SPECIAL FEATURE:
TUTORIAL

Principles and Instrumentation in Time-of-flight Mass Spectrometry

Physical and Instrumental Concepts

Michael Guilhaus
School of Chemistry, The University of New South Wales, Sydney 2052, Australia

The principles of time-of-flight mass spectrometry (TOFMS) are described with a view to understanding the strengths and weaknesses of this method of mass analysis in the context of current applications of mass spectrometry and the more familiar scanning instruments. Fundamental and instrumental factors affecting resolving power, sensitivity and speed are examined. Methods of gating ion populations and the special requirements in the detection and digitisation of the signals in the TOFMS experiment are discussed.

INTRODUCTION

Mass spectrometry is primarily concerned with measuring the mass-to-charge ratio ($m/z$) and abundance of ions moving at high speed in a vacuum. Nearly a century of research in mass spectrometry has established that it is one of the most powerful probes into the structure and composition of matter. Since J. J. Thomson's invention of the magnetic deflection mass spectrometer in the first decade of this century, many descriptions of devices for determining the relative yields of ions as a function of $m/z$ have been published. Currently there is increasing interest in mass spectrometry and this results from the advent of new ionisation methods and the wider use of compact mass spectrometers. New ionisation methods allow the characterisation of a diverse range of polar, ionic or high molecular-weight compounds which previously were not amenable to mass spectrometric analysis. Compact mass spectrometers (bench-top instruments) are increasingly used as detectors for chromatographic separations because of their capacity to identify and quantitate compounds in complex mixtures simultaneously.

TOFMS, an early arrival in the mass spectrometry family, was prominent in the field during the 1960s but was soon displaced by magnetic and quadrupole instruments with their higher sensitivity and mass resolving power. The most significant reason for the failure of TOF to mature was the lack of technologies to facilitate the recording and processing of the mass spectrum in the microsecond time-frame. These facilitating technologies are now emerging along with methods for the ionisation of massive biological molecules and also for fast mixture separation. TOF is highly advantageous in these cases and is therefore reappearing as a prominent mass analyser.

By definition, mass spectrometers are $m/z$ dispersive. In this treatise it will be useful to view ion trajectories in two or more dimensions with at least one being time. TOF mass spectrometers differ fundamentally from scanning instruments in that they involve temporally discrete ion formation and mass dispersion is principally in the time domain rather than along a spatial axis. Whereas the ion optics of spatially mass dispersive instruments can readily be illustrated in a two-dimensional diagram, essentially showing ion trajectories in a plane of the instrument, most TOF optics can be equally well illustrated with a diagram having one spatial dimension and one time dimension. The rays in these diagrams are space-time trajectories and the focal points show ions that are at the same place on the distance axis at the same time. It is more likely than not that ions at this focus are in fact quite defocused on the remaining spatial axes. This is illustrated in Fig. 1 which shows the principal plane of the space-time trajectory as well as one additional spatial dimension. A useful maxim in TOFMS is that 'it matters just as much when the ions are as where they are'.

THE TOFMS EXPERIMENT

The essential principle of TOFMS (Fig. 2) is that a population of ions moving in the same direction and having a distribution of masses but a (more-or-less) constant kinetic energy, will have a corresponding distribution of velocities in which velocity is inversely proportional to the square root of $m/z$. Consider a situation in which ions, under the influence of an external electric field, begin their acceleration from rest at the same time and from the same spatial plane normal to
the acceleration vector. Their arrival times at a target plane (parallel to the plane of origin) will be distributed according to the square root of \( m/z \). In the case of sub-relativistic velocities (as occurs in the keV energy regime) a good description of the TOFMS experiment requires no more than Newtonian physics as shown by the equations that follow. To avoid ambiguity, these equations are developed generally, rather than specifically to mass spectrometry. All fundamental quantities are in SI units as listed in Table 1. It may useful for the reader to note that the constant for converting kg to Da is easily derived from Avogadro's Number \( (N_A) \):

\[
10^{-3}/[N_A] = 1.66 \times 10^{-27} \text{ kg Da}^{-1}
\]

and that the number of electronic charges is obtained by dividing the charge, \( q \), by the charge of the electron \((-e)\). Thus \( q \) appears with \( m \) in many equations rather than \( z \). For positive ions with \( z \) charges, \(-ze\) may be substituted for \( q \).

**Force and acceleration:**

\[
F = Eq
\]

\[
F = ma
\]

\[
a = Eq/m
\]

**Velocity and time to reach it:**

\[
a = \frac{du}{dt}
\]

\[
u = \int Eq/m \, dt
\]

**Position:**

\[
u = u_0 + (Eq/m)t
\]

\[
t_s = \frac{u - u_0}{E} \left( \frac{m}{q} \right)
\]

**Drift Velocity and Accelerating Voltage:**

\[
u = \frac{q}{m} \sqrt{2Es_a/m}
\]

\[
u D = \frac{qE}{m}
\]

**Drift time:**

\[
t_D = \frac{D}{u}
\]

\[
t_D = \frac{D}{\sqrt{2qEs_a/m}}
\]

\[
t_D = \frac{D}{\sqrt{2qV/m}}
\]

**Observed TOF:**

\[
\text{TOF} = t_0 + t_s + t_D + t_a.
\]
Table 1. Fundamental and derived quantities, units and conversion

