JOURNAL OF MASS SPECTROMETRY, VOL. 33, 64–74 (1998)

Spectrum Measurement Using Discrete Detector Arrays

Keith Birkinshaw
Department of Physics, University of Wales Aberystwyth, Aberystwyth, Dyfed SY23 3BZ, UK

Modern spectrometers can produce highly resolved spectra, but this ability has not yet been matched by an ability to measure them efficiently and for instruments such as a spatially dispersive mass spectrometer it is in the development of high-performance focal plane detectors (FPDs) where there are enormous gains in efficiency to be achieved. In this paper two key questions are addressed: (a) how much information does a 1D spectrum or 2D image contain?; and (b) can a given FPD capture all this information? In answering these questions, issues of performance, data quality and technology limitations of 1D and 2D arrays of counters arise naturally. A procedure is evolved for comparing the performance of a detector and a spectrometer which is not limited to spatially dispersive instruments. © 1998 John Wiley & Sons, Ltd.

J. Mass Spectrom. Vol. 33, 64–74 (1998)

KEYWORDS: detector arrays; spectrometer; spectrum measurement; information

INTRODUCTION

Much effort has been invested in designing spectrometers which produce highly resolved spatially dispersed spectra and these are traditionally measured using a single slit detector. This type of detector has the highest resolving power of any currently available detector.[1] However, a single slit can measure only a tiny fraction of a spectrum at any given time and there has been a move towards the use of focal plane detectors (FPDs)[1–3] which can measure a whole section of a spectrum simultaneously. Clearly, the performance of an FPD and the spectrometer to which it is fitted must in some way be matched. A high-performance spectrometer combined with a low-performance FPD is not a satisfactory combination (the term FPD is used here to refer to the combination of a microchannel plate electron multiplier (MCP) mounted in front of a detector array).

One way to view a spatially dispersive instrument is to consider the procedure of producing and measuring a 1D spectrum or 2D image at its basic level of a flow of information from a source to a destination (Fig. 1). The FPD is the interface or 'channel' through which the information flows. This model provides a means of expressing the performance of both the spectrometer and the FPD and hence facilitates a discussion of relative performance and the matching of the two. Within this framework, issues of data quality, performance (resolving power and dynamic range) and technology limitations can be discussed together, not only of 1D but also of 2D FPDs. Extension to 3D and to time dispersion, etc., is also possible.

Clearly, other factors are also of importance to a user, such as reliability, cost, size and power consumption. Only issues relating to the information measured are considered here. The parameters of relevance to this discussion are resolving power, dynamic range and size of the active area of the FPD.

A brief illustration and definition of information are given in the next section and references to texts on information theory are given. In the subsequent section the information output by a mass spectrometer is defined by assuming that spectral peaks have a Gaussian profile, defining a sampling frequency necessary to recover the spectrum accurately and then using information theory to calculate the upper limit of the information output rate. The 'information rate capacity' or 'capacity' (the maximum rate at which the FPD can measure information) of the FPD is also determined. The expressions derived allow the characterization of mass spectrometers and FPDs in terms of their capacity to deliver and transmit information, respectively. It is assumed that the system is noise free. The effect of noise is examined in the Appendix.

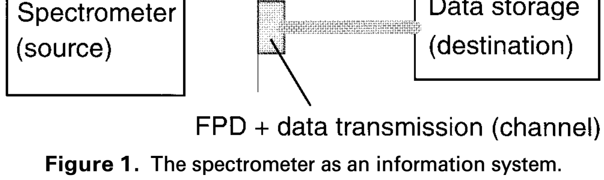

Figure 1. The spectrometer as an information system.

* *Correspondence to*: K. Birkinshaw, Department of Physics, University of Wales Aberystwyth, Dyfed SY23 3BZ, UK
Email: kpi@aber.ac.uk

CCC 1076–5174/98/010064–11 $17.50

Received 30 June 1997
Accepted 3 October 1997

INFORMATION

It is the function of an FPD accurately to record and deliver data produced by a spectrometer to the destination. Information[4,5] is a statistical measure whose value in the present context allows quantification of the performance of an FPD and a spectrometer and hence a comparison of the two. Information can be measured in various units and units of bits are used here. Information rate is the rate at which information is transferred.

Information may be defined as the minimum number of binary digits (binits) required to encode data. Consider a measurement of the number of particles (n) in 1 s incident on a single detector equipped with a sufficiently large counter and assume that this number is transmitted without error to the destination where the count is recorded, i.e. there is no noise. The information (I) obtained from one reading of the counter is

$$I_n = \log_2(1/p_n)$$

where p_n is the probability of measuring n events. If we assume the count is always between 0 and 255 and any value is equally probable within this range, then $p_n = 1/256$ for each reading. The information in a reading is

$$I = \log_2 \frac{1}{(1/256)} = \log_2 256 = \log_2 2^8 = 8 \text{ bits}$$

This is, of course, the number of binits needed to encode the data. Since we read the counter every second, I is the information measured per second. For an array of M such counters,

$$I = \sum_M \log_2 256 = 8M$$

and the average information rate per counter is 8 bits s^{-1}. The number of binits required to transfer information cannot be less than the number of bits of information (without information loss) but the binit rate may greatly exceed the information rate and it may be profitable to compress the former before transmission.

In the following all logarithms are taken to the base 2 and the subscript 2 is omitted. It is noted that $\log_2 A = 3.32 \log_{10} A$.

The above assumption that any reading between 0 and 255 is equally probable will generally result in an overestimate of the information. It can be shown[4,5] that the average information is a maximum when any count between 0 and 255 is equally probable. In the discussion below, this maximum is chosen to characterize the spectrometer and FPD.

INFORMATION OUTPUT OF A SPECTROMETER

We consider a spatially dispersive instrument to consist of three parts: (i) the particle (ion, electron, photon, etc.) production, dispersion and focusing section; this will be referred to for convenience as the spectrometer; (ii) the FPD; and (iii) the data storage section. In other words, we shall regard the spectrometer as the part of the instrument which produces the spectrum at the focal plane and the FPD as a separate item responsible for measuring the spectrum (i.e. detecting the spectrum and delivering the data to the destination). Thus we can ask the rate at which the spectrometer is producing information and whether the FPD has the capacity or ability to measure the information at this rate.

