Vacuum 101 (2014) 410–415

Contents lists available at SciVerse ScienceDirect

Vacuum

journal homepage: www.elsevier.com/locate/vacuum

The physics and technology of quadrupole mass spectrometers

Jonathan H. Batey*

102 Top Road, Kingsley, Frodsham, Cheshire WA6 8BX, UK

ARTICLE INFO

Article history:
Received 16 November 2012
Received in revised form
3 May 2013
Accepted 9 May 2013

Keywords:
Mass spectrometer
Physics
Quadrupole
QMS
Residual gas analysis (RGA)
Partial pressure analysis (PPA)

ABSTRACT

The fundamental principles of quadrupole mass spectrometers (QMS) are discussed, concentrating on instruments used in vacuum applications.

The mass-selective element is the quadrupole mass filter, first described in a 1956 patent. Although originally intended for isotope separation, the concept soon became applied to analytical chemistry and to residual gas analysis (RGA), also known as partial pressure analysis (PPA). The ion optical design of a typical general-purpose quadrupole RGA has altered little since the 1970s – although of course there have been major advances in electronics packaging, data capture, and automation. There is growing interest in miniaturised systems for high-pressure work.

These instruments mostly use electron-impact ionization sources, based on principles established nearly 100 years ago. Those used for vacuum applications are generally of two types: an “open” design, rather like a miniature Bayerd-Alpert gauge, and a “closed” design, generally based on Nier’s work.

The simplest and most robust detector is a Faraday collector; alternatively an electron multiplier is used when the highest sensitivity is required.

Often the quadrupole drive electronics and the amplifier are mounted directly on the vacuum flange, giving a compact assembly. Instrument control and data acquisition is almost always done via a computer interface. This gives a very versatile and powerful system. Nevertheless, to obtain the maximum benefit from any mass spectrometer it is often necessary to keep in mind the underlying physical principles.

1. Introduction

The concepts of mass spectrometry can be traced back to work by Aston at Cambridge and by Dempster at Chicago. Dempster’s 1918 paper [1] clearly shows the vacuum system, ion source, mass analyser and detector – all the basic functionality needed for a mass spectrometer. Dempster used the instrument to study isotopic distributions of volatile metals (Mg, Li, K, Ca, Zn).

In a quadrupole mass spectrometer (QMS) an alternating electric field generates the mass-selective action. The small size of the quadrupole, its relatively low cost, modest power requirements and linear mass scale have all contributed to its widespread acceptance as an instrument for residual gas analysis (RGA) in vacuum systems [2–4].

The quadrupole was first described by Paul & Steinwedel [5], see also Paul & Raether [6]. Although we know the quadrupole primarily as an analytical instrument, it was originally conceived with isotope separation in mind. Finlan et al. [7] describe an experimental system built at Amersham in which the rods were 3 m long with an r_0 (inscribed radius) of 1.35 cm. They report direct observation of the ion beam in the quadrupole, but unfortunately there are no photographs. Von Zahn [8] describes a system that used tensioned molybdenum wires to define the electrode surfaces: the electrodes were 5.82 m long with a hyperbolic profile. The inscribed radius was 3.5 cm. The total tension from the multiplicity of wires amounted to nearly 1000 tons.

Of course most quadrupole instruments are much more compact. Instruments for general-purpose research or for chemical analysis, such as gas chromatography mass spectrometry (GC–MS) or inductively coupled plasma mass spectrometry (ICP-MS) are usually 200 mm to 300 mm in length, with an inscribed radius of 6 mm–7 mm. A mass filter for residual gas analysis (RGA) is typically 100 mm to 150 mm in length with an inscribed radius of around 3.5 mm. It is usually built on a standard UHV-compatible 70 mm flange with electrical connections made via a multi-way feed-through. The drive electronics will usually be very compact, and mount directly onto the vacuum flange. The basic mechanical design of the mass filter and of RGAs in particular has altered little. Much of the commercial development during the past thirty years or so has concentrated on the electronics and data acquisition.

---

* Tel.: +44 1928788706.
E-mail address: jonathan@jhbatey.co.uk.