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
<th>Conversions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>m</td>
<td>mass in kg</td>
<td>kg</td>
<td>1.66 × 10^{-27} kg Da^{-1}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mass in Daltons</td>
<td>Da</td>
<td></td>
</tr>
<tr>
<td>Charge</td>
<td>q</td>
<td>charge in coulombs</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>charge of electron in coulombs</td>
<td>C</td>
<td>-1.60 × 10^{-19} C</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>number of electron charges</td>
<td>C</td>
<td>-q/e</td>
</tr>
<tr>
<td>Time</td>
<td>t</td>
<td>time-of-flight in seconds</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>t_0</td>
<td>t_0</td>
<td>time after t = 0 that ion begins to accelerate</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>t_a</td>
<td>t_a</td>
<td>time that ion is accelerating from u_0 to u_D</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>t_0</td>
<td>t_0</td>
<td>time that ion drifts at constant velocity</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>t_minus</td>
<td>t_minus</td>
<td>turn-around time (2 x time of decelerate to u = 0)</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>t_s</td>
<td>t_s</td>
<td>response time of detection system</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>σ_s^2</td>
<td>σ_s^2</td>
<td>variance of normal distribution of t_s</td>
<td>s^2</td>
<td></td>
</tr>
<tr>
<td>σ_t^2</td>
<td>σ_t^2</td>
<td>variance of time distribution attributable to jitter</td>
<td>s^2</td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>s</td>
<td>distance in metres</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>s_0</td>
<td>s_0</td>
<td>initial position of ion</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>s_a</td>
<td>s_a</td>
<td>distance in E from average s_a to drift region</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>D_d</td>
<td>D_d</td>
<td>drift distance</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>σ_s^2</td>
<td>σ_s^2</td>
<td>variance of normal distribution of s_a</td>
<td>m^2</td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>U</td>
<td>translational energy of ion in joules</td>
<td>J</td>
<td></td>
</tr>
<tr>
<td>U_0</td>
<td>U_0</td>
<td>initial translational energy of ion (at t_0, s_a)</td>
<td>J</td>
<td></td>
</tr>
<tr>
<td>U_d</td>
<td>U_d</td>
<td>drift energy of ion</td>
<td>J</td>
<td></td>
</tr>
<tr>
<td>U*</td>
<td>U*</td>
<td>translational energy of fragment ion formed by decomposition of parent ion having energy U_0</td>
<td>J</td>
<td></td>
</tr>
<tr>
<td>Velocity</td>
<td>u</td>
<td>velocity in metres per second</td>
<td>ms^-1</td>
<td></td>
</tr>
<tr>
<td>u_0</td>
<td>u_0</td>
<td>initial velocity</td>
<td>ms^-1</td>
<td></td>
</tr>
<tr>
<td>u_d</td>
<td>u_d</td>
<td>drift velocity</td>
<td>ms^-1</td>
<td></td>
</tr>
<tr>
<td>σ_u^2</td>
<td>σ_u^2</td>
<td>variance of normal distribution of initial velocity</td>
<td>m^2 s^-2</td>
<td></td>
</tr>
<tr>
<td>Electric field</td>
<td>E</td>
<td>electric-field strength</td>
<td>V m^-1</td>
<td></td>
</tr>
</tbody>
</table>

**MASS RESOLVING POWER**

In most experimental arrangements ions are brought to keV translational energies over a distance of a few millimeters and the time that the ions drift in a field-free region of about 1 m is much larger than the time of acceleration. Equation (15) therefore serves as a useful approximation to determine the approximate flight time of an ion (in the 100 μs time frame for typical conditions). In mass spectrometry it is conventional to measure resolving power by the ratio of \( m/Δm \) where \( Δm \) is a discernable mass difference. In TOFMS it is convenient to work in the time domain. Thus the resolving power \( m/Δm \) can be measured in terms of \( t/Δt \) as follows:

\[
\frac{m}{Δm} = \frac{t}{2Δt}
\]

(17)

The finite time interval, \( Δt \), is usually the full-width at half-maximum (height) of the peak (FWHM). This definition of resolving power gives values which are approximately double those defined by the 10% valley definition which is perhaps more familiar to users of sector mass spectrometers.

Mass resolving power is limited by small differences in measured flight-times for ions of the same mass (typically in the 100 ns time frame). These have their origins in the distributions in initial energy (i.e., molecular speed), position and time of formation of the ions prior to acceleration. Non-ideal acceleration fields and, in some important applications of TOF, collisions, impart additional energy spread. The response time of the detection system as well as instrumental uncertainties in the timing/duration of the gating and detection events contribute additional temporal spread. In most cases these are uncorrelated distributions. Each maps to a contributing arrival-time distribution which can be calculated but not observed separately. The sum of the effects can be obtained by convolution of the individual uncorrelated arrival-time distributions and this can be compared with the observed peak-shape. An understanding of the dependence of the width of the resultant temporal-distributions on key experimental factors such as ion acceleration voltage and drift length allows the best choice of instrumental parameters. In order to realise useful resolving powers for most applications, however, ion optical methods must be used to control and/or compensate for the initial distributions.

**Initial Energy (velocity) of Ions**

**Ions Formed in the Gas-Phase.** Ions generated or sampled from the gas-phase are subject to a Boltzmann distribution of initial velocity \( (u_0) \). The ions initially
moving towards the detector arrive there before the ions which are initially moving away from it. Indeed, the latter ions are first decelerated to zero velocity before being re-accelerated and passing through their original position. The time for this to happen is often referred to as 'the turn-around time' \( t_{-+} \). Assuming ions are all formed at the same time and at the same distance from the detector and setting \( u_0 = -u \) in Eqn (7) gives:

\[
 t_{-+} = \frac{2|u_0|m}{Eq} \tag{18}
\]

Substituting \( u_0 = \sqrt{2U_0/m} \) allows the equality to be expressed in terms of the initial translational energy \( U_0 \):

\[
 t_{-+} = \frac{2\sqrt{2mU_0}}{Eq} \tag{19}
\]

Two ions with the same speed but moving in opposite directions will reach the same final velocity after acceleration. They remain separated by the turn-around time until the detector is reached. This is depicted by the parallel rays in the drift region of Fig. 3. If the drift region is lengthened, this separation in time becomes smaller relative to the total flight time. As is also apparent from Eqn (19), the turn-around time can be decreased by increasing the strength of the accelerating field. As an example consider the turn-around time for an \( m/z \) 100 ion in a source at about 500 K where the mean initial energy of ions is about \( 6.4 \times 10^{-21} \) J (40 meV). In a weak accelerating field of \( 3 \times 10^{-4} \) Vm\(^{-1} \) Eqn (19) calculates \( t_{-+} = 19 \) ns. The resulting peak broadening would significantly limit resolving power. Increasing \( E \) by a factor of 10 reduces the turn-around time to 1.9 ns, which is about the same temporal contribution as made by a good detection system.