Spectrum sampling

Consider a spectrum consisting of a single Gaussian peak. A Fourier transformation of this peak gives a continuous frequency spectrum. According to sampling theory, the sampling frequency (f_s) should be twice the maximum frequency present in the signal (ν_{max}) in order to be able to recover the signal *exactly*. The minimum sampling frequency (f_{smin}) is therefore

$$f_{smin} = 2\nu_{max}$$

All the information present in the spectrum is retained if sampling takes place at a rate $\geqslant f_{smin}$.

However, the Fourier transform of the Gaussian peak has a continuous frequency spectrum and although the high-frequency components are small they are nevertheless present. Unless the incident spectrum profile is limited to frequencies less than or equal to half the sampling frequency, the consequences of sampling at less than f_{smin} must be considered. Consider a spectrum of Gaussian peaks. A single Gaussian has the familiar form

$$y = \exp(-x^2/2\sigma^2)$$

and for $x = 2.5\sigma$, $y = 0.044$. Therefore, a spectrum of Gaussian peaks on a pitch (separation between the centres of adjacent peaks) of 5σ would overlap to give an 8.8% valley—very close to the 10% valley definition of resolution often used in mass spectrometry.

Consider the signal

$$v(x) = \exp[-\pi(cx)^2] \qquad \left(\sigma_x = \frac{1}{\sqrt{2\pi}}\frac{1}{c}\right)$$

where c is a constant. The Fourier transform is

$$X(f) = \frac{1}{b}\exp[-\pi(f/c)^2] \qquad \left(\sigma_f = \frac{1}{\sqrt{2\pi}}c\right)$$

Hence

$$\sigma_f = \frac{1}{2\pi\sigma_x}$$

where b is a constant.

We consider a spectrum of Gaussian peaks with a pitch of $5\sigma_x$ and we sample at a frequency f_s sufficient to correctly sample signal frequencies up to $2.5\sigma_f$. The latter value is somewhat arbitrary but includes most of the frequency spectrum. Therefore,

$$f_s = \text{sampling frequency} = 1/(\text{sample period})$$
$$= 2 \times 2.5\sigma_f = 5/2\pi\sigma_x \approx 0.8/\sigma_x$$

and

$$\text{No. of peaks} = \text{FPD length}/5\sigma_x$$

'Frequency' is used here in the sense of spatial sampling, e.g. in the sense of number of samples per micrometre. Now, if the spectrum is sampled by an FPD the sample period is equivalent to the spatial resolution of the FPD, and since

$$\text{FPD length} = \text{spatial resolution} \times \text{No. of detectors}$$

we have

$$\text{No. of peaks} = \frac{\text{FPD length}}{5\sigma_x} = \frac{\text{FPD length}}{4 \times \text{spatial resolution}} = \frac{\text{No. of detectors}}{4}$$

Hence using an array of 256 detectors, we can measure 64 resolved peaks and we introduce an error due to undersampling of a small fraction of the frequency spectrum. We can therefore say that almost all the information is captured if the spatial resolution is equal to $1.25\sigma_x$. To a resonable approximation this is equivalent to the condition

$$\text{FWHM of spectral peak} \geqslant 2 \times \text{spatial resolution}$$

where FWHM = full width at half-maximum.

If the peaks in a spectrum are better focused giving the same mass range but sharper peaks, then the dispersion must be increased to bring σ_x back to the required value and information is lost (see below).

As an example, we can consider the case of $\sigma_x = 20$ μm, giving $c = 0.02$ μm^{-1} and $\sigma_f = 0.008$ μm^{-1}. We must sample at a frequency of $2 \times 2.5 \times \sigma_f = 5 \times 0.008 = 0.04$ samples μm^{-1} = 1 sample per 25 μm. The Aberystwyth FPD[6] has a spatial resolution of 25 μm and this equates to 1 sample per 25 μm = 1 sample per detector. It can therefore sample peaks with $\sigma_x \geqslant 20$ μm.

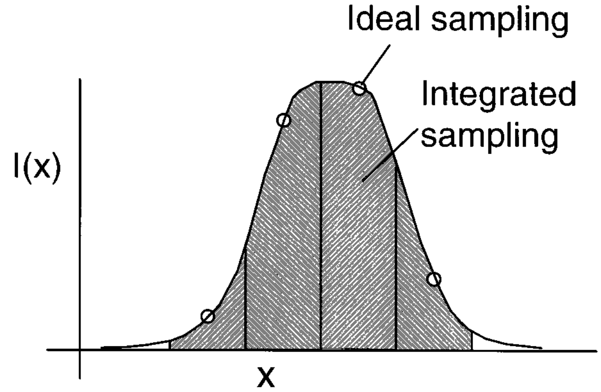

Figure 2. Diagram showing the distinction between ideal and integrated sampling. An ideal sample is taken over an infinitely small value of δx. An integrated sample is taken by a detector site over a section of the peak as shown by the shaded areas.

Spectrum recovery

In the above discussion, it has been emphasized that the exact recovery of a spectrum requires that the incident signal is band limited. Other factors which influence spectrum recovery must also be pointed out. These include:

- The incident spectrum is not measured directly by the array. The spectrum falls on the MCP and each particle in the spectrum initiates an MCP output pulse which is recorded by the array. This results in a widening of the measured peak[1] but in line with other approximations made here it is assumed that there is no spreading of the incident spectrum before measurement by the FPD.
- The spectrum is not sampled at a point (ideal sampling) but rather the area under the ion intensity profile above a detector site is integrated (integrated sampling), as shown in Fig. 2. Calculations were performed using both ideal and integrated sampling and results were compared.

Figure 3(a) shows a hypothetical spectrum consisting of 16 Gaussian peaks (of heights 10, 2, 5, 3, 8, 15, 15, 0.5, 30, 30, 0.1, 3, 10, 17, 10 and 3) over 64 electrodes (solid line) with about a 10% valley between peaks of equal height, i.e. the sampling frequency satisfies the criteria set out above. Figure 3(b) shows the ideally sampled spectrum and from this the original can be simply recovered with good accuracy as shown by the circles in Fig. 3(a). The error in the average peak heights and the error in the peak centroid were found using Mathcad and are given in the following section. Figure 3(c) shows that the lowest intensity peaks are not recovered as accurately as the higher peaks. However, at a higher sampling frequency of 1 sample per 18.75 μm the lowest peaks could be recovered with the better accuracy [Fig. 3(d)]. Since the sampling frequency depends on the spatial resolution, which is fixed, an increase in sampling frequency in practice means increasing σ_x. If σ_x is increased by increasing the dispersion, then information is lost as a smaller fraction of the spectrum is measured by the FPD. If σ_x is increased at the same dispersion, then information is lost due to a loss of resolution (below).

