

Processing APT Spectral Backgrounds for Improved Quantification

D. Haley¹, A. J. London², and M. Moody¹

¹Dept. of Materials, University of Oxford, Parks Rd, Oxford, Oxfordshire, United Kingdom, OX1 3PH

²United Kingdom Atomic Energy Authority, Culham Centre for Fusion Energy, Culham Science Centre, Abingdon, Oxon, OX14 3DB UK

Abstract

We describe a method to estimate background noise in Atom Probe Tomography (APT) mass spectra, and to use this information to enhance both background correction and quantification. Our approach is mathematically general in form for any detector exhibiting Poisson noise with a fixed data acquisition time window, at voltages varying through the experiment. We show that this accurately estimates the background observed in real experiments. The method requires as a minimum the z-coordinate and mass-to-charge-state data as input, and can be applied retrospectively. Further improvements are obtained with additional information such as acquisition voltage. Using this method allows for improved estimation of variance in the background, and more robust quantification, with quantified count limits at parts-per-million concentrations. To demonstrate applications, we show a simple peak detection implementation, which quantitatively suppresses false positives arising from random noise sources. We additionally quantify the detectability of 121-Sb in a standardised doped Si microtip as $(1.5 \times 10^{-5}, 3.8 \times 10^{-5})$ atomic fraction, $\alpha = 0.95$. This technique is applicable to all modes of APT data acquisition, and is highly general in nature, ultimately allowing for improvements in analysing low ionic count species in datasets.

1 Introduction

Atom Probe Tomography (APT) provides a nano-scale view into material microstructures, and is unique in its ability to provide the 3D distribution of all elements across an increasingly wide range of materials systems. More recent applications of APT to materials research and development such as in semiconductors (Tu et al., 2017), geological (Piazolo et al., 2016) and nuclear materials (Dong et al., 2013), has resulted in an increased demand for improved quantification of the datasets acquired using this technique. Research into these materials is often focused on low concentration species, which may form alloying elements, act as dopants, or provide information on e.g. isotopic abundances

in datasets in geological specimens (Saxey et al., 2018) in astronomy applications (Heck et al., 2014), or in isotopic tracing (Haley et al., 2014).

As an element decreases in concentration, it becomes increasingly difficult to ascertain the true amount of said element within a material within a region of interest sampled by an dataset (Haley et al., 2015)(Larson et al., 2014). Indeed further data processing, such as identifying peak overlaps in mass spectra can hinge on the number of counts being accurate (particularly at low counts) (London et al., 2017). Low count data has been identified as a limiting factor in several areas, such as the quantification of species in small regions and/or low-concentration in transistors (Giddings et al., 2018).

However, as with all experimental methods, APT is not free from uncertainty, some of which arises from uncorrelated sources in the detector-specimen system. Statistically it is possible to separate “true counts”, i.e. ions evaporated and detected in a manner that permits correct time-of-flight (TOF) identification, from the contribution of background noise, given appropriate assumptions. Existing methods for “TOF independent” background correction in atom probe are functional (Oltman et al., 2009), but fail to account for changes in experimental conditions, such as voltage, resulting in a over-estimation of the background most noticeable near maximum masses. Furthermore a statistically rigorous treatment of uncertainty in background levels and its consequences for quantification, is currently lacking. We discuss improvements to the background model and show that improvements to quantification are possible.

Here we model the detector and specimen system as having an additive and uncorrelated (to pulsing) noise source that is uniformly distributed in time, and is represented as a Poisson process. This Poisson process, when converted into a histogram, yields a background that is additive to the APT data. We show that this background value can be estimated with only mass-to-charge and spatial information, allowing it to be applied without knowledge of the analysis conditions. In the context of this work, we state the method is “general”, in that there are a small set of assumptions built into the model, and that it can be applied over a large range of problem sets without breaking the assumptions needed. In this case the assumptions needed are that the noise sources are uncorrelated, and thus can be modelled in a Poissonian way. Thus if given a suitable model for the uncorrelated noise component, such as either the “background” noise here, or a model for peak tails, then the uncertainty estimations discussed proceed in the same way.

Using this background model, we can estimate confidence levels for detection. Using this approach we outline a method for the extraction of quantitative thresholds, based upon Skellam distributions (Alvarez, 2007) and modifications thereof (Bityukov et al., 2001)(Zech, 1989), for the estimation of the confidence limits in datasets. This is subsequently demonstrated to have applications to e.g. peak detection, the quantification of background noise, and hence true signal in APT.

1.1 Model

An APT experiment is conducted over a relatively long time scale by microscopy standards - typically datasets are acquired over the period of hours or days. As such, small noise rates in the detection system can add up over time to considerable total background intensity. This background noise in the experiment

can be modelled additively, that is, the contribution of the background noise is superimposed on the true dataset - as such, one must simply subtract it from the observation. To do this however, a suitable statistical model for this noise is required. Background noise, which by its uncorrelated nature is Time-of-Flight (TOF) independent arises from a variety of sources, including the sample and the detector. Such background noise should be distinguished from mass spectra peak overlaps/interference, which are not considered here.

In an APT experiment, the maximum mass-to-charge ratio is derived from the initial potential of the ion, and the rapid conversion of this potential to kinetic energy, as well as some fixed instrument parameters. Specifically, the maximum mass-to-charge, $(M)_{\max}$, is given by the maximum time of flight t_{\max} , the path length traversed by an ion d , the fundamental charge e and the specimen voltage V :

$$M_{\max} = 2eV \left(\frac{t_{\max}}{d} \right)^2 \quad (1)$$

where t_{\max} is given by the inverse of pulse frequency, $\frac{1}{f}$. Noise in the detector, which is continuous in experiment (wall clock) time t_{exp} , is mapped into the TOF spectrum to t_{noise} via a modulo operation $t_{\text{noise}} = \text{mod}(t_{\text{exp}}, \frac{1}{f})$, thus taking values 0 to t_{\max} on each pulse with equal probability in time. This equation assumes that the initial timing delay between the applied pulse to the tip and the timing signal to the detector has been correctly shifted, such that the mass spectrum begins at 0 Da.

If events are randomly occurring in experiment time, this will also randomly occur in the pulsing window as noise hits. Subsequently, we can model random noise events in the detector window, for the entirety of an experiment, as having a Poisson rate of occurrence, λ_d , which thus leads to a TOF independent noise intensity in TOF space t , with expected value given by Equation 2, which takes units of counts/time delta - i.e. counts/ns.

$$I_t(t) = \frac{\lambda_d}{t_{\max}} \quad (2)$$

This noise maps into M space using the prior mass-to-charge relation on the uniform probability distribution to obtain the transformed distribution. Generally, a transform function $y(x)$, applied to a density function $f(x)$ gives a transformed distribution $g(x) = f(x(y)) \left| \frac{dx}{dy} \right|$. In this case t is transformed into m with $\frac{dx}{dy} = \frac{1}{2kt}$ (from $m = kt^2$), where k is a constant of value $\frac{2eV}{d^2}$. It can thus be shown that uniform noise with TOF-space intensity λ_d appears transformed to M space (Equation 1) with intensity $I_m(M)$, for any fixed t_{\max} is thus:

$$I_m(M) = \frac{I_t}{2\sqrt{kM}} = \frac{I'_t}{\sqrt{M}} \quad (3)$$

which is combined with I_t to yield $I'_t = \frac{I_t}{2\sqrt{k}}$. This equation explains the common observation that the background is apparently more “intense” at the low-end of the mass spectrum (despite appearances, this is not the case). This rather arises from a narrowing of the inverse transform that maps the uniform TOF noise into M space, as shown in Figure 1, where the uniform distribution

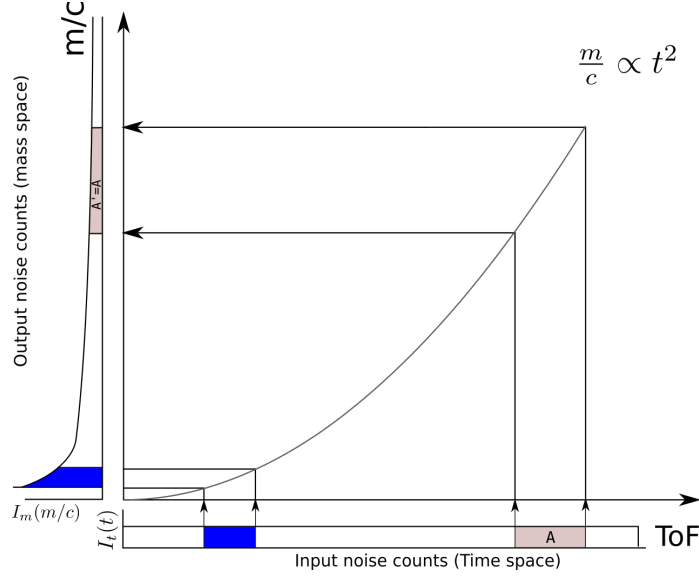


Figure 1: Conversion of an input uniform noise in TOF space into m/c space, which yields the characteristic background shape in APT experiments. Note that identically coloured areas have matching counts, as they define the same time or corresponding mass range, thus the same set of counts lie within these ranges.

in t space is mapped into m space. As the Poisson rate parameter and expected value are the same, we can similarly define an instantaneous rate in m/c space via $\lambda'_d = \frac{\lambda_d}{2\sqrt{k}}$. The difference between λ_d and I_t being the incorporation of pulse rate (Equation 2). Thus λ'_d characterises the noise level independently of pulse-rate (noise arrival rate), and I'_t is the expected value per- m/c bin.