The time for an ion to travel a distance \( s \) from initial position \( s_0 \) can be obtained from the quadratic Eqn (9). Obtaining the roots of this equation and, once again using \( \sqrt{2U_0/m} \) for \( u_0 \) gives:

\[
 t = \frac{-\sqrt{2mU_0}}{Eq} \pm \frac{\sqrt{2m[U_0 + qEs]}}{Eq} \tag{20}
\]

The physical meaning of Eqn (20) is evident in Fig. 3 which shows the space-time trajectory of an ion passing through \( t = t_0 \) with an initial velocity \( u_0 \). Along this trajectory the ion is accelerating at the distance \( s = s_i \) (after \( t_0 \)) and could be imagined to have been travelling in the opposite direction (decelerating) at \( s_i \), before \( t_0 \). The trajectory of the ion before \( t_0 \) is in most cases only imaginary but, usefully, when reflected about \( t_0 \), it gives the trajectory of a corresponding ion at the original position \( (t_0, s_0) \) with an initial velocity of \( -u_0 \). This gives the equation for the time \( t_0 \) to reach the drift velocity:

\[
 t_0 = \frac{\sqrt{2m(U_0 + qEs)}}{Eq} \pm \frac{\sqrt{2mU_0}}{Eq} \tag{21}
\]

Here the turn-around time is apparent in the second term. Setting the distance from \( s_0 \) to the beginning of the drift region as \( s_i \), the drift energy is \( (U_0 + qEs) \) from which the drift velocity is:

\[
 u_0 = \frac{2(U_0 + qEs)}{m} \tag{22}
\]

The drift time is thus:

\[
 t_D = \frac{D}{2} \sqrt{\frac{2m}{(U_0 + qEs)}} \tag{23}
\]

Adding Eqns (21) and (23) and taking out a common factor of \( \sqrt{2m} \) gives the familiar equation published in 1955 by Wiley and McLaren\(^1\) and reproduced by many authors since then:

\[
 t = \frac{(2m)^{1/2}[(U_0 + qEs)^{1/2} + U_0^{1/2}]}{qE} + \frac{(2m)^{1/2}D}{2(U_0 + qEs)^{1/2}} \tag{24}
\]

As stated above, the effect of the turn-around time can be decreased by increasing \( D \). However the spread of arrival times due to differences in the magnitude of initial velocity will increase with drift length. Long drift-lengths also introduce technical difficulties such as the need to increase the size of the detector and the vacuum.

![Figure 3. Parabolic space-time trajectories showing turn-around effect due to ions having initial velocity both away from and towards the drift region.](image)
system. Generally the approach taken to reduce velocity effects in linear TOFMS is to increase $E$.

**Ions Formed from Surface.** Simpler conditions usually prevail for acceleration of ions from a surface. All the ions in the population travel the same distance in the acceleration field from an equipotential surface to an equipotential grid. If $V$ is the difference in the potentials the drift energy is:

$$U_0 = U_0 + qV$$  \hspace{1cm} (25)

and

$$t_0 = \frac{D}{2} \sqrt{\frac{2m}{(U_0 + qV)}}$$  \hspace{1cm} (26)

In most configurations $D \gg s$, $t_0 \gg t_i$ and the ion arrival time spread from $U_0$ is mostly accounted for by the spread in drift time (i.e., $\Delta t_0 \gg \Delta t_i$). Thus, assuming $t \approx t_0$, differentiating with respect to $U_0$ and multiplying by $dU_0$ gives:

$$\frac{dt}{dU_0} = \frac{1}{2} \frac{m}{(U_0 + qV)^{1/2}}$$  \hspace{1cm} (27)

Dividing by:

$$t \approx t_0 = \frac{D}{2} \sqrt{\frac{2m}{(U_0 + qV)}}$$  \hspace{1cm} (28)

and multiplying by 2 gives:

$$\frac{2dt}{t} = \frac{dm}{m} = \frac{dU_0}{U_0 + qV}$$  \hspace{1cm} (29)

or in terms of resolving power:

$$\frac{m}{\Delta m} = \frac{U_0 + qV}{\Delta U_0}$$  \hspace{1cm} (30)

Normally $qV \gg \Delta U_0$ which leads to the simple result that

$$\frac{m}{\Delta m} \approx \frac{qV}{\Delta U_0}.$$  \hspace{1cm} (31)

Thus, when the drift region is much longer than the acceleration region, the resolving power does not depend on the length of the drift region. In practice the time resolution of the detector and digitiser combination places a lower limit on the length of the flight tube in the range 0.5–1.0 m with $V$ (typically 3–30 kV) adjusted to give a $t_0$ of about 100–200 µs for the upper limit of $m/z$. Note that these approximate parameters place the upper limit of $m/z$ at 60 000.

**Energy Focusing.** Wiley and McLaren introduced a technique called time-lag focusing to overcome the energy problem. By confining ionisation to a small distance and allowing a delay between ionisation and ion extraction, the velocity dispersion causes the ion packet to become spread-out. The resulting large spatial distribution (correlated to the initial velocity distribution) is refocused by a spatial focusing technique as described in the next section and shown in Fig. 4. Unfortunately this correction is mass dependent and enhanced resolving power is attainable only over a limited mass range.

More recently, the most common solution to the energy problem has been to make the more energetic ions follow a longer trajectory. This is achieved with some form of retarding field. An early implementation of this approach was to use an electric sector. As shown in Fig. 5(a), the more energetic ions have a larger radius of curvature in this device and therefore take longer to travel through it. The optics are adjusted to minimise the $dt/dU_0$ function. More recent versions of this approach have used a number of electric sectors in a clover-leaf configuration [Fig. 5(b)].

The most successful energy focusing method to date has been the 'reflectron'. Essentially an electrostatic ion mirror, this device creates one or more retarding fields after a drift region. These are orientated to oppose the acceleration field. Ions re-emerge from the device with their velocities reversed. More energetic ions penetrate more deeply and hence take longer to be reflected. Thus the optics can be adjusted to bring ions of different energies to a space-time focus as shown in Fig. 6. Usually the angle of ion entry into the mirror is adjusted slightly away from 90° so that the ions follow a different path after being turned around. This allows the detector to positioned where it will not interfere with the axis of the ions from the ion source. Alternatively an annular detector allows the accelerator and mirror to be coaxial.