0042-207X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.vacuum.2013.05.005

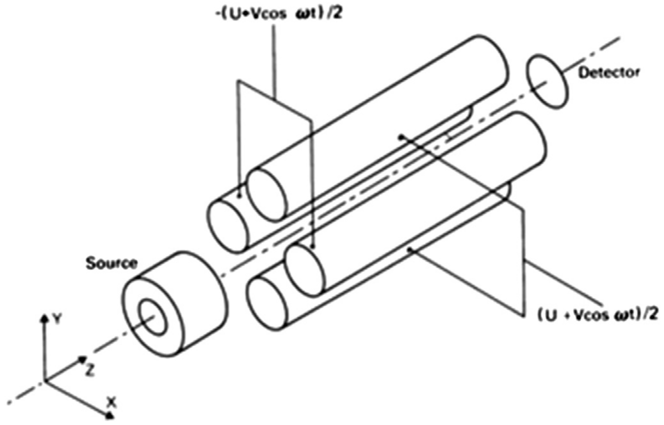


Fig. 1. Quadrupole mass spectrometer for residual gas analysis.

When the author started working with RGA quadrupoles, a data system was something of a novelty. Today it would be inconceivable to manufacture an instrument without one!

Much smaller RGAs have been developed recently [9]. This is largely driven by the desire to operate at relatively high pressure, since small dimensions make for reduced losses by ion/molecule collisions and reactions.

2. The quadrupole mass spectrometer (QMS)

The main functional components of any mass spectrometer are in principle the same as in [1], although usually we would add an inlet system, electronics and a data system for completeness. The main components of a QMS for RGA are shown in Fig. 1.

2.1. Inlet system

Often the RGA is installed in the environment it is to monitor, so there is no inlet system as such. But in coarse vacuum, or for calibration, we may need to install the RGA in a separate differentially pumped environment, so that it can operate at a manageable pressure. If so, we must ensure that the gas composition within the ion source is representative of the composition to be monitored. Therefore the inlet system should be designed so as to avoid fractionation [10].

When the sample is at a pressure higher than the operating pressure for the RGA itself (typically 10^{-6} to 10^{-5} mbar) the sample has to be introduced via a pumping restriction. This might be a valve or a simple aperture as indicated in Fig. 2a. It can be advantageous to operate the ionization volume at a pressure (typically 10^{-4} mbar) that is higher than the pressure in the main RGA housing. Such a “differentially-pumped” source is common in analytical work, see Fig. 2b. The filament is outside the ionization volume and so operates at a reduced pressure. This is beneficial for filament lifetime and reduces unwanted interaction between the filament and the sample which can cause problems if reactive gas species are to be analysed [11].

2.2. Ion source

There are many different types of ion source, but for general-purpose analytical work, electron-impact ionization [12] is usually employed. The ion source may be quite complex – it may incorporate magnets to confine the electron beam [13], and perhaps selectable aperture sizes to change the pressure within the source. But for RGA a simpler arrangement, based on a design [14] intended originally for molecular beam work, usually suffices.

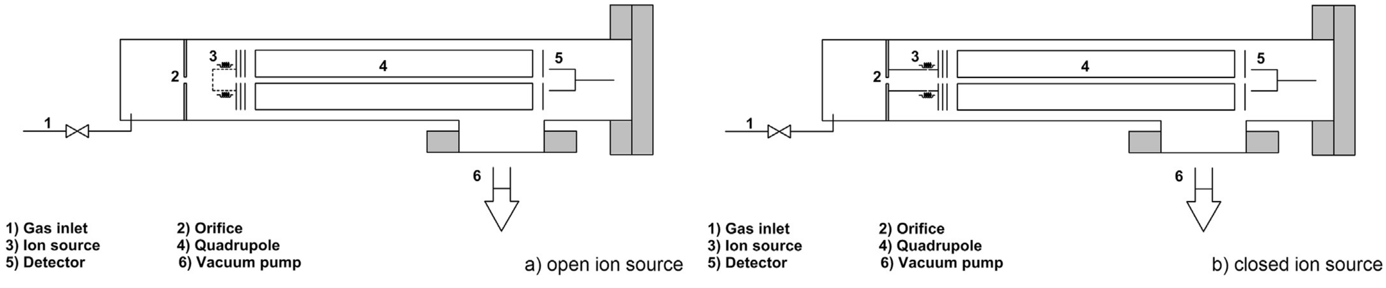
A typical RGA source is shown schematically in Fig. 3. It resembles a small extractor gauge. Electrons are emitted from an electrically-heated filament. It is made from, or coated with, material with a relatively low work function, so that electrons are readily emitted when it is heated.