Note that Equation 1 implies the flight path is a fixed value, however in practice the flight path is often a function of the detector coordinates due to the differing trajectory each ion makes from the tip to the detector. When the bowl-correction is applied to improve mass resolution, the M_{\max} value can be shifted by the differing flight path correction. We measured, for the reflectron fitted LEAP 3000, a difference of up to 6% (extreme values) in the M given by Equation 1 and the bowl-corrected M . This implies if using M_{\max} to estimate t_{\max} , then t_{\max} may be over estimated and the sharp cut-off in TOF is convolved with bowl-correction to give a broader distribution.

It is important to note that any peak height, for any fixed interval of observed signal, undergoes an equal transformation. Thus peak heights for real peaks in M space will also undergo this narrowing effect, appearing higher. However this ultimately does not change noise nor signal counts within this interval, and thus maintains the same Signal-to-Noise Ratio (SNR). This is clear, as the number of objects within two TOF or m/c limits is identical in both spaces - i.e. area is conserved for both signal and noise as shown by the coloured areas in Figure 1).

Equation 3 is however, incomplete. The form of I_m never reaches zero, and integrates to give the following number of counts, where c is a constant of integration:

$$\int I_m dm = I'_t m^{\frac{1}{2}} + c \quad (4)$$

For any closed interval the power is fixed, however for an open interval (i.e. $[0, \text{inf}]$), this equation shows that the background contains an infinite number of counts, which is non-physical. One simple correction is that the background is truncated to lie within $[0, m_{\text{max}}]$ at any given point in experiment time, to represent the finite number of counts of the noise in the spectrum.

However, this simple correction is insufficient by itself, because m_{max} is not fixed but is a function of experiment voltage, as given by Equation 1. However, we can approximate the total accumulated background throughout the experiment to be a sum of smaller instantaneous backgrounds, which approximates integrating I_m over experiment time, effectively a discrete approximation to a double-integral.

A final complication in most current reconstructed datasets is that metadata regarding the experiment may be incomplete in common data formats. Information such as the t_{max} may not be present ¹, preventing a successful post-hoc analysis of the background, as k and subsequently I_m cannot be computed. However, it is still possible to provide an estimate to the background intensity, even without full knowledge of V through the experiment, nor knowing each individual event's apparent time-of-flight. Practically, one can still measure a proxy for λ_d , and by slicing the dataset into chunks which group similar values of V into n groups, it is possible to fit a background value λ'_{d_i} for the i^{th} entry, and sum the result - this is shown in Equation 5, hereafter referred to as an adaptive fit.

$$I_m(M) = \sum_i^n \frac{\lambda'_{d_i}}{2\sqrt{M}}; M \rightarrow [0, M_{\text{max}}] \quad (5)$$

This result can be used to accurately model the behaviour of the background, including the “roll-off” of the background at high masses, which arises due to the voltage change throughout the experiment, which in turn stretches M_{max} . However, with small chunks of data, there may be a bias when estimating M_{max} (Appendix 7).

Whilst λ'_d is an important parameter in the analysis of a particular dataset, it is mostly not suitable for the comparison of differing datasets, as it depends on the length of the dataset itself. Using λ'_d estimate an integrated background intensity and thus SNR is a more appropriate route, as will be discussed.

To reliably estimate background intensity, a region in the mass spectrum free from signal is required, Whilst it is possible to manually identify regions that are presumed to be free from ions, it is less obvious as to how to do this in an automated and non-iterative fashion. Therefore, we identify possible locations of peaks in any atom probe experiment, by brute force. Appendix B provides the details of this computation. The results are summarised in Table 1, and show only a few guaranteed signal-free regions exist, all below 3 Da. However, this computation does not account for effects such as heavy ions “wrapping around” into low mass regions, nor for extreme widths in the H peak.

¹Such as in the EPOS, POS or ATO file formats (Larson et al., 2014)

Table 1: Regions where ions in their natural abundance do not map into mass spectrum, showing the widest 5% of gaps between potential peaks in M space. Values as simulated using $N_m = 3$, $R_m = 0.3$, $r_{sim} = 0.01$, excluding above mass = 100 Da. Details in Appendix B (8).

Start Mass (Da)	End Mass (Da)
0	0.94
1.05	1.45
1.56	1.93
2.02	2.27
2.73	2.93

2 Materials and Methods

We analyse the mass spectrum obtained from the APT analysis of Fe-VC steel, acquired in voltage mode at a temperature of 80 K, pulse frequency of 200 kHz, 15 % pulse-fraction. A constant background subtraction is presented in Figure 2. For later analysis, in order to provide a robust set of data, a standard sample was used, which should have minimal inter-laboratory and inter-operator variance, as compared to samples manufactured within each laboratory. Specifically, test data was acquired using a pre-sharpened silicon needles from a commercially available micro-pillar array “coupon”, which are Sb implanted (CAMECA, Madison, WI, USA). Data was acquired on a LEAP 3000HR in voltage mode, using 1% pulse-fraction at 55 K, frequency 200 kHz, with the voltage held at 3 kV unless otherwise noted.

3 Results

For the Fe-VC steel sample, it can be seen that a constant background correction fits well at low masses (Figure 2(a)), and reasonably up to intermediate masses, when fitting the background using the data between 1.2-1.85 Da (no adaptation).

As the reconstructed data has no metadata for processing, the background intensity parameter cannot be given an exact scaling from the acquired reconstruction. Thus λ'_d for this dataset is given per bin, with a fixed bin M width, here $\lambda'_d = 68100$ at a bin size of 0.02 Da, yielding a total background intensity of 1.20 M counts from 0 to 350 Da. Using an adaptive fit (Figure 2(b)) we find the total counts to be slightly less 1.14 M, giving an SNR of 8.48 (adaptive) and 8.14 (non-adaptive), across the full mass range. As will be subsequently discussed, it is this SNR that limits detectability. Additional results are presented as they arise within the discussion.

Figure 2 (right), shows the background correction for the varying M_{\max} utilising the acquisition voltage data. We find that there is not a large variation in the computed correction, as compared to the version that uses simple z-cuts to divide up the data (Figure 2c). This is likely due to the voltage curve for this dataset, which does not vary over a large range (5.7 - 6.7 kV).

After performing background subtraction, we find that the background computed seems to slightly over-estimate the observed background, likely due to small numerical instabilities at low fit ranges (See supplementary data).

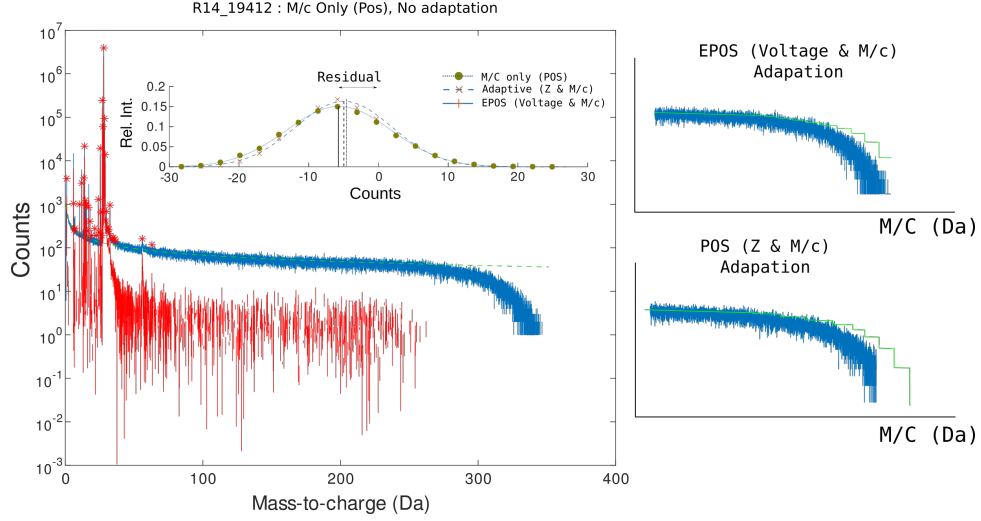


Figure 2: TOF independent background without adaptation (left), blue curve is original data, green is fitted, and red curve is corrected. Residual inaccuracy due to extrapolation over whole m/c range (inset) decreases when Z and/or voltage data is taken into account. Fit to roll-off improves with Z and voltage data (right). Stars on mass spectrum denote automatically detected peak positions.

4 Discussion

The background fits generated can be used to construct estimates for threshold for detectability, to assist peak detection, and to provide improved mass spectra and confidence bounds, as discussed below.

4.1 Simple detectability limit

We subsequently model the residual (“background corrected” spectrum) as a Gaussian of mean μ and variance σ^2 (inset Figure 2) - the mean corresponds to the deviation in predicted (Equation 5) vs the actual background intensity. We can perform a further empirical correction using the mean of this Gaussian ($\mu = -5.68$) recentring the observed mean in this range to 0, which corresponds to an arbitrary reduction in background estimate of approximately $\frac{\mu}{\lambda_d} \approx 0.5\%$, which slightly but also arbitrarily increases the final observed true signal - more robust background statistics are discussed later. This residual is mostly independent of the number of adaptations used when $n > 5$. This is the case when using voltage data from the EPOS file, and using a weighted average voltage to determine M_{\max} . However, the residual changes slightly depending upon where the background is measured.

