9} to 10^{-5} mbar. The instruments were set up with the scan adjusted so that the widths of all the peaks were

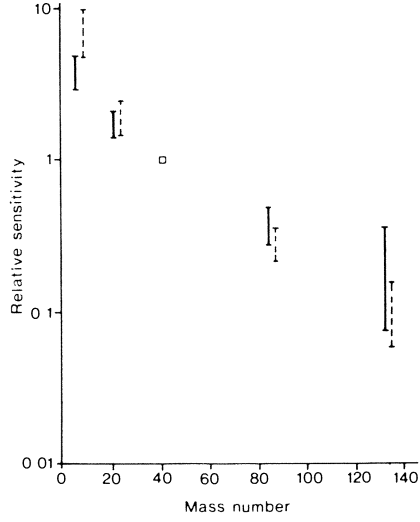


Figure 7.16 Sensitivity spreads for the eight RGA heads for various noble gases, plotted relative to the argon sensitivity. Mass discrimination —— for the Faraday plate detector; – – – multiplier detector. Measurement pressure 10^{-7} mbar for a gas mixture containing 5% each of helium, neon, argon, krypton and xenon with the balance made up of nitrogen. (After Reid and James[21].)

*The simple technique of using singly and doubly charged noble gas ions for comparing sensitivity and mass discrimination over a wide mass range is worth noting. Provided source conditions remain constant, the abundance ratio of singly to doubly charged ions is unlikely to change during a calibration experiment.

approximately 1 amu at 10% peak height over the whole mass range. (For this, as with many other commercial instruments, resolution could be adjusted by offsetting the scan line in addition to the variation in its slope.) The results in Figure 7.16 show not only the severe fall in sensitivity as mass increases, by as much as a factor of 10 for an increase in mass of 100 amu, but also the wide variation from instrument to instrument. Even without the electron multiplier there is a 'spread' of a factor of two in relative sensitivity change over the limited mass range 4–84 amu. It should be noted that these discrepancies reported by Reid and James[21] are in addition to any effects caused by ionization cross-section or cracking pattern variations (see Chapter 6). Further experimental evidence showing the complexity of the relationship between sensitivity and mass for small instruments can be seen from data published by Leybold for their Q200 instrument. Two separate spectra, obtained for the same gas mixture, are presented in Figure 7.17 for this filter, which is approximately 100 mm in length. With the scan line adjusted to give constant sensitivity over the whole mass range, the resolution remains approximately constant (i.e. $M/\Delta M$ constant). However, when ΔM is kept constant (i.e. constant line width), the sensitivity, although constant from 2 to about 40 amu, falls rapidly as mass is increased further. Figure 7.17 illustrates a useful technique for practical operation, i.e. a high resolution for peak identification, followed by a switch to lower resolution for peak height measurement. It also indicates the need for careful calibration over the operating range of the instrument.

In their investigation of RGA characteristics, Reid and James[21] found considerable variations in sensitivity over the pressure range 10^{-8} to 10^{-5} mbar. The non-linearities were similar for all eight filters tested, the sensitivity in every case being maximal at about 10^{-7} mbar, falling by about a factor of two at higher and lower pressures. Similar results have been obtained at Liverpool by Mao *et al.*[22], who measured the characteristics of commercial gauges from four manufacturers. A selection of their results, typical of those obtained from many gauges, is shown in Figures 7.18 and 7.19. These curves indicate similar variations in sensitivity to those reported by Reid and James[21]. The plots in Figure 7.18 show the variations in sensitivity for helium for three types of analyser over the pressure range 10^{-6} to 10^{-4} mbar. To make these measurements, the partial pressure of helium was set and held constant at 10^{-7} mbar. Argon was then introduced to increase the total pressure in the analyser through the range 10^{-6} to 10^{-4} mbar, the helium 'peak height' being monitored carefully as the pressure increased through the two decades. (When the argon pressure was reduced from 10^{-4} to 10^{-6} mbar, the helium signal always retraced the same path.) The results obtained for similar experiments with a fourth instrument are shown in Figure 7.19. In the first experiment a krypton 'trace' was introduced and held constant at about 10^{-7} mbar, with the pressure of the main gas (argon) taken through the range 10^{-7} to 10^{-4} mbar and back to 10^{-7} mbar. The role of the two gases was reversed for the second experiment. This is an accurate technique for