Inaccuracy due to omission of high frequencies. Computations were carried out in which 16 Gaussian functions were added to give the hypothetical spectrum shown in Fig. 3(a). The standard deviation was 20 μm and peaks were separated by 100 μm. This was then sampled in two ways: (i) the intensity was calculated at points spaced at regular intervals of 25 μm (ideal sampling); (ii) the spectrum was sampled by integrating the intensity above the electrodes (integrated sampling).

In general, ideal sampling at the higher frequency gave the better recovered spectrum, as would be expected. When the two smallest peaks are ignored, the results indicate that if the spectrum is sampled by a discrete FPD (i.e. integrated sampling) at a rate of 1 sample per 18.75 μm (0.053 samples μm^{-1}) then the error in a peak centroid is about ± 0.5 μm and the peak height was systematically underestimated by about 3.4% whereas the *relative* peak heights were accurate to

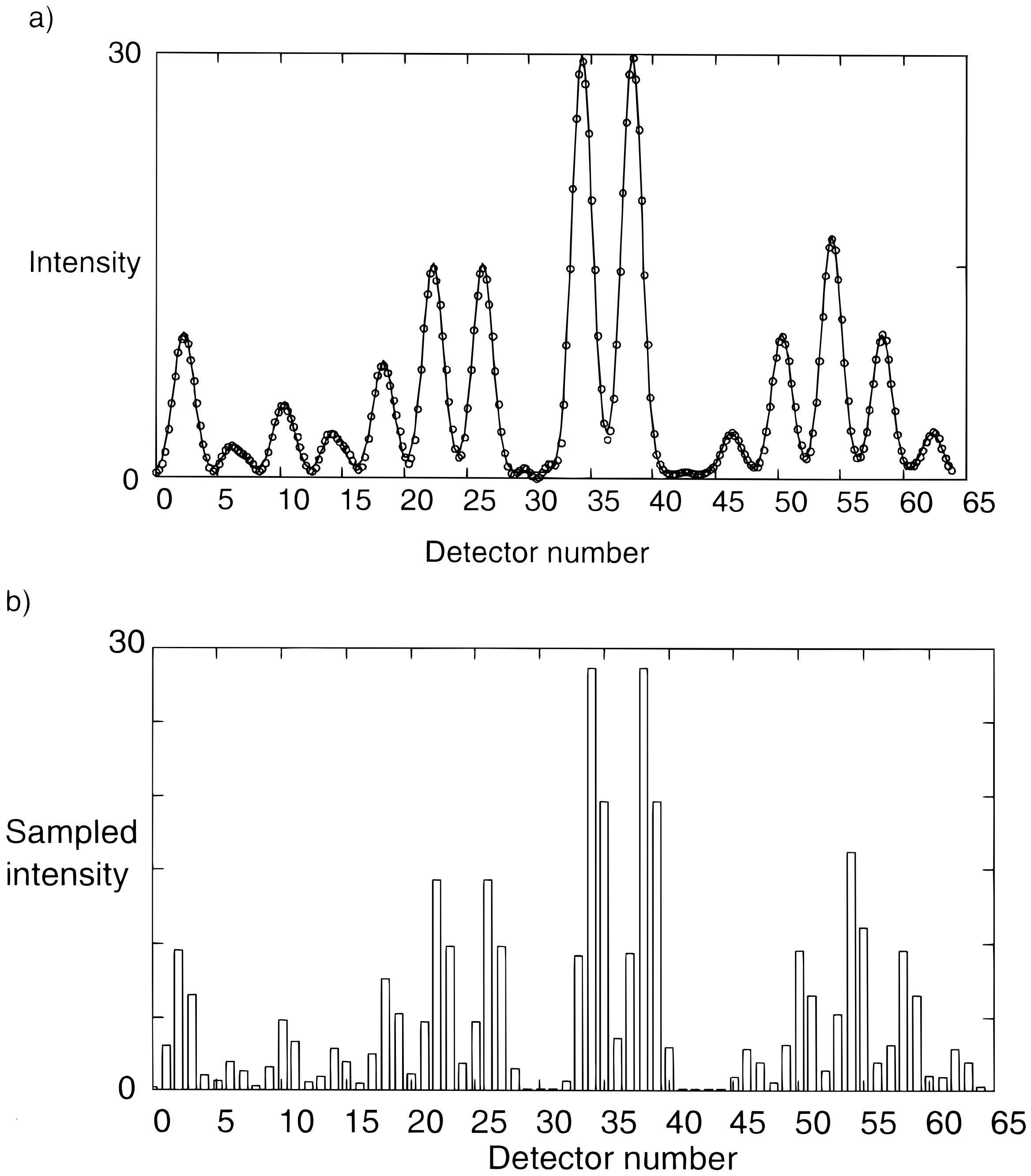

Figure 3. (a) The solid line is a computed incident spectrum composed of Gaussian peaks of height 10, 2, 5, 3, 8, 15, 15, 0.5, 30, 30, 0.1, 3, 10, 17, 10 and 3. The circles show the spectrum recovered from the ideally sampled spectrum given in (b). (c) The lower part of the spectrum shown in (a). It can be seen that the lowest intensity peaks are distorted. (d) As (c) except that the incident spectrum is sampled at intervals of 18.75 μm instead of 25 μm.

better than ±0.5%. At a rate of 1 sample per 25 μm (0.04 samples μm^{-1}) the error in the peak centroids was about ±1.5 μm. The peak heights were underestimated on average by 7.5% (integrated sampling) but the *relative* peak heights were accurate to better than ±3.5%.

For the two smallest peaks the errors were several times larger because of the effect of higher frequency components of the adjacent peaks.

Source information rate

How much information is there in a spectrum? To answer this question we require the parameters indicated in Table 1. Dispersion in 2D is assumed. Units of length are micrometres and time is in seconds.