The electrons are accelerated towards and through a grid to enter the ionization region with enough energy (typically 40 eV–100 eV) to ionize the sample. The electron trajectories are complex, and may traverse the ionization region several times before eventually impacting on the source grid. The emission current (that is, the current of ionizing electrons) is monitored at the source electrode, and is regulated by controlling the heating current that flows through the filament. In effect, the filament and the source electrode behave like the cathode and anode of a thermionic diode valve. Ions formed in the ionization volume are extracted by means of an electric field applied to an aperture plate, and are directed into the mass analyser.

The electron energy is essentially determined by the voltage applied between the filament and the source electrode. It is often set to around 70 eV, which is close to the optimum for most gas species. For some applications it is advantageous to set it lower; this reduces fragmentation and the tendency to form multiply-charged ions, and so simplifies the mass spectrum. For example, when analysing trace levels of water in argon, there is an interfering peak at mass 18 from doubly-ionized ^{36}Ar which is equivalent to around 300 ppm of water. By reducing the electron energy to 40 eV the detection limit of water can be improved by at least a factor of 10.

The sensitivity is proportional to the number of electrons available for ionization, and therefore in principle to the emission current; it is usually expressed in terms of the signal current per unit of pressure. The heating current flowing through the filament is typically 2–3 A and the emission current is typically 1 mA; this can be expected to give a sensitivity of around 10^{-4} A/mbar for a simple gas species such as nitrogen or argon.

A voltage, usually in the range 5–10 V, is applied to the source electrode. The extractor electrode is driven with a negative voltage, typically in the range −50 V to −130 V. This establishes an electric field that penetrates into the ionization region and helps to direct



a) open ion source

b) closed ion source

Fig. 2. Sampling from a higher pressure.

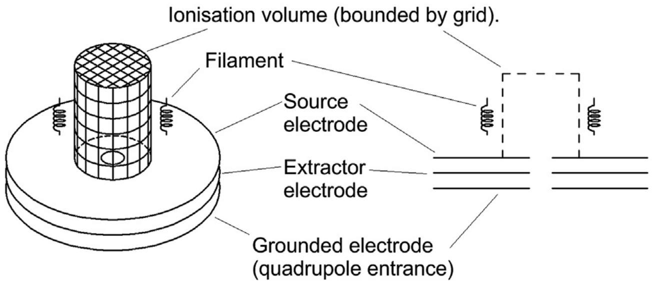


Fig. 3. RGA source.

the ions through the three apertures. This extraction field gives rise to a variation in potential of a few 100 mV over the region where most of the ions are generated. Accordingly ions enter the quadrupole with an energy slightly lower than the source voltage and with a spread of a few 100 mV.

Sometimes the source is operated at a higher voltage. The usual reason for this is to ensure that the ionizing electrons originate in a potential well, so that they cannot escape into the surrounding vacuum environment. For example, the source might operate at +100 V, and the filament at +30 V. This would give an electron energy of 70 eV. It is then necessary to superimpose on the quadrupole rods a voltage to retard the ions back down to a few eV. The voltage is known as the “pole bias” or “field axis”. For the conditions above it would probably be set to about +95 V.

At first sight it may seem an attractive option to raise the emission current in the expectation of higher sensitivity. In practice this is usually unwise, because the density of the electrons will lower the potential locally. When present, this electron space-charge can cause poor linearity at high pressure, because ions tend to be trapped in the resulting potential well [15]. Indeed, it is often best to operate at an emission current substantially lower than 1 mA. Although the sensitivity will be proportionately lower the linearity will be better.

The source shown in Fig. 3 is of rather open construction. This is appropriate when the source is to operate at the same pressure as the gas being sampled, traditionally up to about 10^{-4} mbar, or (more recently) 10^{-2} mbar with special instrumentation [9]. For UHV work outgassing can be a major problem, see sections 2.3 and 3 below.