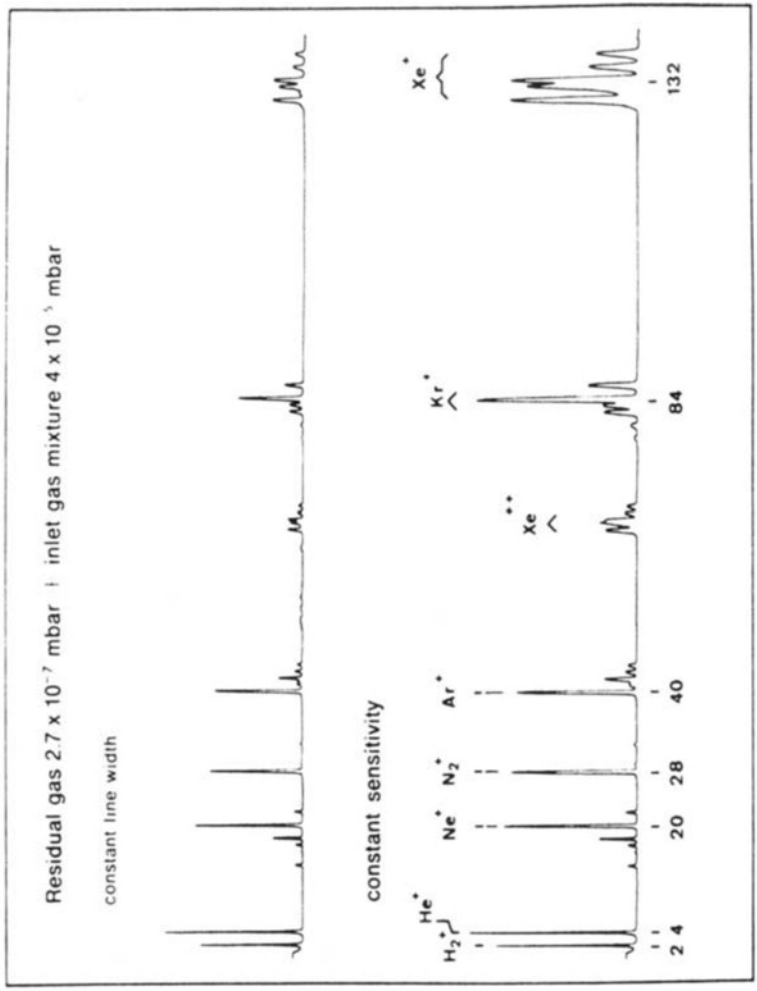


Figure 7.17 Manufacturers' published spectra for commercial RGA operating first at a high resolution (upper curve), and second at a lower resolution to give a constant sensitivity over the whole mass range (lower curve). (Data for Q200 provided by Leybold Ltd.)

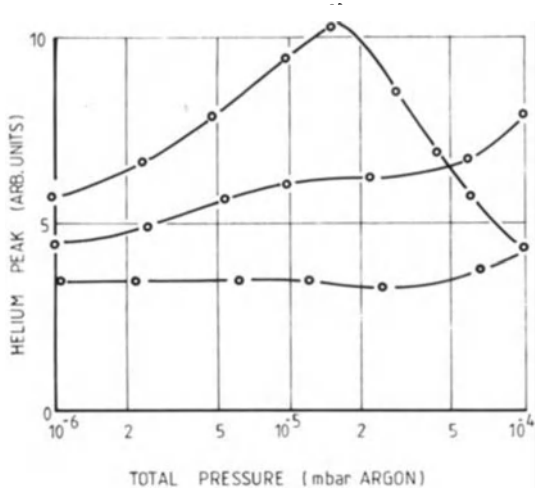


Figure 7.18 Curves showing the variation of sensitivity for helium over two decades of pressure for instruments from three different manufacturers[22]. To obtain these characteristics, the helium peak height was monitored continuously, while the total pressure in the chamber (argon) was varied through the range 10^{-6} to 10^{-4} mbar. The helium partial pressure was maintained constant at 10^{-7} mbar throughout the experiment.

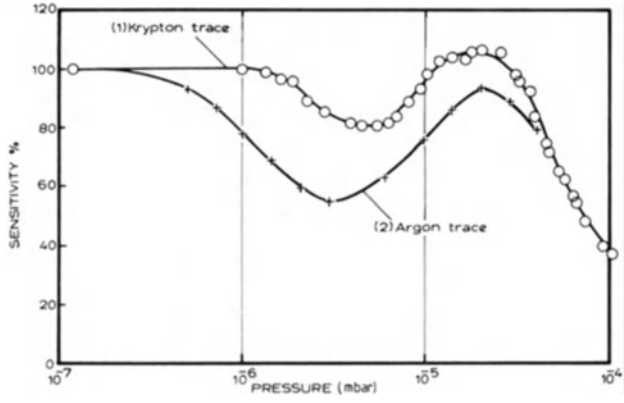


Figure 7.19 Sensitivity variation for argon and krypton over the pressure range 10^{-6} to 10^{-4} mbar[22]. The peak height of the impurity, or trace gas, was measured as a function of the total gas pressure in the vacuum chamber. For experiment (1), krypton was the trace gas with argon the main gas in the chamber; for experiment (2), the role of the two gases was reversed.

measuring variations in sensitivity over a wide range of pressure without the need for sophisticated reference gauges.

Considerable variations in sensitivity have been observed when quadrupoles are operated for long periods even under very good vacuum conditions. Calcatelli[23] and her co-workers for example reported variations in sensitivity of $\pm 20\%$ for argon over a period of 300 days when a commercial

instrument was operated in the range 10^{-8} to 10^{-9} mbar. Rather greater changes have been observed at Liverpool for RGAs operating continuously for long periods under significantly worse vacuum conditions[24]. Results for one particular instrument, which showed better than average performance from the point of view of long-term stability, are reproduced in Figure 7.20. The calibration curves were obtained for argon at approximately equal time intervals over an eight-month period. This analyser operated with the ion source on almost continuously in the residual atmosphere of an oil diffusion pump system (apiezon C) at pressures in the range 1.0 to 10×10^{-7} mbar. In addition, the noble gases nitrogen and hydrogen were admitted to the gauge at a pressure of 10^{-5} mbar for nearly 2000 h.

Precise measurements of the effects of field distortion have been made by Arnold[25], who designed a source and filter system which could be misaligned by small and well-defined amounts. As he expected, Arnold found that any

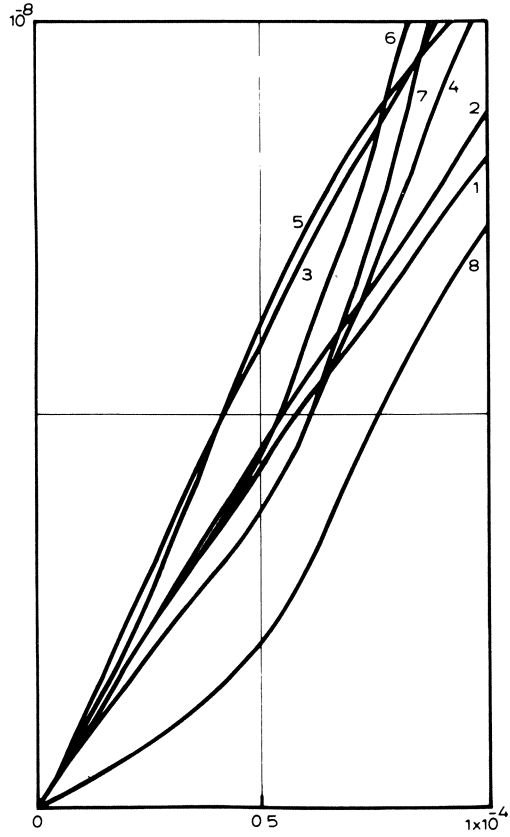


Figure 7.20 Characteristics of commercial RGA measured at approximately equal intervals over a period of eight months. The calibrations are for argon. The numbering on the curves indicates the order in which measurements were made[24].