The mirror has an added advantage in that it increases the drift-length without increasing the size of the instrument. The problem of the turn-around of the ions in the source, being effectively temporal, cannot be corrected by the ion mirror. Thus, strong extraction field is often used with a gaseous ion source. In this case the ions are brought to a sharp spatial focus (see below) located at a short distance along the drift region. Different masses will focus there at different times but, at the respective times, the ion packets will be sharply focused in the direction of the drift velocity. There will be a large energy spread resulting from the energy and spatial spreads in the source but as the temporal and spatial spreads are very small for each $m/z$, the ion mirror can substantially correct for the large velocity spread. The mirror is positioned so that the first spatial focus acts as a pseudo ion source. This is illustrated with the space-time trajectories of Fig. 6.

Due to the lack of the ion turn-around problem, the ion mirror works very well with TOFMS sources which generate ions from surfaces.

**Initial Position of Ions and Spatial Focusing**

An initial spatial distribution of ions maps to an arrival-time distribution due to two opposing factors: (i) ions
initially more distant from the detector spend more time in the accelerator, and (ii) ions initially more distant from the detector have a shorter drift-time because, as they spend more time under the influence of $E$, they reach a higher drift velocity. Setting $U_0 = 0$ in Eqn (24) and differentiating with respect to $s_x$ gives:

$$\frac{dt}{ds_x}(U_0 = 0) = \frac{m}{2qE_s} \times \left(1 - \frac{D}{2s_x}\right).$$

This function reaches a minimum (zero) when $D = 2s_x$. Thus a spatial focus is achieved at a plane located at a distance of $2s_x$ along the drift region. Figure 7 shows the spatial focus principle on a space-time diagram. In a uniform accelerating field ion trajectories are simply parabolas starting at their minima which are coincident
on the time axis but displaced in distance. Tangents to these parabolas intersect at a distance of $2s_0$ in the drift region. This is not a perfect focus as a finite spatial spread, $\Delta s_0$, gives a finite time spread $\Delta t$. However, the spatial focus principle greatly reduces the effect of spatial spreads. For a linear TOFMS with one acceleration field the condition demands a short drift region. If the acceleration region is 10 mm and the drift region is 20 mm the mass spectrum is now in the $5\,\text{ps}$ time window rather than the $100\,\text{ps}$. A spectrum is usually adequately described with 16 kpoints, thus the digitiser resolving power needs to be improved from about 5 ns to about 0.3 ns. Moreover, the maximum repetition rate increases from 10-200 kHz. Signal processing at this rate is at present technologically difficult or impractical.

Acceleration with two or more consecutive electric fields allows the spatial focus plane to be moved to longer distances. In a two stage acceleration system the first field needs to be weaker than the second. The use of a weak acceleration field has adverse consequences when an initial energy spread exists. This is the case in particular with gaseous ion sources. This conflict makes simultaneous energy and spatial focusing difficult without the use of some additional ion-optical devices such as the ion mirror (above) and orthogonal acceleration (below).

### Non-ideal Fields

Ion acceleration regions are most often defined in TOF instruments by fine conducting meshes stretched over support rings. When a potential difference is applied across a pair of these meshes (also referred to as grids) a close to homogenous electric field is created between them. At the edges the homogeneity is lost due to fringe fields. This is overcome by using the central part (perhaps one or two thirds of the diameter) of the meshes. Notwithstanding this, the influence of the fringe field can not be eliminated entirely. Fringe fields cause small deflections in ion trajectories and contribute to a spread of drift velocity. This gives rise to a corresponding arrival-time spread. This is usually small relative to other sources of ion arrival-time spread.

Often overlooked is the lensing effect of the mesh itself. Each hole in the mesh causes local inhomogeneity in the electric fields on either side. Figure 8 shows this effect as a potential function calculated near fine closely spaced wires separating two different electric fields. The 'field penetration' of high to low field regions and vice versa is clearly evident. Ions passing through these regions experience some deflection. The effect is greatly amplified when the preferred trajectory of the ions is not normal to the meshes (see orthogonal acceleration section below). The energy spread from this source can be compensated at least partially by energy focusing devices such as the ion mirror.

### Collisions and Gas-Dynamic Acceleration

When TOFMS is used with desorption techniques such as matrix assisted laser desorption ionisation (MALDI), resolving power decreases significantly as the mass increases. Referring back to Eqn (24) it is clear that TOFMS provides a very simple basis for mass calibration: choosing the mass dependent variable $\sqrt{m/q}$, and assuming that $qE_s \gg U_0$ (i.e., $qE_s \approx U_0 + qE_s$) peaks will occur at positions according to a linear equation: $t = A\sqrt{(m/q)}$ where $A$ is a constant containing the instrument parameters. For very high mass ions in
MALDI the linear $\sqrt{(m/q)}$ calibration breaks down and ions are detected at times earlier than predicted. Both of these factors have been attributed to collisions, in the ion source, between analyte molecules and the large excess of low molecular weight matrix molecules. The desorption process locally creates a rapid expansion of matrix into the vacuum. Collisions between the faster moving and abundant matrix molecules and the slower moving analyte molecules result in a gas-dynamic acceleration of the latter to produce an energy component which increases with mass. It follows that a corresponding component of energy spread also increases with mass. The disparity between relative mass of analyte and matrix can be $10^3$ for large bio polymers and therefore a small component of gas dynamic acceleration can lead to significant resolution loss at high mass.