The measured variance in the residual now allows for the direct computation of simple confidence limits for minimum detection at the per-bin level. For the adaptive case above, $\sigma = 7.4$ counts/bin post-correction, as measured between 100 and 300 Da. This variance can be used to compute significance to the number of corrected counts in each bin, and thus can be used for a somewhat robust peak detection. If we apply this to each bin with a fixed false positive

rate, say 1%, then false positives will occur due to the high number of trials. Thus we must perform a multiple-trial “familywise” correction to confidence, to ensure a fixed false positive rate for the group of tests, e.g. if a false-positive rate is set at 1%, and a trial is performed for each bin in a 200 Da range at 0.01 Da bin width, then $200/0.01 = 20,000$ trials are performed, then 200 false-positives will be expected per spectrum. Pre-adjusting α using a common statistical correction (familywise corrections) accounts for this over the whole test group.

Using a Sidak correction (Cerioli & Farcomeni, 2011), at 0.99 familywise confidence (one-sided Z-score test), gives 19.6 counts for positive detection of a peak (0.01 Da binwidth). So when checking all bins and we identify a bin containing greater than 19.6 counts, we can be 99% confident this is not a random fluctuation, but is rather a real detection. This is a slight improvement over the non-adaptive case, which corresponds to a minimum detection threshold of 20.4 counts. Both these values are post-correction, which corresponds to a “concentration” (more accurately a single-channel signal to overall dataset noise) of 17.1×10^{-6} . This is roughly in line with literature values, of ~ 10 ppm detectability for atom probe (Cerezo et al., 2007)². However, this value should be treated with caution, as detectability is more accurately determined by the relative intensities of other peaks, and the background beneath these³.

4.2 Peak Checking

If the region of m/c space used to estimate the background contains anything other than background counts then the background estimate will be incorrect. A simple test can be performed to give confidence that the background region is signal-free. The computation of intensity in m/c space, I_m , is derived from a Poisson variable as previously stated. For simplicity we assume that we have sufficient counts (≥ 80 total over 10 bins, to ensure mean is well estimated) when peak checking to be in the Gaussian limit. We utilise an Anderson-Darling test (Croarkin & Tobias, 2019) statistic, > 3 , to determine whether the observed background regions of the data is Gaussian, to ensure that any selected region does not contain peaks to guard against the failure of our bootstrapping approach (Appendix B). Such a failure could occur if there are large tails on the observed peaks wider than the mass resolution assumed in the initial bootstrap – this is more likely to be the case in laser mode.

However, the Gaussian assumption described above becomes increasingly invalid, noticeable at < 20 counts, where it may be necessary to estimate the background rate (λ_b) and/or the true signal rate (λ_s) using only few counts. This is the case in practical scenarios, such as in the examination of surfaces or other small regions, where both λ_b and λ_s are not known, or when the species are very dilute - we discuss this limit in the subsequent section.

²This will vary depending on confidence limit, width of actual peak, and for familywise corrections. Here we assume the peak is only 1 bin wide.

³This value does not provide true concentration limits, as the height of other peaks must be accounted for, which are discussed in Section 4.5

4.3 True Signal Confidence Bounds

Often in APT we are interested in accurately measuring the counts of a specific peak after background correction, and thus would like to estimate the confidence interval for the result. The analysis for significance of counts in a given bin in the mass spectrum is simplistic, but not incorrect per-se and can be refined when the assumption of bin-counts being Gaussian distributed is no longer valid. The relation to the “concentration” value reported above and that of Larson (Larson et al., 2014) is not direct as the value quoted here is for a single bin. To compare with Larson, one must rescale the λ_b calculation here by the time window Δt (From Larson et al) to obtain the total background counts, or similarly for I_m . If count limited, merging multiple bins will allow for better detection, however the problem remains the same. Thus changing bin width, assuming the signal area is captured exactly, does not change the signal to noise-ratio, as neither signal counts nor noise counts change over the chosen interval. However, as the peak’s shape is not known a-priori, it cannot be completely bounded, as the peak may continue imperceptibly. Thus signal to-noise changes with range-width, rather than bin width. A more detailed examination is given in Section 4.6.

There are several equations which can be used to model the confidence limits for the true signal λ_s , which are discussed below. The models include (1) Poisson confidence limits, (2) Skellam distribution limits and (3) a modified Skellam model, drawn from other fields such as Health Sciences and High-Energy Particle Physics. The models have several advantages and disadvantages regarding complexity of implementation and correctness, however we claim that these will provide accurate confidence bounds beyond what is available in APT today.

One key method is that of Larson et al. (Larson et al., 2014) which uses integrated values to determine the minimum detectability, referencing the work of Currie et al (Currie, 1968). This work, under certain assumptions, arrives at a threshold level of $14.1\sqrt{\lambda_b}$ for quantitative determination ⁴. There is some uncertainty in the procedure of Currie, due to conflicting assumptions, as noted by Alvarez (Alvarez, 2007) ⁵.

This brings us to our first refinement, Alvarez notes that a more correct model is the Skellam distribution. If the species concentration is known a-priori or if a concentration threshold is specified, then the confidence level required for detecting a true signal from the background corrected intensities can be well modelled using a Skellam distribution. The Skellam distribution gives the probability of a value Z arising from the subtraction of two Poisson distributed variables, X and Y , where the Poisson rate parameters are λ_1 and λ_2 respectively:

$$P(Z|\lambda_1, \lambda_2) = \exp[-(\lambda_1 + \lambda_2)] \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{Z}{2}} I_{|Z|}(2\sqrt{\lambda_1\lambda_2}) \quad (6)$$

Where $I_{|Z|}(\dots)$ is a modified Bessel function of the first kind of order $|Z|$, Z is the support, and λ_1, λ_2 are Poisson rate parameters. We model the arrival of

⁴Table 1, in Currie, Paired observations.

⁵In Currie’s work, Equation 7 assumes normality, however, the paper also invokes the Poisson distribution to equate normal and variance in the derivation (there Table 2). This simplifies the handling of negative values when performing background subtraction, as although $P(\lambda_1) + P(\lambda_2) = P(\lambda_1 + \lambda_2)$ – as per Currie – the converse does not hold $P(\lambda_1) - P(\lambda_2) \neq P(\lambda_1 - \lambda_2)$, as noted by Alvarez.

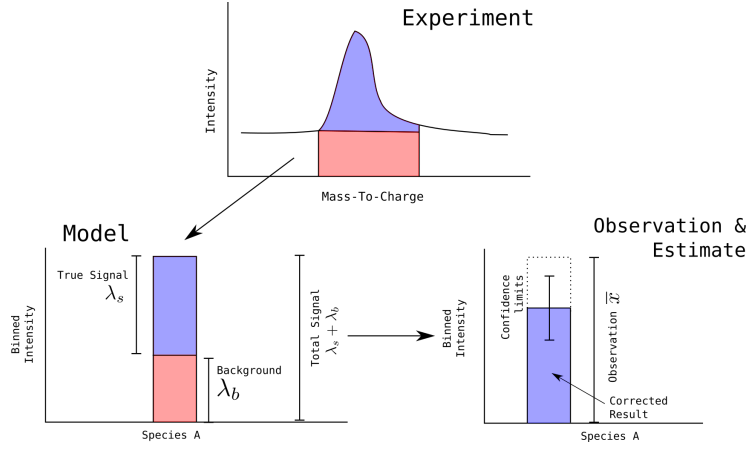


Figure 3: Data conversion from experiment to model and subsequent confidence bounds, showing the key components in the numerical model here.

counts on the detector for both true signal and background as Poisson variables. Figure 3 explains the application of λ_b and λ_s in an APT context, and how this refers to experiment data.

Using the Skellam distribution, or similar distributions, confidence intervals can be estimated from the CDF, and this can give rise to several queries.

- Query 1 : How many counts are required to ensure that a given observation (Z') is significant ($P(Z' > Z) > \alpha$), given a known background and true signal intensity (λ_b, λ_s)?

In practice the true signal intensity will likely not be known, and the query may be difficult to answer in this form.

The background can be reasonably estimated using empty regions of the spectra (as discussed above), such that λ_b is effectively known. It may be that the operator is anticipating a given level of true signal, and thus the Skellam distribution may be used (Equation 6), as shown by the analytical curve in Figure 4. As an example, assume that over a 1 Da interval, the anticipated true signal rate is 15 counts, and the background rate is 20 counts (A little higher than that observed in Si datasets when analysing for Sb), then $P(Z' > Z) > 0.99$ for $\lambda_s = 15, \lambda_b = 20$ occurs at $Z' = 33$ counts. This means that the peak would need to contain 33 or more counts for an operator to be 99% confidence that a peak exists.

To show how this can be used in an experimental context, consider a target species with anticipated concentration of 150 ppm, and assume we plan to collect a dataset of 100 k ions. A typical background is 1×10^{-3} Counts/(Da.Ion) at the expected mass (background occurs before ranging, thus is a normalised total signal value), with a target peak width of 0.2 Da. Our expected ion count will be $150 \text{ ppm} \times 100 \text{ k} = 15$, and the background is $0.2 \times 100,000 \times 10^{-3} = 20$ counts. Thus using Equation 6 ($P < 0.01$) we can provide a confidence bound of $[-2, 33]$ counts in this experiment, or $[-20, 330]$ ppm. Thus with these values we could not be confident that the experiment had detected any material if we detect the expected count, as our confidence interval includes zero. Reducing

Table 2: Approximate lower confidence bound for Skellam distribution (in corrected counts), for known true signal λ_s and background λ_b rates, at $\alpha = 0.99$. Values shown are net observed counts $T = \bar{x} - \lambda_b$ (\bar{x} is observation), $n_s = \frac{T}{\lambda_s}$

λ_b	10		20		30		50		100	
λ_s	T	n_s	T	n_s	T	n_s	T	n_s	T	n_s
1	12	12	16	16	20	20	25	25	35	35
5	17	3.4	21	4.2	24	4.8	29	5.8	39	7.8
10	24	2.4	27	2.70	30	3.00	35	3.50	44	4.40
20	36	1.8	39	1.95	42	2.10	46	2.30	55	1.20
100	127	1.27	130	1.30	134	1.34	136	1.36	141	1.41

the width of these error bounds would require a larger dataset - at 1 M ions ($\lambda_s = 150, \lambda_b = 200$), our bounds would be [95,205] ppm, and we would thus expect positive detection.