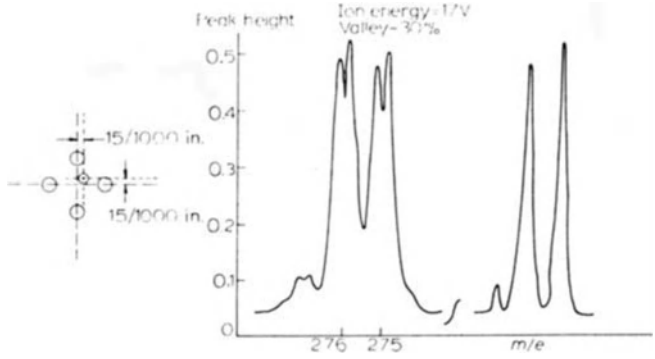


Figure 7.21 Peak splitting observed by Arnold[25]. The degradation of the peak shape is caused by misaligning the source assembly 0.38 mm off axis.

significant mechanical errors caused (i) a loss of resolution and a deterioration in peak shape leading to 'peak splitting', and (ii) a loss of sensitivity. Figure 7.21, reproduced from his report, shows the loss of performance due to setting the ion source 0.015 inches (0.38 mm) off axis. To maintain the original sensitivity, the ion injection energy had to be increased from 8 to 17 eV. Similar deteriorations were found when errors of the same order of magnitude were made to the filter assembly. For example, when one rod was displaced, first by 0.003 inches (0.075 mm) and then by 0.005 inches (0.125 mm), the maximum mass for which unit resolution could be obtained fell from above 300 to 150 and 70 amu respectively. Results in line with those of Arnold[25] have been obtained by Holme *et al.*[26] who, like many other workers[9,27–29], observed peak splitting when any significant mechanical errors were introduced to filter or source. For a similar filter to that used by Arnold[25] (but with shorter rods) Holme and his co-workers[26] found that an increase in ion injection energy of 10 eV was required to maintain sensitivity when the source was displaced by 0.25 mm.

7.5 The use of electron multipliers for signal detection

When the quadrupole mass filter is used for residual gas analysis, the Faraday Cup ion collector and a straightforward electrometer amplifier, with or without a simple modulation technique, are adequate for most applications. The system—small, simple, reliable and inexpensive—allows partial pressures to be measured to below 10^{-10} mbar. The principal limitation is the relatively long time constant which must be 'built into' the output amplifier circuit to suppress unwanted interference when measurements are taken at the lowest pressures. Typically, the response time must be in the range one to ten seconds when measurements are required at 10^{-10} mbar, which sets the maximum spectrum scan rate to about one peak s^{-1} if a true record is to be obtained. For

many applications, sophisticated electronic techniques are used to select only important peaks; this overcomes the limitations set by slow scan rates.

An electron multiplier detection system with a current 'gain' of the order of 10^5 or 10^6 and an effectively zero time constant immediately improves minimum detectable pressure and, at the same time, allows a much lower time constant in the signal output circuit. In practice, the improvement is by a factor of 100 to 1000 in both these parameters. This makes it possible to scan

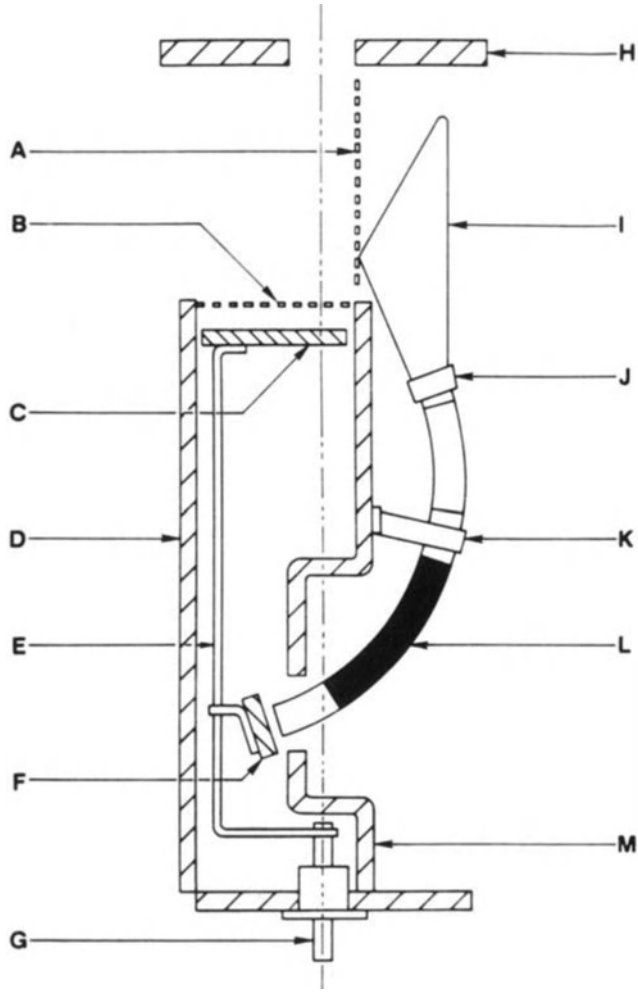


Figure 7.22 Simplified diagram of the channel electron multiplier assembly designed by Reagan *et al.*[30]. *A*, high-voltage; *B*, grounded mesh; *C*, ion collector; *D*, grounded shield; *E*, signal lead; *F*, electron collector; *G*, output connector; *H*, quadrupole exit aperture; *I*, CEM cone, *J*, high-voltage connection; *K*, ground connection; *L*, bias resistor, and *M*, grounded shield.

spectra at the rate of about 10 amu s^{-1} and makes a cathode ray tube display practicable without the need for special storage facilities, even when measurements are made at the lowest pressures. It also allows the fast transfer of information to a computer store.