With these parameters, we define the maximum information from the source (I_{source}) to be

$$I_{source} = S \log\left(\frac{F_{source}}{S}\right)$$

where

$$S = \frac{0.8d}{\sigma_x} \times \frac{0.8e}{\sigma_y}$$

$$= \text{(samples in } x \text{ direction)} \times \text{(samples in } y \text{ direction)}$$

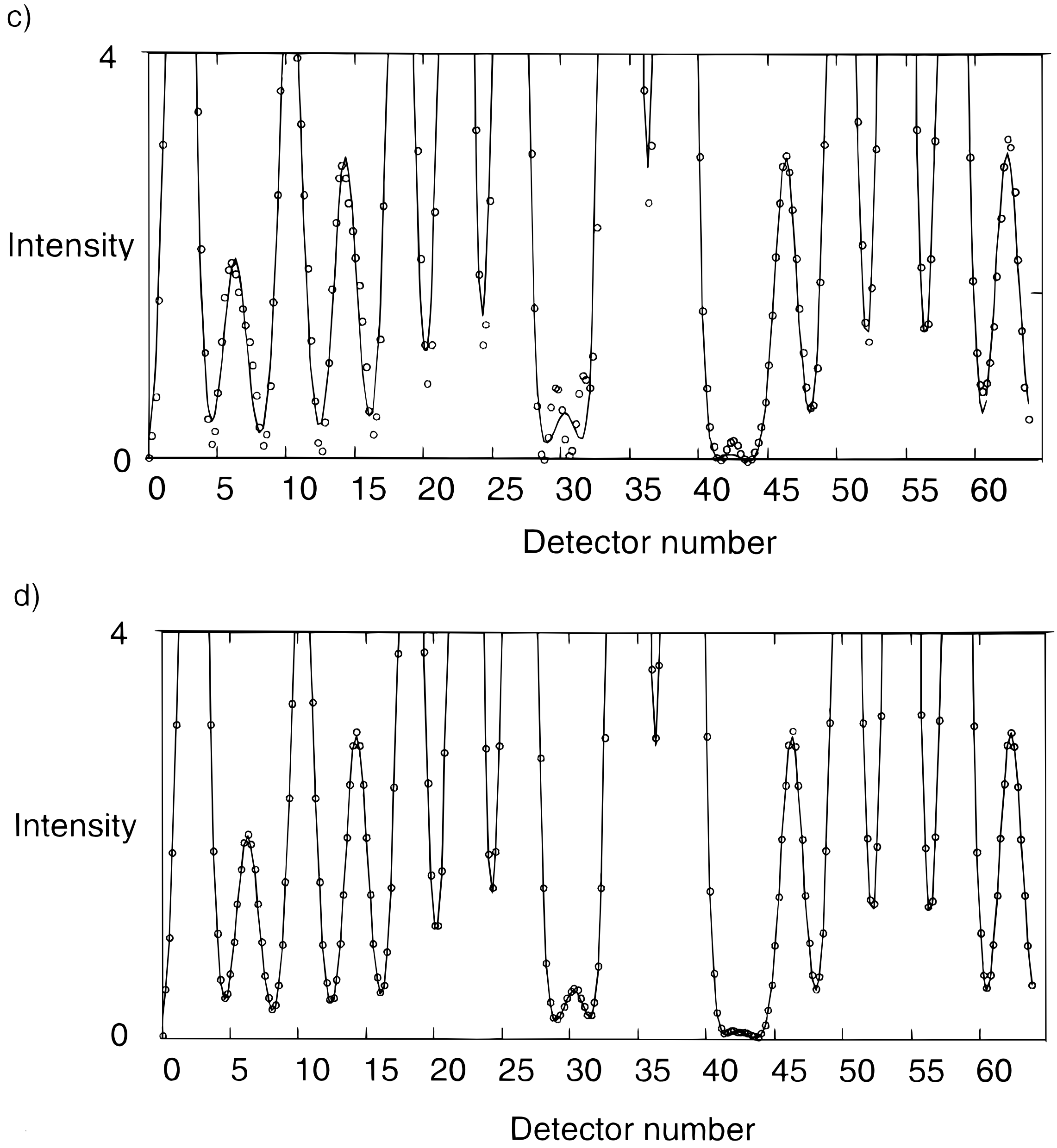

Figure 3.—*Continued*

For a ID spectrum there is one sample in the y direction.

Two assumptions made in this definition are

- the incident spectrum is band limited to frequencies $\leqslant 2.5\sigma_f$, in which case all the information can be recovered from an ideally sampled signal.
- the maximum particle flux on to the focal plane is F_{source}. We assume that the incident flux may vary between 0 and F_{source} and is evenly distributed across the focal plane. At any detector the measured number of counts per second may have any value between 0 and F_{source}/S with equal probability.

Although I_{source} is a maximum if we assume the ion flux to be uniformly distributed across the focal plane (see Information section above), the latter will not generally be the case and is inconsistent with the existence of a spectrum or image. Therefore, I_{source} is an upper limit but as such it provides a useful order of magnitude basis for comparison with the upper limit of an FPD (below).

It can be seen that the maximum information rate in an incident ID spectrum depends on the resolving power of the spectrometer (through σ_x), the ion flux and the area of the focal plane. It should be emphasized that the dimensions of the focal plane ($d \times e$) are indepen-

Table 1. Parameters needed in the calculation of the information in the incident spectrum

Dimensions of the focal plane of the instrument	de (d in x direction and e in y-direction)
Standard deviation of a single peak (assumed to be Gaussian)	σ_x (x direction) σ_y (y direction)
Maximum particle flux falling on the focal plane	F_{source}

dent of the FPD. Ideally, the FPD should span the whole focal plane if it is to capture all the information otherwise some of the information output will be lost.

FPD INFORMATION RATE CAPACITY

We consider 1D and 2D FPDs (Fig. 4) and derive an expression for the amount of information per second which a 2D FPD could measure. The capacity (the maximum information rate) of a 1D FPD then follows simply by setting the number of detectors in the *y* direction equal to unity. It is assumed that the 1D FPD is equipped with *J* binit counters and the 2D array is equipped with *K* binit counters, where $J > K$.

To find the information capacity we proceed as shown in Table 2. We consider a measurement period of 1 s.

In parallel with the case of I_{source}, the value of I_{FPD} depends on the resolving power of the FPD (through *X*, *Y*, *a* and *b*), the dynamic range through the maximum flux (F_{FPD}) allowed by the MCP and the active area of the FPD.

From the equations in Table 2, it can be seen that there can be a trade-off between f_r and *K*. If f_r is doubled then the data accumulation time is halved and *K* is reduced by 1. Therefore, if we wish to have counters with a low number of binits (as may be necessary for a high-resolution array), then counters must be read at a higher rate to ensure that they do not overflow.