However, this query requires a true signal value. As the true signal strength λ_s is often not known a-priori, and is rather being estimated from the dataset, we can re-frame the problem. If poisson variables are independent these variables add to produce a further Poisson distribution with summed rates, so we can define the total observed rate as $\lambda_{sb} = \lambda_s + \lambda_b$.

Thus we can raise the following query, where $\text{Poiss}(x)$ is the Poisson distribution:

- Query 2: Given a known total count ($Y \sim \text{Poiss}(\lambda_{sb})$) and background ($X \sim \text{Poiss}(\lambda_b)$), what are the uncertainty limits in the corrected number of counts, with random variable $Z = Y - X$, e.g. where is $(P(Z > Z') > \alpha)$?

This query is best matched to an actual experiment - where λ_b can be estimated by background noise fitting from a signal-free mass resolution (as discussed previously). The true signal rate itself λ_s is unknown, so we must work with the total signal λ_{sb} , and Z follows a Skellam distribution. We can thus compute $\lambda_s = \lambda_{sb} - \lambda_b$ such that nominally (but not actually, as discussed later) samples of λ_s are Skellam distributed.

Table 2 shows the numerically estimated confidence limit for a chosen confidence level, α , for the Skellam distribution (Equation 6) (Alvarez, 2007) for several selected signal and background intensities. The numerical estimates are provided as direct integration of the Skellam distribution which is not straightforward. For a given background level λ_b , the entries in the table show the minimum level to achieve the stated confidence that the observed corrected signal (observation of $\lambda_{sb} - \lambda_b$) has a given intensity of λ_s or greater.

However, as λ_b and λ_s are not known a-priori, thus confidence intervals derived from the Skellam distribution are not entirely valid, and slightly underestimate the error involved, leading to a potential for under-estimation of the confidence levels. Specifically, the input $\bar{\lambda}_b$ and $\bar{\lambda}_s$ for the inputs are not known with complete certainty, but are rather estimates. This is most problematic at extreme low counts, when the estimator for the Poisson distribution is least accurate.

Figure 4 shows a simulation using estimators $\bar{\lambda}_b, \bar{\lambda}_s$ from a small number of samples from λ_b and λ_s , here 10, i.e. $\bar{\lambda}_s = \text{mean}(S_1, \dots, S_n), S_i \sim \text{Poiss}(\lambda_s)$. As

can be seen, the Skellam distribution is a close fit to the numerically obtained result, and indeed is steeper than the numeric result, which corresponds to an under-reporting of the confidence levels. Note that there is a non-zero probability of the count being negative, due to $X > Y$, which is non-physical. This arises as we assumed λ_{sb} and λ_b were uncorrelated. However, in reality these are increasingly correlated as $\lambda_b \rightarrow \lambda_{sb}$, which is not accounted for in either the analytical distribution nor simulation and marginally decreases the upper confidence bound. Table 3 shows this at differing confidence levels for fixed λ_b , λ_s , as well as the comparison to simulation.

Table 3 also shows a small offset, which increases at increasing confidence levels - either numerical correction factors, or an entirely alternate approach may be required to overcome this limitation, and improve confidence robustness further. Finally, a comparison to the detection limits of Currie is given in Table 4, which shows that the simulation and Skellam distributions are in good agreement at low counts, and offer much tighter bounds than the L_q value of Currie. The comparison is inexact, as Currie arbitrarily selects their detection limit to be $10\text{-}\sigma$ (there Equation 8) for L_q . The method of Bitjukov et al (Bitjukov et al., 2001) is also shown, which can provide even tighter confidence bounds at low counts, due to correct handling of the negative issue discussed above. The selection of a suitable α value however remains up to the operator, as this value is dependent on the problem context.

In the Skellam method for Query 2 we assumed λ_{sb} and λ_b were uncorrelated, which is not fully valid as $\lambda_b \rightarrow \lambda_{sb}$. Bitjukov et al derive a corrected protocol, provided the observed signal is nonzero (there Equation 3.5). Numerical issues may arise however for large observations, due to $(\lambda_s + \lambda_b)^{\bar{x}}$ and $\bar{x}!$ terms, which becomes very large for $\bar{x} > 100$, where \bar{x} is the observation of the total signal. However, at small counts $x < 100$, this method is simple to compute analytically.

$$P(\lambda_s|\bar{x}) \approx \left[\frac{(\lambda_s + \lambda_b)}{\sqrt{2\pi\bar{x}}e} \right]^{\bar{x}} \cdot \frac{1}{\sum_{i=0}^{\bar{x}} \frac{\lambda_b^i}{\sqrt{2\pi i}e^{\frac{i}{e}}}} \quad (7)$$

Bitjukov cites the combined background and signal probability distribution used by Zech (Zech, 1989), which uses Poisson distributions to formulate (Equation 8). This formulation by Zech is numerically stable over a larger range.

$$W(s_i) = \frac{\text{Pois}(s_i; \lambda_s + \lambda_b)}{\text{Pois}_{\text{cdf}}(\bar{x}; \lambda_b)} \quad (8)$$

Here s is the support to compute W , Poiss is either the PMF or CDF for the Poisson, where the first parameter is the function support. W can be accumulated to yield the confidence CDF $c(s)$, where s is the support, s_i for the i th weighting function.

$$c(s) = 1 - \sum_{i=0}^s W(s_i) \quad (9)$$

Numerical root finding can be used to find the solution of $c(s) = \alpha, 1 - \alpha$ to obtain confidence bounds. Figure 4 shows a computation for this equation,

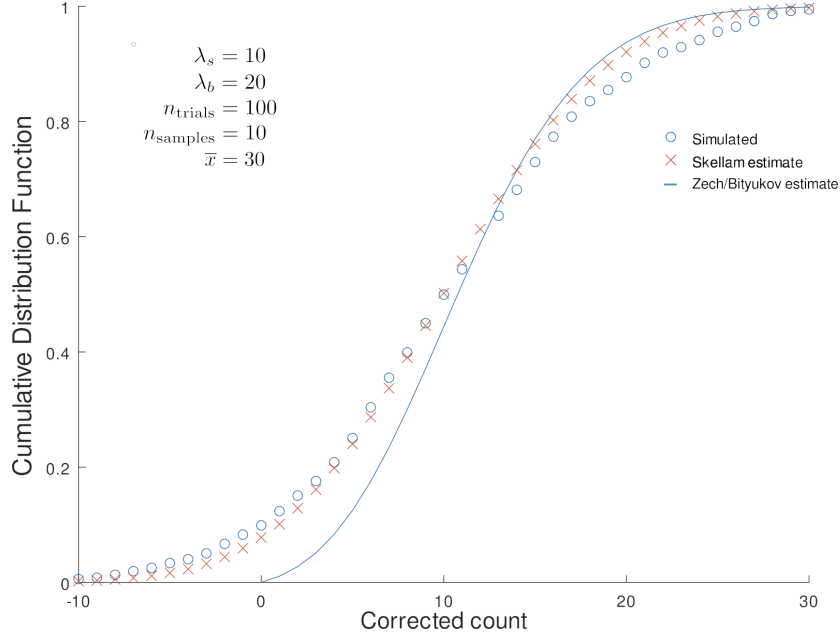


Figure 4: Distribution of analytical and simulation of corrected counts ($P(\lambda_s < Z)$), for a hidden λ_s , λ_b integrating the Skellam distribution (Equation 6, “Skellam estimate”) and using a numerical estimate for the estimated Poisson rates (“Simulated”), and method of Bityukov/Zech also shown for comparison. Note Currie quantification limit is 45 counts.

given \bar{x} as the observed true signal+background value. If \bar{x} is reduced to 10 (matching λ_s in Figure 4, as the Skellam and Bitukov methods take differing parameters), then the Skellam bounds are $[-7, 27]$, whilst the bounds given by Bityukov are $[0, 7.17]$ - recalling that Bityukov has an observation rather than true signal which is the origin of the greater narrowing, as shown in the figure. This method provides the tightest confidence bounds of any of the methods discussed, however it relies on a valid Poissonian assumption.

Finally, a slightly different question may be posed by the operator, which requires an alternative analysis:

- Query 3 : Given an observed signal $\bar{x} \sim \text{Poiss}(\lambda_s)$, and a known background intensity, at what number of counts is this significant ($\bar{x} > \alpha$).

The distinction from query 1 is that here, λ_s is unknown. However, to a first approximation, this query is relatively straightforward, and the confidence bounds on the Poisson distribution itself can be used. The bounds can be obtained from the inverse Poisson CDF, $Q(\lambda_b)$. For example, given an background rate of $\lambda_b = 10$ and $\alpha = 0.99$, the limit for an observation Y being drawn from a random variable $Q' \sim Q$ is $P_{\bar{x}}((Y \sim Q) < 0.01)$ occurs at $Y = 18$ counts. Note that this does not estimate the signal intensity, only determines whether or not this value can be anticipated given a fixed background value. Note that at this rate, the limit of Currie is higher, at 44.58 counts, and thus may cause false negatives. This approach to solving the query however, does not provide any

Table 3: Skellam and simulated counts to estimate λ_s with selected confidence limits $\lambda_s = 10$, $\lambda_b = 20$. The difference arises as λ_s is unknown a-priori, but is estimated from a sample set $n = 10$.