2.3. Filament

Several different materials are used for filament construction, the commonest being tungsten. This is generally satisfactory at pressures below 10^{-6} mbar, but reacts at higher pressures with oxygen, generating CO and CO_2. It also reacts with hydrocarbons, and with water, leading to embrittlement of the filament. There is a tendency to deposit refractory oxides of tungsten onto surfaces of the ionizer. These oxides form insulating layers which are difficult to remove. If the gas mixture changes between oxidising and reducing species the sensitivity can alter a lot, although prolonged operation in an inert atmosphere is possible. It is relatively stable in a halogen-rich environment (because any adsorbed gas evaporates off quickly). It is mechanically robust – at any rate before it has been operated at high temperature. It runs very hot, so it dissipates significant heat in the vacuum system and therefore may cause an unacceptable degree of outgassing.

Thoria-coated iridium filaments are more fragile, but are usually more stable in operation. The thorium oxide emits electrons at a lower temperature, so that there is less outgassing than with tungsten. It is much more stable than tungsten when operated in relatively course vacuum, say above about 10^{-5} mbar. Indeed, the filament does not burn out if the vacuum fails, although operation at atmospheric pressure is not recommended. Arguably there is an environmental concern due to low-level alpha emission from thoria, so there is a tendency to use yttria coating these days. This gives similar performance to thoria, with no health worries.

Table 1
Filament materials.

Material	Operating temperature	Use
Tungsten	2400 K	$P < 10^{-5}$ mbar. Inert gases
Rhenium	2300 K	$P > 10^{-5}$ mbar. Hydrocarbons
Thoria-coated iridium	1900 K	General purpose. Robust
LaB_6	1300 K	Special applications

Rhenium is sometimes used, particularly in vacuum environments rich in hydrocarbons, since it does not form stable oxides or carbides. It is more volatile than tungsten and so it can itself contribute to the residual gas spectrum. Also, rhenium can evaporate onto insulating components in the ionizer, leading to failure of the ionizer.

Finally, lanthanum hexaboride is sometimes used as a filament material. The operating temperature is low, 1300 K, which makes it suitable for analysing products which dissociate at higher temperatures, notably hydrogen. It is also recommended for work with nitrogen oxide. But it is a very brittle material, so the failure rate from mechanical damage is relatively high.

Table 1 summarizes the properties of the various filament materials.

2.4. Quadrupole mass filter (QMF)

The quadrupole depends on the properties of a rather special electric field: the potential distribution is hyperbolic in two perpendicular directions (X,Y) and nominally invariant along the third (Z). This field distribution was first used to confine and shape ion beams using static voltages, but when the field is made to vary with time it becomes mass-selective.

The mathematical theory of the QMF is well-known and is not discussed in detail here. Dawson [16] gives a comprehensive treatment, and there are many others in the literature. From the theoretical treatment emerges the concept of the Mathieu stability plot, an example of which is shown in Fig. 4 (see [17] Fig. 2.5) which

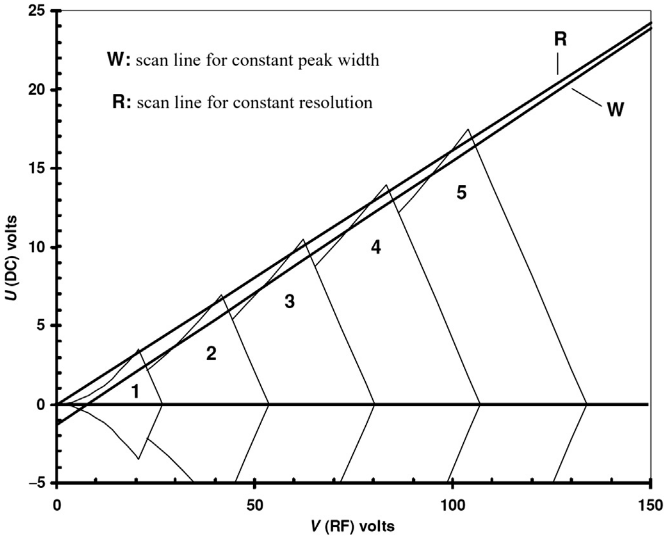


Fig. 4. Mathieu stability plot.

is used to predict the drive voltages required to transmit ions of a particular mass-to-charge ratio. Fig. 5 (see [17] Fig. 2.6) shows conceptual mass spectra generated by varying the applied RF and DC voltages along a "scan line" such as W or R in Fig. 4. More realistic spectra can be generated using ion-optical simulation.