MCP limited count rate

We consider the square 2D array of pulse counters in Fig. 4. To find the maximum rate at which the MCP can deliver pulses to a detector site, we first calculate the number of MCP channels above the array. To a reasonable approximation this is given by ab/P^2, where *ab* is the area of the array, *P* is the pitch of the MCP

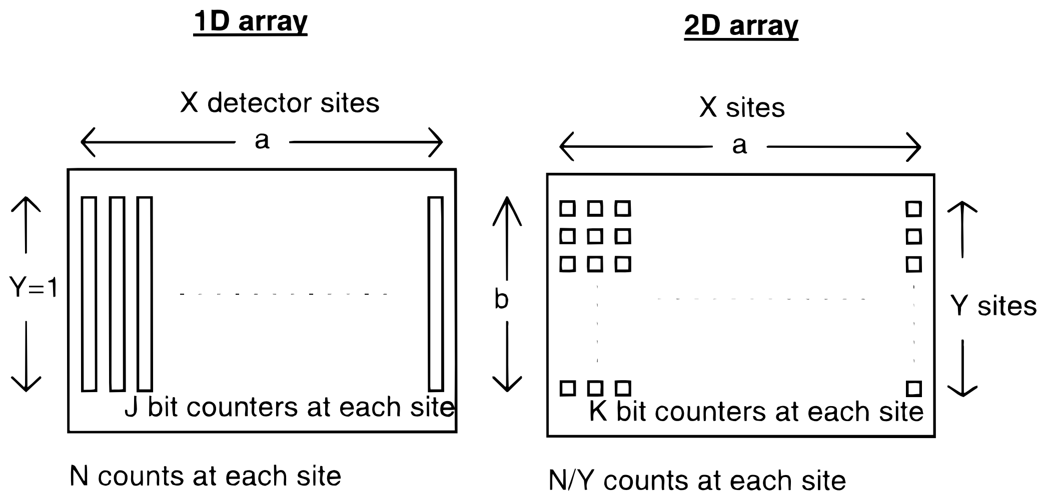

Figure 4. Schematic diagrams of a one-dimensional and a two-dimensional array.

Table 2. Calculation of the information rate capacity of an FPD

Area of a single detector	ab/XY	
Max. particle flux on to FPD[a] P = MCP channel pitch; R = MCP channel recovery time	$F_{FPD} = ab/P^2R$	(1)
Max. number of counts at each site (measured per second)	F_{FPD}/XY	
Number of counter binits needed per site (K)	$K = \log_2\left(\frac{F_{FPD}}{XY}\right)$	(2)
Frequency (f_r) at which the counters must be read[b]	$f_r = F_{FPD}/2^K$	(3)
Information capacity of the FPD[c] (for a 1D FPD, $Y = 1$)	$I_{FPD} = XY \log_2\left(\frac{F_{FPD}}{XY}\right)$	

[a] See the next sub-section.
[b] See the next but one sub-section
[c] The assumptions made in the derivation of I_{FPD} are similar to those made in the derivation of I_{source}. The number of sites is XY and at each site counts between 0 and F_{FPD}/XY can be stored.

channels and P^2 is taken as the average area occupied by an MCP channel. Each channel of the MCP requires a time R to recover between pulses and we take $1/R$ as the maximum rate at which it can deliver pulses. Therefore, the maximum count rate which the MCP can deliver to the array is

$$\text{Max. count rate} = F_{\text{FPD}} = \frac{ab}{P^2 R} \quad (1)$$

and the maximum count rate per site is

$$\text{Max. count rate per site} = \frac{F_{\text{FPD}}}{\text{No. of sites}} = \frac{F_{\text{FPD}}}{XY}$$

where a and b are defined in Fig. 4, P = MCP channel pitch and R = MCP channel recovery time.

In deriving F_{FPD} above, no account is taken of the random time of arrival of particles which limits the particle flux to less than F_{FPD} because of the probability of two particles arriving at the same MCP channel within the recovery time. The number of counter binits needed per site (K) if all counters are to just fill in 1 s is

$$K = \log\left(\frac{F_{\text{FPD}}}{XY}\right) \quad (2)$$

It should be noted that in the case of a single event FPD such as a resistive strip,[1] $X = Y = 1$ and the MCP will not limit the upper count rate. The upper limit in this case is determined by the time taken to measure an MCP pulse and compute its position.

Array limited measurement

The frequency f_r at which the array must be read is determined by the size of the counters. For optimum performance the counters should be read sequentially with no delay between the reading of the final counter and the beginning of the next read cycle.[1] Thus an FPD equipped with K binit counters [where $K = \log(F_{\text{FPD}}/XY)$] may be read in 1 s without counter overflow if the maximum count rate per site is F_{FPD}/XY. In the case of a discrete detector array integrated on silicon, a high resolving power and large counters are incompatible as too much silicon area would be required for a high-resolution array. Large counters would allow a lower read frequency as they could be read more slowly without danger of overflow, but if the count rate is limited by the MCP there is little point in having large counters which will never be filled (unless one-shot spectra are required). As an illustration of this, we consider that in order to read 2^α counts per second from an α binit counter, only α binits per second are needed. To read 2^α counts per second from a β binit counter ($\beta < \alpha$), each of the sites would have to be read $2^\alpha/2^\beta$ times per second. If $\alpha = 8$ and $\beta = 2$, then the read frequency would have to be 64 times faster.

The expression for f_r is found by dividing the maximum count rate per site (F_{FPD}/XY) by the number of counts which can be read in one reading of the counter (2^K) and then multiplying this by the number of counters to be read:

$$f_r = \frac{F_{\text{FPD}}}{2^K} = \frac{ab}{P^2 R \times 2^K} \quad (3)$$

Hence all counters can be read and the read cycle restarted before any counter overflows. Equations (2) and (3) relate the properties of the MCP and the array.

Characteristics of a FPD

Table 3 gives a summary of the equations derived and quantitative results obtained for 1D and 2D FPDs with

Table 3. Summary of equations and results for examples of 1D and 2D FPDs, assuming that the sampling time is 1 s and the FPD spans the spectrometer focal plane