**Metastable Decomposition**

In magnetic sector mass spectrometers, fragment ions from unimolecular and collision-induced decompositions in field-free regions are brought to focus by subsequent sectors at a different mass/energy than the unfragmented parent. This results from the partitioning of momentum/kinetic energy between the products of the reaction. In TOFMS both fragments continue with (nearly) the same velocity and so parent, fragment ion and neutral are detected at the same place in the spectrum. It is well known that some internal energy is converted into translational energy when an ion fragments. In sector mass spectrometers, this energy release gives rise to the characteristic broad peaks resulting from metastable ion decompositions in field-free regions. In the context of TOFMS the effect of the energy release and hence the increased range of velocities of the fragments relative to the parent must be considered as a factor which decreases resolving power. The centre of mass of the products will have a space-time trajectory which is a continuation of the parent’s trajectory if it did not fragment. The trajectories of the fragments, however are spread isotropically in the centre of mass coordinate system and are therefore dispersed in time. The magnitude of this effect can be gauged by calculating the energy release required to give an arrival time spread equal to that due to $\Delta U_0$ in Eqn (31) for a TOF instrument with $m/\Delta m$ (FWHM) = 2000 and a drift energy $U_0 = qV = 3000$ eV. In this calculation the worst case is assumed: where the decomposition occurs at the beginning of the drift region and the reaction axis is aligned with the velocity vector of the parent. To maximise the effect the fragments are of equal mass (i.e., half the mass of the parent). The fragments will have half the energy of the parent and be subject to an energy release of $\Delta U^*$:

$$ t = \frac{qV}{2\Delta t}$$

$$ \Delta U^* = \frac{qV\Delta t}{t}.$$  

Since $qV = 3000$ eV and we set $t/2\Delta t = 2000$, a value of 750 meV is obtained. This is a large energy release which occurs seldom in practice. Moreover, if the reaction axis is not aligned with the drift velocity or if the reaction occurs closer to the detector, an even greater energy release is required to give the $\Delta$ value of this calculation. From this it can be concluded that metastable decompositions in the drift region will not lead to significant loss of resolving power in most cases. The same conclusion is not reached if the fragments of metastable ions experience any further acceleration or deceleration. Because the fragment ion has lower mass and energy than the parent ion, it will take less time to pass through an acceleration region. The velocity of the neutral fragment will of course be unchanged. Thus, as was done on early commercial TOF instruments, a retarding grid near the detector causes peaks for which there are associated metastable processes, to split into three components: neutral, charged fragment and parent. Of more relevance to modern TOF instruments which are used to detect ions with very high $m/z$, post acceleration devices, used to improve the detector efficiency, can cause peak broadening due to metastable fragmentations. Ion mirrors will obviously not reflect neutrals and the lower energy fragments of metastable ions will be defocused and brought to the detector earlier than their parents.

The time between ionisation and metastable fragmentation correlates approximately with the number of degrees of freedom of the parent ion. As molecular size increases so do the number of degrees of freedom. Thus some ionisation methods which produce heavy ions do so without producing many observable fragment ions when TOF is used. This is because the prompt extraction of ions (in techniques such as MALDI) allows less time for ions to fragment in the source. Applications facilitated by high mass TOFMS are often more concerned with structural information contained in fragments of parent ions than maximising the signal for the parent. The long time spent in the drift region usually means that metastable decompositions will eventually take place there. Post source decay (PSD) is a new name for metastable decomposition and these processes can be observed by tuning the ion mirror to direct the lower energy PSD fragments to a detector. This has been shown to greatly increase the amount of structural information in TOFMS spectra of large molecules.

Recently, attempts to perform MS/MS by TOFMS have involved mass selection and collision-induced decomposition of ions in a drift region. While these methods suffer from poor resolving power with respect to mass preselection it is apparent that the potential to exploit PSD processes and obtain structural information for very large ions is an important new aspect of TOFMS.

**Detector Response and Jitter in Electron Optics and Timing Signals**

The temporal response of the detector to the ion arrival event can be an important contributor to peak width. Sensitivity requirements (see Sensitivity section below) demand the use of electron multipliers and microchannel plates (MCPs). Ideally, packets of isobaric ions approach the detector sharply focused in time and therefore, spatially, with a planar distribution. The
detector’s target surface should likewise be planar and parallel to the arriving ion packet. The electromagnetic environment created by the detector must not perturb the motion of the approaching ions to increase the spread in impedance as these cause the signal to ‘ring’, i.e., to reflect back and forth in the transmission line (this is analogous to internal reflection of light at discontinuities in refractive index). The amplifier circuitry must have an extremely fast time-constant in order to respond to the short charge-pulses without significantly broadening them. Equally important in the process of measuring ion flight times is the precise detection of a start event, which usually marks $t = 0$, and a close correlation between this and the associated ion process. This usually involves sensing when a control pulse reaches a certain voltage level. In practice none of these criteria is met perfectly. A pulse detected for a single ion is typically 1–10 ns wide (Fig. 9). This arises largely because of the statistical variation of electron transit times between successive surface interactions in the detector. Such pulse widths are fundamentally limited by space-charge forces: as the electron optics are improved to create a narrow electron arrival time at the anode of the detector, the instantaneous charge density rises creating a force which opposes further narrowing of the pulse. Non-ideality in the other factors described above also contribute. For example, with discrete dynode detectors, there may be ‘jitter’ associated with where the ion impacts on the first dynode because the secondary electrons will then have significantly different distances, hence different transit times, to reach the next dynode. Thus individual pulses may be seen to have less than 5 ns widths, but for multiple ions arriving at precisely the same time, the detected pulses are observed to be centred at different times creating another distribution with a characteristic width. Such distributions are often close to Gaussian, in which case the variances are additive. For example, a jitter distribution characterised by $\sigma_j = 6.0$ ns will add to a single-ion pulse-width with $\sigma_p$ of 4.0 ns giving a net observed detector signal, $\sigma_{det}$ of 7.2 ns:

$$\sigma_{det} = \sqrt{\sigma_j^2 + \sigma_p^2}$$

(35)

Such detector temporal spreads are analogous to those produced by the detector slit in a scanning mass spectrometer. If the temporal spread from the detector is mainly from the electron optics and thus independent of the mass of the ions, the same detector time distribution (plus the temporal effects of the other factors described above) convolutes with the ion arrival-time spread. As these distributions are uncorrelated the variances add. This means that the detector has a greater broadening effect at the low mass end of the spectrum where arrival-time distribution, $\sigma_s$, is small and becomes of similar magnitude to $\sigma_{det}$. This is illustrated in Fig. 9 which shows a method for separating the arrival-time spread and the detector/jitter effects by measurement of mass resolving power over a wide mass range.