α	Skellam	Simulated
0.9	20	21
0.95	22	24
0.99	27	29
0.999	33	39

Table 4: Approximate $\alpha = 0.95$ confidence limits for the detection of a given λ_s , λ_b , in corrected counts. Values are reported are for $\lambda_s \geq x$, where x is a given table entry. Values (rounded up) are reported for Skellam distributions, numerical simulations the approximation of Currie ($L_q, 10\sigma$), and for the derivation of Bitjukov, for a given hidden λ_s and λ_b .

λ_s	λ_b	Skellam	Sim	L_q	Bitjukov/Zech
1	1	6	6	14	4
2	2	9	9	20	5
5	5	12	15	32	6
10	10	21	21	45	8

information on distribution of the background/signal pair - i.e. not all information available is used. The Skellam distribution can be used to provide this, assuming that $\lambda_b = \lambda_b$ and $\lambda_s = \lambda_s$.

Problematically, for a Skellam distributed value, we need to nominate a minimum detection level of interest as λ_s is unknown, however the cutoff cannot be known without nominating λ_s previously. Furthermore if, as in the applications section, one wishes to reject multiple values, the above α confidence will need to be adjusted. If using a multiple-correction method, such as the Sidak correction used in this work, then the number of independent trials that are to be made must be known in advance. Practically, this occurs at each point in the mass spectrum that is being rejected, so the familywise n will be on the order of 200-300 trials, depending on the mass resolution and the channels of interest in the mass spectrum (ranges). As an example, for a per-trial confidence of 0.95, a familywise confidence of 0.9997 is required ($n = 200$). Finally, the error values here are based on numerical models, and provide a minimum error level.

4.4 Test Data

To demonstrate the utility of the above metrics, datasets were acquired to investigate background intensity, rates and the effect on the quantification of low count species. Referring to the Si sample, this was used to generate the tabulated data. The sample was omitted to collect baseline data, as specified by the “No-sample” condition (Table 5).

The background counts for each state were measured, and normalised by the time taken for the experiment, which in turn was estimated from the maximal pulse number and pulse frequency, the results of which are given in Table 5. These values only include the background intensity model, and excludes any

Table 5: List of experiments undertaken for noise testing, sample is identical (where sample present) and in experimental sequence in each voltage case. Note background rate is based upon model. Rates $> 100\%$ is likely due to invalidities in the uncorrelated noise assumption arising from instrumentation. Laser data for comparison purposes. Voltage-mode samples at 1% pulse fraction unless otherwise indicated. Distance is perpendicular from electrode as measured optically. * denotes voltage used before reaching turn-on for the sample.

Condition	Voltage	Total Counts	Background (model) (%Data)	Background (Cts.s ⁻¹)
No-sample	500	9452	102	0.95
Pre-sharp, retracted	3000	9527	105	0.94
Flat-region@100um	3000	9557	106	0.95
Tip@100um	3000	9522	106	0.97
Tip@50um	3000	4696	105	1.11
Tip@20um	2500	4600	101	1.53
Tip@20um	2250	4804	94.2	1.29
Tip@20um	2850*	80612	2.11	21.9
No-Sample	3000	9578	108	0.96
No-Sample	3000	9548	110	0.97
Normal (15% Pulse Frac)	6200	1862391	5.99	59
Laser (0.2 nJ)	8500	45452558	5.43	152.1

non-background data. Note that the background model slightly overestimates total counts, as previously stated, leading to intensities slightly over 100%.

This data was used to estimate a threshold for deviation from the no-sample tests. Assuming Gaussian distribution in the background intensity measured between these experiments, the 50 μm rate was significantly different ($\lambda_{50\mu\text{m}} = 1.11$ counts/s, $P > 0.99$) as compared to the no-sample case (No-Sample group: $\mu = 0.95, \sigma = 0.012$ Counts/s). This in turn implies that the “background” noise is likely arising from the sample, or sample-system interactions. In this case, this is likely due to so-called “DC evaporation” – ion evaporation not correlated to the application of a pulse and hence detection of ions with uncorrelated (randomised) apparent TOF measurements. This is exacerbated when the tip is at the analysis position, even with a reduced voltage.

Examining the case where the standing voltage was raised to near-normal experimental conditions (2850 V - below the turn-on voltage for the sample), then the background fraction drops dramatically, from $\sim 100\%$ in the sample-free datasets, or where the sample is run below a voltage where noticeable evaporation occurs, to just $\sim 2\%$ of the dataset. However, the background rate increases dramatically, again presumably due to DC-evaporation. In comparison, an existing sample at 15% pulse fraction shows a much higher background data fraction of $\sim 6\%$ of the total data, at an approximate noise rate of 59 counts/s. This implies that a significant fraction of the noise in the experiment does not arise from the functioning of the detector or intermediary systems, but instead is uncorrelated ion evaporation from the sample. The laser dataset shows an even higher background count rate, implying additional uncorrelated evaporation beyond that of a voltage mode run, but is not investigated further

here.

To check whether the pulse frequency changes the background, implying adsorption/evaporation out-of-pulse as a noise contributor, the pulse frequency was varied from 50 kHz to 200 kHz in steps of 50 kHz and the SNR (time basis - counts/(counts/s)) computed as shown in Figure 5. Between 50-150 kHz, no clear trend is visible when using an approximate SNR metric, background estimated between 50-60 Da. However, at 200 kHz, a small increase in SNR is visible. To examine if this was significant, historical datasets which matched the parameters of the existing study were identified and the SNR computed to derive a standard error.

It is unclear if there is any change in the range of 50-150 kHz. Under the assumption that errors are similar between the 200 kHz range, and the 50-150 kHz range, there is no detectable difference within this group. However, the 200 kHz data point is different, with a higher SNR. This implies that the adsorption and re-emission of gaseous species within the analysis chamber due to pulsing, but in a TOF-decorrelated manner is unlikely for these experiments, as one would expect a near-linear dependence on frequency (Meier, 2019). If the evaporation was correlated with the initial pulse, then this would appear as non-noise data in the spectrum. It is however possible that re-emission is a continuous process, completely independent of pulsing at the standing voltage.

4.5 Sb-estimation

With the model described above, it is possible to derive probability limits for an observed signal, to determine the confidence bounds of a given observation. In turn this yields whether the observation can be reasonably considered to be detectable ($P(\lambda_s > 0) > P_t$), where P_t is the desired confidence threshold.

Using the 50 kHz data from the previous section, we construct the probability limits for the dopant count and thus concentration within the microtip samples, specifically examining the relatively weak 121-Sb peak. As per the above, λ_s and λ_b are signal and background inputs measured by $\lambda_s \approx \lambda - \lambda_b$, n_{trials} is the number of passes in the simulation, and n_s is the number of samples used to generate $\overline{\lambda_s}$ and $\overline{\lambda_b}$ in the simulation, where the Skellam estimate uses λ_s, λ_b directly.

From Figure 6, we can obtain a probability estimate for the confidence bands in the observed signal, similar to Figure 4. At the $\alpha = (0.05, 0.95)$ levels, this corresponds to confidence bounds of (39, 92), (36, 95) and (51, 88) counts for the Skellam, simulated and Bitjukov/Zech estimates respectively. Using the Skellam data shown, and given the large Si count (2.44 M ions) which is large enough to ignore statistical scatter, this yields a 121-Sb concentration confidence band of ($1.5 \times 10^{-5}, 3.8 \times 10^{-5}$ atomic fraction), based upon counting statistics limits, and the estimated intensity of the background. Notably, the Bitjukov reduces the bounds by $\sim 33\%$ (averaged against Skellam/Simulated), due to the correct incorporation of the non-negative constraint in their derivation which is not present in either the Skellam or simulated approaches. We further note that raw Poissonian confidence bounds do not include the variance from the signal, and thus under-estimates the confidence bounds somewhat. Overall, the method outlined here is a considerable improvement in the precision compared to the Skellam and Poisson methods, and subsequently improves the detection power of the APT system.

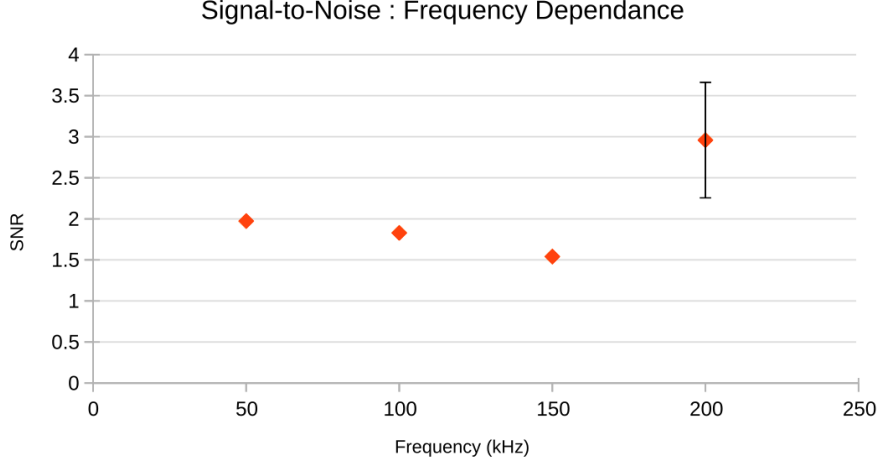


Figure 5: Signal-to-Noise ratio for 121-Sb⁺⁺ range in Si microtip as a function of frequency. Last datapoint includes data from a 3 year period at identical experimental conditions. Error bar is based on standard error ($\pm 2\sigma_e, n = 6$). 50-150 kHz data is significantly different ($P < 0.05$, mean), as compared to 200 kHz.