			$a = b = d = e = 6400$ µm; $X = 256$; $P = 15$ µm; $R = 10^{-2}$ s	
			$Y = 1$	$Y = 256$
Instrument	parameter	2D[a]	1D	2D
Focal plane detector	Count rate capacity of FPD (Hz)[b]	$F_{\text{FPD}} = \frac{ab}{P^2R}$	1.8×10^7	1.8×10^7
	Count rate capacity per site (Hz)	$N = F_{\text{FPD}}/XY$	7.1×10^4	2.8×10^2
	No. of counter binits	$K = \log N$	∼16	∼8
	Read frequency (Hz)[c]	$f_r = F_{\text{FPD}}/2^K$	2.6×10^2	6.6×10^4
	Binit output frequency (binits s^{-1})	$I_{\text{FPD}} = f_r K$	4.1×10^3	5.3×10^5
Spectrometer	I_{source} in bits ($\sigma_x = 20$ µm; $F_{\text{source}} = 1.8 \times 10^7$ Hz)	$S \log(F_{\text{source}}/S)$	4.1×10^3	5.3×10^4
	I_{source} in bits ($\sigma_x = 10$ µm; $F_{\text{source}} = 1.8 \times 10^7$ Hz)	$S \log(F_{\text{source}}/S)$	7.7×10^3	1.6×10^6
	I_{source} in bits ($\sigma_x = 10$ µm; $F_{\text{source}} = 10^8$ Hz)	$S \log(F_{\text{source}}/S)$	9.0×10^3	2.2×10^6
	I_{source} in bits ($\sigma_x = 5$ µm; $F_{\text{source}} = 10^{10}$ Hz)	$S \log(F_{\text{source}}/S)$	2.4×10^4	1.4×10^7

[a] The expressions for a 1D array are obtained by setting $Y = 1$ and $K = J$.
[b] This is determined either by the MCP or the detector circuitry (whichever is the lowest). The expression for the MCP limited rate is given.
[c] f_r gives the minimum read frequency to ensure that all counters are read before any one can fill.

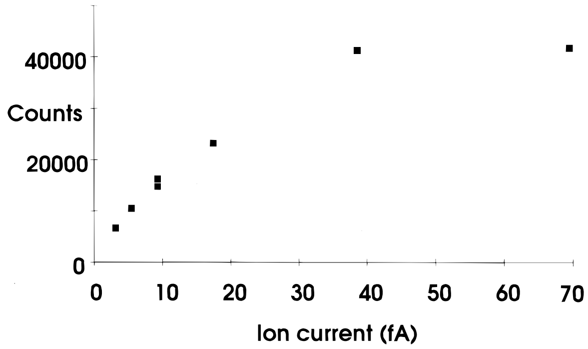

Figure 5. Counts measured in 0.5 s as a function of the ion current. The counts recorded by five detectors were summed to obtain the counts plotted.

the specification indicated. It can be seen that the frequencies at which the arrays must be read are attainable. Therefore, there is no point making the array with more than the indicated number of counter binits per site as this number is sufficient to match the count rate which is limited by the MCP. However, there may be good reason to reduce the number of binits per site and increase f_r (reduce the sample time), as this would allow less circuitry, less silicon area and allow lower cost. If f_r is doubled then K can be reduced by 1. Therefore, instead of the 256×16 binit counters of a 1D FPD we could use 8 binit counters if they are read at $f_r = 6.6 \times 10^4$ Hz (i.e. 256×2^8 Hz).

The information output by the source is shown in Table 3 for four combinations of resolution and ion current. When the spectrum has peaks with $\sigma_x = 5$ μm and a particle flux over the area of the FPD of 10^{10} particles s^{-1} then the maximum information for a 1D spectrum is 2.4×10^4 bits s^{-1} and for a 2D image $I_{source} = 1.4 \times 10^7$ bits s^{-1}. In order to match the specification of the FPD, σ_x must be increased to 20 μm and the particle flux in the focal plane reduced to 1.8×10^7 particles s^{-1} with a loss of a factor of about 6 for the information in the 1D spectrum and a factor of about 26 of the information in the 2D spectrum.

If the FPD is longer in the direction of dispersion for a 1D FPD then the information rate capacity is proportionally larger. If the FPD is wider (in the direction perpendicular to the direction of dispersion) then the information rate capacity will increase as the logarithm of the width.

Table 3 indicates that the information rate capacity of the 2D FPD is more than 100 times greater than that of the 1D FPD. Similarly, the source information for a 2D image is more than 100 times that for a 1D spectrum under the conditions stated. This highlights the problem of measuring a 1D spectrum with a 2D FPD. There is no increase in the amount of information when using the 2D device since there is no dispersion in the Y direction. F_{FPD} is limited by the MCP and the counts recorded on Y sites of the 2D device are identical (to a first approximation) with those recorded on one detector of the 1D device. However, there is a penalty incurred as the read frequency must be more than 100 times greater for the 2D device and the total number of counter binits needed by Y detectors far exceeds the number in one detector in the 1D device. This determines that a large amount of circuitry is needed and inhibits the production of high-resolution 2D arrays of particle *counters*. Where fully parallel 2D data acquisition is required, integrating devices[2,7] such as CCDs are used, although they are not ideally suited to ion counting or the one-dimensional slit geometry of spatially dispersive spectrometers.[1]

By way of illustration, we may ask if a 256×256 2D array of detectors can be made with a spatial resolution of 25 μm in both dimensions using an MCP with the above specification. From Table 3 we see that such a device would require about 8 bits per counter, it could count at a rate of 280 Hz per site and the read frequency would be $f_r = 6.6 \times 10^4$ Hz. Now, it is not possible using present technology to fit an 8 binit counter within an area of 25×25 μm. If we reduce K by 7 bits then we must increase f_r by a factor of 2^7, i.e. 128. This indicates that the read frequency should be 8.4 MHz. This is feasible and tells us that for the low particle flux (the value being limited by the MCP) it is possible to read such a 2D array before any site receives more than one count. In the spirit of these calculations, one should anticipate that this figure is an upper limit on the particle flux which can be accurately measured, not least because the particle flux will not normally be uniformly distributed.

Figure 5 shows results obtained[8] using a 1D discrete detector FPD recently developed at Aberystwyth. A single ion peak covering five detectors was measured by varying the ion current and recording the counts for a fixed time (0.5 s). It can be seen that above about 20 fA the curve begins to deviate strongly from linearity. We can compare this result with that predicted by the above equations as follows: $a = 4800$ μm, $b = 2000$ μm, $P = 15$ μm, $R = 0.01$ s, $X = 192$ and $Y = 1$. Therefore, $F_{FPD} = 4.3 \times 10^6$ Hz, $N = 2.2 \times 10^4$ Hz $= 3.5 \times 10^{-15}$ A, $K = 14.4$ binits and $f_r = 2.0 \times 10^2$ Hz. The Aberystwyth FPD has 8 binits per detector and therefore the read frequency is $f_r = 2.0 \times 10^2 \times 2^{6.4} \approx 1.67 \times 10^4$ Hz.