**Overall Peak Shape**

From the discussion above it is apparent that uncorrelated distributions in the initial position and the velocity, together with temporal spreads from, for example:

(i) ionisation;
(ii) gating;
(iii) jitter and integration detector response times;

combine to create the observed peak shape at a particular $m/z$ in TOFMS. If these contributing distributions are characterised by their respective variances $(\sigma_j^2, \sigma_u^2, \sigma_v^2)$ then the variance of a peak in the TOF mass spectrum ($\sigma_T^2$) can be approximated by an error analysis equation:

$$\sigma_T^2 = \left(\frac{\partial T}{\partial \delta \phi}\right)^2 \sigma_j^2 + \left(\frac{\partial T}{\partial \delta u}\right)^2 \sigma_u^2 + \sigma_v^2$$

(35)

This equation is correct for infinitesimal spreads in $s$ and $u$ but not for typical finite ones. A more useful approach, based on probability theory, is to consider the TOFMS as a transform operator ($T$) containing the relationships for ion flight-times derived earlier in this article. Essentially, for Gaussian spreads in $s_0$, $u_0$ and $t_0$, the relative probability that an ion with a particular $s_0$ and $u_0$ will arrive at a time $\tau$ can be obtained from the product of three exponential terms (Gaussians) whose values diminish as the variables deviate from their respective averages. Numerical integration of the probability function over $s_0$ and $u_0$ gives the peak shape $f(\tau)$ in which $C$ is a scaling constant:

$$f(\tau) = C \int_{s_0} \int_{u_0} \exp \left( \frac{-\{\tau - T(s_0, u_0)\}^2}{2\sigma^2} \right) \times \exp \left( \frac{-(s-s_0)^2}{2\sigma_s^2} \right) \times \exp \left( \frac{-(u-u_0)^2}{2\sigma_u^2} \right) \, ds \, du$$

(36)

One advantage of this type of calculation is that it can reveal the contribution to the peak shape from individual dispersions. The transform operator can be augmented with the equations for extended TOFMS instrument geometries such as those incorporating ion mirror. Fig. 11 shows this 'deconvolution' for a TOF instrument with ions beginning with a small range of velocity but a large range position. This figure shows the interesting result that a Gaussian spatial spread transforms to an asymmetric distribution of flight-time. This is a consequence of spatial focusing where ions on either side of the mean initial position arrive earlier than ions from the mean position.

**SENSITIVITY AND SPEED**

**Comparison with Scanning Mass Analysers**

Unlike quadrupole, ion-trap and most sector mass analysers, TOFMS does not scan the spectrum. Most ions entering the drift region are detected, and even when taking into account a less than 100% analyser transmission, TOF has an intrinsic duty-cycle advantage which increases as does the observed dynamic range in mass (though one must be cautious and consider the efficiency of the ionisation and gating processes as discussed below). Consider a quadrupole mass spectrometer scanning from \( m/z \) 40-1040. Irrespective of the scan-rate, the average dwell-time on each peak in the spectrum must be less than 0.1% of the scan-time. This means that less than 0.1% of the ions generated from the sample are detected! The situation is worse at the high mass end of the spectrum where the inverse relationship between quadrupole transmission and resolving power reduces sample utilisation even further.

The time to scan a wide mass range can be a problem if rapid changes in sample composition are to be moni-
sored. This is the case, for example, in high speed chromatography, desorption and pyrolysis processes. There are fundamental limits to the scan times due to transit times of ions through sectors and quadrupoles, and the ability to control analyser fields becomes technologically difficult beyond about 5 spectra per second. TOFMS is typically far less limited in spectral acquisition rate and signal averaging allows a convenient method to obtain the best compromise between sensitivity and speed (see below).

**Sensitivity/Resolving Power Compromises**

Improving TOFMS focusing, as discussed above, brings a greater number of ions to the detector in a smaller unit of time. This serves to improve sensitivity by increasing the signal-to-noise ratio (S/N). Once optimal focusing is achieved, increasing the number of ions detected, for example by increasing the volume of space in which ions are generated/sampled, may result in more ions but will usually degrade resolving power. This is analogous to opening the source slit on a sector mass spectrometer. The useful concept of phase-space is sometimes used in TOFMS to describe the characteristics of the ion population sampled for the TOF measurement. As well as defining the spatial limits of ion formation it is possible to know or control the temporal and energetic limits of the sampled ions. The spatial axes taken together with energy and time axes define a multidimensional space and a bigger volume of this space contains more ions. Each axis of this space will map to the arrival-time distribution according to the focusing ability of the instrument, and as the volume increases so does the resultant arrival-time distribution. At the extreme, even though many more ions can be detected, the resulting loss of resolving power (i.e., selectivity) can lead to a lowering of the S/N because the chemical noise becomes a significant obstacle. So, as in other mass spectrometers, there is a compromise between sensitivity and resolving power.

The requirement to produce temporally well defined packet of ions in TOFMS brings different problems depending on which of two classes the ion source belongs: (i) those that intrinsically produce bursts of ions; or (ii) those that are continuous in nature.

**Ion Production and Gating with Pulsed Ion Sources**

Ions generated with pulsed energy sources such as pulsed lasers or nuclear fission are readily amenable to TOF analysis. Indeed, such sources are remarkably incompatible with scanning mass analysers because of the slow speed of the scan relative to the temporal size of the ion packets and increasingly because the m/z range extends well beyond that observable with magnetic and quadrupole instruments.

For TOFMS with pulsed sources it is important that the burst of energy is short (ps to ns) in duration. The absorption of the energy to create the ion can be a complex process (e.g., desorption from a surface accompanied by ion/molecule reactions as occurs in MALDI) and it is desirable that this does not add significant temporal width to the ion packet. It is important that the number of ions produced in this process is high but not so high as to produce expansion of the ion packet due to space-charge repulsion within the packet. Hence the amount of energy deposited (e.g., laser power and wavelength) are important factors. The number of ions detectable is crucial to sensitivity. Because of the statistical nature of ion detection, S/N increases with the square root of the number of ions collected. One of the strengths of TOFMS arising from the short time-frame of the measurement, is that signal averaging is easily implemented to improve S/N. Thus if, as in MALDI, one shot of the laser produces fewer ions than necessary to give an adequate S/N, the signal for additional shots can be added. The $\sqrt{N}$ law gives diminishing returns in the long run.