In the above example, our species is dilute, and we consider only a single species in the estimate of concentration confidence, ignoring the uncertainty in the primary peak. The lower and upper count limits are taken as extrema, and the Si counts are considered to be known exactly. This slightly underestimates the true error, for which accurate estimations would need to correctly propagate uncertainty, as the concentration of a dilute species “A” in solute “B” is formed as a ratio $\frac{A}{A+B}$. Thus any Skellam distributed error in A and B would need to be propagated, which is outside the scope of this work, however it is suggested that peak rescaling approaches such as the above (approximating peak shapes from larger peaks in the dataset) may be useful, under the assumption of constant peak shape.

4.6 Effect of range width

There exists an additional concern, which has not been addressed thus far. In the above example, the number of bins that are to be examined is known, however in an actual experiment this is not the case. One simplistic, and incorrect argument could be that increasing the number of bins increases the total available data, and thus minimises error. However, in the limit of a large “range” width, the background subtracted will converge to zero, by repeatedly merging a Skellam distribution with fixed $\lambda_s = \lambda_b$. In the limit as $\lambda_s \rightarrow \infty$ this equates to a “random walk” model, with mean λ_s . For small λ_s , this can be numerically modelled by accumulating subtracted Poisson random variables. Repeating this process can generate a probability distribution function for the error in a given number of bins (Figure 7a), ensuring that the accumulation allows for negative values to prevent a positive bias in a final result.

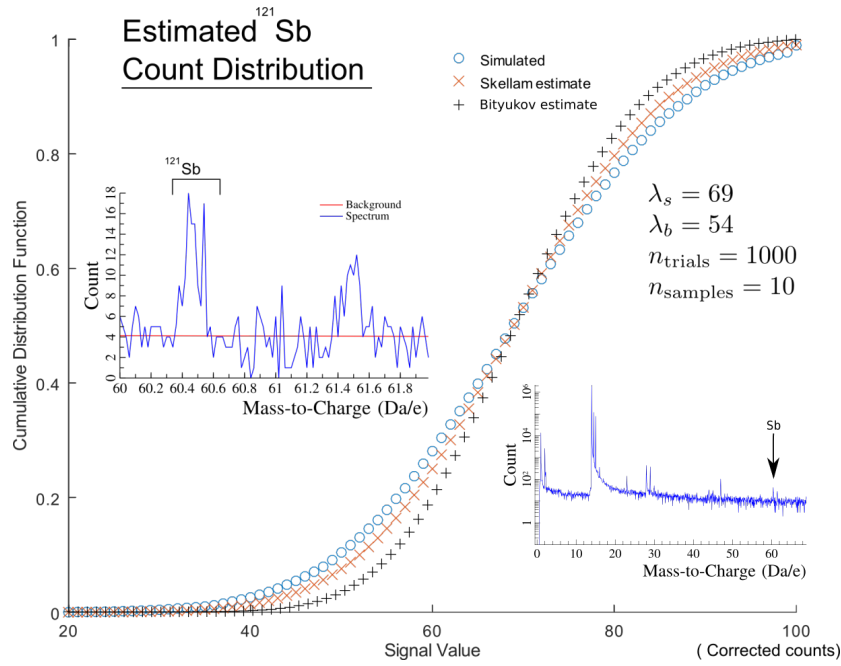
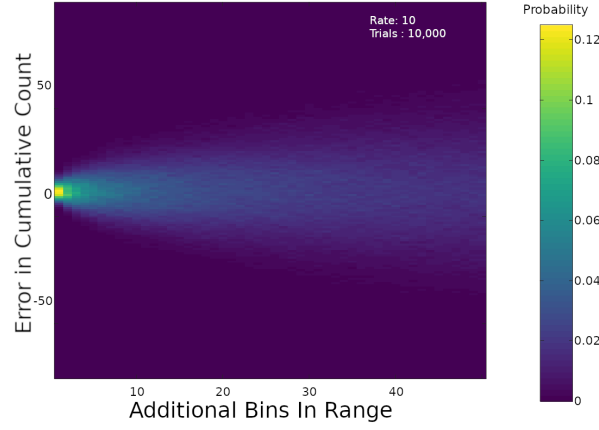
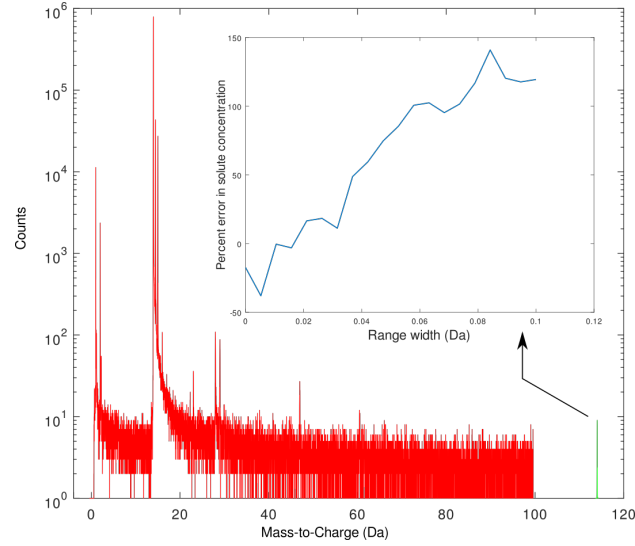


Figure 6: Estimation of ^{121}Sb content in a standardised Si “microtip”, with shown simulated and Skellam-estimates for the cumulative probability function in count space. Signal value is in corrected (background subtracted) counts. Inset (left) shows section of mass spectrum. Note λ_s, λ_b are inputs to generate the shown distribution, obtained from the inset mass spectrum. Inset (right) shows original spectrum, emphasising dilute Sb content.



(a) Integrated counting error for a Skellam random walk.



(b) Modified Si “microtip” spectrum, inset showing concentration as a function of range width.

Figure 7: Effect of “over-inclusion” of bins on count - divergence occurs by a random-walk like process. Values shown in (a) indicate probability estimates of a given error for each bin-count separately. (b) shows modified spectrum from Si microtip and (inset) concentration of green peak vs range width. Green peak generated using a shifted+rescaled Si peak (scale= 1×10^{-5}), with simulated background noise commensurate to real noise levels (not shown) - resulting in inset concentration errors.

From this, one can see that increasing range width to incorporate additional background corrected data, where there exists no true signal, causes the reported value to diverge with increasing range width thus increasing confidence width. There exists a trade-off between maximising SNR, by choosing a narrow range width (Hudson et al., 2011), and maximising counting statistics by maximising range width - simple metrics such as FWHM will fail at small counts. However, to optimise this trade-off effectively would require a quantitative model of peak shape, which is beyond the scope of this work.

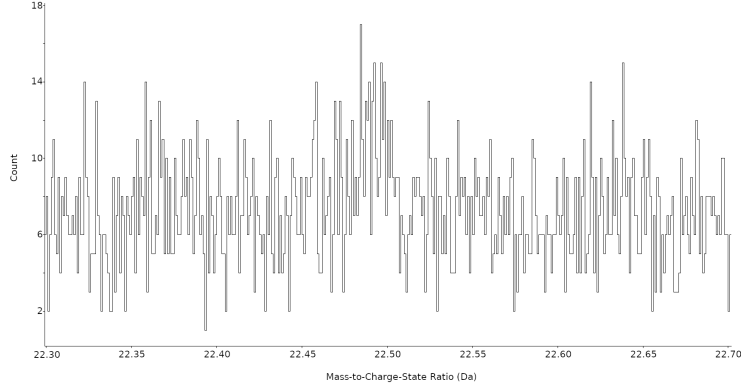
To demonstrate the problem further, we use the Si microtip example, and assume that the peak shape for Sb is identical to that of Si, only scaled down in intensity. This is a conservative approximation as at higher mass the peak will in fact be wider due to the transform shown in Figure 1. By assuming we know this peak shape accurately, we can provide an estimate for the error in the solute concentration measured by adjusting the range width. As the larger peaks will not impact error significantly, we can ignore this, and measure the error as a function of range-width, shown in Figure 7b. Initially, the error is negative, as we have not fully ranged the “true” peak, and as we expand we cross the zero measurement error level. Expanding the range wider worsens our accuracy, as the random walk error arising from ranging noise starts to dominate - this is after background correction, as there is no way to recover the information lost due to the Skellam random walk model shown in Figure 7a. This reinforces the need for accurate peak models in place of specifications for range width - too narrow a range produces error, as does too wide, however without a reference model it is impossible to identify where error is minimised.

4.7 Improving Spectra Correction

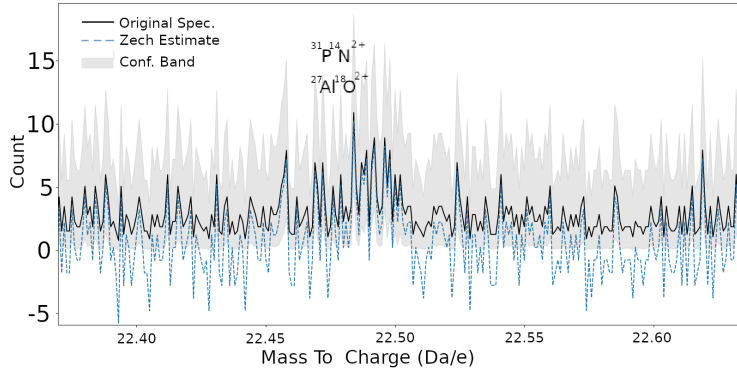
As the prior sections show, we can improve our understanding of the confidence bounds for a given α within a dataset, which has been used to quantify Sb in Si, and to provide robust peak detection. A further simple application is demonstrated which is applicable to any dataset where background estimates from random sources can be made.