For a single electrode, $N = 3.5$ fA and therefore for five electrodes the count rate limit is roughly five times this, i.e. 17.5 fA, in reasonable agreement with the experimental observation of the onset of strong nonlinearity. In Fig. 5, at an ion current of 10 fA (about 6×10^4 ions s^{-1}) the counts accumulated by the five detectors in 0.5 s should ideally be about 3×10^4. The observed counts were about 1.7×10^4, but better agreement would not be expected because the ions in the beam were concentrated at the centre of the five electrodes and because of the approximation made in the model.

CONCLUSIONS

The information available in the focal plane of a spatially dispersive spectrometer has been quantified and an FPD has been characterized in terms of its ability to measure the information. It has been assumed that the spectral peaks have a Gaussian profile. The resolving power, the dynamic range and the size of the FPD are included in this characterization, in addition to properties of the MCP. It has been shown that to ensure spectral peaks are recovered accurately the FWHM of the spectral peak must be twice the spatial resolution or more.

The limiting particle flux has been calculated. For a high-resolution discrete FPD this is determined by the MCP performance. It has been shown that a 2D array of particle *counters* places demands on counter size and read frequency which in some cases be prohibitive. The model developed is applicable to other types of spectrometer and detector.

REFERENCES

1. K. Birkinshaw, *Int. Rev. Phys. Chem.* **15**, 13 (1996).
2. K. Smith, *Exp. Methods Phys. Sci.* **29A**, 253 (1995).
3. L. J. Richter and W. Ho, *Rev. Sci. Instrum.* **57**, 1469 (1986).
4. B. B. Carlson, *Communication Systems*. McGraw-Hill, New York (1986).
5. B. P. Lathi, *Modern Digital and Analogue Communication Systems*. Holt, Rinehart and Wilson (1983).
6. D. P. Langstaff, M. W. Lawton, T. M. McGinnity, D. M. Forbes and K. Birkinshaw, *Meas. Sci. Technol.* **5**, 389 (1994).
7. K. Birkinshaw, *J. Mass Spectrom.* **32**, 795 (1997).
8. K. Birkinshaw and D. P. Langstaff, *Int. J. Mass Spectrom. Ion Processes* **132**, 193 (1994).

APPENDIX

Note on the zero noise model

The assumption is made in the text that in the calculation of I_{source} and I_{FPD} the argument of the logarithm is not less than unity:

$$I_{source} = S \log\left(\frac{F_{source}}{S}\right)$$

$$I_{FPD} = XY \log_2\left(\frac{F_{FPD}}{XY}\right)$$

This assumption is normally valid since both F_{FPD}/XY and F_{source}/S are usually much greater than 1.

In the case of a time-of-flight (TOF) mass spectrometer, the relevant equation for the information may have an argument of less than unity and the following procedure may be used to find the information output. We consider the measurement period (T) of an ion pulse to be divided into n segments, where $n = 0.8T/\sigma_T$ and σ_T is the standard deviation of an ion arrival time peak (assumed constant). Each pulse of ions normally contains far fewer ions than the number of time segments and the probability of two ions arriving in a single segment after accumulation for 1 s is considered to be negligible in the following analysis (although the ions may be bunched together). For N ions per second, the information is

$$\begin{aligned} I_{TOF} = &-(\text{No. of segments containing 0})\log(\text{probability of 0}) \\ &-(\text{No. of segments containing 1})\log(\text{probability of 1}) \\ = &\frac{-0.8T}{\sigma_T}(p_0 \log p_0 + p_1 \log p_1) \text{ bits per pulse} \end{aligned}$$

Now,

$$p_1 = N/n \quad \text{and} \quad p_0 = 1 - p_1$$

Therefore

$$\begin{aligned} I_{TOF} &= \frac{-0.8T}{\sigma_T}[(N/n)\log(N/n) + (1 - N/n)\log(1 - N/n)] \\ &= \frac{0.8T}{\sigma_T}\Omega(N/n) \end{aligned}$$

where $\Omega(N/n)$ is called the 'binary entropy function',[4] with a maximum of unity at $N/n = 0.5$ and zero at $N/n = 0$ or 1 (however, the expression is only valid at low N/n since it was assumed that there is negligible probability of more than one count per segment).

I_{TOF} may be used to characterize the TOF mass spectrometer and the upper limit will be determined by limits on N/n or σ_T. If we take $N/n = 0.1$, $T = 100$ μs and $\sigma_T = 100$ ns, we find

$$I_{TOF} \approx 3.8 \times 10^2 \text{ bits s}^{-1}$$

An expression for I_{TOF} which includes multiple counts per time segment is easily derived. A similar analysis can be carried out for the TOF detector.

Model in the presence of noise

If the possibility of noise is considered then the number of counts received at the destination may not be equal to the number of incident particles. We define the source alphabet $S(a_1, a_2, \ldots, a_j)$ and the destination alphabet $D(b_1, b_2, \ldots, b_k)$, where a_j is the source symbol representing the number of particles incident on a detector and b_k is the destination symbol representing the number of counts recorded at the destination. We can regard $a_1 = 0$ counts, $a_2 = 1$ count up to $a_{256} =$ 255 counts in the present examples. The mutual information $[I(S, D)]$ is a measure of the average quantity of information transmitted across the channel per symbol and is given by

$$I(S, D) = H(D) - H_S(D) \text{ bits per symbol}$$

where $H(D)$ is the destination entropy and $H_S(D)$ is the conditional entropy.[4,5]

The meaning of mutual information can be illustrated as follows. If there were a 1:1 correspondence between the signal sent and the signal received, then all information (uncertainty) in D would be due to information in S. However, if we have a noisy channel then some of the information in D is lost in the channel. The measure of the information which is lost is $H_S(D)$. Hence the information in D which is contained in S is $H(D) - H_S(D)$. This is called the mutual information $I(S, D)$.

In terms of the probabilities of events, we have

$$I(S, D) = -\sum_{k=1}^{T} p_k \log p_k + \sum_{j=1}^{N} \sum_{k=1}^{T} p_{jk} \log[p_j(k)]$$

Now,

$$p_{jk} = p_j(k)p_j$$

Therefore,

$$I(S, D) = -\sum_{k=1}^{T} p_k \log p_k + \sum_{j=1}^{N} \sum_{k=1}^{T} p_j p_j(k) \log[p_j(k)]$$

where
p_j is the probability of a_j being incident;
p_k is the probability of measuring b_k;
p_{jk} is the joint probability of a_j particles being incident and b_k counts being measured;
$p_j(k)$ is the conditional probability that b_k counts are measured given that a_j particles are incident;
N is the size of the source alphabet;
T is the size of the destination alphabet (Fig. A1).