**Ion Production and Gating with Continuous Ion Sources**

Several ion sources of importance in analytical mass spectrometry produce ions continuously. Selecting packets of ions from a continuous stream is by no means straightforward. Electron ionisation is the most widely used continuous ion source and there are two main approaches to gating ions. Firstly, as in the original commercial TOF instruments, ionisation can be performed in the acceleration region of the TOF analyser. The application of extraction pulses in combination with pulsing of the electron beam produces a packet of ions. Wolniki has described a variation of this approach where the ions are accumulated in the source with the introduction of trapping fields. A drawback of these approaches in the context of current mass spectrometry hardware is that they are not readily adaptable to changing the ionisation method. Whereas the modular scanning instruments of today are expected to have interchangeable ion sources such as EI, CI, FAB, ESI etc., such flexibility is not attainable in a standard Wiley and McLaren arrangement. An alternative approach involving remote continuous ionisation followed by gating of the resulting ion beam has been a popular solution. Three general methods have been described.

(i) *In-line Gate.* Most work on the gating of ion beams for TOF has used a configuration in which the ion source, gate and detector are 'in-line'. This usually involves deflection of an energetic ion beam across an aperture. The main drawback here is that the ions, though formed continuously, are sampled only momentarily. As the temporal width of the packet is decreased so too is the number of ions selected relative to the number of ions produced in the ion source. Therefore the approach has an intrinsically strong inverse relationship between sensitivity and resolving power.

(ii) *Orthogonal Acceleration.* As shown in Fig. 12, the ion source, TOF accelerator and detector describe an angle slightly greater than 90°. A low energy collimated ion beam fills a 'wide' acceleration region (about 5–10% the length of the drift region) during its fill-up (field-free) mode. One or more electrodes in the accelerator are pulsed on to create an extraction field which is applied
to be strictly orthogonal to the continuous ion beam. The section of the beam so extracted passes through a grid to at least one further extraction region in which the ions are brought to a kinetic energy about 100 times greater than that of the original ion beam. As soon as the ions have left the first acceleration region the extraction electrodes are relaxed so that the ion beam refills the acceleration region. The component velocity in the original beam direction is conserved and the orthogonal velocity component for isobaric ions is about 10 times the acceleration region. The component velocity in the original beam direction is conserved and the orthogonal velocity component for isobaric ions is about 10 times greater (i.e., approximately \(\sqrt{\text{gain in energy}}\)). A detector parallel to the continuous ion beam is located at the end of the drift region. The drift-time of ions at the high-mass end of the spectrum is thus of the same magnitude as the time they take to traverse the width of the orthogonal acceleration region. Thus ions for the next cycle of extraction are 'accumulating' in the accelerator as the ions of the current cycle are drifting to the detector. This method of gating is potentially more efficient than the in-line approach.

The collimation of the ion beam creates an additional advantage that the ions have minimal velocity spread in the 'TOF direction'. This allows much higher resolving power (hence sensitivity—see above) to be achieved relative to the standard linear TOFMS of the Wiley and McLaren design. Resolving powers of 3000–5000 (FWHM) have been reported for orthogonal accelerations to at least one further extraction region in which the ions are brought to a kinetic energy about 100 times greater than that of the original ion beam. As soon as the ions have left the first acceleration region the extraction electrodes are relaxed so that the ion beam refills the acceleration region. The component velocity in the original beam direction is conserved and the orthogonal velocity component for isobaric ions is about 10 times greater (i.e., approximately \(\sqrt{\text{gain in energy}}\)). A detector parallel to the continuous ion beam is located at the end of the drift region. The drift-time of ions at the high-mass end of the spectrum is thus of the same magnitude as the time they take to traverse the width of the orthogonal acceleration region. Thus ions for the next cycle of extraction are 'accumulating' in the accelerator as the ions of the current cycle are drifting to the detector. This method of gating is potentially more efficient than the in-line approach.

The collimation of the ion beam creates an additional advantage that the ions have minimal velocity spread in the 'TOF direction'. This allows much higher resolving power (hence sensitivity—see above) to be achieved relative to the standard linear TOFMS of the Wiley and McLaren design. Resolving powers of 3000–5000 (FWHM) have been reported for orthogonal acceleration TOF instruments with and without ion mirrors. A reflecting stage in the drift region helps to maintain a high resolving power when beam collimation is difficult to achieve. This appears to be the case when the atmospheric pressure continuous ion sources, such as in electrospray and inductively coupled plasma ionisation, are used.

(iii) Ion-Trap Storage/TOFMS.11 This approach differs from that described by Wollnik7 because ionisation occurs outside the trapping device. An ion beam is directed to a quadrupole ion trap and accumulated. After an accumulation time the ions are ejected into a reflecting TOFMS. This hybrid method has similar duty cycle advantages to Wollnik's device and the orthogonal acceleration method. A reflecting TOFMS is essential in order to attain good resolving power because of the wide energy spread of ions ejected from the trap.

**Analyser Transmission**

Often ignored but quite significant are ion losses due to the limited transparencies of meshes used to define the electric fields in TOFMS. Coarse meshes, (high transmission, e.g., 90%) lead to greater deflection of ions because field penetration between adjacent regions is more significant than with fine meshes (lower transmission, e.g., 50%). Combining three or more meshes with a range of transparency can easily result in an overall transmission of less than 30%. In addition, if the ion trajectories are not perpendicular to the meshes actual transmission of the mesh is less than that specified because of its finite thickness.

Recently attempts to construct accelerators and mirrors without grids have resulted in improved transmissions and resolving powers. However such devices create far from optimal fields near their edges and this can limit the size of the ion source.

**Detector Modes, Noise and Dynamic Range**

As discussed in the section on the effect of the detector on resolving power, TOFMS requires focal plane ion detectors. These are usually discrete dynode electron multipliers (EM) or microchannel plates (MCP) and both can be arranged to give gains in the range \(10^3\)–\(10^5\). At very high \(m/z\) the number of electrons created by collision of an ion with the detector decreases, since the electron yield depends on the momentum of the incoming particle, not its energy. Here post acceleration of ions can improve sensitivity but at the risk of lost resolving power as is common if metastable processes take place in the drift region.