2.5. Detector

After the ions have passed through the mass filter, they must be detected. A recent account may be found in [18]. The signal is usually measured as an analogue current, or as a discrete count rate. A count of 10^6 ions per second corresponds to a current of approximately 1.6×10^{-13} A.

2.5.1. The Faraday cup

The simplest system consists of a flat electrode which collects the ions. The ions are neutralized by transfer of electrons, and the resulting current is equal to the incoming ion current.

In the Faraday cup detector, the electrode is enclosed, except for a small entrance aperture; this ensures that secondary electrons are unlikely to escape before striking the electrode, which might otherwise lead to erroneous readings. A suppressor electrode, operated at a negative potential, can reduce this possibility still further.

Normally, the detected currents are small, and so a sensitive amplifier is needed. Also, the detection system must be well insulated, in order to eliminate leakage currents.

The minimum current that can be reliably detected using a Faraday cup is about 10^{-14} A.

2.5.2. Electron multipliers

Some limitations of the Faraday cup may be overcome by using an electron multiplier. Several types are available.

The principle of operation is that ions emerging from the mass filter are accelerated by a strong electric field, and strike an electrode, or 'dynode' which has a high coefficient of secondary electron emission. Secondary electrons are generated, which are accelerated towards a second dynode. They hit it and generate further secondary electrons. The process repeats until a large number of electrons emerge from the multiplier to strike a Faraday plate, so generating a large signal. The resulting current is further amplified as necessary.

The gain depends on the energy and angle of incidence of the incoming ion, on the nature of the dynode surfaces and on the

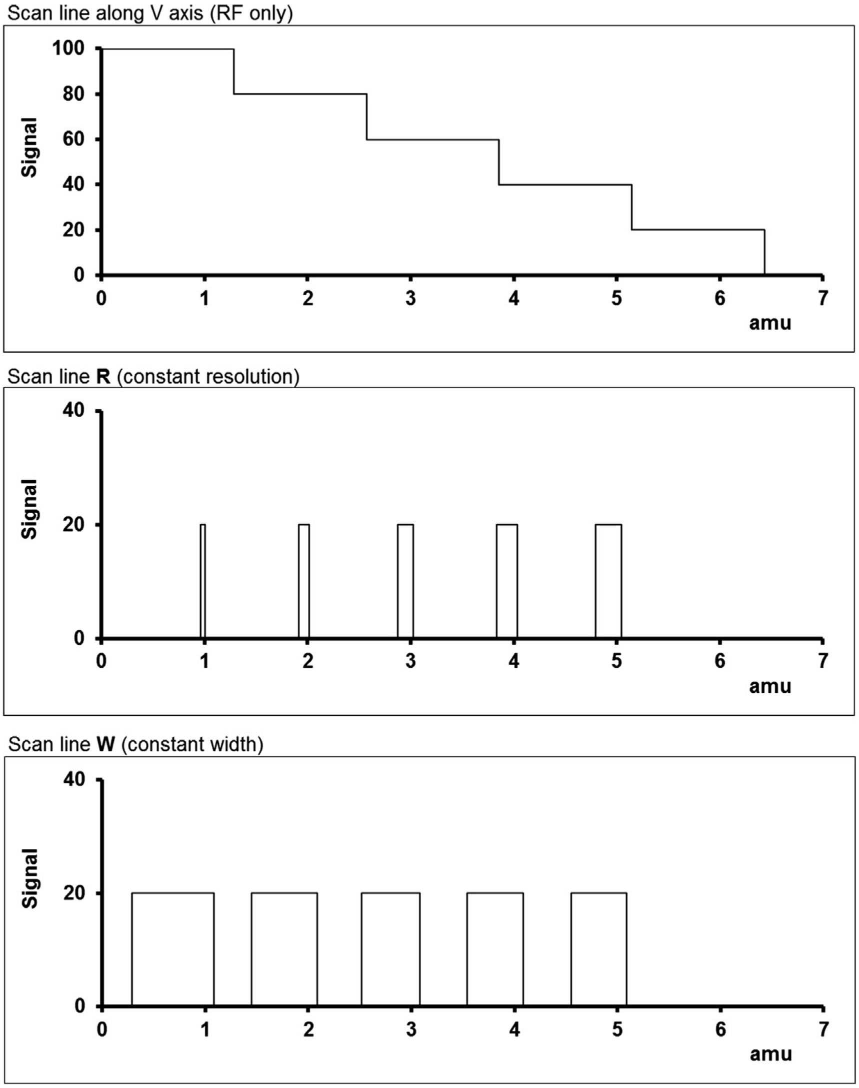


Fig. 5. Conceptual mass spectra.