With no noise present, then if n particles are incident a count of n is measured. In this case

$$N = T$$

$$p_j(k) = 1 \text{ when } j = k$$

$$p_j(k) = 0 \text{ when } j \neq k$$

and the second term in the above summation vanishes. However, if noise is present this may not be the case. As an example we consider that the source outputs one symbol per second each symbol of equal probability (e.g. counts from 0 to 255 with equal probability).

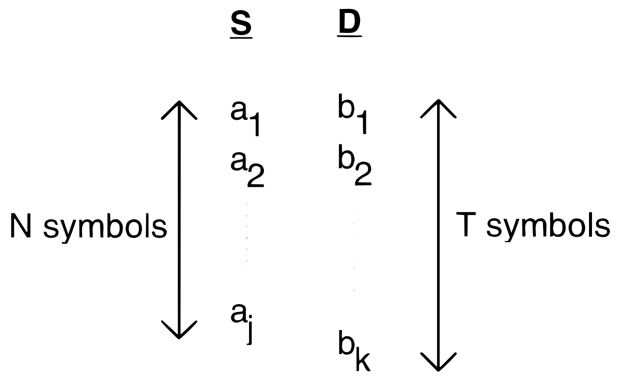

Figure A1. Source (S) and destination (D) alphabets. When the channel is noisy there may be extra symbols in the destination alphabet giving a total of $T > N$ destination symbols.

Therefore,

$$p_j = 1/N$$

We also assume that each destination symbol is received with equal probability:

$$p_k = 1/T$$

Let

$$p_j(k) = \beta \quad \text{for } j = k$$

and

$$p_j(k) = (1 - \beta)/(T - 1) \quad \text{for } j \neq k$$

i.e. we assume that the probability of correct transmission of a count is β and the remaining probability $(1 - \beta)$ is divided equally between the remaining $(T - 1)$ destination symbols. Therefore,

$$H_S(D) = -\frac{1}{N} N\beta \log \beta - \frac{1}{N}(TN - N) \times \frac{1-\beta}{T-1} \log\left(\frac{1-\beta}{T-1}\right)$$

$$= -\beta \log \beta - (1-\beta)\log\left(\frac{1-\beta}{T-1}\right)$$

where the first term refers to the summation for $j = k$ and the second to $j \neq k$. Therefore,

$$H_S(D) = \beta \log\left(\frac{1}{\beta}\right) + (1-\beta)\log\left(\frac{1}{1-\beta}\right) + (1-\beta)\log(T-1)$$

$$= \Omega(\beta) + (1-\beta)\log(T-1)$$

where $\Omega(\beta)$ is the 'binary entropy function',[4] with a maximum of unity at $\beta = 0.5$ and zero at $\beta = 0$ or 1.

We take $T = N = 256$ and consider the following cases:

Case (a): $\beta = 1/256$ (equal probability of any destination symbol regardless of the transmitted symbol, i.e. maximum noise):

$$\Omega(\beta) = 8 - (255/256)\log 255$$

and

$$(1-\beta)\log(T-1) = (255/256)\log 255$$

$$\therefore H_S(D) = 8$$

$$\therefore I(S, D) = 0 \text{ bits per symbol}$$

Case (b): $\beta = 1$ (zero noise):

$$\Omega(\beta) = 0 \quad \text{and} \quad (1-\beta)\log 255 = 0$$

$$\therefore H_S(D) = 0$$

$$\therefore I(S, D) = 8 \text{ bits per symbol}$$

Case (c): $\beta = 0.5$ (50% probability of correct symbol reaching destination):

$$\Omega(\beta) = 1 \quad \text{and} \quad (1-\beta)\log 255 \approx 4$$

$$\therefore\ H_S(D) \approx 5$$

$$\therefore\ I(S, D) \approx 3 \text{ bits per symbol}$$

In summary we have:

			$I(S, D)$ (bits per symbol)
Case (a)	$\beta = 1/256$	all noise	0
Case (b)	$\beta = 1$	zero noise	8
Case (c)	$\beta = 0.5$	50 probability of correct transmission	3

Note that in case (c) we have only 3 bits of information. The source has output 8 bits but 5 bits are lost in transmission. Therefore, in the presence of this noise, the FPD will only transfer a fraction of the information output by the spectrometer, i.e. the value of I_{FPD} will be reduced.

This loss may be understood from another angle. We consider the receipt of b_k at the destination and ask how much of the information transmitted by the source is received. Figure A2 shows that the received symbol could result from the transmission of any source symbol with the probabilities indicated in the present model.

As each of the source symbols is equally likely, the information transmitted by the source is 8 bits per symbol. In the present example each destination symbol is also equally likely (whether there is noise or no noise) and therefore 8 bits of information are received at the destination.

Now, for a noiseless channel,

$$p_j(k) = 1 \text{ when } j = k$$

and

$$p_j = 0 \text{ when } j \neq k$$

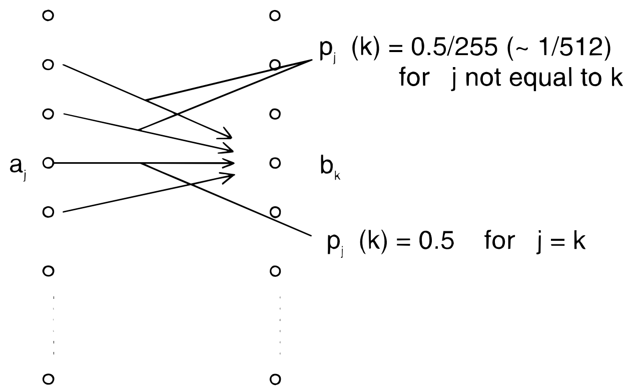

Figure A2. In this example of noisy transmission there is a 50% probability that a transmitted symbol will be received correctly at the destination and an approximately 1/512 probability that one of the other 255 source symbols could have been transmitted.

and therefore the information due to the channel is

$$1 \log 1 + \sum_{255} 0 \log 0 = 0 \text{ bits}$$

For the noisy channel, however,

$$p_j(k) = 0.5 \text{ when } j = k$$

and

$$p_j(k) = 1/512 \text{ when } j \neq k$$

and therefore the information due to the channel is

$$0.5 \log 0.5 + \sum_{255} (1/512)\log 512 = 5 \text{ bits}$$

Therefore, only three of the received bits of information are from the source.

If a more realistic assumption is made that the noise causes a maximum difference of only 2 counts between the incident particles and the received count, then it can be shown that the received information is 6 bits per symbol when $\beta = 0.5$. Other noise can be accommodated in his framework.