To do so, instead of attempting to extract confidence bounds via e.g. the Skellam or Bitjukov/Zech methods, we attempt to better estimate the mean for any given signal bin. We can simply do so by setting $\alpha = 0.5$ to obtain the most probable location within the PDF of Bitjukov/Zech, analogous to a maximum likelihood estimator for the true signal. We can apply this estimator repeatedly to the entire mass spectrum, as shown in Figure 8. This eliminates negative values, which are clearly incorrect, and provides an optimal measure in each bin for the true signal. Note that this should not be summed across repeated bins to obtain total signal - this would bias the calculation incorrectly. Rather data should be accumulated before signal estimation. For the data here, the peak at 22.5 is more visible and clearer than when a naive correction method is employed, this peak may correspond to either AlO^{2+} or PN^{2+} .

This effect is most pronounced at the small counts shown here, where for several bins we can see that within for $\alpha = 0.05$ confidence (no familywise correction), we can be certain that there is true signal within this peak. This can prove useful to an operator, as we cannot perform aggregation of counts to obtain the total true signal for the entire species, without first deciding the range of mass-to-charge values to use. Thus, for small counts and hence dilute



(a) Typically observed mass spectrum, without confidence or most probable estimation, using direct background estimation, and suppressing non-negative values.



(b) Mass spectrum with confidence bands & most probable line. Shaded regions indicate $0.05 < \alpha < 0.95$ confidence bands (Bityukov/Zech model). Note that the most probable corrected version is non-negative, as by definition $\lambda_s > 0$. Confidence bands exceed zero, indicating positive detection vs null (Poisson noise) case (for chosen α).

Figure 8: Comparison of naive correction (direct background subtraction) and most probable estimate for λ_s in a mass spectrum of a steel.

species, this approach may prove useful. For large counts, this approach is not necessary as the naive and most probable methods converge to the same answer.

4.8 Experimental Considerations

Practically, there are several limitations in the presented model. The model assumes that there are identifiable regions of the spectra where ions do not exist, and thus the background can be reasonably estimated. The simulation in Figure 9 shows that the regions where, for a randomly selected material composition, there is an elevated chance of observing a region free from ions. This assumption is reliant on a Gaussian peak shape model, based upon the mass resolution of the instrument. However, it is well known that not only is the Gaussian a poor fit to experimentally observed peak shapes, both in voltage and laser pulsing modes. The peaks in a mass spectrum are typically right-tailed, and thus are more likely to be extended to higher masses. This becomes highly problematic when the sample is pulsed such that the tail of the peak is allowed to extend beyond the maximum time of the pulse window, and appear at a false low-mass value, a problem colloquially referred to as “wrap-around”. In the case of wrap-around to the next TOF window, the model here is unlikely to be suitable for background estimation. Hopefully, applying an Anderson-darling test statistic would fail (Section 4.2), however the robustness of this computation can be improved by utilising several test locations in the mass spectrum, and comparing the computed intensities, at the cost of an increased probability of rejection. This computation can be done with previously acquired datasets, but has not been implemented in the current work.

5 Summary

We have presented a protocol that allows for highly robust computation of background intensities in atom probe spectra, retrospectively applicable to acquired datasets. It is robust in that provided estimates are valid, and that incorrect estimates are statistically precluded. Background rates (Cts/s) and dataset fractions (noise intensity divided by number of collected ions) may be important and readily identifiable metrics for data quality reporting.

We have further quantified the detectability of species in APT datasets, using a confidence interval approach based on combined Poisson distributions. This approach has reduced the quantification below that outlined in prior literature, where we use a 121-Sb implanted in Silicon as an example application. Using the method of Bityukov/Zech, we improve measurement precision by a further $\sim 33\%$ over a Skellam model, which itself can improve over the method of Currie by over 50% - more than halving the number of counts required to quantify “detection”, depending on the chosen confidence level. We recommend the Bityukov/Zech method for low counts, and the Skellam for high counts or where simplicity of implementation is valued. The use of clear confidence bands from these models will be of value in the detection of very dilute species, and extends the detection capabilities of all APT systems, improving the quantitative presentation of confidence data when reporting composition.

Using this protocol we have demonstrated a statistically based method to reduce the false-positive rates in peak detection in atom probe, allowing for

a more aggressive detection of peaks in a given mass spectrum, and better quantification of dilute species - particularly in the face of varying voltages. We additionally demonstrate applications to improve the apparent quality of mass spectra through statistical corrections to raw observations which utilise more information than classical background corrections.

The methods presented here will assist in the development of semi-automated analysis protocols for APT spectra, improving overlap quantification and reducing operator bias, each a considerable source of error in APT, and extends the capabilities of APT through improved analysis of mass-spectral data for almost all datasets.

5.1 Data and code availability

The data used in this paper can be accessed at <https://deposit.ora.ox.ac.uk/datasets/uuid:a1c06908-13e0-49bc-848f-ef2d55a2df51>. The source code used to perform background fitting and peak detection can be accessed at <https://deposit.ora.ox.ac.uk/datasets/uuid:47517d27-13ad-4f18-bd02-70a085693b4f>.

6 Acknowledgements

The authors wish to acknowledge funding by the EPSRC HEMS project, EP/L014742/1. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

- ALVAREZ (2007). Poisson-based detection limit and signal confidence intervals for few total counts, *Journal of Health Physics* **93**, 120–126.
- BITYUKOV, S.I., KRASNIKOV, N.V. & TAPERCHKINA, V.A. (2001). Confidence intervals for poisson distribution parameter, *arXiv hep-ex/0108020v1*.
- CEREZO, A., CLIFTON, P.H., GALTREY, M.J., HUMPHREYS, C.J., KELLY, T.F., LARSON, D.J., LOZANO-PEREZ, S., MARQUIS, E.A., OLIVER, R.A., SHA, G., THOMPSON, K., ZANDBERGEN, M. & ALVIS, R.L. (2007). Atom probe tomography today, *Materials Today* **10**, 36–42.
- CERIOLI, A. & FARCOMENI, A. (2011). Error rates for multivariate outlier detection, *Computational Statistics & Data Analysis* **55**, 544–553.
- CROARKIN, C. & TOBIAS, P.E. (2019). Engineering statistics handbook, Online.
- CURRIE, L. (1968). Limits for qualitative detection and quantitative determination : Application to radiochemistry, *Analytical Chemistry* **3**, 586–593.

- DONG, Y., MOTTA, A. & MARQUIS, E. (2013). Atom probe tomography study of alloying element distributions in zr alloys and their oxides, *Journal of Nuclear Materials*, 270–281.
- GIDDINGS, A.D., KOELLING, S., SHIMIZU, Y., ESTIVILL, R., INOUE, K., VANDERVORST, W. & YEOH, W.K. (2018). Industrial application of atom probe tomography to semiconductor devices, *Scripta Materialia* **148**, 82–90.
- HALEY, D., CHOI, P. & RAABE, D. (2015). Guided mass spectrum labelling in atom probe tomography, *Ultramicroscopy* **159**, 338–345.
- HALEY, D., MERZLIKIN, S., CHOI, P. & RAABE, D. (2014). Atom probe tomography observation of hydrogen in high-mn steel and silver charged via an electrolytic route, *International Journal of Hydrogen Energy* **39**, 12221–12229.
- HECK, P.R., STADERMANN, F.J., ISHEIM, D., AUCIELLO, O., DAULTON, T.L., DAVIS, A.M., ELAM, J.W., FLOSS, C., HILLER, J., LARSON, D.J., LEWIS, J., MANE, A., J., P.M., SAVINA, M.R., SEIDMAN, D.N. & STEPHAN, T. (2014). Atom-probe analyses of nanodiamonds from allende, *Meteoritics & Planetary Science*.
- HUDSON, D., SMITH, G. & GAULT, B. (2011). Optimisation of mass ranging for atom probe microanalysis and application to the corrosion processes in zr alloys, *Ultramicroscopy* **111**, 480–486.
- LARSON, D.J., PROSA, T., ULFIG, R., GEISER, B. & KELLY, T. (2014). *Local electrode atom probe tomography*, Springer.
- LONDON, A.J., HALEY, D. & MOODY, M.P. (2017). Single-ion deconvolution of mass peak overlaps for atom probe microscopy, *Microscopy and Microanalysis* **23**, 300–306.
- MEIER, M. (2019). *Developing Laser-Pulsed Atom Probe Tomography Techniques for 3D imaging of Hydrogen in Steel*, Master’s thesis, Friedrich-Alexander Universität Erlangen-Nürnberg, URL [ArXivsubmit/3177611](https://arxiv.org/abs/submit/3177611).
- OLTMAN, E., ULFIG, R.M. & LARSON, D.J. (2009). Background removal methods applied to atom probe data, *Microscopy and Microanalysis* **15(S2)**, 256–257.
- PIAZOLO, S., LA FONTAINE, A., TRIMBY, P., HARLEY, S., YANG, L., ARMSTRONG, R. & CAIRNEY, J.M. (2016). Deformation-induced trace element redistribution in zircon revealed using atom probe tomography, *Nature communications* **7**, 10490.
- SAXEY, D.W., MOSER, D.E., PIAZOLO, S., REDDY, S. & VALLEY, J. (2018). Atomic worlds : current state and future of atomprobe microscopy in geoscience, *Scripta Materialia* **148**, 115–121.
- TU, Y., HAN, B., SHIMIZU, Y., INOUE, K., FUKUI, Y., YANO, M., TANII, T., SHINADA, T. & NAGAI, Y. (2017). Atom probe tomographic assessment of the distribution of germanium atoms implanted in a silicon matrix through nano-apertures, *Nanotechnology* **28**, 385301.