Reagan *et al.*[30] have described a well-engineered system for mounting a channel electron multiplier (CEM) detector into a standard RGA. The important elements in their design are identified in the schematic diagram in Figure 7.22. The particular multiplier chosen was the model 4775 manufactured by the Galileo Electro-optics Corporation of Massachusetts. Following conventional practice, the multiplier was placed out of 'line of sight' of the ion source, in order to reduce background noise from photon collection[31,32] and also to allow operation in the Faraday Cup (FC) mode. As can be seen from Figure 7.22, the ion beam enters the detector through the quadrupole aperture (*H*) and in the FC mode passes through the grounded mesh (*B*) to strike the collector (*C*). In this mode both grid (*A*) and the cone (*I*) of the multiplier are at ground potential. To operate in the electron multiplier mode, a negative potential of between 1.0 and 3.0 kV is applied to the CEM cone and the mesh (*A*). The ion beam is thus deflected through this mesh onto the cone. The output current from the multiplier is drawn to the collector (*E*), so that in both modes of operation output is via the connector (*G*). By measuring the output current in both the FC and CEM modes at the same pressure, Reagan and her co-workers determined the multiplier gain for the noble gases,

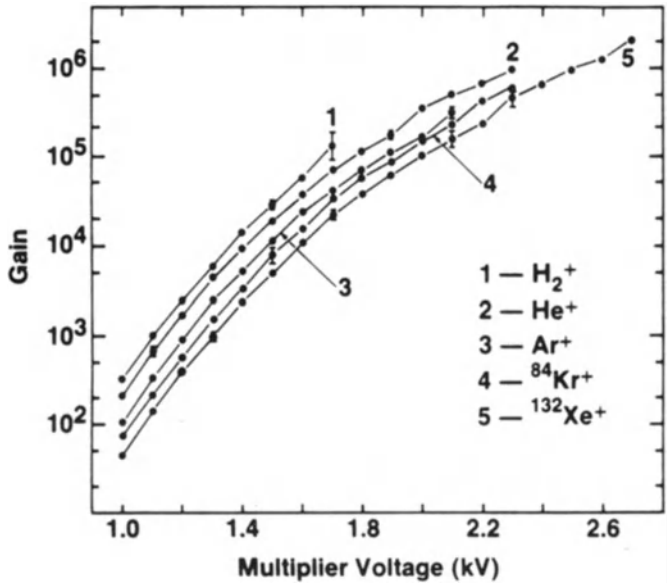


Figure 7.23 Gain of the channel electron multiplier as a function of multiplier voltage measured by Reagan *et al.*[30]. (Model 4775, Galileo Electro-Optics Corp.)

hydrogen, nitrogen and C_2F_6 at various values of multiplier voltage. Results, typical of the four multipliers they tested, are reproduced in Figure 7.23 to show the dependence of gain upon both multiplier voltage and the particular gas molecule. A careful analysis of the results shows gain to fall with increasing mass by approximately

$$D_i = A - B\sqrt{M} \tag{7.16}$$

where D_i is the multiplier gain relative to the '28' signal for nitrogen. With an HT voltage of 1.5 kV, the constants A and B have the values 1.5 and 0.9 respectively. Of all the gases tested, hydrogen was the only one to which the above equation could not be applied.

Electron multipliers enable the useful range of the small RGA to be extended to below 10^{-12} mbar and, at the same time, allow fast recording techniques to be used. The obvious disadvantage, in addition to the added complexity and expense, is the instability of 'multiplier gain' which can vary considerably with respect to both small changes in HT voltage and gas contamination. (See for example reference 16.) There is also the problem of the dependence of gain on ion mass, and to a lesser extent upon the type of gas molecule. These uncertainties and instabilities mean that an effective calibration technique is essential for almost all applications. For analytical work this is not necessarily a disadvantage, since any 'overall' calibration can be designed to cover the whole system; for example, variations in cross-section and filter transmission efficiency as well as the multiplier gain characteristics. Reid and James[21] found large changes in sensitivity over the mass range of their instruments which were not significantly different when a multiplier, rather than a Faraday Cup, was used in the output circuit. Thus for quantitative work with their instruments calibration is essential whether or not a multiplier is used.

7.6 Non-conventional methods of quadrupole operation

(i) Power supplies

Considerations of the basic theory of the quadrupole operation indicate that periodic field variations other than sinusoidal may produce a filtering action. Richards *et al.*[33] demonstrated this experimentally by replacing the $V\cos\omega t$ term in eqn (7.8) by a rectangular wave. With this method of operation no dc component need be generated, since any departure of the wave from square (i.e. when the duty factor $\delta = \varepsilon/\pi$ is not 0.5) is equivalent to applying an additional dc component to the rods. Richards[33] and his co-workers operated their spectrometer by first setting the duty factor to a predetermined level, to give the desired resolution, then observed spectra by steadily increasing the amplitude of the rectangular wave. They deduced that for a satisfactory performance, at resolving powers between 100 and 200, tolerances on the

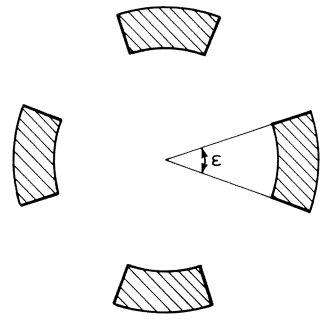


Figure 7.24 Electrode structure proposed by Hayashi and Sakudo[34] to approximate a hyperbolic field.

amplitude, frequency and duty factor of the rectangular wave should all be of the order of 1 part in 10^{+4}. The reason for the lack of development of this technique could be the problems of designing satisfactory power supplies which are more complex than the conventional sinusoidal.

(ii) Filter design

Fields other than true or near hyperbolic can produce a filtering action provided that symmetry is preserved. Hayashi and Sakudo[34] showed that the cylindrical concave structure, with each 'rod' a segment of a circle as illustrated in Figure 7.24, forms the basis of a viable filter. In this structure the inside diameter of the electrodes was 10 mm and the thickness 0.8 mm. The optimum angle ε (Figure 7.24) was calculated and confirmed experimentally to be 40°. In a brief report, Hayashi and Sakudo indicated that their structure gave a performance almost as good as the conventional analyser.