number of dynodes or, in the case of channel multipliers on the aspect ratio of the channel.

Usually the mouth of the multiplier is operated at a high negative potential, typically −1 kV to −3 kV, to give an overall gain in the range 10^4 to 10^8. For an incident ion current of 10^{-18} A, about six ions s^{-1}, the output current from the multiplier can be as high as 10^{-12} A which is readily measured, although analogue filtering with a bandwidth of several seconds will be necessary to obtain a reasonable signal-to-noise ratio.

Besides improved detection limits, the electron multiplier gives improved response time, and hence allows the instrument to be scanned faster. This is because the response of the multiplier is very fast, typically less than 50 ns, and the current which it produces can be measured by a fast electrometer. Three types of multiplier are illustrated in Fig. 6.

The discrete-dynode electron multiplier uses a sequence of separate dynodes, traditionally made of copper/beryllium alloy, which degrades rapidly if exposed to ambient air. More recently air-stable dynode materials have become available. Each dynode contributes a gain of only 2–3 times, but the effect is multiplicative, so a gain of 10^8 is obtainable with a reasonable number of dynodes. Typically 16 to 20 dynodes are used.

In the continuous-dynode electron multiplier, or channel electron multiplier (CEM) the amplification is provided by a continuous surface made from a special glass [19]. The device is in the form of a hollow cylinder with a conductive inner surface, typically with a resistance of 80–100 MΩ. When a high voltage is applied between the ends of the cylinder, a potential gradient is established, and the process of charge amplification is similar to that which occurs in a discrete dynode multiplier, except that the process is distributed along the length of the channel. The main advantages of the CEM are its small size and its tolerance for exposure to ambient air.

The micro-channel plate (MCP) operates in a similar fashion to the CEM, but there are many channels distributed uniformly over the surface [20]. The device is primarily intended for image intensification, but is sometimes used as a low-cost alternative to a CEM. Generally, the glass used to make these devices is of a much higher resistance than that for a CEM, so that the level of output current which can be obtained is lower.

The gain is limited to 10^4, although it is possible to cascade the devices. Arguably the useful lifetime is likely to be lower than that for a CEM, owing to degradation of gain. The outstanding advantage is that because of the short length, 1–2 mm, the transit time is very short, less than 1 ns, so that very rapid events can be monitored. This finds application in time-of-flight mass spectrometry and pulsed laser work.

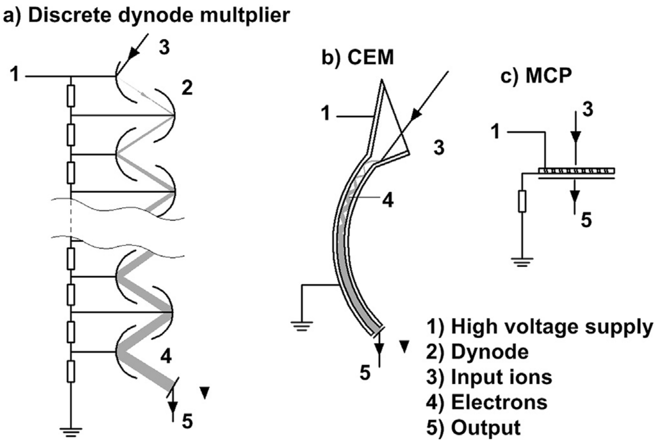


Fig. 6. Electron multipliers.

2.5.3. *Pulse counting*

Usually, the current from the electron multiplier is monitored by an electrometer amplifier. An alternative approach is to count the individual pulses. each of which is generated by one incident ion.