ZECH, G. (1989). Upper limits in experiments with background or measurement errors, *Nuclear Instruments and Methods in Physics Research*.

7 Appendix A: Small samples

Practically, there are small additional biases. Firstly, by utilising a numerical sampling of the uniform probability distribution from the dataset itself to find an estimate of M_{\max} , there is a bias in approximation which decreases the apparent M_{\max} , when taking a fixed number of samples, n_s . This has the effect of causing a small variation in the estimate of the maximum mass-to-charge for each window, which must be strictly less than the true maximum. This however will be small for any large n_s , as for uniform distributions X'_i the mean observed value of $\max(X')$ if X' is uniformly distributed is given by Equation 10, which can be used to correct the estimate to provide an unbiased estimate of M_{\max} .

$$\frac{1}{C} = \text{mean}(\max(X'_1, \dots, X'_{n_s})) = 1 - \frac{1}{n_s + 1} \quad (10)$$

For $n_s = 20$ samples (practically, we usually have a few thousand samples), the estimate is $< 5\%$ from the true value. Therefore this correction is only necessary when the sample size is < 50 .

8 Appendix B: Bootstrapping

The goal of this section is to identify sections within an arbitrary mass spectrum that can be used to initialise “bootstrap” a background fitting routine. To do so, it is necessary to identify regions of the mass spectra that will have a good chance of not containing any ions for any given dataset.

Here, we state that it is highly rare that ions practically have a very low mass. As such, the regions between the 1 and 2 Da peaks, which are often present in a dataset, will usually (but not always) be a good choice for a location to fit the mass-spectrum background, to determine λ'_d . The downside of this approach is that it limits the number of counts that can go into determining this intensity, thus reducing the number of subdivisions that can be made for the adaptive fit. The mass windows below 1 Da may seem a good choice however these can be problematic, as they may not be reliably or consistently within the open time of the detector, due to time-shifting of the zero-time of the pulse, required to synchronise the differing components of the atom probe.

To explore whether this is the only mass region that can be used, we consider the possible combinations of elements, and a given mass width. A small routine was developed to brute-force all the possible ionic combinations’ mass distributions, and these were plotted to identify regions where natural abundances will not fall in a mass spectra. In the simulations here, each ion is given equal weighting. To simulate the mass spectra, we require several parameters - the maximum size of a molecular species in terms of constituent atoms, N_m , the elements to consider, their maximum charge state, the mass resolution of the system R_m , here modelled as a Gaussian $\sigma = R_m$, the domain of the output mass spectrum $[0, m_{\max}]$ and the resolution of the simulated spectra, ΔM . Figure 9 shows a calculation for all natural elements from H to Bi, where the

intensity is the isotopic distribution weighted number of ions at these locations in the spectrum. Whilst there are numerous peaks in the overall combined fingerprint there exists small gaps. We can thus examine this spectra for “breaks” where little intensity maps into the simulated spectrum over a sufficiently wide mass range to be practicable. Each break represents an area where no atom probe experiment will show appreciable intensity at this point in the spectra, provided the simulated spectra is reasonable with respect to the actual experiment (e.g. we ignore long thermal tails in laser mode, which are more intense than the background and may “wrap-around” into the next pulse). We select

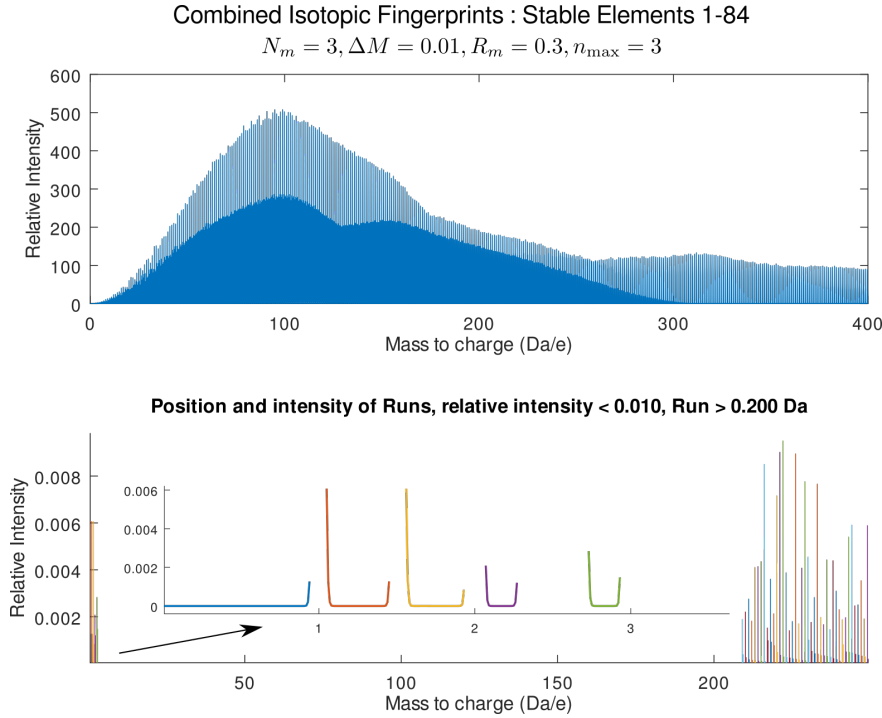


Figure 9: Upper: Combined isotopic fingerprint for all molecular species up to 3 components, all stable elements 1-83. Lower: Regions where there exists a gap in the isotopic fingerprint (note scale change from top) greater than 0.2 Da wide. Such a region can be used to compute background without true signal interference, and does not exist between $\sim 5 - 213$ Da.

the top 5% of the longest breaks, for a threshold of total normalised intensity of $1/5000$. For the parameters given in Figure 9, we find that the longest 5% of breaks (by length) excludes the mass regions of 4.93-213.1 Da. Table 1 shows the mass breaks below 200 Da, as the gaps at the high masses likely arise from considering insufficient numbers of ions in our simulated spectra (i.e. there could exist larger molecules, but we did not include them). Larger molecules will increase the weight of the isotopic fingerprint, and thus we ignore these larger gaps at high masses, as these are simulation artefacts. More specific a-priori information regarding the elements comprising the materials, the ions that are plausible, and the expected composition can provide a more informed approach to the simulation outlined above. Practically, using several available windows

should increase the robustness (reduced false readings) of the protocol, by allowing for a comparison of the background intensity at different masses, at the cost of increased chance of a rejected computation. The above fingerprint is quick to compute, and takes considerably less than a minute in total execution time.

In summary, without further iteration, the M region less than 3 Da is a good starting point to estimate the noise rate λ'_d , and sampling multiple locations is recommended.

9 Appendix C: List of symbols

- I'_t - Normalised intensity in TOF space
- $I_m(M)$ - Intensity of background in mass-to-charge space, m subscript denotes mass-to-charge space
- $I_t(t)$ - Intensity of background in time of flight space
- L_q - Quantification limit of Currie et al.
- P - Probability function
- P_t - Specified confidence value
- Q - Poisson inverse CDF (Quantile function)
- R_m - Mass resolution of simulated peaks when performing search for peak-free zones
- S_i - Sample drawn from Poisson distribution
- T - Net observed counts.
- V - Voltage of ion relative to detector
- X, Y - Random variables sampled from Poisson or Inverse Poisson CDF
- Z - Support for Skellam distribution, in an APT context, this corresponds to subtracted counts
- Z' - Observation variable, in background subtracted counts
- Δt - Window for signal arrival (From Larson et al)
- α - Confidence threshold value
- M - Mass to charge
- M_{\max} - Maximum mass to charge observed
- λ_1 - Placeholder Poisson rate for Skellam distribution (minuend parameter)
- λ_2 - Placeholder Poisson rate for Skellam distribution (subtrahend parameter)

- λ_b - Background Poisson rate in given interval
- λ_d - Arrival rate of ions on a detector across entire experiment
- λ'_d - Transformed Poisson noise rate into M/C space
- λ'_{d_i} - Transformed Poisson noise rate into M/C space, for i^{th} data slice
- λ_s - Signal Poisson rate in given interval
- $\lambda_{sb} = \lambda_s + \lambda_b$
- μ - Gaussian mean
- $\overline{\lambda_{\square}}$ - Estimator for rate λ_{\square}
- σ^2 - Gaussian variance
- c - Constant of integration
- $c(s)$ - Cumulative Distribution Function from Zech calculation
- d - Path length traversed by ion
- e - Fundamental charge on electron
- f - Pulse frequency
- k - System constant combining voltage, flight path and fundamental charge
- N_m - Maximum number of ions in molecule
- r_{sim} - Simulation bin mass step for mass spectrum.
- n - Number of samples
- n_s - Rate normalised observation, in terms of multiples of signal value.
- s_i - Support parameter for Zech confidence calculation
- t - Time of flight for ion/species
- t_{exp} - Experiment wall-clock time
- t_{max} - Maximum time-of-flight observed