(iii) RF-only operation

The operation of the quadrupole with pure RF fields, rather than the conventional steady plus alternating, was first demonstrated by Brinkmann[35] in 1972. He placed a positively biased hemispherically shaped grid to act as a potential barrier at the exit of the filter so that only ions which had gained considerable kinetic energy, of the order of 100 eV, could reach the collector. In this mode of operation a 'spectrum scan' starts at the origin with the operating point moving along the q axis shown in Figure 7.6 (with no dc voltage applied to the rods, $a = 0$). For q confined to the range zero to approximately 0.90, ions follow stable orbits in the filter and reach the exit without a significant gain of kinetic energy. When $q > 0.908$, ions follow unstable paths, moving radially outwards from the filter. In neither case can ions reach the collector. However, as q moves through the interface between stability and instability some ions, on the verge of instability, move close to the rods at the filter exit, thus passing through a region of strong fringing fields. These ions can gain considerable kinetic energy from the fields; sufficient to allow them to traverse the potential

barrier and reach the collector. Thus, because of the extremely narrow range of values of q, for which this energy increase occurs, there is mass selection. Brinkmann[35] confirmed this experimentally, obtaining spectra with a good resolution, noting two problems, however; firstly, the inability to operate at low mass, and secondly, the appearance of relatively long 'tails' on the low mass side of all peaks. Holme[36] showed that, compared with the normal mode of operation, the RF-only technique gave a significantly better resolution at a given sensitivity. Further work at Liverpool[37–40] and Ottawa[41] showed that better performances could be obtained by using spatial, as well as energy, separation at the filter exit. Spatial separation was achieved by using a collector at ground potential in the form of an annular ring with the inner radius only slightly smaller than the value of r_0 (Figure 7.1). A central disc at a negative potential served to collect all ions of low kinetic energy.

Extensive experimental work at Liverpool[37–40] has provided strong support for Dawson's[41] theoretical analysis, in particular the prediction that with ions exposed to only a small number of cycles of alternating field, doublet and triplet peaks would be formed. These multi-peaks have been observed without ambiguity with a separation close to that predicted by theory. The resonances in the stability diagram due to third-order field distortions were demonstrated with a filter having one rod 2% oversize (diameter) to cause the field distortion. Repeated scans were made of the argon spectrum, first with RF only, then with successively increased values of dc offset, the dc being kept constant for each individual scan. This pattern and the results obtained can be seen in Figure 7.25. Of particular note is the satellite peak found in all spectra moving closer to, and eventually merging with, the main peak as the dc is

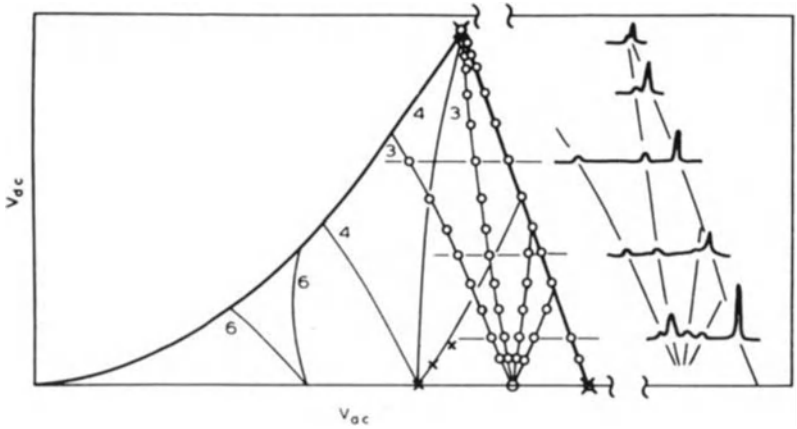


Figure 7.25 Argon spectra Ar^+ obtained for a filter with a serious third-order field disorder[37]. The position of the satellite peaks is indicated by O on the stability diagram. Resonance lines due to third- and fourth-order field distortions as calculated by Dawson are also drawn on stability diagram.

increased. The similarity between the double peak at the apex of the stability triangle and the split peaks observed with the conventional quadrupole, as seen in Figure 7.21, is obvious. The point on the V_{dc}, V_{ac} diagram at which each satellite was generated is marked by a small circle on Figure 7.25[37]. The contour lines on this graph are not smooth lines drawn through the points, but Dawson's[41] calculated resonance lines generated by third-order field distortion. (Note that only energy separation was used in this quadrupole.)

A feature of the characteristics of all RF mode instruments is the relatively constant sensitivity independent of mass (i.e. low mass discrimination) and the much reduced dependence on the contamination of the rods. Applications for this instrument are likely to be in areas where contamination may be a serious problem, and in particular where mass discrimination effects are likely to be important. The main disadvantage is that, although improvements have been made since Brinkmann's[35] early work, it is still not possible to obtain effective peak separation below about 16 amu.

7.7 The monopole mass spectrometer

The pioneering work on the development of the monopole mass spectrometer as an alternative to the quadrupole was carried out by Von Zahn[42] who demonstrated a practical instrument in 1963. Although the principles of operation have been known for some time, relatively little practical work has been carried out on this device, and its particular advantages over the quadrupole have therefore not been fully exploited. Monopole operation is possible because the two planes at right angles, each 45° to the xy axis in the quadrupole field, are always at zero potential when symmetrical potentials are applied to the two pairs of rods. In Figure 7.1 these are the planes aa' and bb'. The hyperbolic field in one quadrant of Figure 7.1 will be unchanged when three of the rods are removed and replaced by a V-shaped plate at zero potential lying along the planes aa' and bb'. This geometry is shown in Figure 7.26. (Because it is appropriate to consider stability in the y direction, it is the rods with the potential $-\phi_0/2$ applied which are significant for monopole operation.)

