Ultramicroscopy ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Ultramicroscopy

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

Guided mass spectrum labelling in atom probe tomography

D. Haley [a,b,*], P. Choi [a], D. Raabe [a]

[a] Max-Planck-Institut für Eisenforschung, Max-Plack Straße 1, Düsseldorf, Germany
[b] Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom

ARTICLE INFO

Article history:
Received 14 August 2014
Received in revised form
26 January 2015
Accepted 2 March 2015

Keywords:
Atom Probe
Data analysis
Standardisation

ABSTRACT

Atom probe tomography (APT) is a valuable near-atomic scale imaging technique, which yields mass spectrographic data. Experimental correctness can often pivot on the identification of peaks within a dataset, this is a manual process where subjectivity and errors can arise. The limitations of manual procedures complicate APT experiments for the operator and furthermore are a barrier to technique standardisation. In this work we explore the capabilities of computer-guided ranging to aid identification and analysis of mass spectra.

We propose a fully robust algorithm for enumeration of the possible identities of detected peak positions, which assists labelling. Furthermore, a simple ranking scheme is developed to allow for evaluation of the likelihood of each possible identity being the likely assignment from the enumerated set. We demonstrate a simple, yet complete work-chain that allows for the conversion of mass-spectra to fully identified APT spectra, with the goal of minimising identification errors, and the inter-operator variance within APT experiments.

This work chain is compared to current procedures via experimental trials with different APT operators, to determine the relative effectiveness and precision of the two approaches. It is found that there is little loss of precision (and occasionally gain) when participants are given computer assistance. We find that in either case, inter-operator precision for ranging varies between 0 and 2 "significant figures" (2σ confidence in the first n digits of the reported value) when reporting compositions. Intra-operator precision is weakly tested and found to vary between 1 and 3 significant figures, depending upon species composition levels. Finally it is suggested that inconsistencies in inter-operator peak labelling may be the largest source of scatter when reporting composition data in APT.

1. Introduction

Atom Probe Tomography (APT) is a powerful technique for obtaining 3D nanostructural data across a very small analysis volume, on the order of tens to hundreds of nanometres in each dimension [1,2]. APT is unique in combining atomic-scale chemical and spatial information, with data in the form of a "point cloud" with an associated mass-to-charge value for each point. The point cloud originates from the 3D reconstruction of discrete 2D detector events recorded during the experiment and can be on the order of 10^8 events.

Labelling of the mass spectrum, much as in any spectrographic technique, is required to assign detected events to a particular atomic species. This step is nominally referred to in APT as "ranging", whereby each spectrum is assigned a separate "range file". Operators select a start and end for each range within the mass spectrum, as shown in Fig. 1.

To estimate how frequently errors can arise in this process, two checks were undertaken on a corpus of 336 manually generated range files, each generated from different datasets. For each file, each range was checked to ensure that the ion label assigned to the range should be spanned; i.e. that the ion should have a peak in the assigned range, for any combination of isotopes. A second check, hereafter referred to as the "side-peak test" was undertaken to ascertain if one peak was assigned to a specific ion, then any theoretically larger peaks for that ion must also be assigned, however not necessarily to an ion of the same type (due to overlaps). Schematics indicating the tests are shown in Fig. 2.

Each check was performed to within a mass window of ± 0.1 amu, with charge states $1^+ \rightarrow 3^+$ (common charge states) being used to generate ion distributions. These checks have the capability to detect incorrect labellings, but are unable to determine the correctness of a given ranging.

From the above corpus, 87 of these files were found to possibly contain one or more errors according to these checks ($\sim$25% of the

* Corresponding author at: Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom.
E-mail address: daniel.haley@materials.ox.ac.uk (D. Haley).

http://dx.doi.org/10.1016/j.ultramic.2015.03.005
0304-3991/

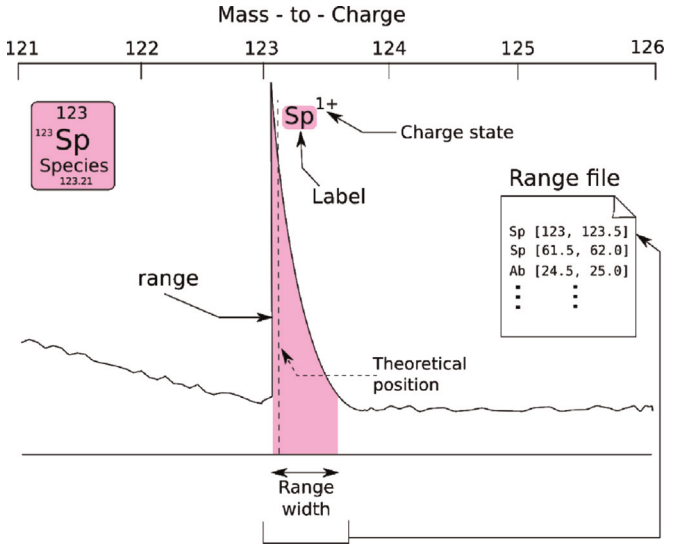

Fig. 1. Illustration of the ranging process: The operator labels a portion of the mass spectrum as a particular species. Selection is based upon isotope's (or combination thereof) theoretical mass and natural abundance.

total). 168 ranges were marked as inconsistent (~2 per file), 12 of which were due to the side peak test, and the remainder due to the labelled species' isotopes not being located within this mass window (direct-peak test). Manual review of a random sample of 20 inconsistent files was conducted. 10 reports were confirmed, 9 were within ± 0.4 AMU of the suggestion and were considered marginal, and one was due to 4+ charge state-thus a false positive.