A single ion produces perhaps 10^7 electrons which constitute a readily detectable pulse. Because the response time of the multiplier can be made quite fast, count rates of up to 10^8 ions s^{-1} can be monitored. The detection limit is set by the background count rate, which is typically less than 1 event s^{-1}, generally representing an order of magnitude improvement over analogue techniques.

2.5.4. *Disadvantages of electron multipliers*

The major disadvantage of an electron multiplier is that the gain deteriorates with time, so it is necessary to check it fairly often. This is most conveniently done with a ‘dual detector‘, in which a Faraday cup is incorporated. The absolute ion current is measured with this and then the applied potential is adjusted to give a convenient gain. In a modern RGA this is done automatically under computer control.

2.5.5. *Choice of detector*

There is no rigid criterion which can be applied to determine which type of detector is best suited to a particular application. Obviously, cost is a significant factor. The following guidelines are suggested. For applications at pressure greater than 10^{-6} mbar a Faraday detector is probably adequate. At pressures in the range 10^{-6} to 10^{-9} mbar, a multiplier is useful to speed up measurements and to detect trace components. At lower pressures, a multiplier is highly desirable. Pulse counting may be considered for special applications.

3. Outgassing and EID

It is important, especially at ultra-high vacuum, that the RGA should give information about the vacuum composition without itself affecting the composition significantly. Therefore outgassing from the analyser must be minimal. At pressures greater than 10^{-6} mbar, this is easily achieved if dead volumes are eliminated and materials are used which are compatible with the vacuum environment. After installing the analyser and evacuating the system it is good practice to bake the vacuum chamber (that is, raise the temperature to a value above the intended operating temperature). This increases the rate of outgassing while the temperature is raised, but on returning to the operating temperature the outgassing rate will be greatly reduced. A modest increase in temperature, say to 80C can be beneficial, but sometimes a much higher temperature is used. The maximum permissible bake-out temperature may be as high as 200C, but it will depend on the materials used in the analyser and perhaps elsewhere in the vacuum system. For the lowest outgassing rates it is advisable to bake source components in a vacuum stove prior to assembly. This accelerates outgassing of gases (notably hydrogen) embedded in the material during manufacture.

A facility for degassing the ion source is often provided on a commercial RGA. This allows the ion source to be raised to a temperature significantly higher than its normal operating temperature, which accelerates the outgassing process.

For operation at pressures lower than 10^{-9} mbar, the source must be designed so as to minimize outgassing. It is often beneficial to degas the source whilst the vacuum system is cooling down after a bakeout to high temperature.

Even after a long bake, peaks are frequently observed, typically at 12, 16, 19 and 35 amu, which are formed by electron-induced desorption (EID) from surfaces within the source, rather than by electron-impact ionization of gaseous species. A peak at mass 19 is particularly common. This has been ascribed EID of fluorine, although the source of the fluorine is often not known, and H_3O^+ is often a more likely possibility. EID peaks can be distinguished from those generated by electron impact ionization because they have a different energy. Fig. 12 of reference [3] shows clearly the difference in spectra obtained when selecting different energy levels using a cylindrical mirror analyser.

4. Instrument control and data acquisition systems

Most modern mass spectrometers are controlled by a computer via an interface such as Ethernet or USB. There will probably be a facility to store and recall sets of tuning parameters, and to tune the instrument automatically.

There will generally be several ways of displaying and recording the data: for example as a scanned mass spectrum (so peak shape can be examined and optimised), as a bar chart (where the signal is recorded only at each nominal mass), or only at selected masses. The signal may be displayed on a linear or logarithmic scale. Auto ranging will often be provided – and maybe even automated switching between Faraday and electron multiplier detectors. The signal may be displayed as a current (amperes) or as a count rate (nominal ions per second), or sometimes more conveniently in units of pressure or as composition (for example percent, ppm or ppb).

Sometimes a mass-spectral library will be included, and perhaps software intended to identify and quantify the main gas species present. But usually the main components can be identified directly from the mass spectrum, and attempts to identify minor components by automated techniques will often prove misleading, due to systematic errors that are hard to quantify. As with any analytical instrument, in-situ calibration can be very helpful, but due to the nature of vacuum systems and processes, it will often not be a practical or economic proposition.

It is often possible to control multiple instruments from one computer, which could be located at some distance from the equipment. Control via the internet is also a possibility.