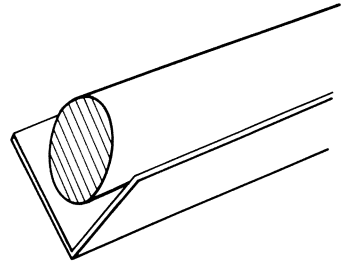


Figure 7.26 Electrode system for the monopole.

The same conditions for stability are applicable for the monopole as for the quadrupole, except that these become a necessary, but not sufficient, requirement for ions to travel along the length of the filter. Because the V electrode intercepts any ions crossing the *aa'* or *bb'* planes, there is the additional requirement that $y > |x|$. This can only be satisfied when the half-wavelength of the fundamental frequency of ion oscillation in the *yz* plane is exactly equal to the distance between the entrance and exit slits. This is illustrated in the grossly simplified sketches in Figure 7.27 which show trajectories when the wavelength is (*a*) less than, (*b*) equal to and (*c*) greater than the length of the instrument. This fundamental frequency is defined by the factor β in eqn (7.28); thus for the wavelength to be long, i.e. frequency low, the constant β must be very close to zero. (For exact focusing to occur, $\beta = 1/N$ where N is the number of cycles of alternating field 'seen' by the ions in their passage through the filter.) This stringent requirement restricts the basic stability zone described in Figure 7.6 to a very small strip along the 'leading edge'; this being indicated by the shading in Figure 7.28. The restricted stable region means that with conventional scanning, resolution is to a first approximation independent of the slope of the scan line, i.e. of the ratio V_{dc}/V_{ac}. This is indicated in Figure 7.28, which shows that a mass scan can be effective along either of the lines OQ or OQ'. (There is in practice a small loss of resolution as the slope of the scan line is reduced.) Thus, the reduction in the dc and RF power supply requirements to scan a given mass range gives the monopole an obvious advantage over the quadrupole. (There is a further gain of a factor of two because only two rather than four electrodes are used.) To set against this advantage, it must remembered that the fundamental wavelength must equal precisely twice the distance between entrance and exit slits; hence the 'focusing' is critically dependent upon the ion injection energy.

Grande *et al.*[43] and Herzog[44] have demonstrated the advantage of the reduced requirement for the power supplies in instruments similar to that designed by Von Zahn. Grande and his co-workers, using a unit giving $V_{ac} =$ 300 V maximum, were able to extend the range of their instrument from 200

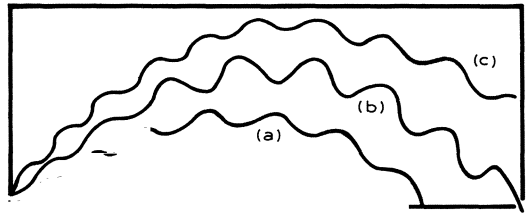


Figure 7.27 Sketch indicating the trajectories of positive ions in the monopole filter. Typical paths are shown when the half wavelength of the fundamental component of oscillation is (*a*) less than, (*b*) equal to and (*c*) greater than the distance between entry and exit orifices. Note: The ions enter and leave the monopole field through small well-defined orifices close to the apex of the '*V*' electrode.

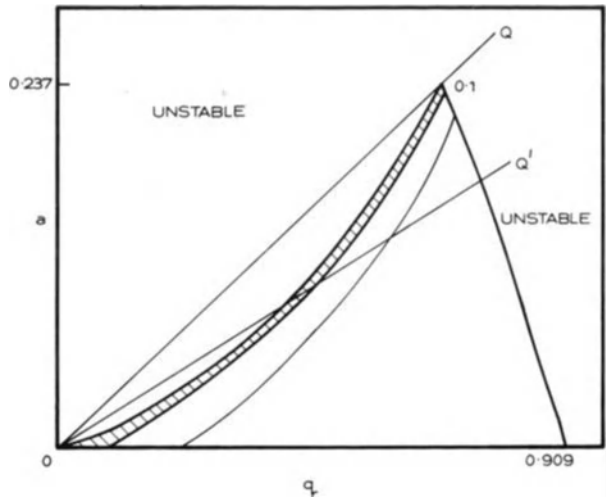


Figure 7.28 The stability zone, indicated by the heavily shaded area, for the monopole filter.

to 300 and then to 600 amu by reducing the slope of the scan line from 0.167 to 0.1 and then to 0.04. At each step the improvement was obtained without loss of sensitivity and with a resolution adequate to cover the whole mass range. Herzog[44] employed the more sophisticated scanning technique of operating along the '0.167 scan line' for the low mass range and then progressively reducing the slope (reducing V_{dc} keeping V_{ac} constant) to cover the higher masses. This gave him the 'best attainable' resolution at low mass.

To set against the real advantages of the simpler power supply demonstrated by the above workers are the stringent requirements for ion injection of the ions through the fringing fields at the entrance orifice. (The resolution is critically dependent upon the velocity of the ions in the z direction, and therefore on the ion injection energy.) As Dawson points out, 'since ion trajectories commence near the V block one might also suspect a greater sensitivity to electrode contamination, that bane of electro-dynamic mass spectrometers. A further limitation is the fact that the peak height is not a monotonic function of ion energy and mass peaks may have different ratios if the ion energy is changed. This is presumably due to variations in the quality of focussing at the exit' (reference 1, p. 46).

7.8 The three-dimensional quadrupole ion trap

If a study is made of ion motion in a three, rather than a two-dimensional quadrupole field, eqn (7.1) must be modified to include the more general case, giving

$$E = E_0(\lambda_1 x + \lambda_2 y + \lambda_3 z) \tag{7.17}$$

where, as previously, E_0 is a constant with the dimensions of field strength independent of position, but not necessarily of time, and λ_1, λ_2 and λ_3 are simple non-dimensional constants. When the Laplace criterion is satisfied, this equation reduces to

$$E = E_0\lambda_1(x + y - 2z) \tag{7.18}$$

and the electrical potential ϕ, obtained by integration, is given by

$$\phi = -\frac{E_0}{2}\lambda_1(x^2 + y^2 - 2z^2) \tag{7.19}$$

or in polar co-ordinates

$$\phi = -\frac{E_0}{2}\lambda_1(r^2 - 2z^2) \tag{7.20}$$

The appropriate electrode structure to set up this field is illustrated in Figure 7.29. This is the three-dimensional rotationally symmetrical quadrupole ion trap with cylindrical symmetry about the z axis. The central ring electrode is a single-sheet hyperboloid and the two end caps a two-sheet hyperboloid. When the potential ϕ_0 is applied between the central ring and the two end caps, the inter-electrode potential is given by

$$\phi = \frac{\phi_0}{2r_0^2}(r^2 - 2z^2) \tag{7.21}$$

with

$$r_0^2 = 2z_0^2$$

(In normal operation the end caps are often grounded and the drive potential applied to the central ring. This does not affect operation.)