Concerns about the appropriateness of ranging go beyond such simple checks, and have previously been tentatively discussed in APT literature [3], and are investigated here. It is the objective of this work to explore computational techniques to aid range selection, identification and validation, and to what extent this can be integrated into a single workflow. Such a workflow, if achievable, could reduce inaccuracies due to discrepancies between operators. Thus, robust methods in this direction can be seen as underpinning standardisation of data analysis of APT, and are the subject of this work.

2. Proposed procedure for mass-to-charge peak labelling

During the analysis of a mass spectrum, operators are performing a *labelling* step, for each peak. Specifically, the task to be performed is that given a peak mass and the available atom types, the peak identity must be as *uniquely* determined insofar as possible.

This is usually based upon some context the analyst has, such as the expected elements. In the identification step, an analyst determines what possible elements can occur at a given mass-to-charge ratio. Common mass spectral techniques have highly complex mass spectra, as these are often used to investigate large-chain organic molecules, such as present in biological systems. These systems often utilise "fingerprint" database techniques to identify molecules based upon models of fragmentation behaviour of large chains [4,5]. Due to the relatively small chains of molecular species often present (usually of size 1) in atom probe mass spectra, such a fingerprinting method has limited applicability in an atom probe context. Here we examine highly robust methods that do not depend upon empirical databases.

Formally, the labelling problem can be expressed as the construction of a mapping from the set of possible species (or combinations thereof), P, to the set of true experimentally obtained (and hidden) species, P'. Note that as there may be multiple candidate solutions, there is often not a one-to-one mapping from P, to P'. In APT, the set P is constructed by building a set of species from elements' isotopes, extracted as a subset from the periodic table, then combining them to produce possible labels. It is desirable to minimise the size of P, as far as possible given any available information.

The set P can be constructed using the following rule: the masses from the isotopes, each having mass m, for each species present in P, must sum to an expected target mass-to-charge, M. The value of M being obtained from the mass spectra itself.

This problem is strongly related to a well explored mathematical problem, known as the *Knapsack Problem*, a particular problem that remains computationally complex [6]. The family of knapsack problems involves attempting to fill a fixed capacity container with a set of discrete weights of different size as completely as possible. In the context of APT the weights are the available isotopes, and the container is the target peak mass.

Optimal solutions to the knapsack problem, from a computational complexity perspective are at best solvable in linear time [7]. However such algorithms only provide one single solution to our posed problem, where for the problem here all solutions that fit the given tolerance must be elucidated. Thus we anticipate that it is improbable that there will be significant improvement, in terms of computational complexity, over a brute force solution.

From this perspective, a brute-force solution must be employed in order to elucidate the set P. To account for the difference of charge state, the knapsack search is repeated N times for each charge state n, with individual masses m/n, accumulating results into P. Pre-processing of the input can be used to preclude

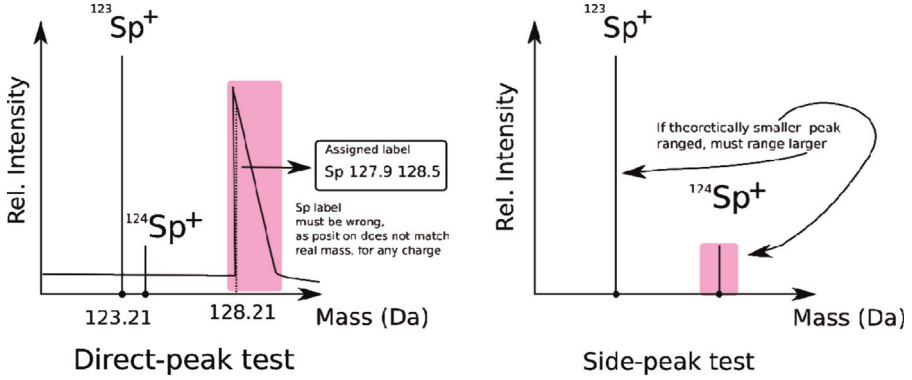

Fig. 2. Peak tests for direct mass existence, and for side-peak existence. Direct peak checks label's theoretical mass lies in the specified range. Side-peak test checks that theoretically larger peaks must be ranged if the smaller peak is ranged. Due to overlaps the identity of the larger peak's label may be different (i.e., something other than Sp^+).

Please cite this article as: D. Haley, et al., Guided mass spectrum labelling in atom probe tomography, Ultramicroscopy (2015), http://dx.doi.org/10.1016/j.ultramic.2015.03.005