Currents produced in TOF detectors allow single ion arrival events to be detected easily as pulses of (usually) 1–5 ns width. As an example, consider a detector with a gain of \(10^6\) and a pulse half-width of 2 ns. A charge of \(q = 10^6e\) Coulombs arrives in a distribution with a width of 2 ns when a single ion is detected. Since charge \(q = \int I\, dt\), the area (total charge, \(q\)) can be used to calculate the instantaneous current (\(q/t\)) at pulse maximum if the pulse shape is known. Assuming a triangular pulse, for which area is equal to FWHM \(\times\) height, the maximum current is:

\[
I_{\text{max}} = \frac{10^6 \times 1.6 \times 10^{-19}}{2 \times 10^{-9}} \text{ A}
\]

\[
I_{\text{max}} = 8 \times 10^{-5} \text{ A}.
\]

For an \(R = 50 \, \Omega\) input impedance, a peak potential (\(V = IR\)):

\[
V_{\text{max}} = 4 \text{ mV}
\]
would be observed on the input stage of the electronics. A high bandwidth (e.g., $B = 500$ MHz) would be required to pass such a fast pulse and this would result in thermal (Johnson) noise higher than in most other instrumental amplifiers. At room temperature ($T = 273$ K) the RMS amplitude of this noise is:

$$N_0 = \sqrt{4kTRB} = 19 \mu V.$$  

(37)

A fundamental upper limit of $S/N = 200$ is indicated here. In practice other sources of noise decrease $S/N$ significantly but such a pulse would be detected on most modern devices with a $S/N$ better than 20.

A key instrumental aspect in TOFMS is how to process the pulses. This is usually done by pulse counting or by fast digitisation of the analog signal produced at the detector. Neither method is entirely satisfactory with currently available commercial devices.

(i) Ion Counting. Ion counting devices are usually based on a multistop time-to-digital converter (TDC). These devices sense the onset of pulses (start and stop events) and store the times of or intervals between these events. Time interval measurements from stop events resulting from many start events are transferred to a computer which reconstructs a histogram of the number of ions detected in consecutive discrete time intervals (bins). This scheme works well as long as there is low probability of coincidence (two or more ions arriving in the same time-bin). When this occurs, only one stop event is detected and the measurement is subject to coincidence losses. The method is therefore 'upper dynamic-range limited'. That is, it is only suitable for relatively low ion arrival rates. However TDCs have a number of redeeming features when used under appropriate conditions: they are remarkably efficient in that null-data is not recorded (i.e., if no ions are detected no data is collected); there is little noise apart from the detector dark current and the chemical noise associated with the spectrometer; the systems are relatively simple and robust; the onset of peaks can be detected with high time-resolution (<1 ns) when a constant-fractio-discriminator (CFD) is used. The CFD detects when a pulse rises to a specified fraction of its maximum height. The rise time of detector pulses is subject to a small relative distribution when compared to the pulse-height distribution so the CFD significantly increases the precision (i.e., time resolution) of measuring ion arrival times.

TDCs with built-in fast digitisers that allow simultaneous recording of the time and height of a pulse have been described but are not yet commercially available. These hybrid devices have the potential to extend upwards the dynamic range of the ion-counting approach.

(ii) Analog Based Methods—Integrating Transient Recorders (ITRs). These are principally based on fast ADC converters (flash-converters). The signal from the detector is digitised at a fixed sampling rate. Significant advances in the speed and memory of these devices in recent years allows signals to be sampled at frequencies up to one or two GHz. Caution must be advised with regard to the meaning of this specification: while devices exist to digitise at this rate, the bandwidth of the input stages are typically half this frequency. Rather than a shortcoming, this is a designed feature of the devices to satisfy the Nyquist condition: there must be at least two sampled points per cycle of a periodic signal or else false signals (aliasing) can result. The bandwidth restriction causes pulse widths to increase and amplitudes to fall. The advantage of ITRs is that coincidence losses are delayed to very high ion arrival rates. The limitations are mainly that the devices are very costly and that they are (currently) limited to 8-bit amplitude resolution in the input stage. The limitation of the resolution of the flash converters is felt in the dynamic range of the detection system. A single ion pulse must give a reasonable analog-to-digital conversion but not so great that a few coincident ions saturate the flash converter. To extend the dynamic range, ITRs are usually used in sum-averaging mode. Several spectra are digitised and the digitisations are summed into (usually) 16 bit memory. Thus 256 ideal pulses, each giving a conversion to 255 ($2^8 - 1$), on a baseline with no noise, gives a sum of 65280 ($\approx 2^{16}$). In practice, however, 16 bit dynamic range is not achieved by these devices. Typically 10–14 bits are achieved because of noise which is digitised along with the signal from the detector.

It is here in dynamic range that TOFMS is most at a disadvantage compared to scanning instruments. Sampling rates of only kHz are needed in a scanning mass spectrometer. At this rate a 16 or 20 bit flash converter can readily provide an excellent dynamic range.

**Speed and Repetition Rate**

It is often stated that TOFMS is fast. While it is true that a spectrum from a single shot can be acquired in about 100 μs or less, it is rarely the case that a single shot will generate sufficient ions to give a good statistical representation of the distribution of $m/z$. The average number of ions produced per shot depends on many factors discussed in this article. Usually signal averaging must be performed until a few thousand ions have been detected. If 100 ions are produced on average in a shot then 100 shots will produce a good spectrum. If 1000 ions are produced in a shot then 10 shots will produce a good spectrum. It is realistic to expect that spectra can be acquired at 10–100 Hz.

The rate at which ion packets can be analysed depends on the flight time of the heaviest ion in the spectrum and upon the method of forming the packets. In laser desorption, for example, this may be limited by the rate at which the laser can be pulsed. In orthogonal TOFMS with a continuous ion source, much higher repetition rates are possible but the upper limit may be imposed by the speed at which the digitiser can signal-average.

**CONCLUSION**

Advances in the technologies of ionisation, ion detection as well as high speed signal and data processing make TOFMS an attractive alternative to conventional mass analysers in specific applications. Development
ion optics and electronics have substantially removed the limitations of early TOF instruments. In this article many of the important instrumental and physical concepts governing the design and performance of TOF instruments have been discussed. There is considerable scope for further development of the facilitating technologies to allow the speed of TOF to be exploited while maintaining a dynamic range commensurate with scanning MS instruments.

REFERENCES


GENERAL READING