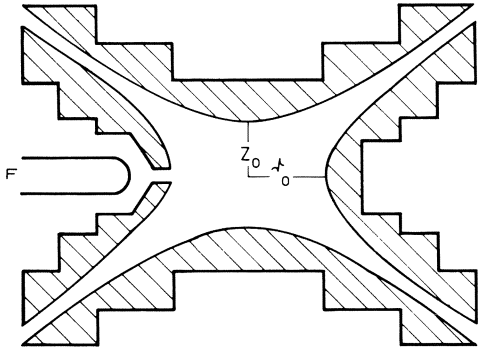


Figure 7.29 Electrode arrangement for the three-dimensional ion trap. The electrodes have circular symmetry about the centre vertical axis. The position of the filament used to inject electrons into the trap is indicated by *F*.

As might be expected, ion motion in the inter-electrode space follows the pattern already discussed for the two-dimensional case; the same stability criteria applying when the combined ac and dc potentials specified by eqn (7.8) are applied. The range of a and q values for which there is a stable solution is also represented by the stability diagram in Figure 7.5. To determine stability in both r and z, the two appropriate diagrams, differing by a factor of -2, must be superimposed (for the two-dimensional case the factor was -1). The resultant stability diagram is plotted in Figure 7.30, showing, as with the conventional two-dimensional filter, the wide range of values of a and q which give stable trajectories.

A number of techniques have been used to adapt this trap for use as a small practical mass spectrometer. In each case, ions are formed inside the trap by a beam of electrons which enters through a hole in one of the electrodes, usually an end cap, as indicated in Figure 7.29. In the early work at Bonn[45] the parameters a and q were set so as to maintain and trap all ions in stable trajectories. The ions were identified by measuring the resonant absorption of power from an auxiliary generator oscillating at 150 kHz due to their oscillations back and forth along the central axis of the trap. Successive ions were brought into resonance by varying the dc component of the applied field. Rettinghaus[46] later described a more sensitive detection technique in which the motion of the stored ions was registered as an induced voltage across a frequency tuned circuit coupled between the end caps. As a voltage was induced only when the frequency of ion oscillation equalled that of the tuned circuit, a mass spectrum could be obtained by operating with zero dc voltage ($a = 0$) and slowly increasing the RF voltage to bring successive ions into resonance.

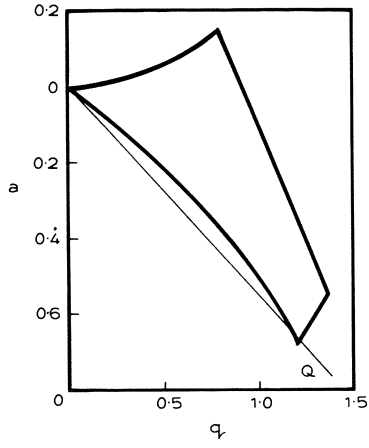


Figure 7.30 Stability diagram for the three-dimensional quadrupole ion trap.

Dawson and Whetten[47–49] made a significant contribution to the technology when they operated the trap as a 'selective store', that is in the manner of a conventional two-dimensional filter. They applied a combined dc and ac voltage to scan along the line OQ as indicated in Figure 7.30. Thus successive groups of ions were brought into resonance and held in the trap for as long as the scan line remained inside the appropriate 'stability zone'. The source was operated in a pulsed mode, the ions being formed and stored in a 5-ms period during which a gate pulse was used to admit electrons to the trap through a hole in the ring electrode. Detection occurred in the following 5 μs period when a voltage was applied to draw ions out through perforations in an end cap on to an electron multiplier. Dawson and Whetten built simple analysers (for example $Z_0 = 8$ mm, $r_0 = 11.35$ mm) capable of operation up to 100 amu with good resolution and able to detect partial pressures down to 10^{-12} mbar. They have published characteristics presenting in considerable detail the behaviour of this type of spectrometer. In particular, they showed that for RGA work a high precision is not necessary in the manufacture and assembly of the electrodes. They built a trap with electrodes made from coarse stainless steel mesh roughly formed into the correct shape by hand ($r_0 \simeq 17$ mm). Despite the imprecise shape of the electrodes, the spectrometer had a half-height resolution of 75, and using a 12-stage EMI electron multiplier with Be–Cu dynodes, partial pressure measurement down to 10^{-12} mbar was limited only by the total pressure achieved in their vacuum system.

Todd[50] and his co-workers have demonstrated a further alternative operating technique using a system geometrically similar to that of Dawson and Whetten, but based upon 'mass selective instability' rather than mass selective storage. This also operates in two phases; during the first, an electron beam is present and ac and dc voltages are applied across the electrodes such that all ions are trapped in the field. After a time interval, typically 1 ms, the electron beam is turned off and the dc and ac voltages and frequency changed either singly or in combination so that trapped ions of consecutive values of m/e become successively unstable. A fraction of these ions pass out through perforations in the electrode structure to impinge on the detector which is either an electron multiplier or a plate collector. Using only a small electrode system ($r_0 \simeq 10$ mm) Todd[50] and his co-workers showed that this technique can give good resolution, allowing the useful mass range to be extended to above 500 amu. They also demonstrated an improved resolution in the presence of a low molecular mass gas at a relatively high pressure (hydrogen or helium at 10^{-3} mbar), this showing the ion trap to be particularly suitable for operation in conjunction with a gas chromatograph.