All this gives great versatility, but it the user should recognise that the performance of the instrument is still dependent on sensible operating parameters. In particular, if the tuning has been set primarily to give the maximum sensitivity, or the smoothest peak shape, it is probable that the analytical performance (particularly linearity) may be compromised.

5. Application of the QMS to vacuum RGA

There are many reasons for wishing to know what gases are present in a vacuum system. Perhaps the commonest application is to investigate why a system fails to achieve its design pressure.

If mass 28 (nitrogen) and mass 32 (oxygen) are present in approximately the ratio 5:1, a vacuum leak is indicated. The quadrupole can be used to locate an external leak by tuning to mass 4 and playing a fine jet of helium over the surface of the vacuum chamber. Sometimes the signal at mass 32 is low, since oxygen is absorbed by a stainless steel surface which has been baked recently. It may help to check the signal at mass 40, due to atmospheric argon.

If there is no leak, the pressure will probably be limited by outgassing of water. Peaks at masses 17 and 18 are then seen approximately in the ratio 3:1. A high-temperature bake will usually cure the problem. Another possible problem is the presence of hydrocarbons, perhaps from an inadequately baffled diffusion pump or an untrapped rotary pump; or due to outgassing from elastomer seals.

In complex vacuum systems, such as particle accelerators, and in many industrial processes, there are particular requirements which can be readily met by the appropriate quadrupole RGA. It is often valuable to monitor the system semi-automatically using a computer-controlled installation and possibly using multiplexed analyser heads. Stored data can be examined and correlated with the performance of the system or process. Evidently, special problems can arise with particular systems for which no rigid guidelines can be given. Further details may be found in the literature [4,21].

6. Conclusions

The principles of the QMS have been established for a long time, and the basic design of modern instruments is broadly similar to those from 30 years ago. Perhaps the most striking recent innovation has been the introduction of physically smaller devices, which can be used at higher pressure than traditional designs.

The near-universal use of computer control and data acquisition gives a very versatile instrument, easily configured to suit a wide range of applications. However, even a small mass spectrometer is a complex device, and unexpected behaviour is not uncommon. The user is strongly advised to study the principles of mass spectrometry and of the mass filter before utilising the QMS – especially if quantitative results are expected.

References

[1] Dempster AJ. Phys Rev 1918;11:316–25.
[2] Hofmann CA. Vacuum 1974;24:65–72.
[3] Batey JH. Vacuum 1987;37:659–68.
[4] Schubert RS. Partial pressure analyzers, analysis, and applications. AVS Press; 2000.
[5] Paul W, Steinwedel H. German patent 944000. 28th June 1956.
[6] Paul W, Raether M. Z Phys 1955;140:262–73.
[7] Finlan MF, Sunderland RF, Todd JFJ. Nucl Instruments Methods 1982;195: 447–56.
[8] von Zahn U. Z Physik 1962;168:129–42.
[9] Holkeboer DH, Karandy TL, Currier FC, Frees LC, Ellefson RE. J Vac Sci Technol A 1998;16:1157–62.
[10] Batey JH. Vacuum 1993;44:639–42.
[11] Batey JH. Int J Mass Spectrom Ion Process 1975;60:117–26.
[12] Herbert CG, Johnstone RAW. Mass spectrometry basics. CRC Press; 2003. p. 13–6.
[13] Nier AO. Rev Sci Instrum 1940;11:212–6.
[14] Brink GO. Rev Sci Instrum 1966;37:857–60.
[15] Cowen MC, Allison W, Batey JH. Meas Sci Technol 1993;4:72–8.
[16] Dawson PH. Quadrupole mass spectrometry. Elsevier; 1976.
[17] Batey JH. Nelms SM, editor. ICP mass spectrometry handbook. Blackwell Publishing Ltd.; 2005. p. 26–40.
[18] Hunter K, Stresau R. Nelms SM, editor. ICP mass spectrometry handbook. Blackwell Publishing Ltd.; 2005. p. 117–46.
[19] Kurz E. Am Lab March 1979;1979:67–74.
[20] Timothy JG. Rev Sci Instrum 1981;52:1131–42.
[21] O'Hanlon J. A user's guide to vacuum technology. Wiley Interscience; 2003.