References

1. Dawson, P.H. (ed.) (1976) *Quadrupole Mass Spectrometry and its Applications.* Elsevier, Amsterdam 1976.
2. Pittaway, L.G. (1974) *Vacuum* **24**, 301.

3. Courant, E.D., Livingston, M.S. and Snyder, H.S. (1952) *Phys. Rev.* **88**, 1190.
4. Blewett, J.P. (1952) *Phys. Rev.* **88**, 1197.
5. Paul, W. and Steinwedel, H. (1953) *Z. Naturforsch* **A8**, 448.
6. Paul, W. and Steinwedel, H. (1956) Ger. Pat. 944, 900; (1960) US Pat. 2, 939, 952.
7. Paul, W., Reinhard, H.P. and von Zahn, U. (1958) *Z. Phys.* **152**, 143.
8. Batey, J.H. (1987) *Vacuum* **37**, 659.
9. Dawson, P.H. (1986) *J. Vac. Sci. Technol.* **A4**, 1709.
10. Brubaker, W.M. (1970) *NASA Rept.* NASW 1298.
11. Holme, A.E., Thatcher, W.J. and Leck, J.H. (1972) *J. Phys. E.* **5**, 429.
12. Pittaway, L.G. (1974) *Vacuum* **24**, 301.
13. Dennison, D.R. (1971) *J. Vac. Sci. Technol.* **8**, 266.
14. Brubaker, W.M. (1968) *Adv. Mass Spectrom.* **4**, 293.
15. Fite, W.L. (1976) *Rev. Sci. Instrum.* **47**, 326.
16. Blanchard, W.R., McCarthy, P.J., Dylla, H.F., La Marche, P.H. and Simpkins, J.E. (1986) *J. Vac. Sci. Technol.* **A4**, 1715.
17. Hu, B. and Qiu, J. (1987) *J. Vac. Sci. Technol.* **A5**, 2657.
18. Holme, A.E., Thatcher, W.T. and Leck, J.H. (1972) *Vacuum* **22**, 327.
19. Watanabe, F. and Ishimaru, H. (1986) *J. Vac. Sci. Technol.* **A4**, 1720.
20. Holme, A.E. (1972) PhD Thesis, University of Liverpool.
21. Reid, R.J. and James, A.P. (1987) *Vacuum* **37**, 339.
22. Mao, F.M., Yang, J.M., Austin, W.E. and Leck, J.H. (1987) *Vacuum* **37**, 335.
23. Calcatelli, A., Bergoglio, M. and Rumiano, G. (1987) *J. Vac. Sci. Technol.* **A5**, 2464.
24. Mao, F.M. and Leck, J.H. (1987) *Vacuum* **37**, 669.
25. Arnold, W. (1970) *J. Vac. Sci. Technol.* **7**, 191.
26. Holme, A.E., Thatcher, W.J. and Leck, J.H. (1972) *J. Phys. E.* **5**, 429.
27. Story, M.S. (1967) *J. Vac. Sci. Technol.* **4**, 326.
28. Munro, D.F. (1967) *Rev. Sci. Instrum.* **38**, 1532.
29. Fairburn, A.R. (1969) *Rev. Sci. Instrum.* **40**, 380.
30. Reagan, N.R. Frees, L.C. and Gray, J.W. (1987), *J. Vac. Sci. Technol.* **A5**, 2389.
31. Goodings, J.M., Jones, J.M. and Parkes, D.A. (1972) *Int. J. Mass Spectrom. Ion Phys.* **9**, 417.
32. Benninghoven, C., Plog, C. and Treitz, N. (1974) *Int. J. Mass Spectrom. Ion Phys.* **13**, 415.
33. Richards, J.A., Huey, R.M. and Hiller, J. (1973) *Int. J. Mass Spectrom. Ion Phys.* **12**, 317.
34. Hayashi, T. and Sakudo, N. (1969) *Proc. Int. Conf. Mass Spectrom.* Kyoto, Japan.
35. Brinkmann, U. (1972) *Int. J. Mass. Spectrom. Ion Phys.* **9**, 161.
36. Holme, A.E. (1976) *Int. J. Mass Spectrom. Ion Phys.* **22**, 1.
37. Holme, A.E., Sayyid, S. and Leck, J.H. (1978) *Int. J. Mass Spectrom. Ion Phys.*, **26**, 191.
38. Ross, D.N. and Leck, J.H. (1983) *Int. J. Mass Spectrom. Ion Phys.*, **49**, 1.
39. Yang, J. and Leck, J.H. (1982) *Vacuum* **32**, 691.
40. Yang, J. and Leck, J.H. (1984) *Int. J. Mass Spectrom. Ion Processes* **60**, 127.
41. Dawson, P.H., Meunier, M. and Tam, Wing-Cheung (1980) *Adv. Mass Spectro.* **8B**, 1629.
42. von Zahn, U. (1963) *Rev. Sci. Instrum.* **34**, 1.
43. Grande, R.E., Watters, R.L. and Hudson, J.B. (1966) *J. Vac. Sci. Technol.* **3**, 329.
44. Herzog, R.F. (1976) in *Quadrupole Mass Spectrometry and its Applications*, ed. Dawson, P.H., Elsevier, Amsterdam, chapter 7.
45. Fischer, E. (1959) *Z. Phys.* **156**, 26.
46. Rettinghaus, G. (1967) *Z. Angew. Phys.* **22**, 321.
47. Dawson, P.H. and Whetten, N.R. (1968) *J. Vac. Sci. Technol.* **5**, 11.
48. Dawson, P.H., Hedman, J. and Whetten, N.R. (1969) *Rev. Sci. Instrum.* **40**, 1444.
49. Dawson, P.H. and Lambert, C. (1974) *Int. J. Mass Spectrom. Ion Phys.* **14**, 339.
50. Stafford, G.C., Kelley, P.E., Syka, J.E.P., Reynolds, W.E. and Todd, J.F.J. (1984) *Int. J. Mass Spectrom. Ion Processes* **60**, 85.

Subject index

Author index